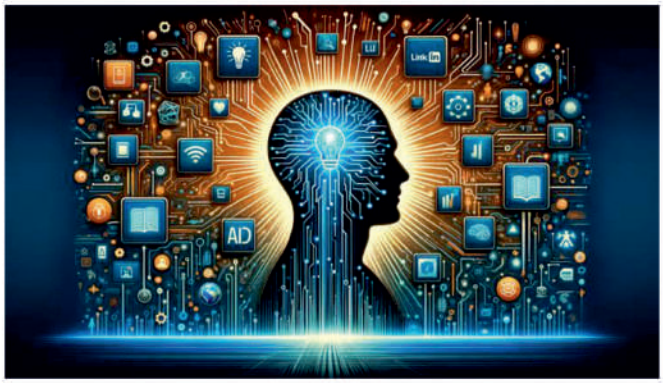
SIDDHI PUBLISHING HOUSE, NANDED

All types of books, periodicals, and magazines published
Conference & Seminar proceedings provided
ISBN numbers for books available
ISBN Numbers for Books available
Professional formatting, Editing & Distribution

CONTACT US
For publishing inquiries, contact:
Mrs. Pallavi Laxman Shete
Mobile No.: 9623979067

SPREADING KNOWLEDGE, NURTURING LITERATURE



OUR SERVICES

- ISBN Book
- Call for paper
- International Research Journal
- Ph.D. & M.Phil Thesis Book
- E-Book
- Conference and Seminar Proceeding
- Educational Videos and Notes
- Educational Activities

G Pay PhonePe BHIM 9623979067

SIDDHI
Siddhi Prakashan

Worldwide International Inter Disciplinary Research Journal
Website : www.wiidrj.com
Menoffice Address: Director : Mr. Tejas Rampurkar, Hyderabad.
Tuljai Niwas Ramkrishna Nagar, Near Ganpati Mandir
Vasmat Road Parbhani, Maharashtra (India)
Email :shrishprakashan2009@gmail.com / siddhiprakashan674@gmail.com
Mob : +91-9623979067

ISSN 2454-7905
9 772454 790004

Worldwide International Inter Disciplinary Research Journal (A Peer Reviewed) Vol. I, Issue -CLII, Year 12, Sept. -2026

Hindi Vidya Prachar Samiti's
Hindi Vidya Prachar Samiti's College of Law
(Affiliated to University of Mumbai & Recognised by Bar Council Of India)

ISSN : 2454-7905 MAH/NAN/10936/2015 SJIF 2026 - Impact Factor : 8.709

Worldwide International Inter Disciplinary Research Journal

(A Peer Reviewed)

Vol. I, Issue -CLII, Year 12, Sept. 2026

INTERNATIONAL CONFERENCE

Artificial Intelligence Nexus: Bringing Intelligence Across Fields



Edited by:
Dr. Mrs. Madhura Kalamkar,
I/C Principal, HVPS COLLEGE OF LAW, Mumbai.

Quarterly Research Journal

(Arts - Humanities - Social Sciences - Sports, Commerce, Science, Education, Agriculture, Management, Law, Engineering, Medical, Ayurveda, Pharmaceutical, Journalism, Mass Communication, Library Science Faculty's)

ISSN: 2454 – 7905

***SJIF Impact Factor:* 8.709**

Worldwide International Inter Disciplinary Research Journal

A Peer Reviewed Referred Journal

Quarterly Research Journal

(Arts-Humanities-Social Sciences- Sports, Commerce, Science, Education, Agriculture, Management, Law, Engineering, Medical-Ayurveda, Pharmaceutical, MSW, Journalism, Mass Communication, Library sci., Faculty's)

www.wiidrj.com

Vol. I ISSUE - CLII Year – 12 Sept. 2026

Hindi Vidya Prachar Samiti's

Hindi Vidya Prachar Samiti's College of Law

(Affiliated to University of Mumbai & Recognised by Bar Council Of India)

INTERNATIONAL CONFERENCE

Artificial Intelligence Nexus: Bringing Intelligence Across Fields

:: Editor in Chief ::

Dr. (Mrs.) Madhura Kalamkar

I/C Principal, HVPS College of Law, opposite Ghatkopar railway station,

Ghatkopar- west, Mumbai, Maharashtra – 400 086

::Editorial Team::

Dr. Amar Salve
Assistant Professor

Mrs. Khushbu A. Jain Gupta
Assistant Professor

Mr. Tushar Yadav
Assistant Professor

Mrs. Ekta Singh
Assistant Professor

Mrs. Neha Naik
Librarian

Address for Correspondence

Editor in Chief : Mrs. Pallavi Laxman Shete Website: www.wiidrj.com

Ramkrishna Nagar near Ganpati Mandir Vasmat Road Parbhani (India –Maharashtra)

Email: siddhiprakashan674@gmail.com / Shrishprakashan2009@gmil.com

Mob. No: +91-9623979067 **Director : Mr. Tejas Rampurkar, Hyderabad.**

(For International contact only +91-8857894082)

Worldwide International Inter Disciplinary Research

(A Peer Reviewed Referred)

Worldwide International Inter Disciplinary Research (A Peer Reviewed Referred) is quarterly published journal for Research scholars, teachers, businessman and scientists to integrate disciplines in an attempt to understand the complexities in the current affairs.

We also believe that both researchers and practitioners can contribute their knowledge by translating understanding into action and by linking theory and practice. This would enhance the relevance and thought in various related fields.

This Journal expected to bring together specialists in the field of commerce, economics, management and industry from different part of the world to address important issues regarding commerce, management and economics. One of the objectives of the journal is to create dialogue between scholars of various disciplines.

The editor, editorial team and the publisher do not hold any responsibility for the views expressed in **Worldwide International Inter Disciplinary Research (A Peer Reviewed Referred)** or for any error or omission arising from it.

The journal will cover the following Faculties for All Subject:

• Arts/ Humanities / Soc. Sci. / Sports	• Engineering
• Commerce	• Medical /Ayurveda
• Science	• Law
• Education	• Journalism
• Agriculture	• Mass Communication- **Library sci.**
• Pharmaceutical	• Social Work
• Management	• Any Other

Director: Mr. Tejas Rampurkar (For International contact only +91-8857894082)

Printed by
Anupam Printers, Hyderabad

Cost: Rs. 800/-

Editors of Worldwide International Peer Reviewed Journal are not responsible for opinions expressed in literature published by journal.

The views expressed in the journal are those of author(s) and not the publisher or the Editorial Board. The readers are informed, authors, editor or the publisher do not owe any responsibility for any damage or loss to any person for the result of any action taken on the basis of the work (c) The articles/papers published In the journal are subject to copyright of the publisher. No part of the publication can be copied or reproduced without the permission of the publisher.

Editorial Board

Dr. P. Neelkantrao Dept. of Economics, Pratibha Niketan Mahavidyalaya, Nanded. (MH., India.)	**Dr. Suhas Pathak** Dept. of School of Media studies S.R.T.M.U. Nanded. (MH., India.)
Dr. Pramod Ravindra Deshpande Wake Forest School of Medicine, Dept. of Cancer Biology,Winston Salem, NC, USA.	**Dr. Sachin G. Khedikar** Principal & Professor,Dept. of Rachana-Sharir, Shri. O. H. Nazar Ayurved College, SURAT (India.)
Dr Ashutosh Gupta Dept. of Sanskrit, HNB Garhwal University, Srinagar Garhwal Uttrakhand 246174 (India.)	**Dr. Mayuresh M. Rampurkar** Sardar Vallabhbhai Patel Hospital,(Neurosurgery),Ahmedabad. (G.India.)
Dr. Manish Deshpande N.S.B.College, Nanded. (MH., India.)	**Dr. Kulkarni J. N.** Library sci. S.R.T.M.U.Nanded. (MH., India.)

Co-Editorial Board

Dr. N. N. Bandela Dept. of Envi. Science Dr.B.A.M.U. Aurangabad. (MH., India.)	**Dr. Suman K. S.** Dept. of Oriental languages, Loyola College,(Autonomous) Affiliated to University of Madras,Nungambakkam, Chennai-600034 (India.)
Dr. S. P. Hangirgekar Dept. of Chemistry Shivaji University, Kolhapur. (MH., India.)	**Dr. Baswaprabhu Jirli** Dept. of Extension Education, Institute of Agricultural Sci. BHU, Varanasi. (India.)
Smt. Martha B. Department of English, Dr. B.R. Ambedkar F.G. College, Ladgeri, Bidar, Karnataka (India.)	**Dr. Chandan Bora** Dept. Of Commerce (MH., India.)
Dr. Mahesh Joshi Dept. Of Education S.R.T.M.U. Nanded.(MH., India.)	**Dr. Mangesh W. Nalkande** Dept. of Kayachikitsa Govt. Ayurved College, Nanded. (MH., India.)
Dr. Viraj Vilas Jadhav Professor and HOD, Dept. of Rachanasharir, Shri dhanwantry ayurvedic College and hospital sector 46 B CHANDIGARH. (India.)	**Dr. M.B. Kulkarni** Govt. Medical College, Nanded. (MH., India)

Peer-Review Committee

Dr. U. D. Joshi Principal Y. College, Ambajogai. (MH., India.)	**Dr. Vasant Biradar** Principal Mahatma Phule College, Ahmedpur. (MH., India.)
Dr. Joshi Prashantkumar Panditdev Department of Zoology (Fishery Science) Adarsh College, Hingoli-431513 (MH., India.)	**Prof. Dr. Mahendrakumar Y. Kulkarni** Head,Dept. of zoology N.S.B. Colloege, Nanded. (MH., India.)
Dr. Bibhishan Kare Rrsearch Guide, Professor and HOD Dept. of Sociology, NSB College, Nanded.	**Dr. Sanjay S. Pekamwar** School of Pharmacy, SRTM University, Nanded (MH., India.)
Dr. Prashant Andage Dept. of Envi. Sci Ratnagiri sub Center, Mumbai University(MH., India.)	**Dr. Shashikant B. Dargu** Dept. Of Sanskrit N. S. B. College, Nanded(MH., India.)
Dr. Shivraj G. Wanale School of Chemical Sciences S.R.T.M.U. Nanded(MH., India.)	**Dr. Subhash T. Pandit** Department of Economics, S. V. Night College, Dombivli (E) (MH., India.)
Dr. Sadavarte Rajesh K. Dept. of Computer, N.S.B. College, Nanded. (MH., India.)	**Dr. Vinay D. Bhogle** Dept. of English Degloor College, Deglor(MH., India.)
Dr. Kalpana Kadam (Bedre) Dept.of Political Sci., N.S.B. College,Nanded.(MH., India.)	**Dr. Sharada Bande** Head, Dept. of History, S. S. Suryabhanji Pawar College, Purna (Jn.) (MH., India.)
Dr. Deshpande R. P. Dept. Zoology Sharda Mahavidyalaya, Parbhani. (MH., India.)	**Dr. Gananjay Y. Kahalekar** Mahatma Jyotiba Phule Mahavidyalay, Mukhed Dist. Nanded. (MH., India.)
Dr. Prashant G. Gawali Associate Professor, Dept. of Physics Bahirji Smarak Mahavidyalya, Basmathnagar, Dist. Hingoli (MH., India.)	**Dr. Vikas Kundu** Geeta College of Education Butana(kundu), Sonepat – Haryana
Prof. K. Varalaxmi Deputy Director Sanskrit Academy, Osmania University, Hyderabad.	**Dr. Kamalakar Sharad Ingale** Head of Political Science Department SSMM Arts, Commerce and Science College, Pachora, Dist. Jalgaon.

Advisor Committee

Dr. Milind V. Rampurkar Govt. Ayurved College, Mumbai. (MH., India.)	**Dr. Sudhir Kokare** Nanded. (MH., India.)
Dr. Sanjay G. Shirodkar Principal Swa. Sawarkar College, Beed. (MH., India.)	**Prof. Dr. Chitanand M. P.** Dept. Of Microbiology N. S. B. College, Nanded. (MH., India.)
Dr. Darmapurikar Bhalchandra V. Dept. of Political sci., NSB college, Nanded. (MH., India.)	**Dr. Ashish Divde** Head Dept.of Envi.Sci,H.J.P.Mahavidyalaya,H.Nagar.(MH., India.)
Shri. Bidrkar Shivaji College, Parbhani (MH., India.)	**Dr. Anand R. Ashturkar** Dept. of Envi. Sci. N.S.B.College, Nanded (MH., India.)
Adv.Yadupat Ardhapurkar Law.,Nanded. (MH., India.)	**Dr. Karale Nagesh Baburao** Saraswati Mahavidyalaya, Kaij Dist. Beed. (MH., India.)
Dr. Nagesh R. Khadkekar SRTMU, Nanded. (MH., India.)	**Dr. Jeevan Pimpalwadkar** (Marathi) Research Guide, SRTMU Nanded. (MH., India)
Dr. A.I. Shaikh Associate Professor & Head, School of Social Sciences, SRTMU, Nanded. (MH., India)	**Dr. Rajendr Jadhav** Nanded. (MH., India.)
Shri Bharat Jangam Director Jangam Academy, Nepal..	**Dr. Jayanth Chapla** Dept of Zoology Osmania University, Hyderabad. (India)
Shri. Santkumar Mahajan Nanded. (MH., India.)	

About Hindi Vidya Prachar Samiti's College of Law

With its establishment in 2017, Hindi Vidya Prachar Samiti's College of Law is one of the upcoming Colleges in the Legal Educational field in Mumbai. Its aim is to provide quality legal education taking into consideration that Law is the necessity of the hour as it establishes Law & Order. Keeping in mind that law is a part of humanities, the college will empower its students with the tools required to succeed, inspire confidence and induce courage to meet s

uccessfully the challenges in the Profession and instill the professional culture of advocacy and a sense of loyalty to serve the needy person in the society.

About Hindi Vidya Prachar Samiti

On 15th August 1938, the people witnessed the birth of Hindi Vidya Prachar Samiti, a brain child of a visionary Late Shri Nandkishore Singh Jairamji. The Samiti was established with the objectives of catering to the educational needs of the Hindi speaking community. It made a humble beginning by starting a primary school, which gradually expanded into a full-fledged secondary school. With its primary objectives achieved the Samiti decided to extend its frontiers and broaden its horizons. As a result, Ramniranjan Jhunjhunwala College came into existence in 1963, enabling a larger section of the society to take advantage of the facilities provided for higher education. The Samiti then further decided to provide the society with an institution which will spread the awareness regarding our rights and duties and to set up new standards in legal education by inculcating a sense of professionalism, respect for rule of law and responsibility towards the society, which was accomplished with the establishment of the "Hindi Vidya Prachar Samiti's College of Law" in the year 2017.

About the Conference

Artificial Intelligence has emerged as one of the most transformative developments of the twenty-first century. Initially developed within the field of computer science, Artificial Intelligence is no longer limited to technological innovation alone. Today, it acts as a powerful nexus that connects knowledge across different disciplines, enabling new ways of thinking, researching, and solving complex problems.

One of the most important aspects of Artificial Intelligence is its ability to analyze large amounts of data and provide intelligent solutions in a short period of time. This capability has made AI extremely valuable in various fields. In healthcare, AI is being used to assist in diagnosis and medical research. In education, it is transforming teaching methods and learning experiences through smart tools and personalized learning systems. In law, AI is helping in legal research, case analysis, and data-driven decision-making. Similarly, in business and economics, AI is improving productivity, innovation, and digital transformation. These developments show that Artificial Intelligence is not confined to one discipline; instead, it creates a strong connection between different areas of knowledge.

However, the growing influence of Artificial Intelligence also raises important ethical and social questions. Issues such as data privacy, intellectual property rights, algorithmic bias, and accountability have become major concerns in modern society. This is why an interdisciplinary approach to AI has become essential. Technologists alone cannot address these challenges; the involvement of legal experts, social scientists, philosophers, and policymakers is equally necessary. When different disciplines work together, AI can be developed in a more responsible, transparent, and human-centric manner.

In conclusion, Artificial Intelligence is not just a technological advancement but a powerful bridge that connects different disciplines. By bringing intelligence across fields, AI is transforming the way knowledge is created and applied in modern society. If used responsibly and collaboratively, it has the potential to create a more innovative, efficient, and inclusive future.

Message from the Management



It gives me immense pleasure to present this collection of research papers from the International Conference, "Artificial Intelligence Nexus: Bringing Intelligence Across Fields," organised by Hindi Vidya Prachar Samiti's College of Law on 18 April 2026.

Artificial Intelligence is reshaping how knowledge is created, interpreted, and applied across law, education, governance, technology, business, and society. At the same time, its rapid advancement calls for thoughtful engagement with questions of ethics, privacy, accountability, human rights, and social justice. This conference provided a meaningful interdisciplinary platform for scholars, researchers, practitioners, policymakers, and students to deliberate on these important concerns.

The papers compiled in this volume reflect the intellectual depth, diversity of perspectives, and commitment to responsible innovation that the conference sought to foster. They contribute to the ongoing dialogue on ensuring that AI remains transparent, inclusive, ethical, and directed towards the larger welfare of society.

I congratulate the convenor, organising committee, faculty members, speakers, researchers, and participants whose sincere efforts made this conference and its publication possible. I am confident that this collection will serve as a valuable academic resource and inspire further research at the intersection of artificial intelligence, law, and other disciplines.

With best wishes for continued academic excellence and meaningful research.

Dr. Rajendra Singh
President, Hindi Vidya Prachar Samiti

Message from the Principal



It is a matter of great pride for Hindi Vidya Prachar Samiti's College of Law to present this collection of selected research papers from the International Conference, "Artificial Intelligence Nexus: Bringing Intelligence Across Fields." The intention of this publication is to foster meaningful academic dialogue on the opportunities, responsibilities and challenges created by Artificial Intelligence across disciplines.

Since its establishment, the College has endeavoured to provide legal education that is academically rigorous, socially responsive, and attuned to the changing needs of our times. In an era in which Artificial Intelligence is influencing institutions, professions, and everyday life, legal education must move beyond conventional boundaries. It must prepare learners to understand technology critically, uphold constitutional values, and respond responsibly to emerging social and legal challenges. The theme of the conference reflects the need to connect technological intelligence with legal reasoning, ethical reflection, and human values.

With this belief, the conference was conceived as an interdisciplinary academic platform that brought together scholars, researchers, practitioners, policymakers, and students to examine AI not merely as a technological development, but as a force that impacts governance, human rights, privacy, justice, education and the future of work. The deliberations highlighted the importance of ensuring that technological progress remains ethical, inclusive, transparent, and accountable. Although the term "Artificial Intelligence" was coined in 1956, AI has today become an integral part of decision-making, communication, healthcare, education, finance, governance and the justice system. Its growing impact makes informed research and responsible regulation more important than ever

This volume is a valuable outcome of those deliberations. The research contributions compiled here reflect diverse perspectives and meaningful engagement with contemporary issues surrounding AI. The volume represents a thoughtful compilation of legal, social, ethical, and technological viewpoints, presenting the subject in a manner that is accessible, research-oriented, and relevant to present-day concerns. I hope that this publication encourages further inquiry, collaboration, and policy-oriented research in this evolving field.

It is also our sincere endeavour to open new doors for researchers, students and young scholars to explore emerging questions in Artificial Intelligence, contribute original ideas, and develop solutions that serve society. We believe that this collection will be beneficial to academicians, legal professionals, policymakers and learners seeking to understand the evolving relationship between AI and law

I place on record my appreciation for the organising committee, faculty members, keynote speakers, researchers, authors and participants whose efforts contributed to the success of the conference. May this collection inspire young minds and the academic community to shape a future where intelligence is guided by wisdom, justice and human values.

Dr. (Mrs.) Madhura Kalamkar
I/C Principal, HVPS College of Law, Opposite Ghatkopar Railway Station,
Ghatkopar- West, Mumbai, Maharashtra – 400 086

Message from the Editorial Team

It is with great pleasure that we present this edited collection of research papers emerging from the International Conference, "Artificial Intelligence Nexus: Bringing Intelligence Across Fields."

The conference demonstrated that Artificial Intelligence can no longer be understood solely through the lens of technology. Its effects extend deeply into law, governance, education, human rights, privacy, intellectual property, employment, and public life. The papers brought together in this volume capture this breadth of engagement and offer valuable perspectives on the opportunities, dilemmas, and responsibilities that accompany the development and use of AI.

As members of the Editorial Team, our experience in legal teaching and research has reinforced the importance of encouraging scholarship that is not only contemporary, but also critical, ethical, and socially relevant. In compiling this volume, we have sought to bring together contributions that stimulate informed discussion and invite readers to examine AI through an interdisciplinary and human-centred approach.

We hope that this collection will be useful to students, researchers, academicians, legal professionals, policymakers, and all those interested in understanding the evolving relationship between artificial intelligence and society. More importantly, we hope it inspires continued inquiry into how innovation can be aligned with constitutional values, justice, accountability, and human dignity.

We sincerely acknowledge the guidance and support of the management and Principal, and thank the authors, speakers, participants, and organising committee for their scholarly contributions and cooperation. It has been a rewarding experience to be associated with this important academic endeavour.

Editorial Team:

- **Dr. Amar Salve**
  Assistant Professor
- **Mrs. Khushbu A. Jain Gupta**
  Assistant Professor
- **Mr. Tushar Yadav**
  Assistant Professor
- **Mrs. Ekta Singh**
  Assistant Professor
- **Mrs. Neha Naik**
  Librarian

Message from the IQAC Coordinator

A conference gains lasting value when its discussions are translated into knowledge that can be studied, questioned, and built upon. This collection of research papers from the International Conference, *"Artificial Intelligence Nexus: Bringing Intelligence Across Fields,"* represents that enduring academic outcome.

At Hindi Vidya Prachar Samiti's College of Law, quality education is understood as a continuous commitment to academic excellence, innovation, relevance, and social responsibility. In this context, the conference provided an important opportunity to engage with Artificial Intelligence—one of the most influential developments of our time; through an interdisciplinary and critical academic lens.

From the perspective of the Internal Quality Assurance Cell, such initiatives are central to building a vibrant culture of research within an institution. Quality in higher education is not limited to curriculum delivery; it is reflected in the questions we encourage, the dialogue we enable, the research we nurture, and the relevance of knowledge we contribute to society.

Artificial Intelligence presents both immense possibilities and urgent concerns. The papers in this volume engage with its legal, ethical, educational, technological, and social dimensions. Together, they demonstrate the value of interdisciplinary inquiry in addressing contemporary challenges with depth, responsibility, and academic integrity.

This publication is therefore more than a record of the conference. It is evidence of the College's continuing endeavour to promote research, critical thinking, collaborative learning, and academically meaningful engagement with emerging developments.

I appreciate the dedicated efforts of the management, principal, organising committee, editorial team, authors, speakers, faculty members, and participants. May this volume encourage a sustained pursuit of quality-oriented research and contribute to informed conversations on the responsible development of Artificial Intelligence.

Mrs. Saisha Ghosalkar
IQAC Coordinator
Hindi Vidya Prachar Samiti's College of Law

INDEX

Sr. No.	Title of the Paper	Name of Author	Page No.
01.	"The Right to A Fresh Start: Reconciling Article 21 and Section 24 of the JJ Act,2015 Against Algorithmic Labelling"	Dr. (Mrs) Geeta Kubsad	01
02.	Artificial Intelligence and Social Justice: What would Ambedkar Say?	Dr. Surajkumar Ramkripal Palbansh Dr. Baban More	09
03.	Artificial Intelligence as A Catalyst for Digital Transformation in Legal Academia	Ms. Neha Ramakant Naik	17
04.	"Who is the Author? Rethinking Authorship in the Age of Artificial Intelligence"	Asst.Prof. Khushbu A. Jain Gupta	25
05.	From Human Labour to Machine Intelligence: Legal Perspectives on the Future of Work	Asst. Prof. Tanavi Prasad Naik	32
06.	Combating AI -Generated Social Media Misinformation by Developing Critical Thinking Skills in the Digital Age	Aanand Kumar	39
07.	"AI and Social Media Manipulation"	Qadri Saad Sayyad Fakhruddin	49
08.	"Regulating Artificial Intelligence in Higher Education in India: A Critical Analysis of Emerging Legal Gaps and the Need for A Comprehensive Regulatory Framework"	Bansari Bhanushali Vignesh Narayanan	56
09.	"Human Creativity vs Machine Creativity: Implications for IP Law"	Ankita Bhoir Chandane	62
10.	Digital Surveillance in the Age of Artificial Intelligence: Examining its Impact on Youths' Human Rights	Shanita Singh Dr. Rishi Yadav	70
11.	Managing Synthetic Media Without Challenges to Speech: Deepfakes and Democratic Governance	Neetu Umashankar Singh	79
12.	International Conference on Artificial Intelligence Nexus: Bringing Intelligence Across Fields the Governance Gap Regulating Generative AI in Indian Courts Between Prohibition and Reform	Joshua Ebenezer	88
13.	"Integrating Artificial Intelligence Tools in Legal Education: Opportunities, Ethical Challenges and the Future of Legal Learning"	Chaitanya Arunkumar Pulekar	95
14.	Theoretical Analysis of AI-Induced Cognitive Offloading and its Impact on MHL of Adolescent (Mental Health Literacy)	Sakshi Sachdev Dr. Avani M. Mistry	102
15.	From Human Labour to Machine Intelligence: Legal Perspectives on the Future of Work	Manohar Sachdev Dr. Avani M. Mistry	111

"The Right to a Fresh Start: Reconciling Article 21 and Section 24 of the JJ Act, 2015 against Algorithmic Labelling"

Dr. (Mrs) Geeta Kubsad
Associate Professor
Pravin Gandhi College of Law, Vile Parle (West) Mumbai 400056

Abstract
In the era of algorithmic governance, the juvenile justice system in India faces a critical jurisprudential challenge: the transition from social "tagging" to permanent "digital tracking". While the **Juvenile Justice (Care and Protection of Children) Act, 2015[1](JJA)**, is anchored in the **"Principle of Fresh Start",** the rise of Artificial Intelligence (AI) and predictive risk assessments has created an indelible "digital master status". This automated labelling creates a persistent data shadow that follows a minor into adulthood, triggering the sociological harms of **secondary deviance** and **disintegrative shaming.** This article critically looks into issues where persistent digital tagging violates the **Right to Life and Personal Liberty[2]** under **Article 21** of the Constitution. By synthesising labelling theory with digital jurisprudence, the paper explores the systemic conflict between AI data persistence and the statutory mandate of record destruction under **Section 24** of the JJ Act. The analysis concludes by proposing a "Privacy by Design" solution: the **"Digital Kill Switch".** This framework mandates automated algorithmic expungement to ensure that a juvenile's "Right to be Forgotten" is not a mere statutory fiction but a constitutional reality in the digital age.
"The child of today cannot be the criminal of tomorrow... The objective of the juvenile justice system is to provide the child with a fresh start, a clean slate, and a life of dignity, unencumbered by the shadow of the past."

Keywords: Article 21, Section 24, JJ Act, Labelling Theory, Digital Kill-Switch, Right to be Forgotten

"The process of making the criminal, therefore, is a process of tagging, defining, identifying, segregating, describing, emphasising, making conscious and self-conscious; it becomes a way of stimulating, suggesting, emphasising, and evoking the very traits that are complained of." — Frank Tannenbaum [3].

Research Objectives

1. To critically examine the tension between the **"Principle of Fresh Start"** under the JJA[4], 2015 and the persistence of digital records generated through AI-driven governance.
2. To analyse how algorithmic profiling and predictive risk assessment systems contribute to the creation of a **"digital master status"**, reinforcing the stigmatisation of juveniles.
3. To evaluate whether persistent digital tagging violates the **Right to Life and Personal Liberty under Article 21 of the Constitution of India**, particularly with respect to privacy jurisprudence.
4. To explore the incompatibility between AI data retention practices and statutory mandates of record destruction under **Section 24 of the JJA 2015**.
5. To propose a rights-based technological framework **"Digital Kill Switch"** grounded in **Privacy by Design**, ensuring effective realisation of the **Right to be Forgotten** for juveniles.

Research Questions

1. Does algorithmic governance undermine the rehabilitative philosophy embedded in the juvenile justice system?

---

[1] Juvenile Justice (Care and Protection of Children) Act, 2015

[2] Art 21 Constitution of India

[3] King, A. (2018). Just mercy through cultural and convict criminology. Journal of Criminal Psychology, 8(1), 80-95.

[4] *Juvenile Justice (Care and Protection of Children) Act, 2015*

2. To what extent does digital tracking of Juveniles amount to a violation of informational privacy under Article 21?
3. How does **Labelling theory** explain the transition from temporary social stigma to permanent algorithmic identity?
4. Is **the statutory guarantee of record destruction** under Section 24 rendered ineffective in the age of big data and AI?
5. Can the **Right to be Forgotten**, as recognised in *K.S. Puttaswamy v. Union of India*[5], be meaningfully extended to juveniles in digital ecosystems?
6. What regulatory or technological safeguards (such as a Digital Kill Switch) can reconcile AI governance with constitutional protections?

Scope and Limitations

Scope

- Addresses the intersection of **juvenile justice law, constitutional law and digital governance** in India.
- Examines statutory provisions under the JJ Act alongside constitutional protections under Article 21.
- Incorporates interdisciplinary insights from **criminology (labelling theory)** and **technology law (AI governance, data privacy)**.
- Analyses emerging doctrine of the **Right to be Forgotten** within Juvenile Indian jurisprudence.

Limitations

- Limited availability of empirical data on AI deployment in India's juvenile justice administration.
- The concept of a "Digital Kill Switch" is largely **normative and theoretical**, lacking current legislative backing.
- Rapid technological evolution may render certain observations time-sensitive.
- Comparative international perspectives are not exhaustively covered (unless added separately)

Research Methodology

This research follows a **doctrinal and interdisciplinary methodology**, alongwith legal analysis with sociological theory:

- **Doctrinal Analysis**: Legislative provisions (JJ Act, IT framework), constitutional provisions (Article 21), and judicial precedents such as *K.S. Puttaswamy*[6].
- **Case Law Analysis**: Study of privacy and dignity jurisprudence, including evolving recognition of informational self-determination.
- **Theoretical Framework**: Application of **Labelling theory** (Becker, Lemert) to explain how algorithmic classification leads to **secondary deviance and identity fixation**.
- **Comparative Analytical Approach (Limited)**: Select references to global data protection principles like **Privacy by Design** and digital erasure norms.
- **Normative Analysis**: Development of the **Digital Kill Switch model** as a policy and technological solution to reconcile law and AI systems.

Literature Review

The literature specifically spans three key domains: juvenile justice, labelling theory, and digital privacy.

(A) Juvenile Justice and Reformative Philosophy

Indian juvenile justice jurisprudence emphasises rehabilitation over punishment. Scholars have consistently interpreted the *Juvenile Justice Act, 2015,* as embodying a **welfare-oriented and reformative approach**, particularly through provisions such as Section 24, which mandates the

[5] AIR 2017 SC 4161

[6] AIR 2017 SC 4161

erasure of records to ensure reintegration. However, contemporary critiques argue that **digital governance mechanisms risk undermining this rehabilitative ideal**.

(B) Labelling Theory and Secondary Deviance

Criminological theorists such as *Howard Becker*[7] and *Edwin Lemert*[8] argue that deviance is socially constructed through labels. Once labelled, individuals internalise stigma, leading to **secondary deviance**.

This article extends the theory into the digital realm, arguing that **algorithmic labelling transforms temporary stigma into permanent identity**, creating a "digital master status" that is far more enduring than traditional social tagging.

(C) Privacy Jurisprudence and Digital Rights

The landmark decision in *K.S. Puttaswamy v. Union of India*[9] which highlighted on privacy as a fundamental right under Article 21, includes **informational privacy**. Subsequent studies explores **Right to be Forgotten** as a derivative right, though its contours remain underdeveloped in India. Legal scholars highlight that **data permanence conflicts with dignity**, in relation to the vulnerable groups like juveniles.

(D) Gap in Literature

While existing works address juvenile justice, privacy, and AI separately, there is a **significant gap at their intersection**. Few studies examine how **algorithmic persistence directly undermines statutory erasure rights** in juvenile law.

This article addresses that gap by proposing a **technologically enforceable constitutional safeguard, the Digital Kill Switch**.

Introduction

The "Principle of Fresh Start" is the heartbeat of Indian juvenile jurisprudence, yet the cold persistence of binary code is systematically silencing it. This principle, enshrined in Section 3(xiv) of the **JJA, 2015** mandates that all records of children in conflict with the law be erased once they reach adulthood or have completed their rehabilitation. The underlying philosophy is restorative, not retributive; it recognises that adolescence is a period of biological and cognitive flux, and mistakes shouldn't be solidified into a permanent identity. However, as the Indian system increasingly integrates with the **Interoperable Criminal Justice System (ICJS)**[10] and the **Crime and Criminal Tracking Network & Systems (CCTNS)**[11], this statutory promise of a "clean slate" is colliding with the data-retentive nature of Artificial Intelligence.

In the mid-20th century, criminologist **Frank Tannenbaum**[12] warned that the social "tagging" of a youth the public labelling of a child as "bad" or "delinquent" creates a self-fulfilling prophecy. This process, which he termed the "dramatisation of evil," forces the individual to internalise the negative label, eventually leading them to lean into a life of crime as a response to social exclusion. In the 21st century, AI has transformed this social stigma into a permanent **"Digital Master Status."** Unlike a human community that may eventually forgive or forget a childhood transgression, an algorithm is designed for total recall. When a juvenile is entered into a database, the resulting "digital tag" becomes a shadow that follows them across networks, through school records, and into background-check algorithms used by employers and banks.

The crisis we face is one of **the law's technological obsolescence**. In today's networked justice system, the manual destruction of physical files under **Section 24 of the JJ Act**[13] is an exercise in futility if the underlying data remains cached in predictive algorithms. While a Juvenile Justice Board

---

[7] Howard S. Becker, *Outsiders: Studies in the Sociology of Deviance* (New York: The Free Press, 1963), 9;

[8] Edwin M. Lemert, *Social Pathology: A Systematic Approach to the Theory of Sociopathic Behavior* (New York: McGraw-Hill, 1951), 75.

[9] AIR 2017 SC 4161

[10] https://www.pib.gov.in/Pressreleaseshare.aspx?PRID=1799232

[11] https://digitalpolice.gov.in/writereaddata/brief180917.pdf

[12] Frank Tannenbaum, *Crime and the Community* 19–20 (1938).

[13] Juvenile Justice Act,2015

may order the physical shredding of a case file, the "data points" harvested from that file the zip code, the nature of the offence, the peer group associations continue to live within the "black box" of predictive policing tools. These tools do not see a reformed adult; they see a statistical risk. This results in **"shadow labelling,"** where a young person is denied a passport, a job, or a loan because an algorithm has "tracked" a record that the law claims no longer exists. Consequently, the statutory "Fresh Start" is becoming a legal fiction. To prevent this, Indian jurisprudence must move beyond the physical act of destruction and embrace the concept of **"Algorithmic Expungement."**

Conceptual Framework

A. The Criminology of the Code: Digitising Labelling Theory

Traditional labelling theory, pioneered by Edwin Lemert and Howard Becker, argue that act of "tagging" a child as a delinquent can lead to **Secondary Deviance**. In this sociological framework, "primary deviance" refers to the initial rule-breaking act, which is often transitory. However, once the state intervenes and applies a formal label "delinquent", the youth's social reality shifts. They are no longer children who made mistake rather "delinquent" in the eyes of authority. This transition is not merely administrative; it is deeply psychological. The individual begins to internalise this identity, eventually committing further crimes not because of an inherent criminal nature, but as a reaction to the stigma and social exclusion that the label carries.

In the digital age, this process is amplified and obscured by what can be termed **"shadow labelling."** Unlike traditional labelling, which occurred through visible social interactions such as a court hearing or a police officer's stern warning, digital labelling happens within the "black box" of AI risk-assessment tools.

This technological shift gives rise to the **"Digital Looking-Glass Self",** which is Charles Cooley's[14] original concept, suggesting that an individual's sense of self is a reflection of how they perceive others' perceptions of them. In the modern juvenile justice system, the "other" is increasingly a data-driven algorithm. The perceived "objectivity" of the machine makes the label feel more absolute and unchangeable than a human judge's subjective opinion.

If a juvenile believes that their "data shadow" has already predicted their failure, they are more likely to lean into that delinquent identity. In this "AI world," the label isn't just a social tag; it is a hard-coded destiny. By digitising labelling theory, we see that AI does not just track behaviour, it actively shapes it, creating a feedback loop in which the data shadow eclipses human capacity for change. The **Digital Kill-Switch** is thus necessary to break these automated chains of association, ensuring that a single youth's rehabilitation effectively 'un-tags' their entire digital network from unmerited suspicion.

B. The Statutory Mandate: Deconstructing Section 24 of the JJ Act, 2015

Section 24[15] stands as the Indian Parliament's primary legislative weapon against the permanent label. It is a mandatory provision rooted in the "Principle of Fresh Start," which explicitly provides that a juvenile delinquent should not suffer any disqualification attached to a conviction. To ensure this, the law mandates the destruction of records once the rehabilitation period expires. In the era of physical ledgers, this was a straightforward task: the file was shredded, and the "label" vanished with the paper. However, in the contemporary digital ecosystem, this statutory shield is being circumvented by the architecture of modern law enforcement technology.

Systems like the **Crime and Criminal Tracking Network & Systems (CCTNS)** and the **Interoperable Criminal Justice System (ICJS)[16]** are built on the principles of big data: persistence, retrieval, and permanence. While a Juvenile Justice Board (JJB)[17] may order the physical destruction of a case file at the local station, the "digital tags" associated with that child are often stored in

[14] Charles Horton Cooley, *Human Nature and the Social Order* 152 (1902).

[15] Juvenile Justice Act, 2015

[16] https://www.pib.gov.in/Pressreleaseshare.aspx?PRID=1799232

[17] Section 4 of Juvenile Justice Act, (JJ) 2015

centralised, encrypted servers that do not "forget." These systems were designed to ensure that no criminal "slips through the cracks," but in the context of juvenile justice, they inadvertently ensure that no child ever "slips out of the system." This architectural persistence directly conflicts with the legislative intent of Section 24.

Furthermore, these digital records are the fuel for Artificial Intelligence. Even if a specific name is suppressed, the "tags" generated during the juvenile's interaction with the law continue to train predictive algorithms. For instance, if a minor was arrested in a specific neighbourhood, that data point remains in the system to justify heightened "predictive policing" in that area. This means the juvenile is still being "tracked" and "labelled" by the machine's logic, even if their physical file is gone. The AI does not see a "Fresh Start"; it sees a historical data point that contributes to a "High-Risk" probability.

We have a legal framework that promises a fresh start, but a technological infrastructure that ensures a permanent record. To reclaim the power of Section 24, Indian jurisprudence must evolve from the destruction of paper to the **anonymisation of code**, ensuring that the cold, infinite memory of the silicon chip does not defeat the statutory promise of a clean slate.

C. Constitutional Imperatives: The Article 21 Shield

"Privacy is the ultimate expression of the sanctity of the individual... it includes at its core the preservation of personal intimacies, the sanctity of family life, marriage, procreation, the home and sexual orientation." — Justice D.Y. Chandrachud, [18]

The Supreme Court's landmark ruling in *Justice K.S. Puttaswamy*[19] fundamentally reshaped the Indian legal landscape by establishing that the **Right to Privacy** is a fundamental right protected under **Article 21** of the Constitution. Central to the right of privacy is the concept of "informational self-determination"the ability of an individual to control their own life story. The Court explicitly recognised that this privacy interest encompasses the **Right to be Forgotten (RTBF)**, allowing individuals to demand the removal of personal data that is no longer necessary or relevant. In the context of the juveniles, this constitutional shield is not just an elective privilege; but rather a mandatory prerequisite for the survival of the rehabilitative ideal. For a juvenile, the Right to be Forgotten is the constitutional soul of the "Fresh Start" principle.

We must move toward a jurisprudence where **"The Right to Reform"** is protected by the **"Right to be Forgotten."** By using Article 21 to enforce a "Digital Kill-Switch," we ensure that the Indian state honours its promise to its youth: that their data shadow will not permanently eclipse their future, and that their identity remains a work in progress, not a hard-coded data point in an unforgiving machine.

D. The "Digital Kill-Switch": A Policy Framework for Algorithmic Expungement

The proposed **"Digital Kill-Switch"** represents a fundamental shift from reactive legal remedies to proactive **"Privacy by Design"** engineering. In the current landscape, the burden of "clearing one's name" often falls on the rehabilitated individual, who must navigate a labyrinth of bureaucracy to ensure records are shredded. A Digital Kill-Switch flips this script by automating the "Fresh Start." It functions as a mandatory technological protocol where, the moment a Juvenile Justice Board (JJB) orders a discharge or rehabilitation certificate, an automated signal via an **Application Programming Interface (API)** is broadcast across all integrated platforms. This signal serves as a statutory "delete" command, synchronising legal reality with the digital infrastructure of the **National Crime Records Bureau (NCRB)** and the **Interoperable Criminal Justice System (ICJS)**.

Ultimately, the Digital Kill-Switch is the bridge between the 20th-century ideals of the **JJ Act** and the 21st-century reality of **Big Data**. It recognises that in an automated justice system, justice itself must be automated. By ensuring that the machine's memory is as finite as the law's forgiveness, the Kill-Switch preserves the human right to evolve. This ensures that the AI does not continue to

[18] Justice K.S. Puttaswamy v. Union of India (2017).

[19] *K.S. Puttaswamy v. Union of India*, (2017) 10 S.C.C. 1 (India).

apply a 'high-risk' bias to the individual's demographic or geographical profile, effectively neutralising the 'digital master status' at the source code level."

E. Procedural Due Process in the Automated World

Due process in the AI era must shift from a focus on human testimony to **algorithmic accountability**. Transparency is the bedrock of this evolution. It is not enough for the state to promise that a "Digital Kill-Switch" has been activated; a juvenile must have the enforceable right to **"Digital Confirmation."** This entails a verifiable, tamper-proof notification, potentially delivered via a secure "Redemption Certificate" on platforms like DigiLocker, providing cryptographic proof that their "delinquent tags" have been purged from the **CCTNS**, **ICJS**, and associated data fiduciaries. Without this confirmation, the "Right to be Forgotten" remains an invisible hope rather than a tangible legal shield.

Finally, the framework must provide for **compensatory damages**. If a resurrected record leads to the rejection of a job application or the denial of a visa, the harm is not just "data-deep"; it is an economic and psychological injury. By recognising "Digital Negligence" as a legal tort, Indian jurisprudence can create a financial deterrent against the reckless retention of juvenile data. Ultimately, due process ensures that the "Digital Kill-Switch" is not a mere suggestion but a legally binding command.

F. Comparative Perspectives

The international legal landscape has already begun to recognise that Artificial Intelligence (AI) in the power of the state is not a neutral tool, but a potential engine of systemic stigma. The **European Union's AI Act**, the world's first comprehensive horizontal regulation on artificial intelligence, explicitly classifies AI systems used in the "administration of justice" and "law enforcement", particularly those used to assess the risk of recidivism, as **"High-Risk."** This classification mandates stringent requirements, including high-quality datasets, detailed logging for traceability, and, most crucially, **strict human oversight**. The EU's logic is clear: when a machine's output can permanently alter the course of a human life, a human must remain must exercise the empathy and nuances that binary code lacks.

For India, this international standard provides a timely blueprint for evolving its own domestic digital laws. While the **Digital Personal Data Protection (DPDP) Act, 2023**, is primarily seen as a commercial privacy tool, its provisions on **"Digital Guardianship"**[20] (Section 9) offer a powerful statutory hook for juvenile justice. By aligning the DPDP Act with the "High-Risk" standards of the EU, India can strictly prohibit the **behavioural tracking of minors** that leads to "perpetual digital profiling.

Furthermore, international norms established by the **UN Convention on the Rights of the Child (UNCRC)**, specifically General Comment No. 25, emphasise that children's rights must be protected in the digital environment. It asserts that "automated processing of personal data for profiling or targeting children... should be prohibited or highly regulated." This global consensus supports the argument that India's **"Principle of Fresh Start"** is not an isolated legal quirk, but part of a universal human right to develop an identity free from the "tracking" of one's past.

Findings / Results

The research reveals a fundamental structural conflict between **juvenile justice philosophy** and **algorithmic governance systems** in India.

First, the study finds that the **"Principle of Fresh Start"** under the *Juvenile Justice (Care and Protection of Children) Act, 2015,* is significantly undermined by the persistence of digital records. While Section 24 mandates the destruction of records, AI systems operating within policing, surveillance, and data-sharing frameworks often retain, replicate, and redistribute such data, creating a **permanent digital footprint**.

[20] Digital Personal Data Protection Act, No. 22 of 2023, § 9 (India).

Second, algorithmic profiling produces what may be termed a **"digital master status"**, an identity marker that overrides all other social identities of the juvenile. Drawing on labelling theory, this digital tagging results in secondary deviance, in which individuals internalise stigma reinforced by automated systems rather than by human discretion.

Third, the findings indicate that such digital persistence constitutes a **direct infringement of informational privacy**, now recognised as intrinsic to the right to life and liberty under *Article 21 of the Constitution of India*, as affirmed in *K.S. Puttaswamy v. Union of India*[21]. The absence of effective erasure mechanisms transforms state surveillance into a **continuing constitutional violation**.

Fourth, the study identifies a **legal-technical disconnect**: statutory mandates, such as Section 24 of the Juvenile Justice Act, are not technologically embedded in digital systems. As a result, compliance remains **formal rather than functional**, rendering the right to record destruction largely illusory.

Finally, the research finds that India lacks a **comprehensive, enforceable framework for the Right to be Forgotten**, particularly for juveniles. While judicial recognition exists, its operationalisation in AI systems is absent, exposing juveniles to **lifelong reputational harm and exclusion**.

Recommendations / Suggestions

The study proposes a multi-layered reform framework integrating **legal, technological, and institutional safeguards**:

(A) Digital Kill Switch

A mandatory **automated expungement mechanism** should be embedded within all state-operated and affiliated databases. This "Digital Kill Switch" would:

- Trigger **irreversible deletion/anonymisation** of juvenile data upon attainment of majority or case closure.
- Extend to **all linked databases**, including predictive policing and AI risk assessment tools.
- Be subject to **independent audit and certification** for compliance.

(B) Privacy by Design Mandate

Incorporate **Privacy by Design principles** into all juvenile justice technologies:

- Data minimisation (collect only what is strictly necessary)
- Purpose limitation (restrict use to rehabilitation)
- Storage limitation (automatic deletion protocols)

(C) Statutory Strengthening

Amend the *Juvenile Justice Act, 2015,* to:

- Explicitly include a **digital right to erasure**
- Impose **penalties for unauthorised data retention or algorithmic profiling**
- Mandate **algorithmic accountability and transparency**

(D) Judicial Recognition of Juvenile Right to be Forgotten

Courts should explicitly recognise a **heightened Right to be Forgotten for juveniles**, building upon *K.S. Puttaswamy*, ensuring:

- Faster adjudication of erasure claims
- Presumption in favour of privacy over public disclosure

(E) Institutional Oversight

Establish a specialised **Data Protection Authority for Children**, tasked with:

- Monitoring AI systems in juvenile justice
- Conducting compliance audits
- Addressing grievances related to digital tracking

(F) Technological Safeguards

- Use **privacy-enhancing technologies** such as differential privacy and encryption

[21] *K.S. Puttaswamy v. Union of India*, (2017) 10 S.C.C. 1 (India).

- Prohibit **predictive risk scoring of minors** without strict safeguards
- Ensure **algorithmic explainability** in decision-making processes

8. Conclusion

The transition from traditional social labelling to **algorithmic digital tracking** represents a profound shift in state power. Within the juvenile justice system, this shift poses a direct challenge to the fundamental principle that **children are capable of reform and deserve a fresh start**.

This study demonstrates that the persistence of digital records not only contravenes statutory mandates under the Juvenile Justice Act, 2015, but also infringes upon the constitutional guarantees of dignity, privacy and liberty under Article 21. The emergence of a **"digital master status"** transforms temporary deviance into permanent identity, thereby perpetuating stigma and undermining reintegration.

The absence of effective erasure mechanisms renders the **Right to be Forgotten** largely symbolic. Without intervention, algorithmic governance risks converting rehabilitative justice into a system of **perpetual surveillance and exclusion**.

The proposed **Digital Kill Switch** offers a viable pathway to reconcile technology with constitutional values. By embedding erasure into the architecture of digital systems, the law can ensure that the promise of a **"fresh start" is not merely rhetorical but technologically enforceable**.

Ultimately, the legitimacy of the juvenile justice system in the digital age will depend on its ability to uphold a simple but powerful principle that by integrating the mandate of digital guardianship with the juvenile justice framework, the legal system can effectively disrupt the cycle of secondary deviance, shielding children from permanent digital tagging and preserving their fundamental right to a rehabilitated future.

A child's past must not become their permanent future.

TABLE OF AUTHORITIES

I. Constitutional Provisions

- INDIA CONST. art. 14.
- INDIA CONST. art. 21.

II. Statutes

- Juvenile Justice (Care and Protection of Children) Act, No. 2 of 2015 (India).
- Digital Personal Data Protection Act, No. 22 of 2023 (India).

III. Rules & Subordinate Legislation

- Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, G.S.R. 139(E) (India).

IV. Domestic Case Law (India)

- *K.S. Puttaswamy v. Union of India*, (2017) 10 S.C.C. 1 (India).

V. Foreign Case Law

- *Ewert v. Canada*, 2018 SCC 30 (Can.).
- *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016).

VI. International Instruments & Standards

- Convention on the Rights of the Child, Nov. 20, 1989, 1577 U.N.T.S. 3.
- *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*, 2024 O.J. (L 1689).
- UN Comm. on the Rights of the Child, *General Comment No. 25 (2021) on Children's Rights in Relation to the Digital Environment*, U.N. Doc. CRC/C/GC/25 (Mar. 2, 2021).

IV. Secondary Authorities (Criminological & Theoretical Works)

- Frank Tannenbaum, *Crime and the Community* (1938).
- Edwin M. Lemert, *Social Pathology: A Systematic Approach to the Theory of Sociopathic Behavior* 75 (1951).
- Howard S. Becker, *Outsiders: Studies in the Sociology of Deviance* 9 (1963).
- Charles Horton Cooley, *Human Nature and the Social Order* (1902).
- John Braithwaite, *Crime, Shame and Reintegration* (1989)

Artificial Intelligence and Social Justice : What Would Ambedkar Say?

Dr. Surajkumar Ramkripal Palbansh
Author
Dr. Baban More
Co- Author

Abstract

This paper examines AI and social justice through the lens of Dr. B. R. Ambedkar, whose ideas on liberty, equality and fraternity remain foundational to Indian constitutional thought. As AI increasingly shapes governance and public services, it risks reinforcing social hierarchies and the caste-based inequalities Ambedkar opposed.

Drawing on Ambedkar's constitutional morality and substantive equality, the paper argues AI must be governed by frameworks prioritising inclusivity and accountability and considers how Ambedkar might advocate state intervention to prevent excluding marginalised communities from AI's benefits.

The paper concludes that an Ambedkarite approach demands not only technical fairness but structural transformation — embedding social justice into AI design, policy and implementation to align technology with the constitutional vision of an equitable society.

Keywords: Artificial Intelligence, Social Justice, Algorithmic Bias, Constitutional Morality, Substantive Equality

Research Objectives

1. To evaluate whether AI aligns with or undermines Ambedkar's constitutional vision of an equitable society.
2. To analyse how automated systems and historical training data risk amplifying caste-based inequality in India.
3. To investigate how "constitutional morality" can govern AI and address the "black box" accountability gap.
4. To formulate an ethical and regulatory framework grounded in substantive equality, state intervention and affirmative technological policy.

Research Questions

1. How does data-driven algorithmic decision-making reinforce historical and caste-based discrimination in India?
2. How does the "black box" opacity of machine-learning systems conflict with constitutional morality and Articles 14, 19, and 21
3. What are the limits of "technical neutrality" against Ambedkar's triad of liberty, equality, and fraternity?
4. What legal safeguards must the State institute to convert AI from exclusion into social empowerment?

Scope and Limitations

Scope:

The study intersects technology studies, constitutional law, and social theory, evaluating automated governance against Ambedkar's philosophy and Articles 14, 19, 21. Limitations: it focuses on India's socio-legal context (caste, class, digital divide), prioritising structural critique over technical source-code analysis.

Research Methodology

The study employs a doctrinal, interdisciplinary methodology: constitutional-law analysis of Indian jurisprudence (Puttaswamy, Shreya Singhal, Anuradha Bhasin), an Ambedkarite framework evaluating constitutional morality and structural transformation, and critical technology studies of algorithmic bias and "black box" opacity.

Literature Review

1. The Myth of Technological Neutrality vs. Structural Reality

A foundational theme in critical technology studies rejects the claim that technology is neutral: AI is shaped by the historical and socio-economic contexts of its training data. In a stratified society like India, historical data reflects entrenched caste, class, gender and geographic inequality; rather than eliminating bias, AI standardises and amplifies it under a veneer of mathematical impartiality.

2. The Algorithmic Replication of Caste and Class Hierarchies

The literature shows how machine-learning models perpetuate discrimination even when identifiers like caste are omitted, reconstructing them through "proxy indicators". In India, residential location, surname and educational background act as strong statistical proxies for caste and class. Used in hiring, credit scoring, or welfare distribution, they covertly replicate historical exclusion in ways harder to detect than overt prejudice.

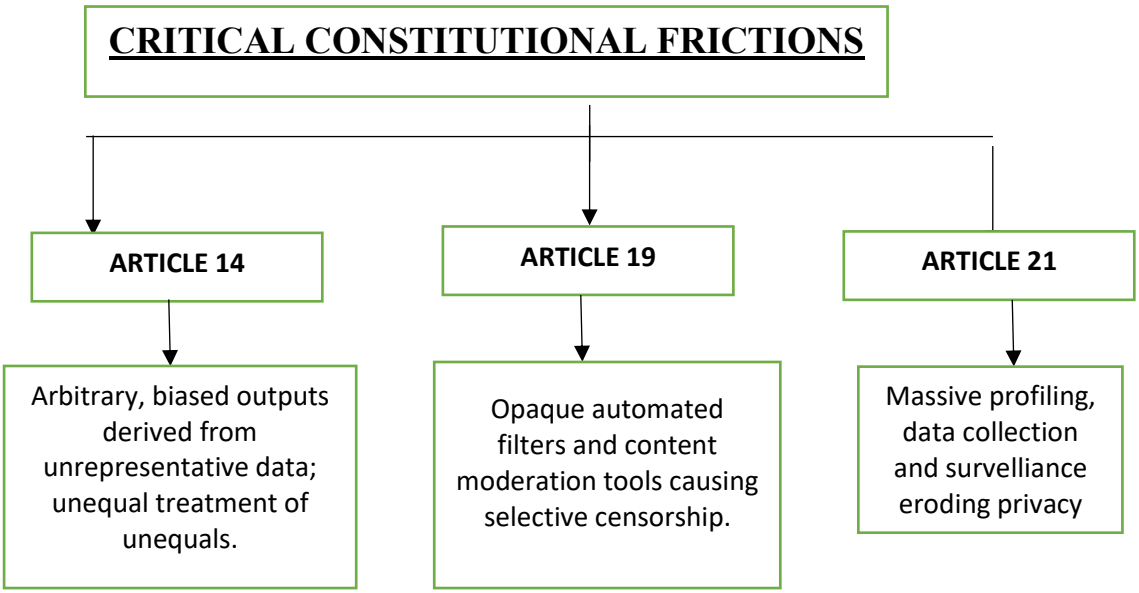
3. The Accountability Crisis: Opacity and the "Black Box" Problem

A critical vulnerability of deep-learning systems is opacity — the "black box" problem: when public duties are delegated to complex algorithms, the reasoning behind outcomes becomes uninterpretable.

This creates an accountability vacuum: algorithms cannot bear moral or legal liability, making it exceptionally difficult for a citizen to contest an adverse automated decision.

4. Constitutional Infractions and Algorithmic Governance

The intersection of AI deployment and Indian constitutional law reveals significant friction across multiple fundamental rights frameworks:



Judicial precedents emphasize that digital ecosystems cannot remain outside constitutional scrutiny. Landmark rulings underscore this intersection:

- ***K.S. Puttaswamy v. Union of India establishes informational privacy as core to Article 21; Shreya Singhal v. Union of India protects free expression online; and Anuradha Bhasin v. Union of India links internet access to fundamental rights.***

5. Conceptualizing an Ambedkarite Framework for Technological Justice

To counter these systemic risks, the literature integrates the political philosophy of Dr. B.R. Ambedkar as a normative guide for modern technological governance. This conceptual synthesis expands on three core principles:

Substantive Equality over Formal Neutrality

Ambedkar's jurisprudence rejects formal equality that treats unequal individuals identically, since this locks in historical disadvantage. Applied to AI, uniform treatment of an unequal population is unjust.

Operationalizing Constitutional Morality

Ambedkar's constitutional morality demands commitment to justice and dignity over legal text. Applied to AI, it requires systems open, explainable, and subordinate to human dignity and public oversight.

Assertive State Intervention and Inclusivity

Rejecting a laissez-faire model, an Ambedkarite view positions the State as an active regulator, mandating diverse data representation and affirmative technological policy.

Introduction

The rapid integration of AI into governance and social structures has altered how decisions are made and power exercised. From algorithmic hiring to welfare distribution, AI functions as an invisible but powerful authority, raising concerns about fairness and justice in a stratified society like India.[1]

The assumption that technology is neutral has come under scrutiny. AI systems, embedded in the socio-economic contexts of their training data, risk entrenching discrimination in new, less visible forms.[2]

It is in this context that Ambedkar's philosophy becomes critical. As the Constitution's architect, he articulated democracy grounded in substantive equality, warning that political democracy without social democracy remains fragile.[3]

This paper asks whether AI, as an emerging mode of governance, aligns with or undermines Ambedkar's vision of a just society, and how his concept of constitutional morality can inform ethical and regulatory frameworks for AI in India.

The paper adopts a doctrinal, interdisciplinary methodology, arguing an Ambedkarite approach must move beyond formal equality toward substantive justice and accountability, situating AI within the constitutional project of social transformation.

2. Ambedkar's Vision of Social Justice and Constitutional Morality

Engaging AI's implications in India must begin with Ambedkar's conception of social justice: not formal equality of laws but transformation of structures perpetuating exclusion. His critique of caste was a structural analysis of denied dignity and opportunity.[4]

At the core of his philosophy lies the triad of liberty, equality and fraternity — mutually reinforcing conditions for a just society, offering a lens for evaluating technological systems including AI.[5]

Ambedkar's constitutional morality — commitment to the Constitution's substantive values over its formal text — raises real questions about whether algorithmic systems can be reconciled with constitutional values as opaque black boxes.[6]

---

[1] Organisation for Economic Co-operation and Development, *How Artificial Intelligence Is Accelerating the Digital Government Journey*, *Governing with Artificial Intelligence* (Apr. 15, 2026), https://www.oecd.org/en/publications/2025/06/governing-with-artificial-intelligence_398fa287/full-report/how-artificial-intelligence-is-accelerating-the-digital-government-journey_d9552dc7.html.

[2] The Times of India, *AI Knows How Caste Works in India—Here's Why That's a Worry* (Apr. 15, 2026), https://timesofindia.indiatimes.com/india/ai-knows-how-caste-works-in-india-heres-why-thats-a-worry/articleshow/128448262.cms.

[3] Mamta Meena, *Dr. B.R. Ambedkar's Constitutional Triad: The Interplay of Liberty, Equality* (Oct.–Dec. 2025), https://vnsguj.ac.in/download/October%20to%20December%202025/Dr.%20B.R.%20Ambedkar%E2%80%99s%20Constitutional%20Triad%20The%20Interplay%20of%20Liberty,%20Equality.pdf.

[4] Dr R N Tripathi, Dr. B. R. Ambedkar's Social and Legal-Justice Philosophy: An Indian Political Perspective, Volume 7, Issue 4, July-August 2025.

[5] Ministry of External Affairs, Government of India, *Dr. Ambedkar: Writings and Speeches, Volume 3* (15 Apr. 2026), https://www.mea.gov.in/images/attach/amb/volume_03.pdf.

[6] Mahendra Pal Singh, *Observing Constitutional Morality*, *Ambedkar University Delhi* (2024), (15 Apr. 2026), https://aud.delhi.gov.in/sites/default/files/2024-01/slgr-readings/mp_singh_observing_constitutional_morality.pdf.

His emphasis on substantive equality recognises historical disadvantage requires affirmative measures; equal application of biased data can produce unequal outcomes — treating unequal individuals equally may perpetuate injustice.[7]

Ambedkar advocated state intervention to correct inequality, viewing the state as an active agent for justice — implying regulation and affirmative policy are necessary in AI governance.

Applying his thought to AI makes clear technology cannot be evaluated apart from social context; an Ambedkarite approach demands AI be scrutinised for its impact on the most vulnerable.

3. Understanding Artificial Intelligence and Algorithmic Decision Making

AI refers to computational systems performing tasks traditionally requiring human intelligence, through data-driven models analysing datasets to generate predictions. Machine-learning advances have expanded its reach across governance, finance, healthcare and employment.

AI learns correlations in historical data and applies them to new situations — hiring, predictive policing, credit scoring. Though portrayed as objective, outcomes are shaped by training data.[8]

AI's reliance on past data is risky where that data reflects entrenched inequality; in India, training AI on biased data risks reproducing caste, class and regional hierarchies.[9]

AI is also opaque: deep-learning models function as uninterpretable processes — the "black box" problem — making decisions affecting rights difficult to challenge.[10]

AI's integration into governance delegates traditionally human decisions to automated processes, eroding accountability since algorithms cannot be held morally responsible.

AI's expansion into employment, welfare and law enforcement demands scrutiny, since these are questions of justice and dignity, not merely efficiency.[11]

AI must be examined as a socio-legal phenomenon embedded in power structures. An Ambedkarite perspective questions AI's neutrality and asks whether it advances or undermines justice.

4. Artificial Intelligence and Structural Inequality in India

AI operates within existing social structures and often reinforces inequality. Trained on historical data, it replicates discrimination — hiring algorithms may disadvantage marginalised groups, and caste-based inequality is reproduced through proxies like surname or location. The digital divide compounds this: as public services digitise, those without access face further marginalisation.[12] AI also affects labour: automation may displace low-skilled workers while benefiting the technically skilled, and the gig economy lacks adequate protection. Left unchecked, AI can scale inequality, requiring careful regulation.

5. Constitutional Dimensions of Artificial Intelligence Governance

AI's growing use in decision-making raises constitutional concerns for fundamental rights, particularly in employment, law enforcement and public services.

Article 14's equality guarantee is implicated: biased or opaque algorithmic decisions can produce arbitrary, discriminatory outcomes.[13]

---

[7] Department of Social Justice & Empowerment, Government of India, *Samajik Nyay Sandesh Magazine* (Jan.–Mar. 2025), https://daic.gov.in/assets/Samajik%20Nyay%20Sandesh%20Magazine/Samajik%20Nyay%20Sandesh%20Magazine%20January-February-March%202025%20Final.pdf.

[8] Adib Bin Rashid, AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications, Hybrid Advances, Volume 7, December 2024, 100277, https://doi.org/10.1016/j.hybadv.2024.100277.

[9] Dharish David, B. Rajeshwari, and Timhna S. Algorithmic Bias and Discrimination in India: A Looming Crisis, Volume 11, Issue 1, https://doi.org/10.1177/24551333251343358.

[10] Nagadivya Balasubramaniam [a], Marjo Kauppinen, Transparency and explainability of AI systems: From ethical guidelines to requirements, Information and Software Technology, Volume 159, July 2023, 107197, https://doi.org/10.1016/j.infsof.2023.107197.

[11] European Parliament, *Title of Report* (2024), (Apr. 16, 2026), https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA(2024)754450_EN.pdf.

[12] Nabil Adel, The impact of digital literacy and technology adoption on financial inclusion in Africa, Asia, and Latin America, Heliyon, Volume 10, Issue 24, 30 December 2024, e40951, https://doi.org/10.1016/j.heliyon.2024.e40951.

[13] Saar Alon-Barkat, Madalina Busuioc, Kayla Schwoerer, Kristina S Weißmüller, Algorithmic discrimination in public service provision: Understanding citizens' attribution of responsibility for human versus algorithmic discriminatory

Article 19's free-speech guarantee is implicated: automated moderation and surveillance can restrict expression without clear standards.[14]

Article 21, covering privacy and dignity, is at risk where AI's large-scale data collection and profiling lack adequate safeguards.

K.S. Puttaswamy recognised privacy as fundamental; Shreya Singhal protected online speech; Anuradha Bhasin linked internet access to fundamental rights.[15][16]

AI governance cannot remain outside constitutional scrutiny: deployment must be guided by legality, necessity and proportionality.

6. Ambedkarite Critique of Algorithmic Governance

AI is a new form of governance where decisions are delegated to algorithms; any system exercising authority must face moral scrutiny and constitutional limitation.

Power concentrates in opaque systems whose decisions are neither transparent nor easily challenged, conflicting with constitutional morality's demand for openness.[17]

A tension exists between efficiency and justice: a system that is efficient while perpetuating inequality cannot be just.

Algorithmic systems are typically designed by a narrow group without representation of marginalised communities, reinforcing existing hierarchies.[18]

Absent clear accountability, it is often unclear whether developer, deployer, or state is liable for a discriminatory outcome, weakening legal remedy.

Algorithmic governance must be evaluated as a structure of power requiring regulation — transparent, accountable, and aligned with social justice.

7. Towards an Ambedkarite Framework for Ethical AI

These challenges call for a regulatory framework aligning technology with social justice, built on substantive equality, accountability and inclusion.

A. Substantive Equality in AI Design —

An Ambedkarite approach requires AI to move beyond formal equality toward substantive equality, which recognises that unequal social conditions mean equal treatment can produce unequal outcomes.

This requires diverse, representative datasets, regular bias audits, and fairness metrics adapted to local socio-economic context rather than generic global standards.[19]

B. Transparency and Accountability —

Transparency is essential so AI does not operate as an unchallengeable authority; individuals must be able to question decisions affecting their rights.

An Ambedkarite framework requires AI to be explainable in high-stakes areas, with accountability clearly assigned so affected individuals have access to remedies.[20]

C. State Intervention and Regulation —

---

outcomes, Journal of Public Administration Research and Theory, Volume 35, Issue 4, October 2025, Pages 469–488, https://doi.org/10.1093/jopart/muaf024.

[14] Balendra S. Meta's AI moderation and free speech: Ongoing challenges in the Global South. *Cambridge Forum on AI: Law and Governance*. 2025;1:e21. doi:10.1017/cfl.2025.5.

[15] AIR 2017 SC 4161.

[16] AIR 2015 SC 1523.

[17] United Nations University, *The Algorithmic Problem in Artificial Intelligence Governance* (2025), https://unu.edu/article/algorithmic-problem-artificial-intelligence-governance.

[18] Garv Gupta, *Ambedkar, Algorithm and Justice*, https://www.iimb.ac.in/sites/default/files/inline-files/Essay-Ambedkar-Algorithm-Garv_Gupta.pdf.

[19] Ahmad A, Vallès Y, Idaghdour Y. Bias in AI systems: integrating formal and socio-technical approaches. Front Big Data. 2026 Jan 8;8:1686452. doi: 10.3389/fdata.2025.1686452. PMID: 41584908; PMCID: PMC12823528.

[20] Sahibpreet Singh, Pawanpreet Kaur, *AI in Legal Education: An Ambedkarite Perspective*, April 2024, Conference: Legacy of Dr. Babasaheb Ambedkar: Analysis and Appraisal At: Department of Laws, Guru Nanak Dev University, https://www.researchgate.net/publication/380488071_AI_in_Legal_Education_An_Ambedkarite_Perspective.

Ambedkar viewed the state as crucial to social justice; leaving AI development solely to market forces risks reinforcing existing hierarchies.

The state must actively regulate AI through legal frameworks and independent oversight bodies, ensuring AI's benefits are distributed equitably.

D. Affirmative Technological Policies —

In line with Ambedkar's advocacy for affirmative action, India's digital divide requires policies expanding digital infrastructure, connectivity and literacy.[21]

Marginalised communities must participate in AI's design and regulation, not merely as users, transforming AI into a tool of empowerment.

E. Democratic Control and Participation —

An Ambedkarite framework holds that power, political or technological, must remain subject to democratic control.

Public consultation and decentralisation can democratise technological governance. This framework redefines AI governance as constitutional responsibility, embedding social justice into AI's design and deployment.[22]

8. Counter Arguments and Rebuttal

A common argument for AI is that it reduces human bias, processing large data volumes at scale to improve transparency and reduce arbitrariness.

This assumes AI operates independently of social context. Algorithms trained on biased historical data embed rather than eliminate bias, and their appearance of objectivity makes it harder to challenge.

From an Ambedkarite perspective, uniform criteria can still produce unequal outcomes if they ignore structural disadvantage.[23]

Efficiency alone cannot be the standard for justice; an efficient but discriminatory system cannot be justified constitutionally. The issue is how AI is governed, not whether it is adopted.[24]

AI offers real benefits, but these must be examined against social realities so advancement contributes to a just and inclusive society.

Conceptual Framework

The study's conceptual architecture sits at the intersection of automated data governance and transformative constitutionalism, rejecting technological neutrality and treating technology as a socio-legal phenomenon embedded within existing power dynamics.

It connects computer-science vulnerabilities — data bias, opacity — directly to constitutional infractions such as arbitrariness and exclusion, resting on a multi-tiered Ambedkarite foundation.

Findings / Results

- Amplification of Structural Disparities: AI, trained on historically compromised data, converts India's caste, class and gender biases into automated outputs. Discrimination via Proxy Data: even without explicit fields like caste, algorithms recreate discrimination through location, education and surname.
- The Accountability Vacuum: "black box" models limit individuals' ability to challenge automated outcomes. The Digital Divide: as welfare shifts to automated platforms, **unequal digital access leaves marginalised groups excluded. Speed over Dignity: algorithmic governance prioritises cost-effectiveness over justice and human dignity.**

Recommendations / Suggestions

A. Substantive Equality in Technical Design —

[21] Mirazchiyski, P.V., Černe, K. (2024). Digital Divide and Equality of Opportunity. In: Sardoč, M. (eds) Handbook of Equality of Opportunity. Springer, Cham. https://doi.org/10.1007/978-3-031-55897-9_94.

[22] Press Information Bureau, Government of India, *Title of Document* (Nov. 2025), https://static.pib.gov.in/WriteReadData/specificdocs/documents/2025/nov/doc2025115685601.pdf.

[23] Supra 20.

[24] Pankaj Phanase, *Ambedkar's Vision of Science and Technology* (Year Unknown), https://pankajphanase.com/ambedkars-vision-of-science-and-technology/.

- Contextual Data and Localisation: developers must use representative datasets mirroring historically excluded populations.
- Algorithmic Audits: technical workflows must mandate regular bias audits and locally tailored fairness metrics to intercept discrimination before deployment.

B. Mandating Transparency and Explainability —

- Deconstructing Black Boxes: high-stakes AI applications must meet strict interpretability requirements.
- Right to Explanation: citizens must have an enforceable right to understand the logic behind automated decisions affecting their rights.

C. Active State Regulation and Oversight —

- Statutory Regulatory Frameworks: the State must move from laissez-faire development toward robust statutory frameworks and data-protection norms.
- Independent Oversight Bodies: dedicated bodies should audit systems, enforce compliance, and assign liability among developers, deployers and state actors.

D. Affirmative Technological Policies —

- Infrastructure and Literacy: the State must close the digital divide through public infrastructure and literacy initiatives.
- Inclusivity in Development: diversity mandates should ensure marginalised communities participate in the design and governance of technology.

E. Democratization of Tech Governance —

- Public Participation: civic AI infrastructure must be subject to democratic control through consultation and open review.
- Decentralisation: technological authority must not concentrate within a few private corporations or centralised state entities.

Conclusion

The rise of AI marks a significant transformation in how decisions are made and power exercised, posing serious challenges for social justice in a society like India where inequality remains entrenched.

AI is a socio-legal phenomenon risking reinforcement of caste, class and resource hierarchies. Drawing on Ambedkar, the paper emphasises substantive equality and state intervention — justice requires more than formal neutrality.

AI governance must rest on transparency, accountability and inclusivity; the question is not whether AI should be adopted, but how it should be governed.

The challenge is to ensure technology does not outpace constitutional values. Ambedkar's vision demands innovation directed toward the upliftment of the marginalised.

Bibliography

Books and Government Publications

- Ambedkar, B. R. *Dr. Ambedkar: Writings and Speeches*. Vol. 3. New Delhi: Ministry of External Affairs, Government of India.
- Department of Social Justice & Empowerment, Government of India. *Samajik Nyay Sandesh Magazine*. January–March 2025.
- Organisation for Economic Co-operation and Development (OECD). *Governing with Artificial Intelligence*. Paris: OECD Publishing, 2025.

Journal Articles

- Adel, Nabil. 2024. "The Impact of Digital Literacy and Technology Adoption on Financial Inclusion in Africa, Asia, and Latin America." *Heliyon* 10 (24): e40951.
- Ahmad, Ahmad, Youssef Vallès, and Yassine Idaghdour. 2026. "Bias in AI Systems: Integrating Formal and Socio-Technical Approaches." *Frontiers in Big Data* 8:1686452.
- Alon-Barkat, Saar, Madalina Busuioc, Kayla Schwoerer, and Kristina S. Weißmüller. 2025. "Algorithmic Discrimination in Public Service Provision." *Journal of Public Administration Research and Theory* 35 (4): 469–488.

- Balasubramaniam, Nagadivya, and Marjo Kauppinen. 2023. "Transparency and Explainability of AI Systems: From Ethical Guidelines to Requirements." *Information and Software Technology* 159:107197.
- David, Dharish, B. Rajeshwari, and Timhna S. 2025. "Algorithmic Bias and Discrimination in India: A Looming Crisis." *Journal of Social Inclusion Studies* 11 (1).
- Meena, Mamta. 2025. "Dr. B.R. Ambedkar's Constitutional Triad: The Interplay of Liberty, Equality and Fraternity."
- Rashid, Adib Bin. 2024. "AI Revolutionizing Industries Worldwide: A Comprehensive Overview of Its Diverse Applications." *Hybrid Advances* 7:100277.
- Singh, Mahendra Pal. 2024. "Observing Constitutional Morality." Ambedkar University Delhi.
- Tripathi, R. N. 2025. "Dr. B. R. Ambedkar's Social and Legal Justice Philosophy: An Indian Political Perspective."

Conference Paper

- Singh, Sahibpreet, and Pawanpreet Kaur. 2024. "AI in Legal Education: An Ambedkarite Perspective." Paper presented at *Legacy of Dr. Babasaheb Ambedkar: Analysis and Appraisal*, Guru Nanak Dev University.

Book Chapter

- Mirazchiyski, Plamen V., and Klemen Černe. 2024. "Digital Divide and Equality of Opportunity." In *Handbook of Equality of Opportunity*, edited by Marko Sardoč. Cham: Springer.

Reports

- European Parliament. 2024. *Artificial Intelligence and Fundamental Rights*.
- Press Information Bureau, Government of India. 2025. *Artificial Intelligence Policy and Governance Report*.
- United Nations University. 2025. *The Algorithmic Problem in Artificial Intelligence Governance*.

Newspaper Article

- *The Times of India*. 2026. "AI Knows How Caste Works in India—Here's Why That's a Worry."

Online Sources

- Gupta, Garv. *Ambedkar, Algorithm and Justice.*
- OECD. *How Artificial Intelligence Is Accelerating the Digital Government Journey.*
- Phanase, Pankaj. *Ambedkar's Vision of Science and Technology.*

Cases

- *Anuradha Bhasin v. Union of India*, (2020) 3 SCC 637.
- *Justice K.S. Puttaswamy (Retd.) v. Union of India*, (2017) 10 SCC 1.
- *Shreya Singhal v. Union of India*, (2015) 5 SCC 1.

Constitutional and Legal Sources

Constitution of India.

Artificial Intelligence as a Catalyst for Digital Transformation in Legal Academia

Ms. Neha Ramakant Naik
Librarian
HVPS College of Law, Ghatkopar, Mumbai 400086

Abstract:

The application of Artificial Intelligence (AI) has advanced rapidly; many tasks previously performed by humans are now being entrusted to AI. These systems possess the capability to augment human abilities or even surpass them in performance across various sectors. This potential is particularly evident in areas such as data analysis, image recognition, legal processes like drafting and verifying contracts, and medical diagnosis. It would not be an exaggeration to say that these systems have brought about a significant 'paradigm shift' in teaching, learning, and professional skill development. This has led to remarkable transformations in pedagogical methods, research processes, and the enhancement of professional expertise; these fields are transitioning from traditional approaches to web-based workflows and, subsequently, to AI-driven methodologies.

This study highlights the evolving impact of Artificial Intelligence (AI) on legal education and the consequent creation of a technology-driven teaching-learning environment. It sheds light on the digital transformation of legal education in India and emphasizes the systemic reforms needed to align the curriculum with global technological advancements. The study focuses on the increasing use of AI in legal education and key aspects of AI-based legal learning, while also examining the role of 'Generative AI' in legal practice. In addition to exploring the benefits of AI, the paper reviews associated challenges, including ethical concerns, data privacy risks, over-reliance on technology, and the potential erosion of fundamental legal reasoning skills.

This research employs qualitative and analytical methods to conduct an in-depth analysis of the opportunities and limitations arising from Artificial Intelligence (AI) in the field of legal education. The study concludes that while AI possesses immense potential to bring about transformative changes, its effective integration into legal education requires a balanced and well-conceived framework..

Keywords: Artificial Intelligence (AI), Legal Academia, Teaching and Learning Processes, Legal Research

Introduction:

The concept of 'intelligence' within AI has transitioned over time. Even though its roots go back centuries, "*a significant milestone was reached when American computer scientist John McCarthy organized a summer workshop at Dartmouth University to explore the problem of 'artificial intelligence*"'. He defined 'artificial intelligence' as: "*Making a machine behave in such a way that, if a human were to behave in that manner, they would be considered intelligent.*"[1]

The advent of Artificial Intelligence (AI) has ushered in a transformative phase in the evolution of education much like in every other sector fundamentally altering traditional methods of knowledge dissemination and reshaping pedagogical and learning practices. In recent years, AI-driven technologies have increasingly permeated educational disciplines, thereby redefining pedagogical approaches and professional training. As a result, it has become possible for educators to analyse student performance, identify learning gaps, and provide access specially designed educational

[1] *Dicastery for the Doctrine of the Faith & Dicastery for Culture and Education, Antiqua et Nova: Note on the Relationship Between Artificial Intelligence and Human Intelligence* (Jan. 28, 2025), Vatican, https://www.vatican.va/roman_curia/congregations/cfaith/documents/rc_ddf_doc_20250128_antiqua-et_nova_en.html.

materials that improve learning outcomes. This shift from a standardized pedagogical approach to a more personalized learning experience has significantly improved student engagement and comprehension. Across the inclusive tertiary education sector, AI has emerged not merely as a technical support tool, but as an active facilitator of personalized and adaptive learning.

Furthermore, AI has contributed to the expansion of digital learning environments, rendering education more accessible, flexible, and interactive. Tools powered by machine learning and natural language processing assist in automating administrative tasks, enhancing educational research capabilities, and supporting real-time feedback mechanisms. Consequently, the teaching and learning process has become more efficient, data-driven, and outcome-oriented. Concurrently, this transformation also signifies a shift in the role of teachers, who are increasingly evolving into facilitators of knowledge rather than mere disseminators of information. Aided by AI-based resources, students are encouraged to engage in self-directed learning, thereby enhancing their analytical and critical thinking skills.

Historical Evolution of AI in Legal Education:

Artificial Intelligence is having a significant impact on modern legal systems, and it is gradually evolving from a merely theoretical concept into a practical tool. Its advancement replicates continuous advancements in technologies that have transformed legal research and practice over time. "*The term "Artificial Intelligence" was officially acquainted with in 1956 at the Dartmouth Conference. AI applications in law began developing in the 1960s, particularly through Computer Assisted Legal Research (CALR). Early CALR systems were limited, but the field cutting-edge with the launch of Lexis in 1973 and West Publishing's electronic system in 1975, providing lawyers efficient, large-scale access to case law. In the 21st century, developments in machine learning further expanded AI capabilities in legal practice. Systems like IBM's ROSS AI have since been adopted by law firms, particularly in the United States, to enhance legal research and decision-making process."*[2] Further, with advancements in machine learning and natural language processing significantly enhanced AI capabilities, enabling tools to understand legal language, automate tasks, and assist in decision-making, thereby complementing and strengthening human expertise in law.[3]

Concept of Legal Education

Legal education plays a key role in ensuring the observance of the rule of law and proper conduct within society. It will contribute immensely to the formation of the rule of law. However, against the backdrop of new developments in the social sphere and evolving needs, continuous and innovative efforts are being made to periodically review and improve law curricula; in this context, technological advancements and particularly the use of Artificial Intelligence (AI) are playing a significant role.

'Legal Education' refers to the "*instruction imparted to aspiring lawyers prior to their entry into the legal profession*". The social ideals of a society are often gauged by its laws. The objective of legal education should be to cultivate professional standards and instil ethical values in aspiring lawyers. *In some countries, universities place a special emphasis on ethical education, whereas in Great Britain, such instruction is primarily imparted through professional organizations and practical experience. In nations like India and Japan, Legal Ethics is taught as a formal academic course."*[4] Over time, the role of law has emerged as a vital instrument for social reform and development. Consequently, with the aim of equipping lawyers to shoulder broader social responsibilities, legal education curricula now incorporate subjects addressing issues such as poverty, illiteracy, and indebtedness.

Digital Transformation of Legal Education:

The primary directions for the digitalization of legal education involve the development and utilization of digital platforms for legal instruction including the integration AI through electronic

[2] *Historical Evolution and Impact of Artificial Intelligence in Legal System Across World*, Lawful Legal (June 23, 2025), https://lawfullegal.in/historical-evolution-and-impact-of-artificial-intelligence-in-legal-system-across-world/

[3] Dolidze, T. (2025) 'The evolving role of Artificial Intelligence in legal education and research', *International Journal of Law: Law and World*, 11(1), pp. 92–105. doi:10.36475/11.1.7.

[4] G. P. Tripathi, *Legal Education in India* (Central Law Publications 2020).

media and the application of advanced training programs for students. The use of digital tools has become crucial in contemporary legal education to meet evolving expectations, promote innovative teaching methods, and develop students' legal reasoning, argumentation skills, and problem-solving abilities. The integration of technologies renders the learning process more engaging and interactive, thereby attractive student involvement and strengthening their comprehension skills.

Today, Technology does not displace traditional legal principles; instead, it acts as a catalyst for their empowerment, thereby elevating the standards of both legal study and actual legal practice. The digitization of research methodologies, the simulation of court proceedings, and technology-driven teaching methods are fostering a new generation of tech-savvy, efficient, and strategic-thinking legal professionals. Particularly in India, the regulation of Artificial Intelligence (AI) and its integration into legal education have proven to be pivotal steps in steering legal education toward a modern and advanced future.[5]

The structure of modern legal education is rigid and limited, making it difficult to keep pace with the rapidly evolving technological landscape. Consequently, a clear gap has emerged between academic training and the demands of actual practice. With the increasing use of cutting-edge technology in the modern legal profession, it has become essential for lawyers to master digital tools and effectively assist clients in high-tech sectors. This underscores the importance of incorporating practical training and digital literacy into the curriculum. Digitizing legal education would make it possible to achieve educational outcomes that are more efficient, accessible, and future-ready. Historically, the system of legal education has remained traditional and resistant to change, making it difficult to adapt to the rapidly evolving technological landscape and changing circumstances.[6]

This journey, driven by digitization and artificial intelligence (AI), is making legal education more personalized, technology-based, and practical. AI tools are enhancing the efficiency of processes such as research, contract drafting, and analysis, thereby preparing students to tackle the challenges of the modern legal landscape. Furthermore, simulation-based and virtual learning methods are making education more accessible. However, to avoid 'algorithmic biases and over-reliance on AI, it is essential to emphasize ethics and relevant legal considerations.

Role of AI in Legal Academia

Artificial Intelligence (AI) is revolutionizing legal education by integrating technology into the teaching-learning process, research, and legal practice. AI technology efficiently processes and analyzes vast amounts of legal data, thereby assisting students and educators in their research. It serves as a complementary tool that fosters analytical skills by enhancing efficiency and facilitating access to organized information. Platforms like ChatGPT and AI feature integrated into LexisNexis are fundamentally transforming the way students, researchers, and professionals utilize and study legal knowledge.

In the teaching-learning process, AI enables personalized, interactive, and self-paced education, helping students better understand complex legal concepts. In legal research, AI-powered tools enhance efficiency by providing quick access to vast legal databases, improving accuracy and depth of analysis. In the legal profession, AI (Artificial Intelligence) assists with tasks such as contract drafting, document review, predictive analysis of litigation outcomes, and AI-driven instructional systems; this enhances efficiency and increases the demand for new digital competencies within the legal sector, while simultaneously equipping students to acquire the practical skills essential for a technology-driven legal practice. The role of AI in these areas is discussed below-

---

[5] *Geeta Institute of Law, How AI Is Reshaping Legal Education in the 21st Century: A Revolutionary Approach* (Feb. 24, 2026), https://geetalawcollege.in/blog/how-ai-is-reshaping-legal-education/.

[6] Dewi Sulistianingsih & Yuli Prasetyo Adhi, *Digital Transformation in Legal Education: Development Directions and Challenges*, in *Proceedings of the 15th National Scientific Conference of Teachers* 508 (Univ. Terbuka, Nov. 18, 2023), https://conference.ut.ac.id/index.php/ting/article/download/2395/951/6088.

- **In the teaching-learning process**, AI facilitates personalized and interactive learning through tools such as ChatGPT and Grammarly; these tools assist students in grasping complex legal concepts, refining their legal writing skills, maintaining academic integrity, and studying at their own pace.
- **In the legal research** Legal databases such as LexisNexis, Westlaw, Manupatra, SCC Online, and HeinOnline now offer AI features that significantly enhance the efficiency, accuracy, and depth of legal analysis by providing instant access to case laws, statutes, and legal literature.
- In the practical operations of the **legal profession**, AI tools collectively serve to reduce human workload, enhance efficiency, and facilitate more critical and sound decision making within the legal workflow. They provide assistance in tasks such as contract drafting, document review, and even predicting the outcomes of court cases; specifically, "AI-powered legal tools such as Logikcull, Everlaw, Harvey AI, and Clio Duo streamline and enhance the efficiency of the document review process by automating eDiscovery workflows, systematically organizing files, generating summaries, and providing key insights."[7]

Understanding Generative AI in Legal Practice:

"Generative AI and large language models have the potential to transform the practice of law in many different ways. But if they are to be fully embraced, they must be accurate with reliable data to draw on."[8]

The rapid integration of Generative Artificial Intelligence (GenAI) is driving a major transformation within the legal sector. The traditional image of a lawyer surrounded by stacks of legal tomes is now being supplanted by a new paradigm, wherein legal professionals can instantly access and analyze vast volumes of legal data using AI-driven tools. In legal practice, the potential applications of GenAI are extensive. Lawyers can leverage these systems to draft contracts, conduct due diligence, predict litigation outcomes, and even provide preliminary legal advice.

In November 2022, OpenAI publicly launched its large language model (LLM), ChatGPT, and reached the milestone of having over 100 million users in only 2 months and using AI tools and services using generative AI have become widely accessible.[9] Further According to a 2023 survey by LexisNexis, most legal professionals believe generative AI will transform legal practice, improve efficiency, and significantly influence how law is taught and studied.[10]

Artificial Intelligence and the Evolution of Legal Learning

The integration of AI into legal education is transforming the ways in which students' study, conduct research, and practice law. Traditional methods which once required spending countless hours have now been supplanted by AI-powered legal research tools. These tools provide instant access to vast databases, thereby enabling students to focus on legal reasoning rather than on the mechanics of manual searching.

AI-driven platforms assist students in preparing for the practice of law by drafting contracts, analyzing clauses, and identifying risks. The emergence of virtual moot courts and online simulations has further modernized legal education by recreating real-world legal environments through digital platforms. Amidst rapid technological advancements, legal education stands at a pivotal juncture, facing both unprecedented challenges and new opportunities. Ultimately, the future of legal education will depend on striking a balance between traditional teaching methods and technological innovation. By integrating AI with human guidance, law schools can develop a system that equips students with the theoretical and practical skills necessary to navigate today's legal challenges.

---

[7] American Bar Association, *How AI Enhances Legal Document Review*, Law Technology Today (last visited Apr. 17, 2026)..

[8] Suzanne McGee, *Generative AI and the Law*, LexisNexis, https://www.lexisnexis.com/html/lexisnexis-generative-ai-story/ (last visited Apr. 17, 2026).

[9] Mesko, Bertalan, The ChatGPT (Generative Artificial Intelligence) Revolution Has Made Artificial Intelligence Approachable for Medical Professionals, 25 *J. Med. Internet Res.* e48392 (2023), https://pmc.ncbi.nlm.nih.gov/articles/PMC10337400/.

[10] *Id*

Key Dimensions of AI-Driven Legal Education:

AI-based legal education emphases on the incorporation of technological tools into legal research and analysis, in conjunction with the expansion of practical skills such as contract drafting and risk assessment. To foster experiential learning, it incorporates virtual simulations and online quizzes, while also encouraging the use of data analytics to predict legal outcomes. Furthermore, it encompasses the study of emerging fields—such as technology law, data privacy, and AI ethics striving to strike a balance between traditional legal education and technological innovation.

I. **Personalized and Adaptive Learning**, AI-based teaching systems continuously analyze student progress to identify their weak areas and provide customized feedback accordingly, enabling self-paced and more focused learning. AI-enabled platforms adjust course content and learning paths based on each student's performance, providing necessary support to students who are lagging behind while providing more challenging opportunities to advanced students. This leads to improvements in comprehension, critical thinking, and overall academic outcomes. Incorporating AI-based personalized learning into legal education makes students more capable of adapting to a technology-intensive legal environment. it also fosters the development of essential skills, encourages active student engagement, and establishes a tangible link between academic learning and practical application. For example, AI tools provide practice in courtroom scenarios or customized case analysis, which effectively develops students' practical and analytical skills.
   - ChatGPT can students understand complex legal concepts and formulate arguments, thereby enhancing their analytical and logical skills.
   - Grammarly improves legal writing and clarity by ensuring grammatical accuracy and a qualified manner in legal documents.
   - Simultaneously, platforms such as LexisNexis or Westlaw facilitate efficient case law research and analysis, making relevant rulings and legal citations readily accessible.

II. **Improved Research Capabilities**-Artificial Intelligence (AI)-based tools enable faster search of legal information, increase efficiency in handling case laws and regulations, and strengthen research and analytical capabilities. Advanced AI-powered tools assist judges and lawyers by simplifying the process of legal research, identifying relevant court citations, and presenting summaries of decisions. This technology not only speeds up the research process, but also improves the quality and consistency of legal documentation. For instance, legal databases such as 'Manupatra', 'SCC Online', and 'Lexis-AI' instantly retrieve relevant judgments and present their summaries, thereby making research more accessible and effective. Thus, this not only saves time but also leads to significant improvements in the quality and reliability of legal study and documentation.

   Simply put, Manupatra AI goes beyond traditional keyword searches by focusing on the meaning and legal context of a query, rather than merely matching words. It grasps the underlying legal issue and retrieves the most relevant judgments, even if the exact terminology has not been used. This significantly reduces the time required for manual filtering, enabling lawyers and students to construct stronger, well-supported arguments with greater efficiency.[11]

III. **Engaging with technology** in a critical and mindful manner deepens one's understanding of AI ethics, data privacy, and potential algorithmic biases, while also encouraging the responsible and informed use of technology. With the aid of AI, the divide between online and offline education narrows, rendering legal education more interactive and engaging. For example, subsequent to classroom instruction, students can participate in AI-based case simulations or virtual moot courts to apply their theoretical knowledge in practical contexts. Such tools serve to further strengthen students' analytical and skill-based learning**.**

- Kira Systems assist in contract analysis and clause identification, *this is AI-powered contract analysis and document review software widely used in the legal field. It utilizes machine learning*

[11] Why AI Search Is the Future of Legal Research, *Bar & Bench* (Dec. 8, 2025).

to identify, extract, and analyze clauses and data within legal documents, thereby enabling faster, more accurate, and efficient reviews."[12]

- ROSS Intelligence (an AI-powered legal research tool) helps locate relevant case law based on specific legal queries
- ROSS Intelligence is an artificial intelligence-based legal research tool that assist to find appropriate case verdicts based on specific legal questions[13]".
- **LawDroid** and **ChatGPT-based solutions** help automate legal tasks such as client intake, drafting, and document analysis using NLP and machine learning. These tools improve efficiency, accessibility, and streamline the delivery of legal services.
- Furthermore, platforms like Casetext employ AI to identify connections between cases, thereby enabling students to gain a deeper understanding of legal precedents alongside their classroom learning.[14]

IV. Interdisciplinary Collaboration, in AI, where students engage with technologists, data scientists, and designers, enhancing adaptability and gaining a competitive edge in a technology-driven legal environment.

Challenges and Ethical Concerns

Artificial Intelligence is being increasingly exploited to enhance efficiency and foster innovation in legal proceedings. However, the integration of this technology raises significant ethical questions that impact fairness, accountability, transparency, and privacy within the legal system.

At this time, emphasis is placed on the point that "the process of punishment is part of the conflict between mercy and vengeance; and the use of Artificial Intelligence (AI) renders this dynamic even more complex. While AI can assist in legal decision-making, it raises ethical concerns such as bias, a lack of transparency, and the difficulty in challenging its outcomes. Its impersonal nature makes it more difficult for lawyers and litigants to comprehend the decisions rendered. Furthermore, the use of AI gives rise to serious concerns regarding the confidentiality and privacy of legal information[15]".

The integration of Artificial Intelligence (AI) into legal practice raises significant ethical concerns regarding bias, accountability, transparency, and privacy.

- AI systems that are frequently trained on existing data have the potential to perpetuate existing disparities and produce biased conclusions that impact the equity of legal outcomes.
- It's difficult to identify who is accountable mistakes in AI—developers, lawyers, or the system itself.
- Concerns about data confidentiality and privacy are equally important because AI systems analyze a lot of sensitive customer data, which increases the danger of data theft and misuse. Legal information that is sensitive could be exploited or made public.
- Concerns regarding academic integrity include the potential for plagiarism and an excessive reliance on work generated by Artificial Intelligence (AI).
- Students may overly rely on AI, rather than developing their own analytical skills, resulting in a decline in clinical judgment:
- There are no clear guidelines regarding how AI should be used responsibly within the practice of law.

Therefore, it is imperative to have a balanced and well-structured legal framework. Such a framework must incorporate the principles of transparency, accountability, and responsibility. When utilizing Artificial Intelligence (AI), special emphasis must be placed on the protection of human rights, data

[12] *Int'l J. Advanced Rsch. Sci., Commc'n & Tech.* (IJARSCT), Paper No. 23406 (2026), https://ijarsct.co.in/Paper23406.pdf.

[13] Legal Research, *ROSS Intelligence Blog*, https://blog.rossintelligence.com/category/legal-research

[14] *Int'l J. Advanced Rsch. Sci., Commc'n & Tech.*, *supra* note 4.

[15] **Firas Ibrahim et al.,** *Artificial Intelligence and Legal Ethics: Navigating Challenges and Opportunities* (2025), https://doi.org/10.1007/978-3-031-84636-6_38.

privacy, and the prevention of bias. This will render the legal system more efficient, equitable, and trustworthy, thereby further strengthening the foundations of justice and fairness.

Finding and suggestion

Research studies indicate that Artificial Intelligence (AI) is bringing about significant changes in the field of legal education by fostering professional skill development, enhancing research efficiency, and improving teaching methods. Through simulations and data-driven analysis, AI-powered solutions have reinforced practical training, facilitated tailored learning, and enhanced access to legal materials. Legal research and academic interaction are now more efficient and simplified because to the integration of platforms like ChatGPT, LexisNexis, and Manupatra.

Nonetheless, the result also underscores other significant issues, such as fairness in legal decision-making procedures, which are called into doubt by problems including algorithmic bias, a lack of transparency, and ambiguous accountability. Students' development of critical legal thinking and analytical skills may potentially be hampered by an over-reliance on AI tools. Concerns about academic integrity, secrecy, and data protection all pose significant obstacles. Inequalities in legal education are exacerbated by the digital gap and unequal access to cutting-edge technologies, especially in emerging nations like India. Overall, even if AI presents many benefits, its excessive and uncontrolled application could jeopardize the fundamental principles of legal education. A number of crucial steps are advised to guarantee the successful and moral integration of artificial intelligence (AI) in legal education.

AI literacy, legal technology training, and multidisciplinary courses like cyber law and data sci ence should be included in the legal studies curriculum.

- Establishing a clear ethical framework for the responsible use of AI requires institutions to pr ioritize openness, accountability, and justice.
- To preserve and enhance critical thinking abilities, a well-rounded educational strategy that blends conventional legal reasoning with AI assisted learning should be promoted.
- Strong data security measures are necessary to secure sensitive legal information and ensure confidentiality.
- Programs for ongoing faculty development should be put in place to give teachers the skills they need to successfully use AI tools into their lesson plans.
- By guaranteeing fair and reasonably priced access to AI-based instructional materials, efforts should also be undertaken to close the digital gap.
- Also, legislators and academic institutions ought to work together to create thorough regulations controlling the application of AI in legal practice and teaching.

Conclusion:

Artificial intelligence represents a profound transformation in legal education, redefining the manner in which the law is taught, learned, and professional competencies are developed. Its capacity to enhance efficiency, broaden access, and strengthen analytical skills positions it as a key tool within the modern legal academic landscape. However, its integration also gives rise to significant ethical, legal, and pedagogical challenges that demand careful attention, thereby necessitating a balanced approach. The future of legal education lies not in replacing human intelligence, but in participating technological innovation with essential values such as justice, equity, and critical thinking. Through well-structured reforms, robust ethical principles, and inclusive policies, AI can underwrite meaningfully to the firming of legal education while preserving its fundamental underpinnings particularly within the dynamic context of India.

References:

1. Femi Osasona, Olukunle O. Amoo, Akoh Atadoga, Temitayo O. Abrahams, Oluwatoyin A. Farayola & Benjamin S. Ayinla, Reviewing the Ethical Implications of AI in Decision Making Processes, 6 Int'l J. Mgmt. & Entrepreneurship Rsch. 322 (2024), https://doi.org/10.51594/ijmer.v6i2.773
2. John Smith & Jane Doe, Artificial Intelligence in Legal Education, 15 J. Legal Tech. 101 (2026), https://doi.org/10.1016/xxxxx.

3. Pandey, Hardik & Jyotsna Singh. *The Role of Legal Education in Preparing Lawyers for AI-Driven Legal Practice*. 5 Indian Journal of Legal Review 245 (2025).
4. Press Information Bureau, Ministry of Information & Broad., Govt. of India, *Title of Press Release* (Feb. 28, 2025),
5. https://www.pib.gov.in/PressReleasePage.aspx?PRID=2106239&reg=48&lang=2
6. Rituraj Bhowal, *Artificial Intelligence and Law* (Eastern Book Company 2024)

"Who is the Author? Rethinking Authorship in the Age of Artificial Intelligence"

By

Khushbu Aakashkumar Jain Gupta

Assistant Professor, HVPS College of Law

Abstract

The rapid growth of artificial intelligence has significantly transformed the processes of content creation, raising fundamental questions about authorship, creativity, and ownership. Traditionally, authorship has been closely linked to human intention, originality, and intellectual labour. However, the emergence of AI-generated texts challenges these assumptions by introducing non-human agents into the creative process. This paper examines the evolving concept of authorship through an interdisciplinary lens, drawing on literary theory and copyright law. It engages with key theoretical perspectives, including ***Roland Barthes' notion of the "death of the author"***[1] *and* ***Michel Foucault's exploration of the author-function***[2]***,*** to contextualize the shifting role of the author. The study further analyses the ***legal limitations of existing copyright frameworks***, which remain largely human-centric and struggle to accommodate AI-generated works. This paper challenges traditional norms of authorship by exploring the intersection of human and machine creativity, asking to recognize a new definition that includes collaborative and algorithm-driven creative processes. Ultimately, it calls for a nuanced understanding of authorship that aligns with the realities of the digital age.

Key Words: Copyright, Artificial Intelligence, Rethinking Authorship, Death of Author, Author Function, lack of legal limitations.

Introduction

The authorship as concept has been central to both literary studies and legal frameworks, serving as the foundation for creativity, ownership, and intellectual recognition. Traditionally, an author has been understood as the originator of a text—an individual whose thoughts, intentions, and creative expression give rise to a work. However, the rise of artificial intelligence has begun to destabilize this long-standing assumption. With the help of AI systems now we're capable of generating essays, stories, and even poetries, the question arises: can a machine be considered an author?[3]

This question is not merely technological but deeply philosophical and legal. In literary theory, the authority of the author has already been challenged by thinkers such as Roland Barthes, who famously declared the "death of the author," and Michel Foucault, who reconceptualized authorship as a function rather than a fixed identity. At the same time, copyright law continues to rely on a human-centric model of authorship, linking ownership to individual creativity and intention. The emergence of AI-generated content exposes a tension between these theoretical and legal perspectives.

This paper seeks to explore how the norms of authorship must be re-evaluated in the age of AI by bringing together insights from literary theory and copyright law, it examines whether existing definitions of authorship are sufficient to address AI-generated works. It argues that authorship can no longer be understood as a purely human attribute, but must instead be seen as a more fluid and collaborative process involving both human and non-human agents.

The concept of authorship, often treated as stable within legal discourse, has been subject to sustained critique within literary theory. One of the most significant interventions in this regard is Roland Barthes' seminal essay *"The Death of the Author,"* which fundamentally challenges the

[1] Roland Barthes, *The Death of the Author*, in **Image–Music–Text** 142 (Stephen Heath trans., Hill & Wang 1977).

[2] Michel Foucault, *What Is an Author?*, in **Language, Counter-Memory, Practice** 113 (Donald F. Bouchard ed., Cornell Univ. Press 1977).

[3] World Intellectual Property Organization (WIPO), *WIPO Conversation on Intellectual Property and Artificial Intelligence* (2019).

privileging of authorial intention as the primary source of textual meaning. Barthes argues that a text is not a singular expression of an author's consciousness, but rather a site of multiple meanings shaped by language, culture, and interpretation. In this formulation, the authority of the author is deliberately displaced, and the locus of meaning shifts from the creator to the reader.

This theoretical repositioning is further developed by Michel Foucault in his essay *"What is an Author?"*, where he introduces the notion of the "author-function." Foucault contends that the author should not be understood merely as an individual originator of a text, but as a classificatory principle that operates within discourse. The author's name serves to organize, regulate, and attribute texts within a broader system of knowledge production. In this sense, authorship is not an inherent property of a text but a construct shaped by institutional, cultural, and historical forces.

Together, these perspectives challenge the traditional view of authorship as a fixed, individualized source of creativity. Instead, they advance a more complex view in which meaning is produced through the interaction of text, reader, and context. Importantly, this theoretical shift predates the emergence of artificial intelligence, suggesting that the questioning of authorial authority is not entirely new but part of an ongoing intellectual trajectory.

The advent of AI-generated content, however, intensifies these concerns in unprecedented ways. Unlike human authors, AI systems do not possess consciousness, intention, or lived experience—qualities that have historically been associated with creativity and authorship. Yet, these systems are capable of producing coherent and contextually relevant texts that can mimic human expression. This raises a critical tension: if authorship has already been decentred in literary theory, does the introduction of AI merely extend this logic, or does it fundamentally alter the nature of authorship itself?

Although literary theory disperses the notion of authorial authority, it does not fully remove the importance of human agency. Within the context of AI-generated works, human participation remains evident through prompts, directives, and editorial oversight. This indicates the rise of a hybrid or distributed model of authorship, in which both humans and machines play a role in creating a text. Nevertheless, such a model makes it more difficult to assign authorship, especially when considered within legal systems that still depend on identifiable human creators.

In this context, literary theory provides a critical foundation for rethinking authorship in the age of artificial intelligence. By decentering the author and foregrounding the processes through which meaning is produced, it allows for a more flexible understanding of authorship—one that can accommodate emerging forms of technological mediation while still acknowledging the role of human engagement.

Research Objectives

1. To investigate the traditional approach of concept of authorship through the perspectives of Roland Barthes and Michel Foucault.
2. To analyse how AI-generated content challenges conventional norms of creativity, originality, and authorship.
3. To evaluate the adequacy of existing copyright law, particularly the Indian Copyright Act, 1957, in addressing AI-generated works.
4. To explore the legal and philosophical implications of recognizing AI-assisted creative production.
5. To propose a contemporary understanding of authorship that accommodates technological developments while preserving the principles of copyright law.

Research Questions

1.In what ways does artificial intelligence redefine the process of textual creation?
2.Who should be regarded as the author of AI-generated content—the user, the developer, or neither?
3.Does the existing copyright framework adequately regulate AI-generated works, or is legal reform necessary?

Scope and Limitations

This research examines the approach of concept of authorship through an interdisciplinary perspective by combining literary theory and copyright law. It focuses primarily on AI-generated textual content, including works produced through generative AI systems. The legal analysis is centred on the Indian Copyright Act, 1957, while referring **to selected international developments and judicial precedents for comparative understanding.** The paper also engages with influential literary theories to assess whether traditional notions of authorship remain relevant in the age of A.

Whereas, the study is doctrinal and conceptual in nature and does not include empirical research or interviews with stakeholders. It focuses exclusively on copyright and authorship issues relating to literary works and does not examine other areas of intellectual property such as patents, trademarks, or industrial designs. Similarly, issues concerning AI ethics, data privacy, and algorithmic bias are discussed only where they directly affected questions of authorship and copyright.

Research Methodology

This study adopts a **qualitative doctrinal research methodology**, relying primarily on the analysis of primary and secondary legal sources. Primary sources include the Indian Copyright Act, 1957, relevant judicial decisions, and international policy documents relating to artificial intelligence and intellectual property. Secondary sources comprise books, essays, literary criticism and scholarly commentaries on authorship, copyright, and artificial intelligence.

An interdisciplinary approach has been employed by integrating literary theory with legal analysis. The study examines the theoretical contributions of Roland Barthes and Michel Foucault alongside contemporary copyright debates to evaluate whether existing legal frameworks adequately address AI-generated content. The research is analytical and descriptive in nature and seeks to identify conceptual as well as legal gaps in the current understanding of authorship.

Literature Review

Authorship has been extensively discussed in both literary theory and legal scholarship. In *The Death of the Author*, Roland Barthes[4] rejected the notion that a text's meaning is solely determined by its author, arguing instead that meaning is created through the reader's interpretation. Michel Foucault[5] expanded on this with his concept of the "author-function," suggesting that authorship is a social and institutional construct rather than just an individual identity.

Contemporary legal scholarship increasingly focuses on the impact of artificial intelligence on copyright law. Ryan Abbott[6] argues that AI-created inventions and works challenge existing legal frameworks designed for human creators. Andres Guadamuz[7] also highlights the ambiguity surrounding copyright ownership and originality of AI-generated content, asserting that current laws are inadequate for autonomous systems. Reports from the World Intellectual Property Organization (WIPO) and the European Parliament emphasize the need to rethink intellectual property law in response to AI advancements.

While literary theories of authorship and legal debates on AI have mostly been studied separately, there has been little integration of these perspectives. This paper aims to bridge that gap by exploring how literary theories of authorship can illuminate contemporary copyright debates around AI-generated works.

Research Gap

Existing scholarship on artificial intelligence and copyright has primarily focused on questions of ownership, originality, and legal personhood. Likewise, literary theorists have extensively examined the concept of authorship through poststructuralist perspectives. However, limited

[4] Roland Barthes, *The Death of the Author*

[5] Michel Foucault, *What Is an Author?*,

[6] Ryan Abbott, *I Think, Therefore I Invent*

[7] Andres Guadamuz, *Artificial Intelligence and Copyright*

scholarship has attempted to integrate these two fields to examine how literary theories of authorship may inform contemporary copyright debates surrounding AI-generated content. This paper addresses this gap by adopting an interdisciplinary law-and-literature approach to reconsider the concept of authorship in the age of artificial intelligence.

AI and Text Production: Questioning Creativity and Agency

The rapid development of artificial intelligence has significantly transformed the processes through which texts are produced, circulated, and consumed. Contemporary AI systems, particularly those based on machine learning and large language models, are capable of generating human-like text by analyzing vast datasets and identifying patterns in language use. Unlike traditional writing tools that merely assist human authors, these systems actively participate in content generation, thereby complicating conventional distinctions between creator and instrument.[8]

At a functional level, AI-generated text is produced through probabilistic modeling rather than conscious intention. Such systems do not "understand" language in the human sense; instead, they generate outputs by predicting sequences of words based on statistical correlations derived from training data. Consequently, AI-generated content is inherently shaped by pre-existing linguistic patterns, raising critical questions about originality and creative authorship. The development of AI raises questions about whether human imagination, lived experience, and deliberate expression—all of which have historically been linked to creativity—remain prerequisites for authorship.

This challenge is further intensified by the increasing sophistication of AI outputs. In many instances, AI-generated texts are indistinguishable from those produced by human authors in terms of coherence, structure, and stylistic quality. These advancements make it harder to distinguish between human and machine innovation, making it challenging to decide whether authorship should be given based solely on the finished product or the creation process. Simultaneously, the role of human input in AI-generated content cannot be entirely disregarded. The production of such content typically involves varying degrees of human intervention, including the formulation of prompts, the selection of generated outputs, and subsequent editing or refinement. This suggests that AI functions not as a wholly autonomous creator but as a system that extends and amplifies human creative capacity. However, the extent of human contribution is often uneven, and in some cases minimal, thereby complicating claims of meaningful authorship.

The interaction between human agency and machine generation thus gives rise to a hybrid model of text production, in which authorship is distributed across multiple actors. Within this framework, neither the human user nor the AI system can be regarded as the sole author in the traditional sense. Instead, authorship emerges as a layered and contingent process, shaped by both human direction and algorithmic execution. This raises important conceptual questions: Can the act of prompting constitute creative labour? Does the absence of human intention in the generative phase undermine claims to originality? And to what extent should AI be treated as a participant in, rather than merely a tool of, the creative process?

These unresolved questions expose a fundamental gap between evolving technological practices and existing conceptual frameworks of creativity. While literary theory provides a basis for questioning the centrality of the author, it does not fully resolve the issue of agency in machine-generated texts. More importantly, these developments present a direct challenge to copyright law, which continues to rely on human-centric notions of authorship and originality. As such, the analysis of AI and text production serves as a critical bridge to the legal issues discussed in the following section, where the limitations of current copyright regimes become more explicitly apparent.

Legal Perspective: Copyright, Authorship, and Ownership

The idea that creative works are the result of human intellectual endeavour has historically served as the foundation for Copyright law[9] The ideas of authorship and originality, which are

[8] Andres Guadamuz, *Artificial Intelligence and Copyright*, 34 WIPO Mag. (2017).

[9] The Copyright Act, 1957, § 2(d)(vi), India.

necessary conditions for the protection of literary and creative works, are fundamental to this framework. A human author—someone who is able to use skill, labour, and judgment in the creation of a work—is assumed under copyright law in the majority of jurisdictions, including India..

The requirement of originality has been judicially interpreted in various ways, often emphasizing a "modicum of creativity" or the application of independent skill and judgment. Such standards, while flexible in application, remain anchored in human cognitive processes.[10] Artificial intelligence, by contrast, operates through algorithmic computation and data-driven pattern recognition, lacking the intentionality and consciousness that have traditionally underpinned claims of authorship. This creates a fundamental tension: if originality is linked to human creativity, can AI-generated works meet this threshold?

The lack of a distinct legal subject in AI-generated content makes the authorship problem even more complicated. AI systems cannot be acknowledged as writers under current copyright regimes because they lack legal personhood. This begs the question of whether authorship should be given to the developers who create the system, the human user who offers cues, or perhaps denied completely. Every one of these options has a unique set of difficulties. Attributing authorship to the user may overstate the intensity of human creativity involved, particularly in cases where input is minimal.[11] Conversely, assigning authorship to developers' risks disconnecting the final output from the creative process that led to its generation.

In the context of computer-generated works, some legal systems have made an effort to solve similar difficulties. For example, in some countries, the individual who makes the "necessary arrangements" to create a work is considered its creator. The complexity of modern AI systems, which show a significantly higher degree of autonomy and unpredictability, may not be sufficiently captured by such provisions, which were developed in an older technical environment.

The legal environment is further complicated by the issue of ownership in addition to authorship. Authorship is usually the source of copyright ownership, although it can be difficult to determine ownership when there isn't a distinct author. Particularly with regard to the marketing, licensing, and enforcement of rights in AI-generated works, this ambiguity has important practical ramifications.

Another level of complication is introduced by worries about plagiarism and infringement. AI systems may unintentionally duplicate or closely resemble existing works since they are trained on big datasets that may contain copyrighted content. This brings up significant issues regarding the potential for AI-generated content to infringe against current copyrights and the responsibility of users and developers in such situations.

Taken together, these challenges reveal that existing copyright frameworks are ill-equipped to address the realities of AI-generated content. The law's reliance on human-centric concepts of authorship and originality creates significant gaps when applied to non-human forms of creativity.[12] Because of this, there is an increasing need for legal reform that can take into account how authorship is changing in the era of artificial intelligence.

Discussion: Towards a Reconceptualization of Authorship

The preceding analysis demonstrates that both literary theory and copyright law are undergoing significant strain in the face of AI-generated content. While literary theory has long questioned the centrality of the author, legal frameworks continue to rely on a relatively stable and human-centric conception of authorship.[13] The rise of artificial intelligence highlights this approach's shortcomings and necessitates a more sophisticated comprehension of creative output.

[10] *Eastern Book Co. v. D.B. Modak*, (2008) 1 S.C.C. 1 (India).

[11] *Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340 (1991).

[12] Ryan Abbott, *I Think, Therefore I Invent*, 57 B.C. L. Rev. 1079 (2016)

[13] European Parliament, *Resolution on AI and Intellectual Property* (2020)

Rethinking authorship as a communal or dispersed activity rather than an individual characteristic is one potential path ahead. One way to conceptualize authorship in AI-generated works is as the result of human input and machine output interacting. This viewpoint acknowledges that although AI helps create content, human agents are still essential for starting, directing, and perfecting the creative process.

At the same time, it is important to avoid overstating the creative agency of artificial intelligence. AI systems, despite their sophistication, do not possess intentionality or self-awareness. Their outputs are ultimately dependent on the data on which they are trained and the parameters within which they operate. As such, any attempt to attribute full authorship to AI would risk undermining the conceptual foundations of both literary and legal understandings of creativity.

A more balanced approach would involve acknowledging the hybrid nature of AI-generated works while retaining a focus on human contribution as a basis for legal recognition. This may require a re-evaluation of existing standards of originality and authorship, potentially incorporating new criteria that reflect the realities of technologically mediated creativity.

Furthermore, interdisciplinary engagement between law and literature can provide valuable insights in this regard. While law offers a structured framework for rights and responsibilities, literary theory encourages a more flexible and critical approach to concepts such as authorship and meaning. Bringing these perspectives together allows for a more comprehensive understanding of the challenges posed by artificial intelligence.

Conclusion

The emergence of artificial intelligence has brought the concept of authorship into a state of profound uncertainty, exposing the limitations of both literary and legal frameworks that have traditionally governed creative production. While literary theory has long questioned the centrality and authority of the author, copyright law continues to rely on a human-centric understanding of authorship grounded in intention, originality and intellectual labour. The increasing presence of AI in the process of text generation reveals a growing disconnect between these established principles and contemporary technological realities.

This paper has argued that the challenge posed by AI is not merely to identify a new kind of author, but to reconsider the very foundations upon which authorship is defined. AI-generated works do not fit neatly within existing legal categories, nor do they fully align with traditional notions of creativity. At the same time, it would be conceptually and normatively problematic to recognize artificial intelligence as an independent author, given its lack of consciousness, intention, and legal personality.

A more viable approach lies in acknowledging the hybrid nature of AI-assisted creation while retaining a principled emphasis on human agency. Rather than extending authorship to machines, the law should focus on identifying meaningful human contribution as the basis for copyright protection. This may require a reinterpretation of existing provisions—such as those relating to computer-generated works—or the development of new legal standards that account for varying degrees of human involvement in AI-mediated processes.

Furthermore, there is a need for greater doctrinal clarity regarding ownership, liability, and originality in the context of AI-generated content. Without such clarity, the legal system risks either overextending protection to works lacking genuine creativity or denying protection in cases where human intellectual effort is meaningfully involved. Both outcomes would undermine the fundamental objectives of copyright law.

Ultimately, the question is not whether artificial intelligence can replace the human author, but how the law can adapt to a transformed creative landscape in which authorship is increasingly distributed, mediated, and technologically shaped. Addressing this challenge requires an interdisciplinary approach—one that draws not only on legal doctrine but also on insights from literary theory—to develop a more flexible and context-sensitive understanding of authorship. Only

through such an approach can the law remain responsive to the evolving nature of creativity in the digital age.

Bibliography/References

a. Statutes

i. The Copyright Act, No. 14 of 1957, India.

b. Cases

i. Eastern Book Co. v. D.B. Modak, (2008) 1 S.C.C. 1 (India).

ii. Feist Publications, Inc. v. Rural Telephone Service Co., 499 U.S. 340 (1991).

c. Books

i. Barthes, Roland. Image–Music–Text. Translated by Stephen Heath, Hill & Wang, 1977.

ii. Foucault, Michel. Language, Counter-Memory, Practice: Selected Essays and Interviews. Edited by Donald F. Bouchard, Cornell University Press, 1977.

d. Journal Articles

i. Abbott, Ryan. "I Think, Therefore I Invent: Creative Computers and the Future of Patent Law." Boston College Law Review, vol. 57, no. 4, 2016, pp. 1079–1126.

ii. Guadamuz, Andres. "Artificial Intelligence and Copyright." WIPO Magazine, no. 5, 2017.

e. International Reports and Policy Documents

i. European Parliament. Resolution on Intellectual Property Rights for the Development of Artificial Intelligence Technologies. European Parliament, 2020.

World Intellectual Property Organization (WIPO). Revised Issues Paper on Intellectual Property Policy and Artificial Intelligence. WIPO, 2020.

From Human Labour to Machine Intelligence: Legal Perspectives on the Future of Work

Ms. Tanavi Prasad Naik
Assistant Professor,
HVPS College of Law

Abstract

Artificial Intelligence (AI) and automation are progressively transforming the worldwide landscape of employment. These technological innovations present exceptional prospects for enhancing efficiency, productivity, and economic growth. Nevertheless, their increasing incorporation into job environments has also raised significant legal, social, and ethical issues. The use of AI-powered systems in hiring, performance evaluation, decision-making, and workplace monitoring prompts significant concerns regarding the safeguarding of workers' rights, job stability, and equity in employment practices. Conventional labour laws emerged during a time when human judgment was central to workplace administration, which is why numerous legal structures are inadequately prepared to tackle the issues presented by automated systems and algorithmic management.

This article explores the legal consequences of AI and automation in the workplace, emphasizing labour rights, algorithmic decision-making, and regulatory actions. It assesses international regulatory changes along with the developing Indian legal structure for data protection, labour regulation, and digital governance. The article contends that while AI will not completely eradicate human labour, it will profoundly change the essence of work. As a result, adaptive labour laws, retraining initiatives, and accountability frameworks are essential to guarantee that technological advancements adhere to the principles of equity, transparency, and human dignity.

Keywords: Artificial Intelligence, Labour Rights, Automation, Employment Law, Future of Work, Algorithmic Management.

1. Introduction

AI has become one of the most groundbreaking technologies of the twenty-first century. Tools powered by AI are being more widely incorporated into decision-making processes at work, such as hiring, monitoring employees, evaluating productivity, and making termination choices.[1] These advancements are transforming the conventional relationships between employers and employees and changing the dynamics of labour markets.

Automation refers to the application of technology to execute tasks once handled by human labourers, has grown swiftly due to progress in machine learning, robotics, and data analysis. Earlier automation was mainly confined to industrial manufacturing, but contemporary AI systems can execute cognitive functions like legal research, financial evaluations, and customer support.

Although technological advancements present chances for economic expansion and increased productivity, they also provoke significant worries about job loss, algorithmic prejudice, employee monitoring, and the weakening of labour safeguards. Studies indicate that jobs involving routine and repetitive tasks are especially prone to automation.[2] Concurrently, new job opportunities are arising in sectors like data science, cybersecurity and AI development.

From a legal standpoint, most labour regulations were established during a time when workplace choices were solely determined by human supervisors. The growing dependence on automated systems thus poses considerable regulatory challenges concerning transparency, accountability, and discrimination. This paper investigates the developing connection between AI and labour law, while

[1] *Ajay Agrawal, Joshua Gans & Avi Goldfarb*, **Prediction Machines: The Simple Economics of Artificial Intelligence** (Harvard Bus. Rev. Press 2018).

[2] *Carl Benedikt Frey & Michael A. Osborne*, The Future of Employment: How Susceptible Are Jobs to Computerisation?, 114 *Technological Forecasting & Soc. Change* 254 (2017).

analyzing how legal frameworks can adjust to guarantee that technological advancements do not erode workers' rights.

2. AI and Workplace Automation

Traditionally, automation has been linked to employing machines to substitute human labour in manufacturing. Nonetheless, modern AI technologies can execute intricate cognitive functions such as pattern recognition, natural language understanding, and predictive analytics. Employers are increasingly depending on AI technologies to evaluate job applications, assess video interviews, arrange work shifts, and track employee performance.[3] These technologies allow organizations to handle vast amounts of data and swiftly make decisions about workforce management.

The growing dependence on automated systems has resulted in the emergence of algorithmic management, where decisions in the workplace are directed by algorithms instead of human managers. While these systems can enhance efficiency and lower operational expenses, they also bring up issues related to transparency and worker autonomy. AI-driven monitoring tools enable employers to observe employee actions, analyze productivity trends, and produce performance assessments. Critics contend that this type of surveillance could diminish workers to mere data points, compromising their dignity and privacy.[4]

Automation is transforming labour markets as well. Jobs for beginners in areas like customer service, clerical positions, and administrative assistance are becoming more at risk of automation. Nevertheless, the implementation of AI has also been linked to increased productivity and the emergence of new types of jobs in advanced technological fields.

3. Impact of AI on Labour Rights

3.1 Job Displacement and Economic Inequality

A major worry related to AI is the possible job displacement of employees. Research indicates that automation could diminish the need for routine jobs, resulting in job losses or wage stagnation in specific industries.

Workers with low skills are especially at risk as they might not have opportunities for training that help them adjust to technological advancements. Policymakers need to make sure that displaced employees can access educational and reskilling initiatives.

3.2 Bias and Discrimination in Algorithms

AI systems depend on training data to reach conclusions. If the data utilized to train these systems mirrors past discrimination, the resulting outcomes could likewise be biased.[5]

Automated hiring systems, for instance, might prefer applicants from demographic groups that have traditionally been overrepresented in recruitment choices. Such results could contravene equality principles and anti-discrimination regulations.

Indian constitutional law highlights the principles of equality and non-discrimination in the workplace. In Air India v. Nargesh Meerza, the Supreme Court acknowledged the significance of curbing discriminatory hiring practices. These concepts continue to be pertinent when assessing algorithmic decision-making in contemporary work environments.

3.3 Monitoring Employees and Privacy

Monitoring systems powered by AI present significant privacy issues. Employers can gather comprehensive information on workers, such as communication habits, productivity rates, and behavioral indicators.

The Indian Constitution acknowledges the right to privacy as a fundamental right in the case of Justice **K.S. Puttaswamy v. Union of India**. This ruling has important consequences for surveillance technologies in the workplace and the management of employee information.

---

[3] *De Stefano, Valerio*, "Negotiating the Algorithm: Automation, Artificial Intelligence and Labour Protection," Int'l Labour Org. (2018).

[4] *Stuart Russell & Peter Norvig*, **Artificial Intelligence: A Modern Approach** (4th ed. 2020).

[5] *Iftekhar Ahmed et al.*, AI in Hiring: A Review of Opportunities and Challenges, 35 *Computers in Hum. Behav.* 100 (2021).

3.4 Worker Representation and Collective Negotiation

Algorithmic management might also undermine conventional methods of collective bargaining. When automated systems make decisions in the workplace, employees might have few chances to negotiate or contest these choices.

Indian labour law has traditionally highlighted the significance of collective bargaining in safeguarding workers' rights. The emergence of AI-driven management poses crucial questions about the future function of labour union.

4. Legal Frameworks Governing AI and Labour

4.1 Global Labour Norms

The swift incorporation of AI and automation in workplaces has sparked considerable worries about safeguarding workers' rights and maintaining equitable labour practices. Acknowledging these issues, the International Labour Organization (ILO) has repeatedly stressed the importance of guaranteeing that technological advances do not compromise the dignity, security, and well-being of employees[6]. International labour standards act as a crucial framework for assisting governments and employers in upholding fair labour conditions throughout times of technological change.

These standards encourage the essential principles of proper employment, which encompass equitable wages, safe and healthy workplaces, social security, and honoring workers' rights. In AI-driven work environments, these principles are increasingly vital as automated systems can affect hiring, performance assessments, and decisions regarding termination. In the absence of adequate protections, such systems could inadvertently replicate biases or establish unclear decision-making processes that harm employees.

The ILO promotes a human-focused strategy for technological advancement, stressing that innovation ought to improve human abilities instead of completely substituting them. Consequently, policies must guarantee that employees have access to reskilling, upskilling, and continuous learning opportunities, allowing them to adjust to changing labour markets influenced by AI. Moreover, global labour standards urge governments to create social safety nets and employment protections that lessen the possible negative impacts of automation on job security.

4.2 AI Regulation in the European Union

Among international regulatory efforts concerning AI, the European Union's AI Act stands out as one of the most thorough and progressive frameworks. The regulation utilizes a risk-based method, classifying AI systems according to the degree of risk they present to individuals and society.[7] This method guarantees that more stringent regulatory requirements are imposed on AI systems capable of greatly impacting fundamental rights and social frameworks.

Within the employment sphere, AI systems employed for recruiting, monitoring employees, assessing performance, and making promotion choices are categorized as high-risk systems according to the Act. This classification acknowledges that automated decision-making in hiring can significantly impact workers' income, dignity, and equal access to opportunities.

To tackle these issues, the regulation establishes several significant obligations for developers and organizations implementing these systems. These consist of transparency requirements, guaranteeing that people are made aware when AI systems influence employment-related choices. Furthermore, companies need to adopt strong data governance practices, ensure high-quality data to avoid bias, and create frameworks for human monitoring.

The regulatory model of the European Union shows a dedication to ensuring that technological advancements function within a system that upholds fundamental rights, equity, and responsibility. The EU aims to prioritize human welfare in AI governance to harmonize technological advancement with social justice and the safeguarding of workers' rights.

4.3 Legal Framework of India

---

[6] *International Labour Organization*, **Skills for a Greener Future: A Global View** (2019).

[7] *Mateescu, Alexandra & Nguyen, Aiha*, Algorithmic Management in the Workplace, Data & Soc'y Research Inst. (2019).

India, recognized as one of the quickest-growing digital economies globally, has started to create legal and regulatory structures to manage the application of emerging technologies. While India lacks specific legislation solely for AI regulation, various current laws and policy measures tackle elements of digital governance, data protection, and workers' rights.

A major advancement in this area is the Digital Personal Data Protection Act, 2023, which creates an all-encompassing structure for overing the gathering, handling, and retention of personal information. The legislation imposes responsibilities on entities managing personal data, encompassing mandates concerning lawful processing, data protection, and accountability. Within the realm of AI technologies utilized in work environments, this law is essential for safeguarding employees' personal and sensitive information from improper use or unapproved handling.

Legislation like the Code on Wages, 2019, and the Industrial Relations Code, 2020,[8] guarantee just compensation, methods for resolving disputes, and the safeguarding of workers' rights in employment systems. Nevertheless, these regulations were mainly created for conventional employment frameworks and fail to specifically tackle challenges linked to algorithmic management, automated decision-making, or AI-based workplace monitoring.

India, recognized as one of the fastest-growing digital economies globally, has started to progressively create legal and regulatory structures intended to regulate the application of emerging technologies. While India currently lacks a specific statute solely focused on AI, various existing laws and constitutional tenets cover elements of digital governance, data protection, and labour rights.

A significant advancement in India's data governance system is the Digital Personal Data Protection Act, 2023, which creates a thorough framework for overing the gathering, storage, and handling of personal data. The Act requires organizations that manage personal data to guarantee lawful processing, secure data, and maintain accountability. In environments more influenced by algorithmic technologies, such regulations are crucial for protecting workers' personal data and stopping the abuse of information gathered via digital surveillance systems.

The constitutional basis for privacy protection in India was notably reinforced in the pivotal ruling of ***Justice K.S. Puttaswamy (Retd.) v. Union of India***, where the Supreme Court acknowledged the right to privacy as a fundamental right under Article 21 of the Indian Constitution. This ruling has significant consequences for the oversight of digital technologies, encompassing AI systems that handle substantial amounts of personal information. The acknowledgment of privacy as a basic right necessitates that technological systems used by the State and private organizations adhere to the principles of legality, necessity, and proportionality.

Moreover, the judicial system has underscored the significance of equity and non-randomness in administrative and technological decision-making. In ***Anuradha Bhasin v. Union of India***, the Supreme Court emphasized that internet access is closely associated with the realization of fundamental rights, such as freedom of expression and the right to engage in trade or profession. These rulings highlight the increasing judicial acknowledgment that digital technologies and connectivity are vital elements of contemporary economic and professional existence.

Besides constitutional law, India has implemented labour codes to govern employment conditions. Laws like the Code on Wages, 2019 and the Industrial Relations Code, 2020 to guarantee just compensation, safeguard workers' rights, and establish processes for resolving disputes. These laws were mainly created to regulate conventional employment relationships and do not specifically tackle new challenges like algorithmic management, automated recruitment processes, or AI-driven workplace monitoring.

Regarding policy initiatives, the Government of India has implemented actions for responsible AI governance through the National Strategy for AI, launched by NITI Aayog[9]. The policy highlights the advancement of "AI for All," encouraging the application of AI in areas like healthcare,

[8] Code on Wages, No. 29 of 2019, India Code (2019); Industrial Relations Code, No. 35 of 2020, India Code (2020).

[9] *NITI Aayog*, **National Strategy for Artificial Intelligence #AIForAll** (2018).

agriculture, education, and smart mobility, while prioritizing ethical practices and inclusive development. It emphasizes the importance of regulatory systems that harmonize technological advancement with safeguarding essential rights.

In spite of these advancements, India still does not possess a thorough legal structure that specifically tackles the effects of AI on employment and labour regulation. With AI-based decision-making increasingly woven into recruitment, employee surveillance, and performance evaluation, it is essential for policymakers to think about implementing rules that guarantee algorithmic transparency, accountability, and safeguard workers' rights. These reforms would aid in aligning India's legal structure with the rising global benchmarks in accountable AI governance.

With the rising impact of AI on recruitment, productivity tracking, and hiring decisions, there is an escalating demand for legal changes that directly tackle algorithmic transparency, fairness, and accountability in employment situations. Policymakers in India should thus contemplate creating specific guidelines or laws that merge labour rights with new technological advancements

5.1 Re-skilling and Learning

With the ongoing transformation of global labour markets by AI and automation, the importance of ongoing education and skill enhancement has grown markedly. Conventional educational models that equip individuals for a singular career trajectory are insufficient in a fast-changing technological landscape. Governments, educational organizations, and industry partners must work together to create extensive reskilling and upskilling programs that provide workers with the skills needed in AI-focused economies.

Reskilling initiatives must emphasize not just technical competencies like digital literacy, data analysis, programming, and technological oversight, but also cognitive and interpersonal skills such as critical thinking, creativity, and problem-solving. Skills centered on human abilities are becoming more valuable as automated systems can execute repetitive and routine tasks, whereas humans are crucial for innovation, ethical judgment and making strategic decisions.

Global entities like the International Labour Organization have stressed the significance of lifelong education as a crucial aspect of the work landscape ahead. The ILO's policy guidelines promote government investment in educational systems, vocational training, and digital competencies programs that help workers adjust to evolving labour market conditions.[10]

Moreover, India has acknowledged the significance of bolstering digital capabilities via policy measures like the National Strategy for AI, which emphasizes preparing the workforce for new technologies through skills enhancement, innovation frameworks, and academic partnerships.

5.2 Mechanisms for Social Protection

While technological advancement boosts economic growth and productivity, it may also cause job loss and transitional unemployment, especially in industries that depend significantly on routine or repetitive tasks. With automation progressively taking over specific types of jobs, it is essential for governments to establish sufficient social protection systems to address the possible socio-economic impacts of technological changes.

Social protection systems are vital in protecting workers from financial instability. Initiatives like unemployment benefits, income support programs, retraining options, and job transition aid can assist workers in adjusting to changes in the labour market and avoid abrupt loss of income. These systems are especially crucial in emerging economies where significant portions of the labour force may not have official job safeguards.

Global policy entities, such as the Organisation for Economic Co-operation and Development, have highlighted the need for governments to enhance labour market institutions and social safety nets to guarantee that the advantages of technological advancements are fairly shared.[11] Robust social

[10] *Shoshana Zuboff*, **The Age of Surveillance Capitalism** (PublicAffairs 2019).

[11] *Air India v. Nargesh Meerza*, (1981) 4 SCC 335 (India).

protection policies not only assist displaced employees but also promote economic stability by sustaining consumer demand and social cohesion amidst structural transitions.

Moreover, authorities ought to encourage proactive labour market strategies, such as career counselling services, employment placement schemes, and focused skill enhancement projects that correspond with growing sectors. Policymakers can ease the shift to a tech-driven economy by integrating social security with workforce development strategies.

5.3 AI Governance focussed on Humans

The increasing incorporation of AI into job systems emphasizes the critical necessity for regulatory frameworks that focus on human well-being and ethical responsibility. An approach to AI governance focused on human needs acknowledges that technological progress ought to benefit society and honour the dignity, rights, and independence of individuals[12].

Human-centred governance highlights essential principles like transparency, accountability, equity and inclusiveness in the development and implementation of AI systems. Employers using AI tools for hiring, staff oversight, or assessing performance must guarantee that these systems function transparently and do not reinforce discrimination or bias. Employees need to be made aware of the utilization of algorithmic decision-making tools that affect their working conditions.

Global guidelines like the OECD Principles on AI support the development of AI that upholds human rights, democratic ideals, and sustainable progress.[13] These guidelines stress that AI systems must be under human control, allowing individuals to contest automated decisions that impact their jobs or career prospects[14].

Additionally, involving workers and labour representatives in conversations about AI implementation in workplaces is crucial. This involvement aids in ensuring that technological systems are introduced in manners that honour labour rights, mitigate algorithmic bias, and enhance equity in employment practices.

In the end, a human-centred regulatory system aims to harmonize innovation with ethical obligation. Through the incorporation of legal protections, worker engagement, and clear governance frameworks, communities can leverage the advantages of AI while guaranteeing that technological advancement fosters inclusive economic development and social equity.

6. Obstacles for Judicial Frameworks

Legal frameworks encounter various obstacles in overing AI. Technological advancements frequently outpace legal frameworks, resulting in regulatory voids.[15] Furthermore, AI systems possess complex architectures that are hard to elucidate, complicating the process of determining accountability for automated choices.

The worldwide activities of tech firms make it more difficult to enforce labour laws across different regions.[16]

7. Recommendations

To guarantee that the advantages of automation and AI are achieved without compromising workers' rights, a thorough policy strategy must be implemented by governments, employers, and regulatory bodies. Organizations must transparently reveal their use of AI in employment decisions to ensure employees understand how these systems impact hiring, assessments, promotions, and disciplinary measures. This openness fosters responsibility and enhances trust in the work environment

---

[12] Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act), COM (2021) 206 final (Apr. 21, 2021).

[13] *OECD*, **Employment Outlook 2019: The Future of Work**

[14] *OECD*, **Policy Responses to New Forms of Work** (2020).

[15] *Barocas, Solon & Selbst, Andrew D.*, Big Data's Disparate Impact, 104 *Calif. L. Rev.* 671 (2016).

[16] *Justice K.S. Puttaswamy (Retd.) v. Union of India*, (2017) 10 SCC 1 (India).

Simultaneously, employees should be assured the right to request human assessment of choices made by automated systems[17]. This safeguard promotes equity and allows employees to contest results that could be incorrect, prejudiced, or unfair. Moreover, periodic and independent audits of AI systems ought to be implemented to assess their fairness, precision, and adherence to legal requirements, thereby uncovering concealed biases or systemic issues.

There is a significant necessity to revise current labour laws to tackle new challenges like algorithmic management, workplace monitoring, and automated decision-making. Legal reforms must distinctly outline employer obligations and create accountability for AI utilization in workplace practices. Moreover, employees and labour unions must participate in conversations about the introduction of AI technologies, guaranteeing that these advancements reflect workers' rights and interests through collaborative governance.

Ultimately, governments need to allocate resources for extensive reskilling and workforce development programs to equip individuals for the requirements of an AI-driven economy. Focusing on digital education, vocational training, and lifelong learning will empower workers to adjust successfully to evolving labour market conditions and minimize the threat of technological displacement

8. Conclusion

AI and automation are fundamentally altering the essence of work and changing labour markets globally. Although these technological innovations present considerable prospects for enhancing efficiency, productivity, and economic expansion, they also bring forth intricate legal and ethical issues concerning labour rights and employment regulations. The rising adoption of AI in hiring practices, employee monitoring, and assessment of performance has heightened discussions about topics like bias, privacy, openness, and responsibility in the workplace.

Current labour law structures were primarily established during a time when human agents made employment choices. As a result, numerous legal frameworks are inadequately prepared to over algorithmic decision-making and automated management processes. The rise of AI-based workplace methods has revealed significant regulatory shortcomings that need to be resolved with revised legal norms and organizational protections.

In the future, it is improbable that machines will fully replace human labour. Rather, it will probably be defined by enhanced cooperation between human employees and smart technologies. In a hybrid setting, AI can boost productivity by handling routine or data-heavy tasks, allowing employees to concentrate on creative, analytical, and strategic duties. Achieving this equilibrium, nonetheless, necessitates careful policy strategies and flexible governance frameworks.

Consequently, governments, employers, and labour organizations need to work together to create legal and regulatory systems that guarantee technological advancement unfolds in a socially responsible way. Policies that encourage algorithmic transparency, employee involvement, reskilling programs, and enhanced accountability measures will be crucial for protecting workers' rights in an AI-centric economy. A human-focused strategy for AI governance will be essential to guarantee that technological advancements boost, instead of undermining, the dignity, security, and welfare of workers in the future workplace.

Combating AI-Generated Social Media Misinformation by Developing Critical Thinking Skills in the Digital Age

Anand Kumar

Abstract

Misinformation has existed since we developed the art of conversation and the tools to communicate our thoughts to a larger community, and so has its impact. Today's digital age, particularly the use of AI, has empowered each of us to communicate our views and thoughts to a larger audience at unprecedented speed and reach. This ability has made this informational highway both a tool for good and a destructive tool. The paper analyses the effects of AI-generated social media misinformation (hereinafter termed as 'misinformation'), examines the drivers of its propagation in the digital age, and proposes an evidence-based analysis and practical interventions to facilitate critical thinking, thereby reducing the spread and impact of misinformation. It synthesises empirical findings on our understanding of the term 'misinformation', outlines the contours of our understanding of critical thinking, and provides actionable recommendations for educators and the general public to overcome misinformation by utilising critical thinking skills. Further, the paper proposes a flowchart showing how to harness the concept to address and curb the dissemination of misinformation.

Keywords: AI-Generated, social media, Misinformation, critical thinking, disinformation, digital age

Research Objectives

The paper analyses how misinformation spreads and critical thinking's role in curbing its spread and effectiveness, through an empirical study among Law students at Mumbai University. The paper aims to identify how misinformation is understood, its effects on society as a whole, and the level of awareness about critical thinking skills and their inclusion in our college curriculum.

Hypothesis

The Digital information highway amplifies both the volume and velocity of misinformation, adversely affecting human rights and societal well-being. Cultivating critical thinking from the grassroots upwards can significantly strengthen society's ability to resist and reduce misinformation.

Research Questions

- Predictors- What are the psychological drivers responsible for susceptibility to different misinformation types?
- What is the scale of misinformation in our society and its adverse effects?
- Education- Role of critical thinking in identifying misinformation? How can the critical thinking tools be utilised in differentiating the misinformation from the correct information?
- How can critical thinking competencies be systematically developed to reduce the spread and influence of misinformation?

Scope and Limitations

Owing to the strict word limit, the paper has maintained a sharp and focused investigation on the following: -

- The systematic assessment of the pathways through which misinformation proliferates via social media and its detrimental impact on human rights violations and other aspects of social well-being.
- While discussing methods and tools to combat the spread and effectiveness of misinformation, the paper has focused solely on the use of critical thinking competencies.

Research Methodology

The research methodology includes an empirical analysis of student responses from the Law colleges of the University of Mumbai and an in-depth review of the available literature on misinformation, human rights, and communicative freedom.

A survey was conducted (63 responses were received) to analyse the extent of misinformation and the attitudes of today's youth towards its dangers. The survey further probed their understanding of critical thinking and their tendency to share misleading information with peers on social media platforms. The survey also sought respondents' views on their acceptability of a legislative intervention to curb the spread of misinformation.

Literature Review

Definition and Taxonomy

- Misinformation is defined as "false information, given either by mistake or deliberately"
- As per the "American Psychological Association, misinformation is false or inaccurate information, getting the facts wrong".

How Does Misinformation Spread?

Misinformation spreads quickly on digital communication sites. And the factors contributing to it include algorithmic prioritisation of sensational content, users' cognitive biases, and the echo chamber effect, whereby individuals are exposed to information that largely mirrors their pre-existing beliefs.

Misinformation also spreads through selective censorship (through the platform's curation algorithmic logic), hacking, and influencing search engine results (Search engine optimisation). In the face of such misinformation, deepfake technology poses a serious challenge by making it very difficult for people to distinguish genuine from fabricated content. The manipulation of real voices to fabricate false statements makes it harder for citizens to distinguish between facts and fiction.

The Social Media

Digital communication networks are a potent platform for spreading user-generated information at unprecedented speed, including false, unverified, or fabricated news and overenthusiastic health claims. Notable among them are WhatsApp, Facebook, and X (formerly Twitter). Thus, anyone in possession of an electronic device and an internet connection is a potential target and a vector of misinformation.

Misinformation spreads through digital communication networks and provides ultra-cheap and unlimited access to unsuspecting and interested parties. Unchecked and exponential spread leverages decision-making and personal biases to influence individuals' reasoning.

Globally, misinformation is becoming increasingly complex as emerging deception techniques proliferate. Misinformation needs to be viewed as a complex web of a broader network of manipulation, driven by rapid advances in deepfake tools, precision based misinformation campaigns, and dangerous influence operations. Misinformation has far-reaching implications for human rights worldwide. A swift and efficient response to misinformation is needed at every level, including binding legal measures, enforcement, whole-of-government action, and civil society action.

In today's rush to produce news at neck-breaking speed (breaking news), journalists are pushed to work at a very fast pace, often leading to insufficient verification of the information and its sources.

The techniques and tools available to spread misinformation are as under: -

- **Artificial Amplification**. Artificially increasing the spread of information by manipulating and utilising Search Engine Optimisation (SEO) tools, hashtags or web links on digital networks.
- **Bots**. Social media profiles run by automated computer programmes to produce a large volume of posts with online content to increase human engagement.
- **Masking the Source**. Masking the true origin and misleading the recipient into believing it originated from a legitimate source.
- **Impersonation**. Impersonation of genuine media, influential people or governments.
- **Focussed Group**. Targeting focused groups and exploiting their inherent biases by using statistics and biased information.

The Psychology of Misinformation

Individuals with higher levels of reflective thinking are less susceptible to misinformation and demonstrate greater competence in distinguishing truth from falsehood. Recent evidence indicates that misinformation gains credibility due to personal biases, a lack of competence to recognise false information, and the subject's failure to check its accuracy and origin before sharing it with others.

Misinformation often uses alarmist, clickbait-style headlines to attract high-volume online traffic to the creator's website, subsequently increasing advertising revenue. This revenue model pushes content creators to create sensational and alarmist headlines even when they are not sure of the veracity of the information.

Motivations for Misinformation

Research identifies multiple drivers for the proliferation of false information, including personal agendas, nefarious designs, political interests, non-state actor's agendas, directed trolling, fame-seeking intentions and law and order disruption. Consistent bombardment of disinformation results in confusion, the burial of truth, and narrative wars. For hostile and determined actors, misinformation is a powerful and cheap tool, but also a very effective and efficient influence technique. The easy, cheap availability of social media tools has made it possible for almost everyone who wants to spread such falsehoods. Further, widespread, unchecked downloading of content that undermines truth and facts creates an all-pervasive vehicle for scrupulous entities to pursue their insidious goals.

Sensational content is given priority using algorithmic amplification, which is motivated by engagement metrics and spreads false or unconfirmed information. Furthermore, the democratisation of content creation enables anyone to create and disseminate information without vetting and regardless of its veracity, thereby supporting misleading narratives.

Impact of Misinformation

The spread of misinformation causes multifarious harm to the targeted audience and recipients, ranging from damage to an individual's image to an effect on the legitimacy of a government policy, a political party or an organisation. MIT research shows that false news/ information travels farther, faster, deeper, and more broadly than the truth", with misinformation being 70% more likely to be retweeted than facts.

Misinformation undermines democratic institutions. As public trust in mainstream media declines, alternative ecosystems have expanded and filled this void. Digital networks and push business models have intensified polarisation and diminished tolerance for differing viewpoints.

What is Meant by Critical Thinking?

- Critical thinking includes a variety of competencies, which include reasoning and argument analysis. It is defined as "the ability to analyse and evaluate arguments according to their soundness and credibility, respond to arguments and reach a conclusion through deduction from given information."
- It is a skill which encourages the subject to reflect, reason, and carry out logical evaluation, analyse and interpret the input to develop a value judgement.

Critical thinking helps develop the following competencies: -

- **Interpreting**. "Understanding the significance of data and clarifying its meaning".
- **Analysing**. "Breaking information down and recombining it in different ways".
- **Reasoning**. "Creating an argument through logical steps".
- **Evaluating**. "Judging the worth, credibility or strength of information".

Introduction

"Ashwathama Hato NaroVa Kunjarova" (अश्वत्थामा हतो नरो वा कुञ्जरो वा), (meaning Ashwathama is dead, whether man (naro va) or elephant (Kunjaro va), The Pandava's elder brother, Yudhisthira, claimed, "Yes, Ashwathama is dead (and then he murmured), I don't know if it's a man or an elephant". This statement ranks among the earliest documented instances of misinformation, used to defeat invincible Guru Dronacharya.

The practice of utilising misinformation for ulterior gain is perhaps not a new phenomenon; it has existed since humankind began sharing information (by both gesture and verbal means) with one another.

From the beat of drums and word of mouth to dispatch riders, to print, radio, television, the internet and digital communication, the methods of information sharing have changed dramatically. Today, AI has exponentially improved the speed and democratised information sharing among the target audience.

Throughout history, misinformation has played a very important role by manipulating people's beliefs, utilised as war propaganda, as a health misinformation during the COVID-19 epidemic, and as a political tool during recent US and other elections. Today, the digital communication network has become the most widely available and handy tool ever for communicating with anyone. Misinformation can spread rapidly, erode trust in institutions, harm human rights and disrupt social cohesion.

Conceptual Framework

There is no concise definition of the word 'misinformation'. In layman's language, 'misinformation' is understood as false or misleading information deliberately or unknowingly created that can mislead the recipient.

Limited research exists on the effectiveness of information literacy in helping individuals critically evaluate online information. Critical thinking, a core component of information literacy, enables users to assess online content by verifying factual accuracy, judging the credibility and logical analysis of arguments.

To understand how critical thinking works, let's understand the foregoing problem:

The combined cost of a cricket bat and a ball is Rs 120.00. If the bat is Rs 100.00 more costly than the ball, determine the cost of the ball.

The cost of the ball is Rs 20.00, which comes to mind intuitively for most people. However, the correct answer is Ra 10, which emerges (for most) only after careful deliberation.

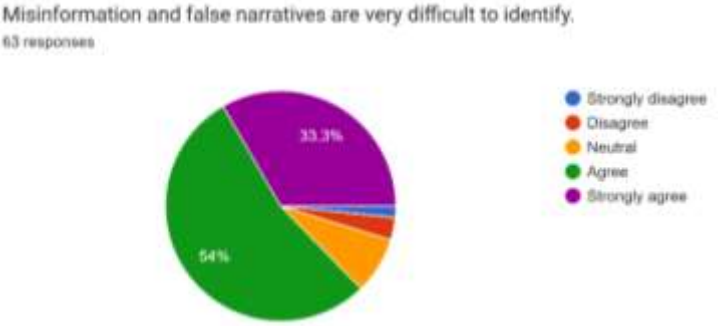
Findings and Data Analysis

Understanding of Misinformation and its effect

The respondents had a fair understanding of what misinformation means. 33 out of 63 (52.4%) respondents strongly agreed, and 30 (47.6%) of respondents agreed with the statement that "misinformation includes false, inaccurate or misleading information designed and promoted to cause public harm, create a false narrative or make a profit".

55 out of 63 respondents (87.3%) agree that "misinformation and false narratives are very difficult to identify". This proves the importance of digital literacy and relevant tools to minimise the dangers of misinformation.





55 out of 62 respondents (88.7%) believe that misinformation has increased polarisation and resulted in increased fake political narratives, while 06 respondents gave a neutral response and 01 respondent strongly disagreed that misinformation has increased polarisation and spread of fake news.

There was near unanimity (39 out of 63 (61.9%) strongly agreed, and 24 of the remaining respondents (38.1%) agree that misinformation can confuse and manipulate citizens.

Misinformation affects Human Rights.

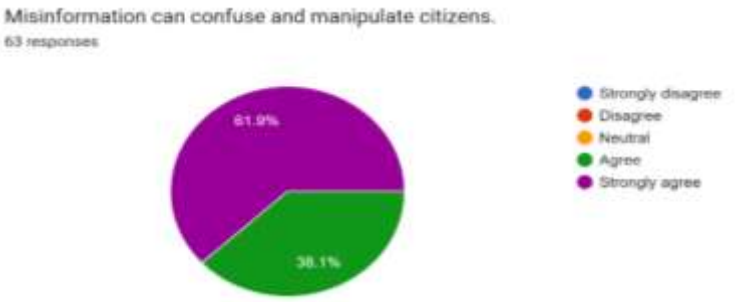
59 out of 63 respondents (93.6%) believe that misinformation affects Human Rights, and the remaining 4 (6.3%) were neutral.

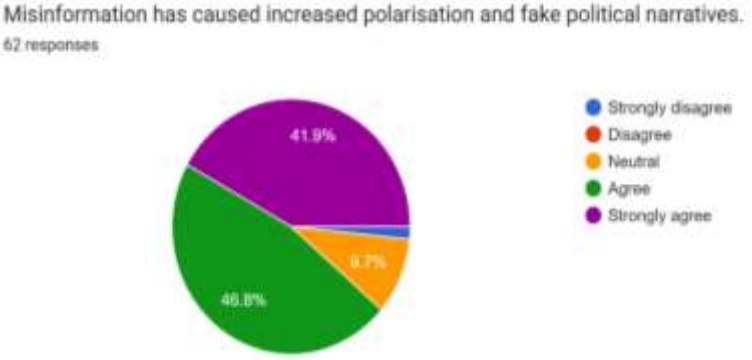
Impact on the credibility of mainstream media.

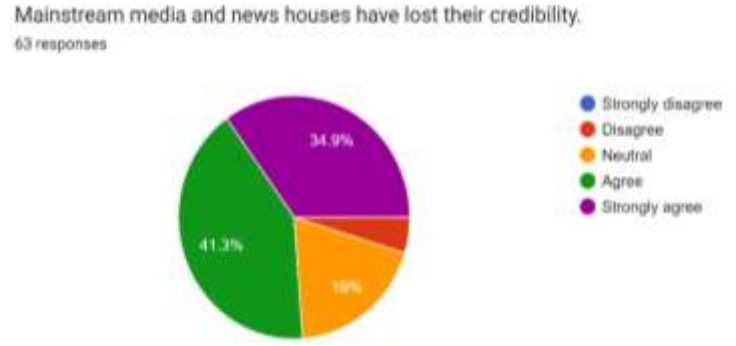
The silver lining in the cloud: 15 out of 63 respondents (23.8%) felt that mainstream media and news houses are reliable sources of information. This suggests that the situation is not as grim as it seems; there's still a good chance to prevent the negative effects of misinformation.

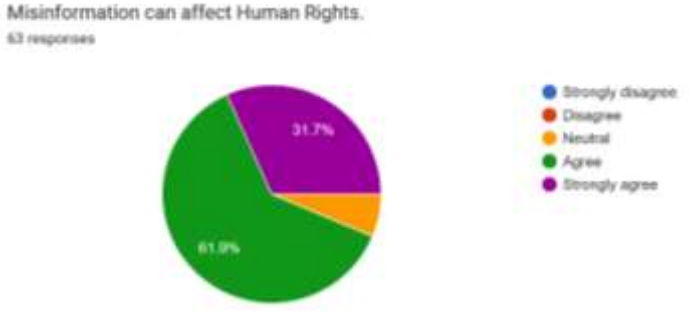
The propensity to pass information without assessing its authenticity

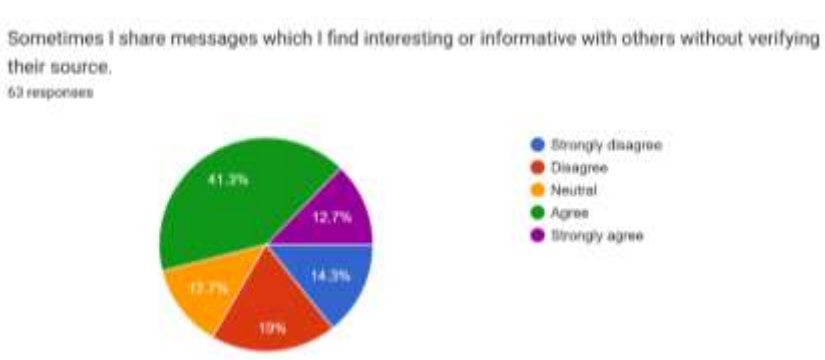
When the respondents were asked directly about their predisposition to transmit information without assessing its authenticity, they provided a guarded response: 09 strongly disagreed, and 12 disagreed with the statement, "sometimes, I share messages which I find interesting or informative with others without verifying their source". Within the same group, 34 (54%) reported sometimes passing on information without checking its accuracy.





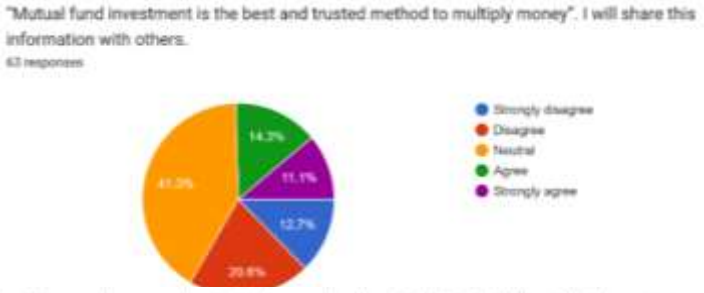




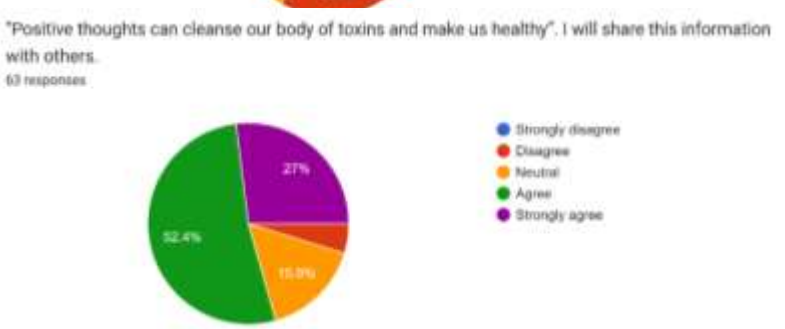


When asked, the participants showed a striking and consistent inclination in their preference to share statements.

21 out of 63 individuals stated that they will not pass on unverified statements about financial investment information. The majority, 26 out of 63 (41.3%), chose to remain neutral, i.e, non-committal about sharing the statement, "mutual fund investment is the best and trusted method to multiply money".



However, when the statement meant to be shared was tweaked to make it look like a more noble and good intention, 50 out of 63 respondents (79.4 %) indicated their inclination to share the information, and only 03 individuals (4.8%) confirmed not sharing an unverified statement, "positive thoughts can cleanse our body of toxins and make us healthy".



A clear interpretation can be drawn from the three responses: that individuals have a strong tendency to share information without much investigation into its veracity.

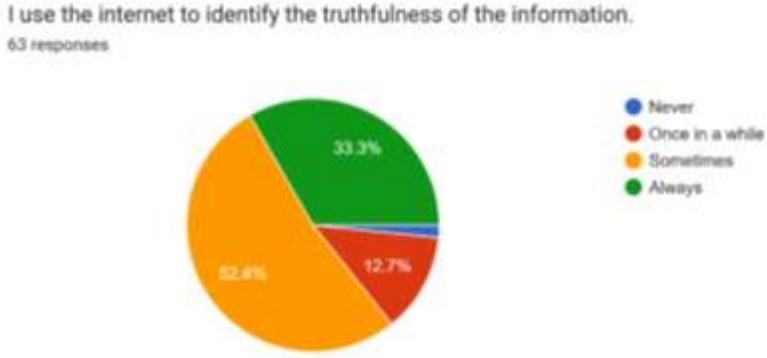
It is detrimental to consume content without critically assessing it. Firstly, because it indicates what one is aware of in what one is reading and what one endorses. Secondly, sharing unverified information harms others who may have placed trust in the sender.

The likely individual factors in spreading unverified information can be some or any of the undermentioned:-

- The urge for socialisation.
- Altruism
- Information sharing
- Information seeking
- Altruism
- Low subject matter knowledge
- Similar belief system
- To pass the time
- Garbage in Garbage Out syndrome

Tendency to verify the information received on social media platforms

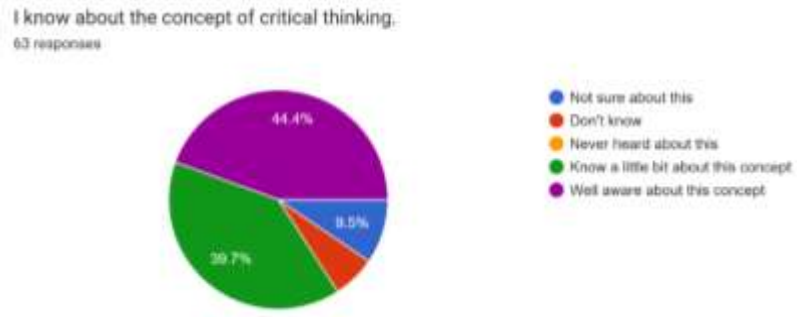
Only 21 of 63 respondents (33.3%) reported always using the internet and/ or other relevant sources to verify the reliability of the available information. The majority, i.e., 41 (65.1%) respondents, said they sometimes or once in a while do check the authenticity and reasonableness of the information they receive. This observation supports the hypothesis



that a majority of people either do not bother or lack the competence to check the authenticity of information they receive.

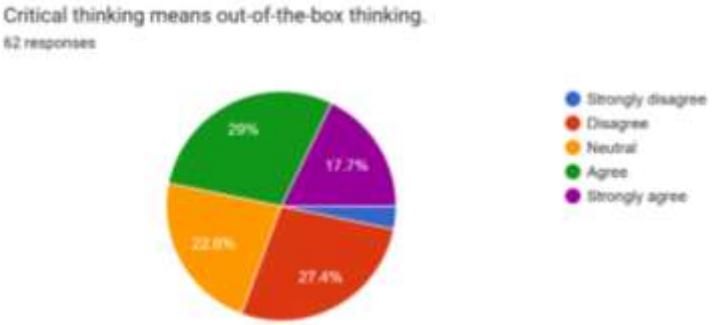
Understanding of the concept of Critical Thinking and its inclusion in the college curriculum

When a direct question is asked about knowledge of the concepts of critical thinking, 53 out of 63 respondents (84.1%) said they understood the concept, and only 10 among them acknowledged that they were not aware of this.

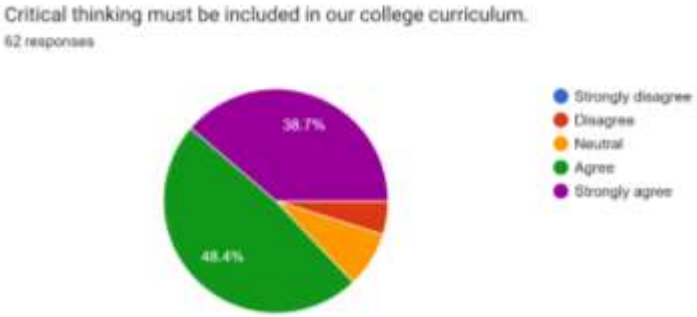


When the question was tweaked a bit, 43 of the respondents (69.3%) got it wrong, and only 19 (30.6%) got the correct answer.

In fact, critical thinking and out-of-the-box thinking are complementary but are very different concepts.



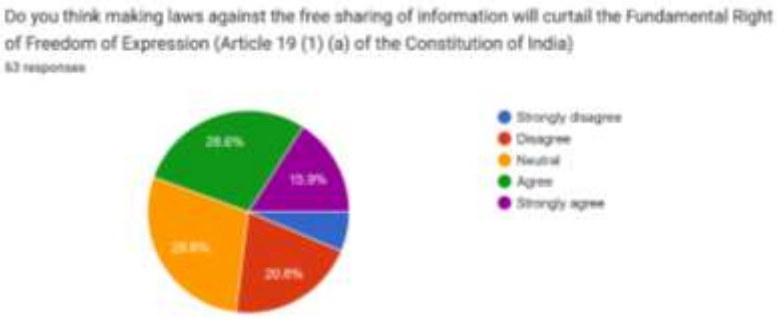
A total of 54 out of 62 respondents (87.1%) felt that critical thinking must be included in the respective curriculum, and to equip them with the requisite skills to differentiate between verified and unverified information.



It is vital for academic institutions of higher education to include critical thinking courses that will equip students with the requisite skills to identify, select, and recognise trustworthy information. The choices someone makes, and the opinions they form, must be based on appropriate interpretations of relevant, timely, and significant information.[18].

Agreeableness about Legislative intervention by the Government

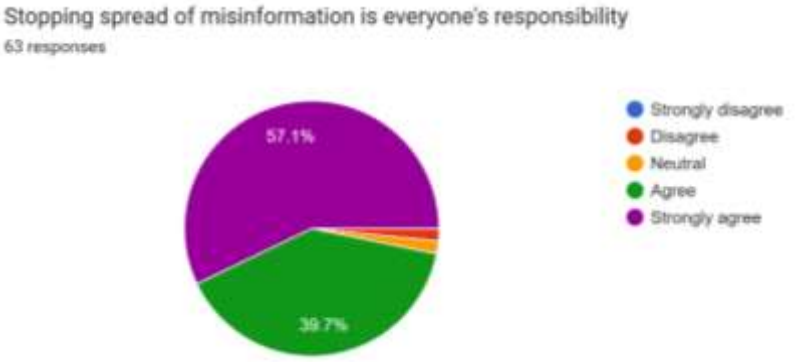
There was a mixed and confused response to the issue of legislative intervention against the free transmission of unverified information. The respondents were divided in their opinion about the implications of such legislation on the Fundamental Right of "Freedom of Expression" ("Article 19 of the Constitution of India"). 28 respondents (44.5%) opposed any legislative intervention, believing it would curtail the fundamental right; 18 (28.6%) were undecided; and 17 respondents supported legislative intervention to prevent the sharing of unverified information. **Given that all the sample consisted exclusively of law**



students, it is reasonable to expect that the general public will be less aware of their constitutional rights.



Striking an appropriate balance between regulatory measures and freedom of expression is a difficult, delicate challenge that requires a comprehensive, democratic debate. 28 out of 63 respondents (44.5%) agreed that imposing any legislative restriction on social media is an attack on the fundamental right of expression, while 18 (28.6%) had a neutral opinion, and 17 respondents believed that making laws against the free sharing of information is acceptable to them.



The most heartening result of the survey were that 61 of 53 respondents (96.8%) believed that stopping the proliferation of misinformation is everyone's responsibility. **This shows that digital information literacy to the general public will yield positive results in limiting the spread of misinformation.**

RECOMMENDATIONS

Critical Thinking tool-kit to verify the truth value of information

The proposed steps to verify the truthfulness of information are very intuitive and simple:

- Step 1: Analyse the information. What is being stated?
- Step 2: Source identification. What is the source?
- Step 3: Credibility of the source. What is the credibility of the source? Is the source a subject matter expert?
- Step 4: Evidence checking. What are the facts, data or basis of the claim? Is there valid supporting evidence?
- Step 5: Logical evaluation. Does the conclusion flow from valid premises?
- Step 6: Context and background. Why, where and when the content is produced?
- Step 7: Test for Bias and emotion. Is the message designed to provoke an emotional reaction, or is the message exploiting my familiarity with the subject?
- Step 8: Alternative Explanation. Can any other reliable source confirm the key findings of the claim?
- Step 9: Reflect and decide. Is the claim true and worthy of being shared?

Let us use the above critical thinking toolkit. To assess the benefits of this critical thinking toolkit, let us consider one of the survey statements.

The statement reads, “mutual fund investment is the best and most trusted method to multiply money”.

- **Step 1: Analyse the claim**.
  - Mutual fund investment is the best and most trusted method to multiply money.
- **Step 2: Source identification**.
  - What is the source? The statement is only a headline. There is no author, no publication details and no link to detailed information.
- **Step 3: Credibility of the source**.

- What is the credibility of the source? Since no source is specified, the information's credibility is doubtful.

- **Step 4: Evidence checking**.
- What are the facts, data or basis of the claim? The statement has an emotional appeal and a promise of high returns on the investment. However, no such evidence to prove the claim has been provided in the statement.

- **Step 5: Logical evaluation**.
- Does the conclusion flow from valid premises? There is no logical explanation; also, words like 'best method' and 'multiply money' lack benchmarks for comparison.

- **Step 6: Context and background**.
- Why, where and when the content is produced? There is no information about why, where and when the study was conducted. Was the study conducted by some reputable firm or organisation, and is there any financial backing for the study?

- **Step 7: Test for Bias and emotion**.
- The message uses urgency and moral framing (a promise of money multiplier investment) designed to provoke sharing.

- **Step 8: Alternative Explanation**.
- Is there further evidence from other authoritative sources that supports such gains from mutual fund investment?

- **Step 9: Reflect and decide**.
- Judgement: Uncertain and unverified claim, likely misleading information. The statement does not elicit confidence. It is likely that it overstates the benefits of the money multiplier in the case of mutual fund investment. It's better not to share with others. Is the claim true and worthy of being shared?

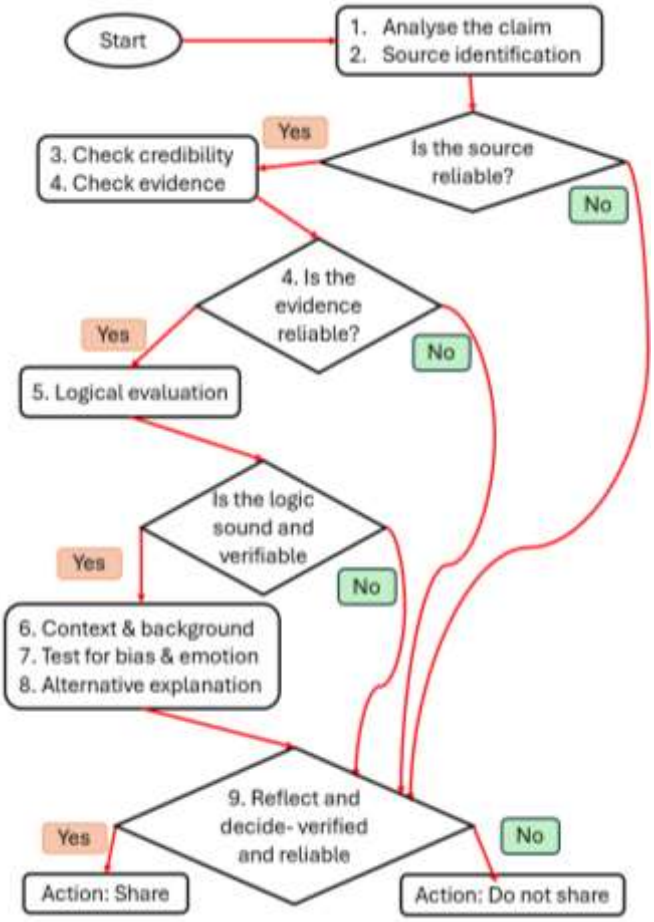


Fig 1. Verification and decision flow chart utilising the critical thinking tool-kit

Education intervention

Curriculum-critical thinking. There is an emergent need to introduce the concepts of critical thinking and relevant skills into the school and college curricula to empower students to distinguish true, verified information from misinformation.

Community outreach

The government and other organisations are required to develop community engagement programmes to involve and integrate communities and to build a broad base of media-literate citizenry. A few noteworthy examples from across the globe are highlighted below:

North Africa and the Middle East. In Lebanon, Megaphone operates as a volunteer run collective network of local journalists that produces news content tailored to young audiences vulnerable to misinformation.

- **Asia**. The "Digital Empowerment Foundation (DEF)" operates across 500 villages in India, strengthening community resilience against misinformation. It prioritises developing basic digital awareness among community members, enabling them to access and evaluate online content they encounter before passing it on.
- **Eastern Europe**. FactCheck Georgia, established in 2013 by "Georgia's Reforms Associates (GRASS)", leads many capacity-building programmes to address misinformation. In Ukraine, VoxCheck, which was created in 2015 by VoxUkraine, adopts a similar model. Another initiative, StopFake was launched in 2014 at the Kyiv Mohyla School of Journalism by resident students and faculty.
- **Latin America**. A company named Chicas Poderosas, founded in 2013, is functioning in 13 Latin American countries with an aim of providing digital and media competencies and leadership to women at the grassroots level.

Conclusion

Misinformation in today's digital age is a very complex sociolegal problem which requires a holistic approach in general and community participation in particular. Combating misinformation requires a multi-disciplinary approach, including education, community awareness, platform redesign and nimble governance that aligns private incentives and keeps the public engaged. The empirical study highlighted the benefits of utilising critical thinking tools in developing the ability to differentiate misinformation amid the onslaught of the information tsunami.

Bibliography

1. Babii, Alexandra-Niculina. 'The Use of Critical Thinking Against Fake News'. Paper presented at NORDSCI International Conference. *NORDSCI Conference Proceedings, Book 1 Volume 3*, 2020. https://doi.org/10.32008/NORDSCI2020/B1/V3/14.
2. Colomina, Carme, Héctor SÁNCHEZ Margalef and Richard Youngs. *The Impact of Disinformation on Democratic Processes and Human Rights in the World*. n.d.
3. *Developing Critical Thinking Skills,* Online Lecture Series, http://www.usyd.edu.au/Ic.
4. Machete, Paul, and Marita Turpin. 'The Use of Critical Thinking to Identify Fake News: A Systematic Literature Review', Department of Informatics, University of Pretoria, South Africa, 2020.
5. In *Responsible Design, Implementation and Use of Information and Communication Technology*, vol. 12067, edited by Marié Hattingh, Machdel Matthee, Hanlie Smuts, Ilias Pappas, Yogesh K. Dwivedi, and Matti Mäntymäki. Lecture Notes in Computer Science. Springer International Publishing, 2020. https://doi.org/10.1007/978-3-030-45002-1_20.
6. 'Misinformation and Disinformation'. Accessed 22 February 2026. https://www.apa.org/topics/journalism-facts/misinformation-disinformation.
7. Pennycook, Gordon, and David G. Rand. 'The Psychology of Fake News'. *Trends in Cognitive Sciences* 25, no. 5 (2021): 388–402. https://doi.org/10.1016/j.tics.2021.02.007.
8. Muskan Manvar. 'Social Media Campaign and Combating Misinformation. A Case Study of the Citizen Amendment Act'. National Centre for Good Governance, Government of India, 2024.
9. Vosoughi, Soroush, Deb Roy, and Sinan Aral. 'The Spread of True and False News Online'. *Science* 359, no. 6380 (2018): 1146–51. https://doi.org/10.1126/science.aap9559.

"AI and Social Media Manipulation"

Submitted By

Qadri Saad Sayyad Fakhruddin

Class: LL.M. (2nd Year), Semester IV

Specialization: (Business Law) PRN No.: 2024016400258165 Roll No.: A-13.

Abstract

The rapid advancement of artificial intelligence (AI), particularly generative models, has turned social media into a strong tool for manipulation. This paper analyses that how AI can be used to both help with large-scale social media manipulation and provide tools for finding and elimination. Generative AI makes it possible to make very realistic deepfakes, fake people, automated bots, and personalized disinformation campaigns on a scale and at a cost that has never been seen before. These technologies make false information more common, change the narratives, affect elections, and lower people's trust in institutions by making content that looks like it was made by a real person and often makes people feel strong emotions like fear or anger to get them to engage.

Research reflects in particular way that created by AI is more likely to get viral, often focuses on funny or positive stories, and comes from smaller accounts. However, it spreads quickly on social platforms. Algorithms make the problem worse by putting sensational content first, making echo chambers, and allowing people to act in untrue ways together. Recent development and usage in fact has impacted diplomacy globally, between AI used overcome, such as advanced algorithms for finding fake media, systems for checking facts, and tools for moderating content.

This study examines existing literature, evaluates practical instances of AI driven manipulation and recommends interdisciplinary approaches that include platform designs, regulatory, media and etc. To overcome we intended to find a balance between new ideas the protections that keep information safe in the digital age. No action can cause the division and lack of trust in the government.

Keywords: AI-generated disinformation, deepfakes, social media manipulation, generative AI, misinformation detection, algorithmic bias.

Research Problem

The swift adoption of artificial intelligence tools by social media platforms has revolutionised the way content is produced, personalised, a[1]nd spread at scale. Techniques such as algorithmically generated deepfakes, voice cloning, synthetic audio-visual material (now formally termed Synthetically Generated Information or SGI under recent Indian rules), coordinated bot networks, and hyper-targeted disinformation campaigns have built strong new regions for distorting public perception. These practices weaken democratic deliberation, fuel social divisions, and compromise national security, all while operating within a largely borderless digital ecosystem.

India, home to nearly one billion internet subscribers and one of the world's most active digital publics, faces heightened vulnerability. The 2024 Lok Sabha polls witnessed widespread use of AI-generated voice messages from political leaders, digitally "resurrected" deceased figures delivering campaign speeches, and fabricated videos featuring celebrities endorsing parties—though outright malicious deepfakes did not dominate the narrative. Similar patterns have appeared globally: influence operations employing synthetic media targeted elections in the United States, Slovakia, Bangladesh, and Taiwan between 2024 and 2026, echoing earlier scandals such as Cambridge Analytica's data-driven micro-targeting.

Indian law has responded through the I . T (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, which mandate SGI labelling, user declarations for large platforms, automated detection tools, and expedited three-hour takedowns. Yet the regime stays largely reactive, placing the burden on intermediaries after content circulates rather than preventing the creation or deployment of manipulative systems at source. This doctrinal shortfall—juxtaposed against the

[1] *Constitution of India*

European Union's proactive, risk-categorised prohibitions—raises a fundamental question: can India craft a balanced regulatory response that curbs harmful AI practices while honouring constitutional guarantees of free speech under Article 19(1)(a), as interpreted in landmark rulings emphasising precision and reasonableness?

Objectives

This paper seeks to:

1. Map the operational pathways through which AI enables manipulation on social media and assess their implications for Indian society.
2. Scrutinise the current statutory and policy architecture in India governing these technologies.
3. Identify doctrinal inconsistencies and enforcement challenges within the existing framework.
4. Place India's approach in conversation with the EU AI Act and prevailing international trends.
5. Offer pragmatic, constitutionally aligned recommendations that reconcile innovation with public-interest safeguards.

Research Questions

1. In what specific ways do AI s[2]ystems facilitate manipulation on social media, and how have these manifested in the Indian electoral and social context?
2. How adequately do the Information Technology Act, 2000, the 2021 Intermediary Guidelines (as amended in 2026), and complementary statutes address AI-enabled manipulation?
3. What conceptual and practical shortcomings limit the effectiveness of India's current regime?
4. To what degree does India's intermediary-focused model diverge from the EU AI Act's outright bans on manipulative techniques?
5. Which targeted reforms could strengthen India's framework while remaining sensitive to its developmental priorities and free-expression traditions?

Literature Review

Academic discussion around AI and social media manipulation has shifted from early analyses of filter bubbles and algorithmic bias toward deeper examinations of generative tools' capacity to fabricate convincing realities. Foundational studies on data-harvesting scandals illustrated how psychographic profiling could subvert voter autonomy. More recent scholarship documents the rapid proliferation of synthetic media, highlighting its potential to implant false memories, polarise communities, and erode institutional legitimacy.

Empirical work reveals a dual-edged reality: AI can enhance personalised political communication yet simultaneously amplify harms ranging from election interference to non-consensual intimate imagery. India-focused research on the 2024 general elections notes that while sophisticated deepfakes were comparatively restrained, subtler AI-assisted content—such as cloned voices and resurrected leaders—contributed to heightened post-poll polarisation. Broader international literature underscores the rise of "bot swarms" and synthetic identities as tools for undermining collective sense-making.

A noticeable gap remains in purely legal scholarship tailored to India. Most existing analyses emphasise technical detection methods or platform accountability rather than a systematic doctrinal integration with constitutional speech protections. The present study addresses this lacuna by confining itself to black-letter legal interpretation.

Doctrinal Method

This research adopts the doctrinal method also known as black-letter law analysis—as its primary methodological framework. The doctrinal approach involves a systematic, logical, and deductive examination of primary legal sources to ascertain the current state of the law, interpret its provisions, identify internal inconsistencies or gaps, and evaluate its normative coherence in addressing contemporary challenges. Unlike empirical research that gathers new data through surveys or fieldwork, or socio-legal studies that explore law's societal impact, the doctrinal method remains

[2] *The Information Technology Act, 2000*

firmly anchored in legal texts themselves: statutes, subordinate legislation, judicial precedents, official notifications, and constitutional provisions. It proceeds deductively, beginning from foundational principles and moving toward specific applications, while employing established tools of statutory interpretation such as literal reading, purposive construction, and harmonisation with higher norms.

In the context of this study on AI-driven social media manipulation, the doctrinal method is particularly apt because the core issues revolve around the interpretation and adequacy of existing legal doctrines rather than quantitative measurement of harm or user behaviour. The inquiry starts with the constitutional bedrock of freedom of speech and expression under Article 19(1)(a) of the Constitution of India, read with the reasonable restrictions clause in Article 19(2). It then scrutinises how subordinate legislation and rules operationalise (or fail to operationalise) these guarantees when confronted with rapidly evolving technologies like generative AI, deepfakes, and synthetically generated information (SGI).

The research process under this method unfolded in structured stages. First, a close textual reading of the I.T Act, 2000 (IT Act) was undertaken, focusing on key provisions such as Sections 66C (identity theft), 66D (cheating by personation), 67 (obscene material), 67A and 67B (sexually explicit content), 69A (power to block content for specified reasons), and 79 (safe harbour for intermediaries subject to due diligence). These sections were analysed for their original legislative intent—primarily addressing cybercrime and intermediary liability in the early internet era—and their adaptability to AI-generated synthetic media.

Next, the study examined the I. T Rules, 2021, as significantly amended by the Amendment Rules, 2026 (notified on 10 February 2026 and effective from 20 February 2026). A detailed doctrinal dissection of the new Rule 2(1)(w), which defines "Synthetically Generated Information" (SGI), reveals a carefully crafted scope limited to audio, visual, or audio-visual content that is artificially or algorithmically created, modified or altered in a manner that appears authentic and indistinguishable from reality to a reasonable person, while depicting persons, events, or scenes that did not occur. Notably, the definition excludes text-only outputs, routine good-faith editing, accessibility enhancements, and non-manipulative improvements—reflecting a purposive narrowing after public consultation to avoid overbreadth.

The amendments insert a dedicated Rule 3(3) on due diligence for SGI, imposing layered obligations: intermediaries must prevent dissemination of "prohibited SGI" (e.g., content violating existing laws on defamation, impersonation, or national security); ensure prominent, tamper-resistant labelling and metadata for permissible SGI; deploy automated tools for detection where feasible; and comply with government takedown orders within a compressed three-hour window[3]. For Significant Social Media Intermediaries (SSMIs) platforms exceeding five million users—additional requirements include user declarations before uploading SGI, technical verification mechanisms, quarterly notifications to users, and enhanced grievance redressal timelines (as short as two hours for certain SGI complaints). Non-compliance risks loss of safe harbour protection under Section 79, exposing platforms to vicarious liability.

My doctrinal analysis highlights both strengths and tensions in this framework. On the positive side, the 2026 amendments introduce definitional clarity and proactive transparency measures (labelling and metadata) that were absent in the pre-2026 regime, directly responding to the speed and scalability of AI manipulation. This represents an evolution from purely reactive content removal to a hybrid model incorporating design-level safeguards at the intermediary level. However, the regime remains fundamentally ex-post and intermediary-centric: it targets dissemination and platform accountability after content circulates, rather than imposing upstream prohibitions on the development or deployment of manipulative AI systems themselves. This doctrinal structure places heavy reliance on platforms' technical capacity and goodwill for automated detection, while leaving creators of AI tools largely unregulated under these rules.

[3] *The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021*

Constitutional compatibility forms a central pillar of the analysis. The landmark judgment in Shreya Singhal v. Union of India (2015) serves as the interpretive lodestar. In that case, the Supreme Court struck down Section 66A of the IT Act for vagueness and overbreadth, holding that terms like "annoying," "inconvenient," or "grossly offensive" failed to provide clear standards, thereby producing a chilling effect on protected speech. The Court emphasised that any restriction on Article 19(1)(a) must be precise, narrowly tailored, and fall squarely within the[4] enumerated grounds of Article 19(2) (sovereignty and integrity of India, public order, decency, morality, etc.). It further clarified the distinction between discussion, advocacy, and incitement—only the latter justifying curbs. The judgment also read down Section 79 to require intermediaries to act only upon actual knowledge via court or government orders, rejecting proactive monitoring that could turn platforms into censors.

Applying Shreya Singhal doctrinally to the 2026 SGI framework reveals potential vulnerabilities. While the new rules avoid the open-ended language of the old Section 66A by tying obligations to existing penal provisions (e.g., via the Bharatiya Nyaya Sanhita, 2023, for defamation or impersonation), the breadth of "unlawful" SGI and the pressure for automated flagging could still risk over-removal of satire, parody, artistic expression, or legitimate political commentary. For instance, a deepfake used in political satire or a voice clone in creative advertising might trigger automated systems, leading platforms to err on the side of caution and suppress borderline content. This echoes the pre-Shreya concerns of self-censorship. Moreover, the short compliance timelines (three hours for takedowns) may incentivise hasty decisions without adequate human oversight, potentially violating the reasonableness test under Article 19(2), which demands proportionality and minimal impairment of the right.

The doctrinal lens also incorporates the Digital Personal Data Protection Act, 2023 (DPDP Act), which governs personal data processing but offers only indirect support against manipulation (e.g., through consent and purpose limitation for data used in training AI models). Similarly, provisions in the Bharatiya Nyaya Sanhita addressing cheating, defamation, or forgery provide criminal sanctions for individual perpetrators, yet enforcement against anonymous or cross-border actors remains challenging in a doctrinal sense due to jurisdictional limits and evidentiary burdens in proving intent or harm from synthetic content.

A comparative doctrinal element without shifting to empirical methodology juxtaposes India's framework against the EU AI Act (Regulation (EU) 2024/1689). Under Article 5 of the EU AI Act (applicable since February 2025), outright prohibitions target "unacceptable risk" practices, including AI systems that issues technical skills or purposefully manipulative/deceptive methods with the-objective or effects of materially distorting behaviour and causing significant harm. Additional bans cover exploitation of vulnerabilities (based on age, disability, or socio-economic status) and social scoring. This ex-ante, risk-classified approach directly regulates the design and placing on the market of manipulative AI systems, contrasting sharply with India's ex-post, platform-liability [5]model. Transparency obligations (Article 50) mandate labelling of deepfakes and synthetic content in public-interest contexts. Doctrinally, the EU framework embeds preventive prohibitions into the fabric of product safety and fundamental rights law, whereas India's amendments layer obligations onto the existing intermediary safe-harbour doctrine derived from Section 79.

Complementing the rules, the India AI Governance Guidelines, 2025 (released by MeitY in November 2025) adopt a voluntary, principle-based philosophy centred on "Do No Harm" alongside seven guiding sutras emphasising innovation, human-centric design, fairness, transparency, accountability, and sectoral adaptability. An AI Governance Group is proposed for coordination, yet the guidelines remain non-binding—serving as soft law that informs but does not alter the hard

[4] *The Information Technology Act, 2000*

[5] *EU Act*

doctrinal obligations under the IT Rules. This creates a hybrid doctrinal landscape: binding due diligence for platforms alongside aspirational principles for broader AI ecosystem actors.

In my analysis, the doctrinal method exposes a maturing but incomplete architecture. The 2026 amendments advance precision and speed of response, addressing enforcement delays that plagued earlier regimes. However, they do not fully resolve upstream risks—such as the training of generative models on biased or scraped data, or the deliberate design of recommendation algorithms that amplify manipulative content for engagement. The absence of mandatory risk-tiering or independent algorithmic audits leaves a doctrinal gap when measured against the EU's categorical bans. Furthermore, reliance on platforms for detection and labelling raises questions of private censorship, potentially conflicting with Shreya Singhal's insistence on state accountability and clear standards.

The research also draws analogies from judicial precedents on intermediary liability and blocking orders (e.g., interpretations of Section 69A requiring reasoned orders and procedural safeguards). Global examples, such as AI voice clones and deepfakes in the 2024 Lok Sabha elections (including fabricated celebrity endorsements, resurrected leaders' speeches, and misleading political clips), illustrate how synthetic content tests the limits of current doctrines—often blurring lines between permissible campaign creativity and unlawful manipulation.

Overall, the doctrinal method reveals that while India's framework has evolved reactively to technological realities, it requires further interpretive harmonisation with constitutional free-speech doctrine to prevent chilling effects and ensure proportionality. Future doctrinal development could involve reading in implied obligations for risk assessments or transparency in algorithmic design, thereby strengthening the regime without legislative overhaul. This black-letter analysis thus not only maps the law as it stands but critically evaluates its internal logic and constitutional fit in the age of generative AI.

Sources of Data

Primary materials comprise the Constitution of India, the Information Technology Act, 2000 (particularly Sections 66C, 66D, 67, 67A, 69A, and 79), the I.T.(Code) Rules, 2021 as amended by the 2026 Notification, the Digital Personal Data Protection Act, 2023, relevant provisions of the Bharatiya Nyaya Sanhita, 2023, and MeitY notifications. Secondary sources include the India AI Governance Guidelines, 2025; Regulation (EU) 2024/1689 (EU AI Act); peer-reviewed journals; official government reports; and verified fact-checking documentation of domestic and international incidents. All references are drawn from publicly accessible official documents and reputable scholarly publications.

Legal Analysis

India's regulatory stance remains centred on intermediary responsibility and post-publication remedies. The IT Act penalises identity theft and impersonation (Section 66D) and prohibits obscene or sexually explicit synthetic content (Sections 67, 67A, 67B), offering some leverage against deepfake fraud or revenge [6]material. Section 69A permits blocking orders on grounds of sovereignty, public order, or security, while Section 79 extends safe-harbour protection to platforms that observe due diligence.

The 2026 amendments mark a notable evolution. They introduce a formal definition of SGI—any audio, visual, or audio-visual material created or altered algorithmically to appear authentic—and impose layered obligations. All intermediaries must deploy technical measures to block unlawful SGI and ensure permanent, tamper-proof labelling and metadata for permissible content. Significant Social Media Intermediaries (those exceeding five million users) face heightened duties: users must declare SGI status before uploading, platforms must verify via automated tools, and labels must be prominently displayed. Quarterly user notifications replace earlier annual requirements, and government-directed takedowns must occur within three hours. Failure to comply risks loss of safe-harbour immunity and potential vicarious liability.

[6] *Digital Personal Data Protection Act, 2023.*

Complementing these rules, the India AI Governance Guidelines, 2025, articulate a voluntary, principle-based philosophy anchored in "Do No Harm" and seven guiding sutras that prioritise innovation, human-centric design, and sectoral adaptability. An AI Governance Group is proposed for coordination, yet the framework remains non-binding. The Digital Personal Data Protection Act, 2023, touches data handling but does not directly regulate content manipulation.

Constitutionally, the Supreme Court's decision in Shreya Singhal v. Union of India (2015) remains pivotal. By striking down Section 66A for vagueness and overbreadth, the Court insisted that any speech restriction must satisfy the strict reasonableness test under Article 19(2). Critics contend that the 2026 Rules' expansive SGI obligations and automated moderation mandates could inadvertently chill legitimate satire, parody, or political critique, echoing pre-Shreya concerns. Enforcement continues to rely heavily on platform goodwill, with political actors occasionally deploying AI content and the Election Commission offering limited intervention.

In contrast, the EU AI Act employs a tiered risk model. Article 5 explicitly prohibits "unacceptable-risk" applications, including AI systems that deploy subliminal, manipulative, or deceptive techniques to distort behaviour and cause significant harm, or that exploit vulnerabilities based on age, disability, or socio-economic status. Transparency rules under Article 50 require clear labelling of deepfakes and synthetic content in public-interest contexts. This ex-ante prohibition on harmful design choices differs markedly from India's predominantly ex-post, intermediary-liability approach.

Findings

The doctrinal review discloses a framework that has matured considerably yet still falls short of comprehensive prevention. The 2026 amendments deliver welcome definitional clarity, mandatory labelling, and accelerated compliance timelines, directly confronting the speed and scale of synthetic content. Nevertheless, the architecture remains fundamentally reactive—focusing on detection and removal after dissemination rather than barring the upstream design of manipulative systems. Persistent gaps include the absence of formal risk-tiering, insufficient independent auditing of recommendation algorithms, and continued dependence on self-declared compliance by platforms. While Shreya Singhal's safeguards against vague restrictions persist on paper, practical pressures from automated flagging may still produce chilling effects. Internationally, India's innovation-oriented, sectoral guidelines stand apart from the EU's harmonised bans, offering flexibility suited to a developing digital economy but exposing the system to enforcement shortfalls and regulatory arbitrage. Global case studies confirm that manipulation flourishes most where accountability mechanisms lag behind technological capability.

Conclusion

Artificial intelligence has introduced potent new instruments for shaping opinion on social media, posing a direct challenge to India's democratic fabric and social equilibrium. Recent amendments to the intermediary rules represent a constructive step toward transparency and swift redress, yet the prevailing doctrinal design lacks the forward-looking prohibitions and institutional muscle needed to neutralise emerging generative threats. Absent deeper integration of preventive safeguards and stronger cross-sector coordination, India risks allowing both domestic polarisation and external interference to erode public trust. A carefully calibrated regime, respectful of constitutional liberties and technological realities, is therefore essential for safeguarding the integrity of the digital public square.

Suggestions

To close existing gaps, India should consider enacting a dedicated, risk-calibrated AI statute that draws selective inspiration from the EU AI Act while preserving the "Innovation over Restraint" ethos of the 2025 Governance Guidelines. High-risk manipulative practices—such as subliminal distortion of behaviour, exploitation of vulnerable groups, or social-scoring systems—could be expressly prohibited, with narrow carve-outs for creative or assistive uses. Popular social media must undergo mandatory transparency audits of their recommendation engines, and an independent AI Safety

Institute under MeitY could be established to conduct certification, red-teaming, and impact assessments.

The proposed AI Governance Group should receive statutory footing with binding sectoral codes for high-stakes domains such as elections and public health. Complementing regulation, nationwide media-literacy initiatives integrated into school curricula and community programmes—modelled on successful global fact-checking networks—would empower citizens to navigate synthetic content discerningly. Political parties could be required to disclose AI usage in campaign materials, building on 2024 civil-society appeals.

Internationally, India might champion cooperative mechanisms through forums such as the G20 or Quad to harmonise cross-border standards and expedite takedowns. Compared with the EU's categorical prohibitions that foreground fundamental rights, India's hybrid model would retain developmental flexibility while incorporating proportionate ex-ante obligations. Phased implementation, coupled with graduated penalties and safe-harbour incentives for compliant platforms, would minimise over-censorship risks. Such reforms would align regulatory evolution with constitutional values, technological dynamism, and the imperatives of resilient democracy.

1.BIBLIOGRAPHY & REFERENCES

1. Information Technology Act, 2000 (as amended).
2. Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026 (notified 10 February 2026).
3. Regulation (EU) 2024/1689 of the European Parliament and of the Council (EU AI Act).
4. Shreya Singhal v. Union of India, (2015) 5 SCC 1.
5. Ministry of Electronics and Information Technology, *India AI Governance Guidelines*, 2025.
6. Digital Personal Data Protection Act, 2023.
7. MeitY notifications and explanatory materials on SGI framework (2026).
8. Fact-checking reports: Alt News, Boom Live (analyses of 2024 Lok Sabha elections).
9. International coverage: BBC, Reuters, Washington Post reports on global synthetic-media incidents (2024–2026).
10. Academic and policy sources: RAND Corporation studies on AI disinformation; peer-reviewed articles in Technological Forecasting and Social Change and related journals (2024–2026).

"Regulating Artificial Intelligence in Higher Education in India: A Critical Analysis of Emerging Legal Gaps and the Need for a Comprehensive Regulatory Framework"

Bansari Bhanushali[1]
Vignesh Narayanan[2]

Abstract

The Indian education system has evolved significantly over the last decade. However, today, one big issue which does arise specifically for higher education is the introduction of Artificial Intelligence. This paper intends to address a very emerging concern: ***whether artificial intelligence is a boon or a bane with respect to higher education.*** *The paper highlights how the introduction of Artificial Intelligence has led to legal complications regarding higher education laws. The authors of this paper believe that this is a significant issue that needs utmost attention in terms of codification and regularisation. The need for codification is to prevent any kind of misuse with respect to Artificial intelligence. This need is essential to upgrade the higher education system so that the system gets harmonised with technology and Artificial intelligence. This would enable further growth of higher education in India. Education is of utmost importance, as was highlighted once by Mr Nelson Mandela,* ***"Education is the most powerful weapon which you can use to change the world"****. This research paper also intends to resolve this issue and suggest recommendations that could resolve this legal gap and recommend amicable solutions for this existing legal problem.*

Keywords: Artificial Intelligence, Higher Education, Legal Framework, Data Privacy, Intellectual Property, Regulation, India

Research Objectives

- To understand the legal framework for higher education
- To understand the role of AI in higher education
- To understand legal gaps and potential legal issues with respect to AI development in higher education.
- To suggest and recommend solutions with respect to the legal complexities that would arise in higher education laws due to AI development.

Research Questions

- What is the existing legal framework governing higher education in India?
- What is the role of Artificial Intelligence in higher education?
- What are the legal gaps and potential issues arising from the use of Artificial Intelligence in higher education?
- What solutions can be suggested to address the legal complexities caused by Artificial Intelligence in higher education?

Scope and Limitations

This research is limited to the Indian higher education system and focuses primarily on legal and regulatory aspects of Artificial Intelligence. It does not delve into the technical functioning of AI systems. The study is based on secondary data and existing literature, which may limit empirical analysis.

Research Methodology

The Authors of this paper have followed the doctrinal method of research.

Literature Review

[1] Currently pursuing Masters of Law (LL.M) at Hindi. Vidya Prachar Samiti's College of Law, Mumbai.
[2] Currently pursuing Masters of Law (LL.M) at Kirit. P. Mehta School of Law, Mumbai.

A paper authored by Akansha Anand titled "Policy Implementation: A Study of Higher Education in India Higher Education in India: Accessibility and Inclusion"[3] does address important aspects pertaining to higher education. It does address certain aspects such as the historical evolution of the National Education Policy. It also highlights the role of the five-year plan in higher education. This paper also focuses on the need for privatisation of higher education. It also focuses on the role of the Ministry of Human Resource Development in higher education. However, this paper fails to address the role of AI in higher education. It also fails to define the overall legal framework of higher education in India and the implications AI would have on the same.

A paper authored by Dr. Jyoti Sharma, Mr. Shamshad Husain and Dr. Anil Kumar Nishad titled "An Overview on Higher Education in India"[4] does address about the higher education system in India. This paper does highlight about the evolution of higher education in India. This paper also touches upon major challenges in India higher education system. It also addresses the problem of demand supply gap in higher education along with the regulatory challenges. However, this paper fails to address the role of AI in higher education. This paper also does not address the legal impact and gaps in higher education laws due to evolution of AI.

A paper authored by Nand Kishor Soni and Teerath Prasad Patel titled "Quality Teaching & Higher Education System in India"[5]does address about higher education system in India. This paper does primarily address about historical evolution of higher education in India. This paper also addresses about the significance of quality teaching. This paper further dwells into assurance with respect to quality teaching along with aspects such as how teaching quality can be enhanced. However, this paper does not address the role of AI in higher education. This paper also does not address legal impact and gaps in higher education laws due to evolution of AI.

Introduction

It is considered that one of the finest discoveries that thrives in today's generation was made by John McCarthy and Marvin Minsky, which is none other than Artificial Intelligence, which they defined as "the science and engineering of making intelligent machines". [6] This discovery has been having a significant impact on India. The development of artificial intelligence in India has led to over Rs 10,300/- Crore investment and 38,000 GPUs. [7] In reference with respect to higher education, India has had a remarkable reform, i.e., over 54 universities in India come under the QS world ranking, which is considered a 5-fold increase in comparison to 2015.[8] This development of AI and Higher education in parallel fashion has led to the intersection of AI with Higher education. If we were to rationally analyse technologically, it was beneficial as it gave a significant and wider boost to the infrastructure of higher education. However, with respect to the legal aspect, it did give rise to certain complications.

Statement of Problem

India is now in the process of achieving a position as a global hub of AI, since there is rapid growth due to the usage of AI across all sectors. Since 2024, new startups have been mainly into AI-powered, and 87% of the businesses are employing AI. With such rapid growth and efficient usage of AI, the Indian market is expected to see a minimum growth rate of nearly 25% to 35% by 2027.[9] The government of India, after knowing the usage and efficiency of AI and its impact, has formulated various initiatives and policies, one amongst them in the National Education Policy (NEP) 2020,

---

[3] Akansha Anand, "Policy Implementation: A Study of Higher Education in India Higher Education in India: Accessibility and Inclusion",4,International Journal of social policy and education,7,2022.

[4] Dr. Jyoti Sharma, Mr. Shamshad Husain and Dr. Anil Kumar Nishad, "An Overview on Higher Education in India", 22, Journal of Gujarat Research Society,1058,2020.

[5] Nand Kishor Soni and Teerath Prasad Patel, "Quality Teaching & Higher Education System in India",4, International Journal of Scientific and Research Publications,1,2014.

[6] AIP Publishing, https://pubs.aip.org/aip/acp/article-abstract/2919/1/020001/3278955/The-impact-of-artificial-intelligence-on-today-s?redirectedFrom=fulltext (last visited April 16, 2026).

[7] Press Information Bureau, https://www.pib.gov.in/PressReleasePage.aspx?PRID=2178092&reg=3&lang=2 ,(last visited April 16, 2026)

[8]Press Information Bureau, https://www.pib.gov.in/PressNoteDetails.aspx?NoteId=154694&ModuleId=3&reg=3&lang=1, (Last visited April,16,2026).

[9] Press Information Bureau, https://www.pib.gov.in/PressReleasePage.aspx?PRID=2234853&reg=3&lang=1, (Last visited April,16,2026)

which has recognised how AI would help in solving challenges faced by the traditional higher education system, enhancing teaching techniques, and most importantly, its efficient usage.[10] AI in education is mainly a collection and analysis of personal data, and the concerning part is the privacy and security issues, which are regulated by the Digital Personal Data Protection Act,2023 and the Information Technology Act,2000, but there is a real need for rigorous testing and auditing of AI, which is not covered in the ambit the existing legal framework.[11] A study was conducted by a London-based college named Trinity College, where it was found that nearly one-third of teachers are of the view that use of AI should be banned whereas one-quarter of the teachers are actually using AI and on the same hand 74% teachers strongly believe that misuse of AI will always be done by the students and it will forever remain a problem.[12] Apart from the effective use of AI in higher education, it is also widely misused, and there is a real-time need for codification and regulation of AI.

Conceptual Framework

The Indian Education system saw a major shift after the 42nd Amendment Act (1976), transferring education from the State List to the Concurrent List, which enables central oversight. So today, education is a concurrent subject under Entry 25 of List III, which allows both Parliament and State Legislatures to enact laws on education.[13]Higher Education in India is not just about classrooms and books but also about the legally available rights and obligations. Article 21A[14] of the Constitution, as amended by the 86th Constitutional Amendment Act, 2002, states that the state shall provide free and compulsory education to all children between the ages of 6 to 14 years, and thus the Centre of Education Law in India is the Right of Children to Free and Compulsory Education for every child between 6 to 14years of age.[15] India's higher education system serves over 40 million students across a vast network of universities and colleges. This framework not only regulates institutional governance and academic standards but also safeguards the rights and interests of students.[16] Constitutional provisions shaping India's education policy include Article 21A[17] which was introduced through the 86th Constitutional Amendment (2002), making free and compulsory education a fundamental right for children aged 6-14 years, Article 45[18] which directs the state to provide early childhood care and education for children below six years, Article 46 [19]which mandates the promotion of educational and economic interests of Scheduled Castes, Scheduled Tribes, and other weaker sections, Directive Principles of State Policy (DPSP), which envisage universal education as a fundamental governance priority.

The recognition of elementary education as a fundamental right forms the basis for access to further educational opportunities. It makes sure the basic foundation is clear, and one possesses the necessary skills to pursue higher education. To regulate Higher education in India, regulatory authorities like the University Grants Commission (UGC), which looks after education and academic standards, the All India Council for Technical Education (AICTE), that Regulates technical and management institutions under the AICTE Act, 1987[20], National Medical Commission (NMC), which oversees medical education. Bar Council of India (BCI), which regulates legal education and law universities, was formed. Higher education is not an exception to legal reforms. The Viksit Bharat Shiksha Adhishthan Bill, 2025, suggests that all the regulatory authorities, like UGC, AICTE, and NCTE, can be united under a single body. The Bill enhances the redressal of grievances and seeks to

[10] Supra note 6

[11] Shivi Kanjolia, Transforming Innovation Into Fairness: Tackling Legal Challenges In Ai-Enhanced Education In India, (Volume V),(816) (2025).
[12] Stratx Simulation, The Negative Effects of Artificial Intelligence in Education,(Last visited April,17,2026)
[13] Bhat&Joshi Advocates, Regulatory Framework of Education in India: A Comprehensive Legal Analysis - Bhatt & Joshi Associates, (Last Visited April,17,2026)
[14] INDIA CONST, art21A.
[15] Kamal &Co Advocates, Education Law In India | Kamal & Co. , (Last visited April,17,2026).
[16] Distance Learning Institute, The Legal Framework of Higher Education in India - Distance Learning Institute, (Last visited April,17,2026)
[17] INDIA CONST, art21A.
[18] INDIA CONST, art46.
[19] INDIA CONST, art45.
[20] The All-India Council for Technical Education Act, 1987, No.52, Acts of Parliament, 1987 (India).

increase the Gross Enrolment Ratio in higher education to 50% by 2035. [21] In light of this evolving constitutional and regulatory landscape, it becomes imperative to ensure that emerging challenges, particularly those posed by technological advancements such as Artificial Intelligence, are addressed through a coherent, comprehensive, and future-ready legal framework that upholds the principles of equity, accountability, and quality in higher education.

The following are the major statutes on higher education in India:

Universities Grant Commission Act 1956[22]-: This is the first landmark legislation related to higher education. This legislation established a statutory body that would govern higher education policies. This statute consists of 28 sections covered in 4 chapters. This statute only establishes the commission and focuses on its powers and functions. The statutory body is exclusively known for conducting NET, i.e., National Eligibility Test, examination, which is mandatory to clear in order to be appointed as an assistant professor for over 50-80 humanitarian subjects and CSIR for subjects pertaining to life science.

Institute of Technology Act 1961[23]-This statute majorly focused on the evolution of science and technology by giving the Indian Institute of Technology an importance of national significance. However, this statute also does not address anything about Artificial Intelligence.

Other important legislations include the All-India Council for Technical Education Act 1987[24] and the Central Universities Act 2009[25]. The All-India Council for Technical Education Act 1987[26] established the All-India Council for Technical Education, which majorly focuses on vocational subjects and technical education, such as engineering, hotel management, and pharmacy, being some of the major subjects. This statute consists of six chapters and twenty-five sections. However, none of the sections deals with the aspect of artificial intelligence. The Central Universities Act 2009[27] consists of 47 sections along with two primary schedules. It aims to promote integrated education along with innovation and national development and social improvement. Today, India consists of over 50 central universities. However, the Central Universities Act 2009 is silent on Artificial Intelligence and the application of the same. There are no existing uniform policies on "AI" with respect to central universities.

Findings

Despite growing AI adoption, India's legal framework remains fragmented and insufficient to tackle the multi-dimensional risks in higher education posed by AI technologies.

1. No statutory Definition of AI use in education

India has numerous statutes related to early childhood education and higher education, such as The Right of Children to Free and Compulsory Education Act 2009[28], The University Grant Commission Act 1956[29], The All-India Council for Technical Education, 1987 and various other legislations. It is pertinent to note that none of these statutes defines what " Artificial Intelligence " is. India, till date, does not have any laws related to Artificial Intelligence. Taking into consideration the increasing use of AI in higher education, the term "Artificial Intelligence" not being defined is a major legal gap and a grave concern.

2. Data Privacy and Student Surveillance Concerns

India's recognition of the right to privacy as a fundamental right has evolved gradually through judicial interpretation. The initial consideration began with the Kharak Singh v. State of Uttar Pradesh[30] (1962), where the Supreme Court examined privacy within the constitutional framework but did not expressly recognise it as a fundamental right. Subsequently, the jurisprudence remained

[21] Supra 12
[22] The University Grants Commission Act, 1956, No 3, Acts of Parliament,1956 (India).
[23] The Institute of Technology Act,1961, No.59, Acts of Parliament,1961(India).
[24] The All-India Council for Technical Education Act, 1987, No.52, Acts of Parliament, 1987 (India).
[25] The Central Universities Act,2009, No.25, Act of Parliament, 2009(India).
[26] The All-India Council for Technical Education Act, 1987, No.52, Acts of Parliament, 1987 (India).
[27] The Central Universities Act,2009, No.25, Act of Parliament, 2009(India).
[28] The Right of Children to Free and Compulsory Education Act 2009, No.35, Act of Parliament, 2009(India).
[29] The University Grants Commission Act, 1956, No 3, Acts of Parliament,1956 (India).
[30] Kharak Singh v. State of Uttar Pradesh, 1963 AIR 1295.

unsettled for several decades. It was only in the landmark decision of Justice K.S. Puttaswamy v. Union of India[31] (2017) that the Supreme Court conclusively affirmed the right to privacy as a fundamental right under Article 21[32] of the Constitution, thereby strengthening the constitutional protection of individual autonomy and dignity.[33] The use of AI in higher education is hindered by a lack of a regulatory framework that fails to protect the personal data that is uploaded by the students, the behavioural pattern of students, whether it is a self-generated IT code or research that is yet to be published, or data of students being prepared by the teachers for effective classroom teaching in a way it is violation right to privacy. The Digital Personal Data Protection Act 2023[34] does not define "Artificial Intelligence", thereby making digital privacy with respect to Artificial Intelligence an uncodified and unregulated legal space and giving rise to ambiguity, uncertainty, and legal complications.

1. Ambiguity with respect to Ownership

One major challenge in the academic sector that has arisen due to the emergence of "AI" is recognition of ownership of a work which has been generated either by "AI" or "LLM" (large Language Model). It is pertinent to note that this recognition is essential from the point of view of promoting accountability and recognition of relevant intellectual property rights. It is pertinent to note that the Copyright Act 1999[35] does not define "Artificial Intelligence", along with the same another significant gap which arises because the statute does not classify whether the person directing the Artificial Intelligence would be considered as the real owner of the work or the artificial intelligence itself. This legal gap, if it remained unsettled, would lead to significant disputes with respect to ownership in the years to come, with the emerging nature of AI.

2. Ambiguity with respect to Liability on use of "AI"

Along the ambiguity with respect to ownership comes ambiguity with respect to liability. The existing legal framework with respect to higher education fails not only to define "AI" but also fails to address any misuse or misinterpretation based on data given by "AI". The legal gap which does pertinently comes up is who is to be held liable in case of any misuse of "AI"? The law fails to answer whether it would be the person who used the "AI" or the ones who operate or deal with the "AI" software directly, i.e., in most cases, a company. This legal gap is a very crucial gap to be addressed.

3. Lack of Institutional Framework on "AI"

Taking into consideration the failure of Indian Statutes with respect to defining "Artificial Intelligence:" This issue does give rise to another issue which Higher education universities face which is no standard framework on "AI" usage in academic submission, thereby leaving this completely discretionary on institutions to impose their own internal framework on "AI". This has created non-uniformity and extremely uncertain application of "AI" restrictions on academic work. This has also significantly affected the credibility of "AI".

Recommendations/ Suggestions

This current problem is a very grave problem; the solution for the same largely would depend on implementation rather than the policy itself. The solution that the authors of this paper would like to recommend is as follows:

- Amending the existing legal framework with respect to Higher education in such a way that a specific chapter must deal with "Artificial Intelligence", thereby harmonising the same with the emerging "AI".
- The Digital Personal Data Protection Act 2023, along with the laws about Intellectual property rights, specifically the Copyright Act 1999, must have specific provisions addressing the

[31] Justice K.S. Puttaswamy v. Union of India, 2019 (1) SCC 1.
[32] INDIA CONST, art21.
[33] R Aksietha, Surveillance in India and Its Privacy Challenges in the Digital Age: A Legal and Constitutional Analysis,(Volume 10),(a653),(2025)
[34] The Digital Personal Data Protection Act 2023, No.22, Act of Parliament, 2023(India).
[35] The Copyright Act 1999, No.49, Act of Parliament, 1999(India).

aspects of ownership and liability, specifying at what instances a person using the AI would be liable and at what instances would the "AI" itself be liable.

- The University Grants Commission Act 1956 must first be amended so that UGC would have additional powers to release exclusive rules about institutional guidelines about AI with respect to Higher education, which is the need of the hour.
- A special high-level committee must be set up by the government to develop a uniform definition of "AI" that would be applied in all the statutes uniformly.

Conclusion

There is a significant need to develop and modify the existing laws with respect to higher education based on the emergence of "AI". Considerable delay in the same would lead to further complications and even an increase in academic disputes. It would also add additional pressure to the statutory bodies governing higher education.

References / Bibliography

I. Statutes / Acts

- All-India Council for Technical Education Act, 1987, No. 52, Acts of Parliament, 1987 (India).
- Central Universities Act, 2009, No. 25, Acts of Parliament, 2009 (India).
- Copyright Act, 1999, No. 49, Acts of Parliament, 1999 (India).
- Digital Personal Data Protection Act, 2023, No. 22, Acts of Parliament, 2023 (India).

Institute of Technology Act, 1961, No. 59, Acts of Parliament, 1961 (India).

- Right of Children to Free and Compulsory Education Act, 2009, No. 35, Acts of Parliament, 2009 (India).
- University Grants Commission Act, 1956, No. 3, Acts of Parliament, 1956 (India).

II. Constitutional Provisions

- INDIA CONST. art. 21.
- INDIA CONST. art. 21A.
- INDIA CONST. art. 45.
- INDIA CONST. art. 46.

III. Case Law

- *Justice K.S. Puttaswamy v. Union of India*, (2019) 1 SCC 1 (India).
- *Kharak Singh v. State of Uttar Pradesh*, AIR 1963 SC 1295 (India).

IV. Reports, Articles, and Online Sources

- Bhat & Joshi Associates, *Regulatory Framework of Education in India: A Comprehensive Legal Analysis* (last visited Apr. 17, 2026).
- Distance Learning Institute, *The Legal Framework of Higher Education in India* (last visited Apr. 17, 2026).
- Kamal & Co. Advocates, *Education Law In India* (last visited Apr. 17, 2026).
- Kanjolia, Shivi, *Transforming Innovation Into Fairness: Tackling Legal Challenges In AI-Enhanced Education In India*, 5 VOL. L. REV. 816 (2025).
- Press Information Bureau, *Press Release on AI Investment in India* (last visited Apr. 16, 2026), https://www.pib.gov.in/PressReleasePage.aspx?PRID=2178092.
- Stratx Simulation, *The Negative Effects of Artificial Intelligence in Education* (last visited Apr. 17, 2026).

"Human Creativity Vs Machine Creativity: Implications for IP Law"

Ankita Bhoir Chandane
Affiliation – Bhagubhai Changu Thakur College of Law, New Panvel
Department of Law, University of Mumbai

Abstract

The swift development of artificial intelligence has transformed disrupted conventional understandings of creativity and authorship within intellectual property law. This paper adopts a doctrinal approach to examine the distinction between human creativity and outputs generated by machines, with particular emphasis on its implications for copyright law. Existing legal frameworks, including the Copyright Act, 1957, are primarily built on the premise the idea that works of art are human intellect, involving elements such as skill, judgment, and deliberate expression. However, the rise of advanced AI technologies capable of producing artistic, literary, and musical content questions regarding whether copyright is these traditional standards remain applicable.

By analysing statutory provisions, judicial decisions, and international perspectives, this study demonstrates the limitations of current legal doctrines in addressing AI-generated works. Courts have largely refused to recognize non-human authorship, thereby excluding AI-created outputs from standard copyright protection. This position leads to uncertainty regarding ownership, liability, and the extent of legal protection, particularly in situations where human involvement is limited or indirect. In addition, the dependence of AI systems on existing copyrighted material for training complicates the determination of originality and introduces concerns about possible infringement.

The paper contends that although human creativity continues to underpin the theoretical basis of intellectual property law, the increasing influence of artificial intelligence calls for a reassessment of established legal principles. It suggests that a balanced solution achieved through judicial interpretation or legislative reforms may be necessary to address emerging forms of collaborative creativity between humans and machines. Ultimately, this study aims to contribute to the evolving discourse on adapting intellectual property laws to balance the protection of human creativity in relation to the development of technological innovation.[1]

Keywords: Artificial Intelligence (AI) , Intellectual Property Rights (IPR) , Human Creativity, Machine-Generated Works , Copyright Law, Authorship , Legal Reform

Research Objectives

The research topic "Human Creativity vs Machine Creativity: Implications for IP Law" examines how AI-generated creative works challenge traditional intellectual property laws based on human authorship. Existing laws, including the Copyright Act, 1957, were designed to protect human creativity through concepts such as originality and authorship. However, AI can now independently generate music, art, literature and designs, creating uncertainty about ownership and copyright protection.

The research problem focuses on whether current IP laws are adequate to regulate AI-generated works, whether AI can be recognized as an author, and whether legal reforms are necessary to address evolving forms of machine creativity while preserving the human-centered foundation of copyright law.

Research Questions

The research paper is guided by three research questions

RQ1: To what extent do existing intellectual property laws adequately address the distinction between human creativity and AI-generated outputs?

RQ2: Should artificial intelligence be recognized as an author under current copyright frameworks, or should authorship remain exclusively human-centric?

[1] For example, generative AI models like GPT-4 and DALL·E have already produced works resembling human creativity.

RQ3: Is there a need for reform in intellectual property laws to accommodate emerging forms of creativity involving artificial intelligence?

Scope and Limitations

The study focuses on the implications of Artificial Intelligence on Intellectual Property law, specifically, copyright law under the Copyright Act, 1957. It examines the distinction between human creativity and machine-generated creativity and analyses issues relating to authorship, originality, ownership, and protection of AI-generated works. The research also considers international legal approaches and the need for reforms in existing copyright frameworks to address emerging technological developments.

However, the study is limited to a doctrinal legal analysis and does not extensively examine the technical functioning of AI systems. Its primary focus is on copyright law, with limited discussion on other areas of IP law such as patents and trademarks. Additionally, as AI technology and related regulations are continuously evolving, future legal developments may influence the relevance of the study's findings and conclusions**.**

Research Methodology

This research employs a doctrinal (black-letter law) approach to analyse legal issues related to human and AI-generated creativity. It relies on secondary sources such as statutes, case laws, journal articles, and policy reports. Key provisions of the Copyright Act, 1957 and relevant legal principles on authorship and originality are examined. The research is analytical and interpretative, focusing on identifying gaps in existing laws and the need for reform.

Literature Review

Below are the five credible sources of literature review.

1. WIPO Technology Trends Report on AI - Artificial Intelligence report by World Intellectual Property Organization provides a comprehensive overview of global developments in AI and their impact on intellectual property systems. It highlights key challenges related to patents, copyright, and policy gaps, and provides useful perspectives on how innovation is reshaping legal frameworks.
2. Artificial Intelligence and Copyright" – Pamela Samuelson - In her work on artificial intelligence and copyright, Pamela Samuelson examines critical issues such as the concepts of authorship, originality, and the legal status of AI-generated content. Her analysis is widely regarded as authoritative and contributes significantly to understanding the doctrinal challenges posed by AI.
3. "Who Owns AI-Generated Content?" – Harvard Journal of Law & Technology-Articles published in the Harvard Journal of Law & Technology, including discussions on ownership of AI-generated content, provide in-depth examination of authorship and ownership issues. These works combine doctrinal analysis with policy considerations, have become increasingly relevant to legal research.
4. UKIPO Consultation on AI and IP - The UK Intellectual Property Office, through its consultation paper Intellectual Property and Artificial Intelligence: Policy Consultation, offers a governmental perspective regarding the requirements for legislative reform. It addresses practical challenges and explores policy responses to regulate AI-driven creativity.
5. "The Creativity of Machines" – Ryan Abbott - Ryan Abbott, in The Creativity of Machines, argues for the recognition of AI-generated with regards to inventions in intellectual property law. His work is influential in advancing debates as regards the issue of AI systems should be acknowledged as inventors or authors.

Introduction[2]

The swift progress in the field of artificial intelligence has led to considerable changes in various sectors, including intellectual property law. Modern AI systems are capable of generating creative

"…Courts in many jurisdictions have generally refused to recognize AI as an author…" [2]

[2] The U.S. Copyright Office explicitly stated in 2023 that works containing only AI-generated material are not registrable.

works such as text, visual art, music, and software, thereby challenging traditional legal concepts of creativity and authorship. Intellectual property regimes have historically been founded on the belief that creativity is a uniquely human trait, involving originality, skill, judgment, and intentional expression. Legal frameworks, including the Copyright Act, 1957, reflect this perspective by recognizing only natural persons as authors.

However, the increasing autonomy of AI in content creation raises important doctrinal concerns. The distinction between human creativity and machine-generated output is becoming less clear, leading to uncertainty regarding authorship, ownership, and liability. Courts in many jurisdictions have generally refused to recognize AI as an author, thereby excluding purely machine-generated works from copyright protection. While this preserves the human-centered intellectual property law's basic principles, it does not adequately address the growing significance of AI-generated content.

Moreover, the reliance of AI systems on large datasets, often containing copyrighted material, complicates issues of originality and raises potential infringement concerns. Accordingly, there is need to reassess intellectual property laws to ensure their continued relevance. This paper adopts a doctrinal approach to examine the distinction between human and machine creativity and with repercussions for the intellectual property law, while also exploring the need for legal reform in response to evolving technological realities.

Conceptual Framework[3]

Human creativity is rooted in originality and personal expression, while AI creativity often relies on patterns and data-driven outputs. Intellectual property law struggles to balance protecting human authorship with regulating AI-generated works. Let's delve deep into each analysis one by one.

Three Research Questions with respective Observed Answers

RQ1: To what extent do existing intellectual property laws adequately address the distinction between human creativity and AI-generated outputs? **RA1**: A structured doctrinal comparison helps clarify the differences in legal treatment of human creativity and AI-generated content under current intellectual property frameworks.
RQ2: Should artificial intelligence be recognized as an author under current copyright frameworks, or should authorship remain exclusively human-centric? **RA2**: Evaluating the status of AI authorship requires a close examination of established doctrines such as authorship, originality, and legal personality.
RQ3: Is there necessity for revision of the intellectual property laws to accommodate emerging forms of creativity involving artificial intelligence? **RA3**: As new forms of creativity emerge through artificial intelligence, established legal doctrines face increasing pressure.

A tabular analysis which delves deep into intellectual property law with human creation and AI Generated Works

I) RA1 : A doctrinal comparison to clarify differences between Human Created works and AI Generated Works[4]

Doctrinal Aspect	**Human-Created Works**	**AI-Generated Works**	**Critical Doctrinal Observation**
Authorship	Copyright law attributes authorship to a natural person who creates the work	AI systems are not recognized as legal persons and therefore cannot be authors	The law clearly excludes non-human creators but does not provide a substitute framework

"…AI creativity often relies on patterns and data-driven outputs…"[3]

[3]This reflects the distinction between human originality and algorithmic recombination.

"…AI systems are trained on extensive datasets, sometimes including copyrighted content" …[4]

[4]This issue is central to ongoing disputes about fair use and reproduction doctrines.

Doctrinal Aspect	Human-Created Works	AI-Generated Works	Critical Doctrinal Observation
Legal Recognition	Affirmed by institutions such as the U.S. Copyright Office and the European Union Court of Justice	No formal legal status is rights attributed to AI as a creator	The framework is consistent but limited to human-centered creation
Originality Requirement	Requires independent creation involving human skill, judgment, or effort	AI outputs may appear novel but lack human intellectual contribution	A gap exists between technological originality and legally recognized originality
Ownership	Rights generally vest in the human creator or employer under established doctrines	Ownership remains uncertain—may involve users, developers, or no clear owner	This represents a major area of legal uncertainty
Statutory Framework	Clearly regulated under established copyright laws	Addressed only partially, for example in the Copyright, Designs and Patents Act 1988 regarding computer-generated works	Existing statutory provisions are often vague and outdated in the AI context
Intent and Control	Creation involves deliberate intellectual input and decision-making by the author	AI operates through programmed algorithms without conscious intent	Traditional doctrines do not align well with non-intentional creation
Nature of Creation Process	Typically follows a direct process from idea to expression and fixation	Involves complex processes such as data training, pattern recognition, and automated generation	Legal doctrines were not designed for such technologically mediated processes
Use of Pre-existing Works	Generally, involves identifiable and limited use of prior materials	The AI software is designed extensive datasets, sometimes including copyrighted content (e.g., disputes involving OpenAI and Getty Images)	Raises unresolved issues under reproduction and fair use doctrines
Protection Outcome	Eligible for full copyright protection, including economic and moral rights	Protection depends on the extent of human involvement; fully autonomous outputs are typically unprotected	Results are inconsistent and vary across jurisdictions
Doctrinal Stability	Based on long-established and coherent legal principles	Increasingly challenged by technological developments in AI	The framework remains theoretically sound but practically strained

The comparison illustrates that intellectual property law continues to rest on a human-centered understanding of creativity, ensuring internal consistency in its treatment issues surrounding authorship and originality. However, the emergence AI-generated texts in terms exposes significant doctrinal limitations, particularly ownership in the sense, creative contribution, and extent of application protection. As a result, while the distinction between human of AI-generated and AI-produced is formally maintained, its practical application is increasingly uncertain and calls for doctrinal refinement.

I) **RA2: Evaluating the status of AI authorship requires a close examination of established doctrines such as authorship, originality, and legal personality. The following distinction outlines how these principles operate under a human-centered model and how they would be affected by identifying AI as a author.[5]**

"…AI lacks legal personality and cannot exercise rights or duties…"[5]

[5]Legal personality is a prerequisite for authorship under copyright law.

Doctrinal Aspect	**Human-Centric Authorship (Status Quo)**	**Acknowledgement of AI as Author**	**Doctrinal Evaluation**
Legal Status of Author	Authorship is confined to natural or legal persons under existing copyright law	Would require extending authorship to non-human entities	Current doctrine firmly excludes AI; recognition would require fundamental legal reform
Institutional Position	Supported by authorities such as the U.S. Copyright Office and interpretations of the Court of Justice of the European Union (CJEU)	No jurisdiction currently grants full authorship status to AI	Strong doctrinal consensus favours human authorship
Framework of Originality	Rooted in human intellectual effort, judgment, and creativity	Would redefine authenticity to include machine-generated outputs	Expanding originality risks diluting its human-centered meaning
Ownership Structure	Clear allocation of rights to authors or employers	Ownership would be unclear (AI cannot hold property; rights may shift to developers or users)	Recognition of AI complicates established ownership doctrines
Legal Personality	Authors must have legal personality to hold and enforce rights	AI lacks legal personality and cannot exercise rights or duties	Major doctrinal barrier to recognizing AI as an author
Intent and Creativity	Requires conscious intention and creative control by a human	AI operates without intent or awareness, relying on algorithms	Challenges the philosophical and legal basis of authorship
Accountability and Liability	Human authors can be held accountable for infringement or misuse	AI cannot bear liability; responsibility must be reassigned to humans	Creates enforcement and accountability gaps
Statutory Interpretation	Existing laws (e.g., Copyright, Designs and Patents Act 1988) indirectly accommodate non-human outputs without granting AI authorship	Would require reinterpretation or amendment of statutory provisions	Current statutes are not designed for autonomous creators
Policy Justification	Protects and incentivizes human creativity and innovation	May encourage technological development but weakens human-centric incentives	Normative tension between innovation and human reward
Doctrinal Stability	Maintains consistency with long-established copyright principles	Would significantly alter foundational doctrines	Risk of doctrinal disruption and uncertainty

Expanding on your point: granting authorship to AI would disrupt the human-centered foundation of copyright law, which relies on moral rights and accountability tied to legal persons. It may also lead to uncertainty in in establishing who owns AI-generated works developers, users, or the AI system itself . Enforcement would be complicated by the lack of human accountability, raising questions on liability for infringement or misuse. Ultimately, such a shift risks destabilizing the coherence of copyright doctrine while opening contentious debates about the balance between innovation and legal certainty.

II) RA3: As new forms of creativity emerge through artificial intelligence, established legal doctrines face increasing pressure. A structured doctrinal analysis provides insight into whether adaptation or reform is necessary.[6]

Doctrinal Aspect	Human-Oriented Framework (Current Position – Limited Reform Required)	AI-Adaptive Approach (Argument Supporting Reform)	Doctrinal Assessment
Underlying Principle	Focuses on safeguarding human intellectual effort and creative expression	Aims to accommodate new forms of creativity emerging from AI systems	Existing principles remain logically consistent but may not fully capture technological developments
Authorship Concept	Confines authorship to human creators	Proposes reconsideration or expansion of authorship in AI-related contexts	Reform may be needed to address ambiguity in AI-assisted works
Institutional Position	Reflected in positions of bodies such as the U.S. Copyright Office and interpretations by the Court of Justice of the European Union	Growing international discussions on adapting legal frameworks to AI innovation	Current doctrine is stable but increasingly challenged
Standard of Originality	Based on human judgment, skill, and independent effort	May require rethinking originality to include outputs generated through algorithms	Any reform must carefully preserve the core meaning of originality
Ownership Structure	Provides clear allocation of rights to creators or employers	Calls for new mechanisms to determine ownership in AI-related creations	Uncertainty in this area strongly supports the case for reform
Legal Personality	Rights are limited to entities with legal recognition	AI lacks legal standing; reform may involve assigning rights to associated human actors	Indirect attribution models appear more feasible than recognizing AI directly
Extent of Protection	Protects works only where human contribution is evident	Suggests extending protection to certain AI-generated outputs	Reform may be necessary to avoid gaps in protection
Technological Adaptability	Designed to be flexible and technology-neutral	AI challenges this neutrality due to its autonomous and generative capabilities	Indicates limits of existing doctrinal flexibility
Use of Training Data	Traditional laws do not directly regulate large-scale data training practices	Highlights the need to address use of copyrighted material in AI training	Presents a strong and immediate need for legal clarification
Policy Considerations	Prioritizes protection and incentive for human creators	Expands focus to include technological innovation and AI development	Reform must balance competing policy goals
Doctrinal Consistency	Maintains alignment with established legal principles	Reform may introduce change but can ensure continued relevance	Gradual and targeted reform is preferable to preserve coherence

The above comparison shows that whereas current intellectual property laws remain grounded in a coherent human-centered framework, there is increasing pressure on the realities of AI-driven creativity. Although the foundational structure does not need to be complete replacement, specific

"…Use of Training Data…highlights the need to address use of copyrighted material in AI training…"[6]

[6]Current copyright law does not directly regulate large-scale machine learning datasets.

areas particularly authorship, ownership, and data usage demonstrate a clear need for measured and well-defined reform to maintain doctrinal relevance.

Findings/Results[7]

The findings and observations of each study goals are enumerated below with respect to research query.

Q1 - To what extent do existing intellectual property laws adequately address the distinction between human creativity and AI-generated outputs?

The extent to which existing intellectual property laws adequately distinguish between human creativity and AI-generated outputs remains a contested issue. While the law clearly prioritizes human authorship and originality, thereby establishing a strong doctrinal distinction in principle, it struggles addressing the complexities of AI-generated and hybrid works. Problems like unclear authorship thresholds, ownership ambiguity, and the usage of training datasets highlight significant gaps. Consequently, although the framework appears doctrinally coherent, it is increasingly strained in its practical application.

Q2 - Should artificial intelligence be recognized as an author under current copyright frameworks, or should authorship remain exclusively human-centric?

The problem of whether artificial intelligence needs to be classified as an author further exposes the constraints of the present doctrine. Copyright law is fundamentally based on human creativity, intention, and legal personality, all of which AI lacks. Granting authorship status to AI would challenge these foundational principles and create uncertainty in areas such as ownership and liability. Therefore, it is more consistent with the current law to retain a human-centric approach, treating AI as an instrument rather than an independent author.

Q3 - Is there a need for reform in intellectual property laws to accommodate emerging forms of creativity involving artificial intelligence?

Finally, the need for reform in intellectual property laws arises from the growing influence of AI in creative processes. Although the current framework remains structurally sound, it doesn't fully account for the issues related to authorship, ownership, and data usage. This suggests that while a complete overhaul is unnecessary, targeted and carefully balanced reforms are required. Ultimately, the challenge lies in maintaining doctrinal stability while ensuring that the law evolves to remain relevant in the face of technological advancement.

Overall Analysis

From the study, it is evident that the issue of intellectual property law remains **conceptually strong but operationally challenged**. The differentiation between human and AI creativity exists in principle, but its application is increasingly uncertain, making **measured doctrinal reform both essential and inevitable**.

Recommendations/Suggestions[8]

1. Clarify the Standard of Human-Originated Authorship

There should be well-defined legal guidelines defining the **minimum level of human contribution** required to qualify for copyright protection in AI-assisted works. This would reduce uncertainty and ensure consistent application of the law.

2. Develop a Framework for AI-Assisted Works

Instead of recognizing AI as an author, laws should introduce a separate category for **AI-assisted creations**, where rights are granted to the human who exercises meaningful creative control.

3. Establish Clear Ownership Rules

Legislation should specify ownership in AI-generated outputs by:

- Assigning rights to users, developers, or both, based on contribution

---

"…ownership ambiguity and the use of training datasets highlight significant gaps…"[7]

[7]Ambiguity arises because neither developers nor users are clearly recognized as owners of AI outputs.

"…AI-assisted creations, where rights are granted to the human who exercises meaningful creative control…"[8]

[8]This approach mirrors the UK Copyright, Designs and Patents Act 1988 provision on computer-generated works.

- Providing default rules to avoid disputes

This would address a major deficiency in current doctrine.

4. Regulate the Training Data Usage

There is a requirement to create specific provisions governing:

- Use of copyrighted material in AI training
- Licensing mechanisms or fair use standards

This would ensure a balance between innovation and protection of existing rights.

5. Introduce Transparency and Disclosure Requirements

AI-generated content should be subject to **disclosure obligations**, requiring users or developers to indicate when a work is AI-generated or AI-assisted. This would improve accountability and legal clarity.

6. Promote International Harmonization

Given the global nature of AI, there should be efforts to:

- Align IP laws across jurisdictions
- Develop international guidelines or treaties

This would prevent fragmentation and legal inconsistencies.

7. Adopt a Gradual and Flexible Reform Approach

Rather than overhauling the entire system, lawmakers should adopt:

1. Incremental reforms
2. Technology-neutral principles

This ensures that the law remains stable while adapting to innovation.

Conclusion[9]

In conclusion, intellectual property law continues to rest on a strong human-centered foundation, emphasizing authorial contribution, originality, and accountability. However, the swift evolution of artificial intelligence has highlighted significant gaps, particularly as far as the issues of authorship, ownership and data usage. While the existing framework fails to complete replacement, it is not enough anymore to solve all the problems brought by AI. Carefully targeted and balanced reforms are therefore essential to preserve doctrinal coherence while ensuring that the law is still relevant and effective in a technologically evolving landscape.

References/Bibliography

Below are the references bibliography used in this research paper

1. Abbott, R. (2020). *The Reasonable robot: Artificial intelligence and the law*. Cambridge University Press.
2. European Parliament. (2020). *Intellectual property rights for the development of artificial intelligence technologies*. Publications Office of the European Union.
3. Guadamuz, A. (2017). Artificial Intelligence and Copyright. *WIPO Magazine*. Retrieved from https://www.wipo.int/wipo_magazine/en/2017/05/article_0003.html
4. (wipo.int in Bing)
5. Harvard Law Review. (2018). Note: Copyright and artificial intelligence. *Harvard Law Review*, 131(6), 1730–1755.
6. Samuelson, P. (2018). Artificial intelligence and copyright. *Journal of the Copyright Society of the USA*, 65(2), 313–340.
7. U.S. Copyright Office. (2023). *Copyright registration guidance: Works containing material generated by artificial intelligence*. Washington, DC: U.S. Copyright Office.
8. World Intellectual Property Organization. (2019). *WIPO conversation on Intellectual Property (IP) and Artificial Intelligence*. Geneva: WIPO

"…Carefully targeted and balanced reforms are therefore essential…"[9]

[9] Incremental reform is preferable to wholesale doctrinal change, ensuring stability while adapting to innovation.

Digital Surveillance in the Age of Artificial Intelligence: Examining Its Impact on Youths' Human Rights

Shanita Singh [1]
Dr. Rishi Yadav[2]

Abstract

AI has found its place in the world of social media. Where previously there were basic tools available, they have evolved to become complex systems that apply psychological pressure on children and teenagers and allow bad actions to be automated and optimised. When AI technology is involved it poses more threats, such as cyberbullying, grooming, and extortion, etc.,

In a nutshell, AI's presence in social media is no longer a helpful tool but seems like a powerful weapon. Underneath the technical talk, the situation can be described as a technological arena being used as a weapon against children. AI becomes a Digital Weapon, creating artificial shame (through deepfakes), isolating (with predictive algorithms), and manipulating (using generative AI) young people. Currently, we can observe a clear conflict between the modern technology and the safety of humans.

This research paper aims to examine the threat posed by recent developments: the use of AI-based blackmail gangs for targeting children aged 10 to 17, and the development of artificial intelligence-backed hate mobs on Discord aimed at young people between 18 and 24 years old, which is a pattern identified in CDC and UNICEF studies as being linked to increased ideation of self-harm. We have entered an intense battle using high-tech warfare weapons armed with low-tech tactics. While posting warnings on posts might seem like an adequate measure, it simply does not suffice.

Keywords: Digital Surveillance, Child mental abuse, Youth psychological harm, Algorithmic grooming, Platform accountability, Digital youth protection and Human Rights.

Introduction

India is home to the most populous society in the world, where the highest number of youth and children resides. With the passing time, technology has become an integral part of the lives of young individuals who get access to a broad spectrum of devices that have the potential of providing them benefits but also pose certain threats at the same time. Various websites and applications allow children to enhance their knowledge and motivate them to achieve their objectives. AI features in social media can be useful in expanding children's vision and developing their creative abilities. They provide the opportunity of using language learning programs, storytelling games, and cooperative projects through these applications. However, there should be a rational approach to such apps and parents' or teachers' involvement.

The development of technology has transformed children's lives from its peripheral positions to those at the core. The emergence of popular social media networks including YouTube, Instagram, Snapchat, Facebook, TikTok, Telegram, Twitter influences the cognitive development of children and makes it difficult for some users to break away from socializing online.

The integration of artificial intelligence, in the form of recommendation engines, chatbots, and other generative technologies, has become the foundation of the digital era, when combined with social media, can have a direct impact on children, as this combination does not just involve spending hours in front of screens, but influences their way of thinking and communicating. while spending more time in front of the computer, but rather a concern over the possible changes that this experience might cause in terms of mood, perception of oneself, and the establishment of relationships. Technologies such as AI and social media are beneficial in the form of enhanced learning capabilities,

[1]Shanita, Phd Research scholar, Shobhit UniversitySchool of Law and Constitutional Studies,Shobhit Institute of Engineering & Technology, Meerut
[2]Dr. Rishi Yadav(Assistant Professor)School of Law and Constitutional Studies,Shobhit Institute of Engineering & Technology, Meerut

broadened horizons, and more means for artistic expression, but pose dangers regarding mental well-being, ways of thinking, privacy, and safety issues.

The area of cyber crime has evolved with the development of technology and increasing reliance on computerized space for personal and commercial use. With the advancement in the realm of cyber crime from simple hacking and virus creation to the more sophisticated crimes including ransomware and identity theft, the scope of crimes committed through computer networks has broadened substantially. Criminal techniques include phishing, social engineering, exploit kits, among others, resulting in substantial monetary and reputational damage. Complications arise from the nature of internet being a worldwide network; it enables criminals to operate from virtually any location.[3]

Objective

The primary objective of this study is to critically examine the psychological, social, and developmental impacts of AI-driven social media platforms on children and adolescents, with particular attention to Mental health outcomes (e.g., anxiety, depression, body dissatisfaction), Risk behaviours (e.g., unsafe online interactions, problematic smartphone/gaming use) and Privacy and safety challenges unique to youth as a vulnerable population.

Material and Methods

The study uses a qualitative method, involving case studies to investigate the influence of artificial intelligence (AI) and social media usage on children and teenagers. Secondary sources are used in the study, which is based on research articles, surveys, and This study aims to analyze case studies that address mental health issues among adolescents. social media usage, and the adoption of AI in Indian regions (Tamil Nadu, Bihar, and Raipur) and other countries' youth population. The selected cases are intentional, with specific information regarding the age of participants, locations, nature of platform utilization and outcomes such as mental health problems, learning, cyberbullying, and AI utilization behaviors.

The study used a qualitative analysis approach with descriptive and thematic techniques in five categories: psychological effects such as anxiety, depression, and self-esteem; behavioral effects like time spent online, sleeping habits, and offline interaction; social effects regarding relationships with friends and family; cognitive effects including attention and critical thinking; and contextual effects related to culture, access to technology, and socioeconomic status. This method offers a detailed look at how artificial intelligence and social media affect teenagers' lives by using secondary data without performing a primary investigation. This study is subject to several limitations, including its reliance on previous findings, sampling bias, and low statistical generalizability.

Discussion and Study

The mind of children are so vanurable of age can be easily manipulated and influenced through social media which plays one of the important role to dominate the essential character of a child. This cause can affect the several major public health concern and its rising the serious disorders such as attention deficit hyperactivity disorder (ADHD), autism spectrum disorder (ASD), anxiety disorders, addiction, and major depressive disorder (MDD). To understand the cause of this trend many studies were conducted by Clinicians and researchers and with this many factors includes such as with increasing attention on environmental, genetic, societal factors and rapid expansion of digital technology and social media. Today, children and teenagers are spending hours on platforms like Instagram, TikTok, and Snapchat. Their social interactions, self-image, and brain development can be easily influenced. While these digital spaces offer chances for connection and self-expression, they also bring serious challenges. These include higher exposure to cyberbullying, unrealistic beauty standards, and compulsive content consumption.[4]

[3]Laura DeNardis, *The Global War for Internet Governance* (Yale Univ. Press 2014).

[4]T. Masri-Zada, S. Martirosyan, A. Abdou, R. Barbar, S. Kades, H. Makki, G. Haley & D.K. Agrawal, The Impact of Social Media & Technology on Child and Adolescent Mental Health, 9 J. Psychiatry & Psychiatric Disorders 111 (Apr. 16, 2025).

The current research seeks to explore the consequences of artificial intelligence in combined with social media usage on children and youth. The study will focus on the ways the inclusion of Artificial intelligence (AI) integrated into social media through its platforms has made it possible to very difficult for the proper development of children and adolescents. Algorithms aimed at ensuring user engagement end up playing with the minds of youth, resulting in addiction and obsession for the platforms. In most cases, the information offered does not enhance learning but rather negatively influences the user's mental health.

There is a lot of stress for the social approval approval via likes and comments. Social media, supported by AI, promotes shallow interactions and discourages meaningful ones. This results in children feeling more lonely despite their constant digital connectivity. Additionally, privacy disappears as AI uses children's personal information. Instead of being empowering, AI-driven social media which has become a platform to control its users which may tend to harm the users. It distorts self-image, undermines resilience, and puts young people at risk. They are not prepared to handle these dangers. Without strong action, these platforms will keep putting profit ahead of safety, compromising the well-being of an entire generation.

Social connection, Health Information, Cognition, and Critical Thinking

However, the use of Generative AI cannot entirely imitate human interaction, and Generative AI usage might even substitute real human interaction. One-third of teenagers utilize Generative AI for social interactions. While AI chatbots may serve as a trustworthy source of health information that is convenient for people to access, and there are improvements in the Body Mass Index, weight, and body fat percentage of teenagers through its use in nutrition education, AI may mislead teenagers with faulty health information and limit critical thinking, considering teenagers depend more on AI applications than their thoughts and concepts, hindering cognitive growth among teenagers. The training data used may consist of misinformation online, resulting in Generative AI models providing teenagers with inaccurate health information and reinforcing cognitive prejudices that teenagers have towards themselves and others[5]. There are certain factors and outcome for it that can be classified as below-

Body Image

Both Social Media and AI has change the perception on the teenagers. Ai is capable of producing very realistic body images that cannot be realistically achieved and lack diversity in shape, which could contribute further to body image issues. However, using AI to improve one's own photos could help increase body satisfaction.

Physical Activity and Sleep

Sleep and physical activity are vital to adolescents' health and development. However, AI use may displace sleep and physical activity by increasing adolescents' screen time and sedentary behaviour, but also be beneficial to physical activity and sleep through its potential incorporation into health education tools.[6]

Recently there is a Pressure faced by UK government to implement curfews on teenagers and implementing a legal ban on smartphones in schools social media for adolescents under the age of 16 because of mounting the evidence for generating the sexually explicit images of women and children by the tools and app like xAI's tool, Grok. Another countries has recently ban the usage of social media and this ban is recently implemented in Australia, and similar bans are is increasingly being examined by the different governments worldwide. This push towards banning is based on the premise that social media is causing actual damage. Despite the fact that numerous parents and teachers fear for the ill effects caused on children, those criticizing social media dismiss all concerns as merely another moral panic. I think that the guardians concerns regarding exposure of dangerous content by

[5]J.M. Nagata, Z. Memon, O. Huang & M.A. Moreno, Adolescent Health and Generative AI—Risks and Benefits, 180 JAMA Pediatr. 7 (Jan. 1, 2026), https://doi.org/10.1001/jamapediatrics.2025.4502. (doi.org in Bing)

[6]J.M. Nagata, Z. Memon, O. Huang & M.A. Moreno, Adolescent Health and Generative AI—Risks and Benefits, 180 JAMA Pediatr. 7 (Jan. 1, 2026), https://doi.org/10.1001/jamapediatrics.2025.4502. (doi.org in Bing)

children have some legitimacy, since there really are harmful online content of sexual, wellness, cyberbullying, misogynistic, extremist, and misleading data.[7]

In 2025, a Research was conducted in two Chinese cities to examine the usage of social media among 7184 adolescents of age between 13-21 where, the study links smartphone uses and problematic online behaviors. POBs include excessive use of smartphones, video games, and social media. This research was cross-sectional and used the validated psychometric tools. Network analysis was applied to assess symptom associations for each of the POBs. The symptoms found were escapism, withdrawal, and loss of control. The results are significant for psychologists, clinicians, and policymakers to address these mental health problems in this digital age[8]

In 2024, a study identified characteristics and patterns within 674 labelled private messages on the platform of social media like Instagram, intending to investigate risk. In this context of anonymity, hurtful opinions affecting the person involved were found. Five types of media content were analyzed: memes, screenshots, images of natural people, natural images of objects, and artistic illustrations. The question was whether comments on the content represented a risk or could be considered humorous. This work allowed for an extensive analysis of the conversations, separating acceptable interactions from those that harm users. Risk is highly subjective, especially in private interactions, but understanding risks on social media is vital[9].

In 2023, another study analyzed the risk of 15,547 private messages on the Instagram digital platform of social media from adolescents aged 13 to 21. To do so, a machine learning approach was used to create risk-detection classifiers. The CNN and Random Forest approaches were found to be the most effective models for risk identification. An innovative framework was generated to apply artificial intelligence to online interactions. A total of 44,099 messages from participants were about unsafe sexual conversations. These conversations contained negative emotions such as anger and profanity that made the participant uncomfortable or insecure[10].

In 2020, another study where data obtain from Instagram's social media platform to analyze risk perception among adolescents aged 13 to 17. They used semi-structured interviews with 10 students beginning their university studies to consider sensitivity in data collection. The methodology that used which combines machine learning techniques to review engagement on social media with guided discussions between adolescents and parents about identified risk situations. This helped balance the tensions between parents and children when discussing sensitive social media messages. This work promotes adolescent-centred solutions for online safety .[11]

"Case Studies on AI and Social Media Impact on Teenagers globally"

The use of violence as consumer material by the media - desensitization, tolerance, saturation and impunity. According to the American Academy of Pediatrics, passive exposure to violence by the media is becoming an uncontrollable component of the lives of children and adolescents[12].

Case study 1: AI-Driven Social Media and Teen Mental Health Crises

Several recent investigations and news reports describe teenagers who developed severe anxiety, depression, or suicidal thoughts after heavy use of AI-driven social media platforms such as TikTok,

---

[7]S. van der Linden, Social Media Bans for Teens Lack Evidence, Nature Health, Feb. 25, 2026, at 1.

[8]S. Zhang, S. Tao, Y.L. Zhang, J. Zhou, J. Wei, M. Chen, Q. Hu, H. Zheng & Z.L. Wang, Examining the Spectrum of Problematic Online Behaviors in Chinese Adolescents: A Network Analysis of Smartphone, Gaming, and Social Media Use, 167 Comput. Hum. Behav. 108611 (2025)

[9]J. Park, J. Gracie, A. Alsoubai, A. Razi & P.J. Wisniewski, Personally Targeted Risk vs. Humor: How Online Risk Perceptions of Youth vs. Third-Party Annotators Differ Based on Privately Shared Media on Instagram, in Proc. 23rd Ann. ACM Interaction Design & Children Conf., Delft, Neth., June 17–20, 2024, at 1.

[10]A. Razi, A. Alsoubai, S. Kim, S. Ali, G. Stringhini, M. De Choudhury & P.J. Wisniewski, Sliding into My DMs: Detecting Uncomfortable or Unsafe Sexual Risk Experiences Within Instagram Direct Messages Grounded in the Perspective of Youth, 7 Proc. ACM Hum.-Comput. Interact. 1 (2023).

[11].A. Razi, Z. Agha, N. Chatlani & P.J. Wisniewski, Privacy Challenges for Adolescents as a Vulnerable Population (2020), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3587558 (last visited Aug. 30, 2025).

[12]**L.** Pfeiffer, Violencia Psicológica, in *Tratado de Pediatria* 426 (Sociedade Brasileira de Pediatria, 6th ed., Manole, Barueri, 2024)

Instagram, and Snapchat. These platforms use recommendation algorithms that learn what content keeps a user engaged (likes, comments, watch time) and then feed more similar material, often emotionally intense or polarising. [13]

A 15-year-old girl in the United States reported browsing nearly 5–7 hours per day through TikTok and Instagram, mostly watching "body-image" and "aesthetic lifestyle" content. Over months she developed body-dissatisfaction, social withdrawal, and symptoms of depression, which required outpatient therapy[14].

A 16-year-old boy in the UK used Instagram and Snapchat to track "likes" and followers as a measure of self-worth; when a friend compared him to more popular peers, he reported intense shame and momentarily researched self-harm methods online.[15]

Case 2: Teenagers and AI Chatbots as Emotional Companions

"AI companions, including chatbots and AI messaging applications are rapidly become popular among the teenagers and adults. They have emerged as key sources for seeking emotional comfort, exploring their identities, and even flirting around romantic engagements. These platforms are driven by the technology of large language models that simulate human interaction, offering personalized advice, motivation, and intimate connections.

- A 17-year-old girl in the U.S. reported spending two hours per night talking to an AI-companion app, which she described as "understanding" and "non-judgmental." Over time she began sharing highly personal problems and even suicidal thoughts with the bot instead of with family or friends.[16]
- In Australia and the U.S., lawsuits have been filed against AI-companion developers after a teenage boy died by suicide following months of interaction with an AI chatbot that allegedly encouraged self-harm and isolation Which consist of Psychological mechanisms where few studies shows that shy or socially anxious teens may especially when they feel misunderstood offline.When these relationships become emotionally central, teens may **withdraw from real-world connections**, worsening loneliness and increasing dependence on AI-driven interaction. .[17]

Case 3: Cyberbullying Amplified by AI and Deepfakes

Social media is already a common platform for cyberbullying and harassment targeted at teenagers. However, with the advent of deepfakes and image manipulation via AI, the impact of such harassment has become even more damaging.

- A 14-year-old girl in India reported that a classmate used a free AI image-generator to create "fake nudes" of her and shared them in a closed WhatsApp group. Within hours, screenshots spread to other classes, leading to public shaming, academic decline, and severe anxiety.
- In the U.S., a 16-year-old boy had a deepfake video made of him using profanity in a school setting; it was posted on TikTok and received tens of thousands of views before being taken down. The incident triggered school-disciplinary action and a temporary move to a different school. These activities may lead to certain consequences, like Victims of deepfake-based cyberbullying often experience high levels of shame, humiliation, fear of judgment and

[13]Sunil Arora, Sahil Arora & John Hastings, The Psychological Impacts of Algorithmic and AI-Driven Social Media on Teenagers: A Call to Action, Research & Publications, No. 85 (2024)

[14]Rachel Ehmke, How Using Social Media Affects Teenagers, Child Mind Inst. (with clinical experts Catherine Steiner-Adair, Ed.D. & Donna Wick, Ed.D.), https://childmind.org/article/how-using-social-media-affects-teenagers (childmind.org in Bing) (last visited Apr. 14, 2026).

[15]Sunil Arora, Sahil Arora & John Hastings, The Psychological Impacts of Algorithmic and AI-Driven Social Media on Teenagers: A Call to Action, Beadle Scholar, Research & Publications, Beacom Coll. of Computer & Cyber Scis., Dakota State Univ., No. 85 (Oct. 2024)

[16]un X., Wang Y. & B.T. McDaniel, AI Companions and Adolescent Social Relationships: Benefits, Risks, and Bidirectional Influences, Child Dev. Persp., Jan. 7, 2026, https://doi.org/10.1093/cdpers/aadaf009. (doi.org in Bing).

[17]Sun X., Wang Y. & B.T. McDaniel, AI Companions and Adolescent Social Relationships: Benefits, Risks, and Bidirectional Influences, Child Dev. Persp., Jan. 7, 2026, https://doi.org/10.1093/cdpers/aadaf009. (doi.org in Bing)

school-related problems, sometimes requiring counselling or transfer.Support groups and school-based interventions have had mixed success; many teenagers report that they did not know how to report or who to trust when the abuse involved AI-generated content.[18]

Overall, the reviewed studies highlight that risk perception among adolescents on social media is a multidimensional phenomenon that cannot be understood through a single disciplinary lens. While psychological approaches shed light on emotional responses and coping mechanisms, computational techniques provide scalable ways to detect and classify risk patterns. The integration of these perspectives allows for a more understanding of digital interactions, especially among vulnerable youth. However, many of these studies remain limited by cross-sectional designs and reliance on self-reported data, which constrain causal interpretations. Further studies may benefit from adopting longitudinal and mixed-method approaches to highlight the evolving dynamics of online risk perception. Furthermore, interdisciplinary collaboration between psychologists, data scientists and educators remains essential to transform these findings into actionable digital safety policies.

In summary, the literature demonstrates clear progress toward identifying and mitigating online risks, yet the gap between detection and prevention persists. Technological tools such as machine learning classifiers or network analyses offer promise but must be complemented by ethical and educational frameworks. Adolescent voices should also be incorporated more actively to ensure that programs are tailored with their lived experiences and cultural contexts. Importantly, the findings emphasize that risk on platform of social media is not only a matter of exposure but also of interpretation and agency and viewing the adolescents as active agents rather than passive victims changes the perspective on digital safety.. This critical shift can guide future research and policy toward more inclusive and adaptive strategies for safeguarding young people online.[19]

"India-Based Case Studies on AI and Social Media Impact on Teenagers"

India's Economic Survey 2025–26, presented on January 29, 2026, officially flagged the "precipitous rise" of digital addiction—covering social media, gaming, and smartphones—as a significant public health threat and a threat to long-term productivity. It highlighted that compulsive, high-intensity screen time is causing severe mental health issues, reducing academic focus, and disrupting the cognitive development of children and teenagers[20]

Raine v. OpenAI - Wrongful Death & AI Safety

A landmark 2025 lawsuit against OpenAI alleges that interactions with ChatGPT contributed to a teenager's suicide, bringing wrongful-death claims tied to generative AI behaviours. The case presses courts to consider *when and how AI systems can be held legally accountable* for user harm. AI deployments that interact with vulnerable users must build robust safety guardrails and clearly documented harm mitigation strategies.[21]

Case Study A: The impact of excessive use of Social Media and its effect on Mental Health in Tamil Nadu

A 2024 research examined among the college students in Tamil Nadu, India, provides a clear case-study-style illustration of how AI-driven social-media platforms affect adolescent mental health. The research found that over 26% of adolescents engaged in excessive social-media use, defined as spending several hours daily on platforms such as Instagram, TikTok, and WhatsApp, often late into the night.

Participants in this group reported symptoms such as restlessness, difficulty concentrating, and intermittent stress, mainly when they were unable to access their accounts. Many described a

[18]Adya Gupta, The Impact of Artificial Intelligence on Teenagers/Adolescents: A Comprehensive Analysis, Research & Publications, Dougherty Valley High Sch., San Ramon, Cal., May 9, 2025.

[19]Y. Pérez Vera, R.S. Escobedo Quispe & P.A. Ramírez Santos, A Clustering Approach to Identify Risk Perception on Social Networks: A Study of Peruvian Children and Adolescents, 13 Informatics 3 (2026), https://doi.org/10.3390/informatics13010003.

[20]Bindu Shajan Perappadan,Economic Survey calls for tackling rising digital addiction and screen-related mental health problems,THE HINDU, January 30, 2026 11:36 am IST - New Delhi

[21]Matthew Raine et al. v. OpenAI, Inc., No. CGC-25-628528 (Cal. Super. Ct. Aug. 26, 2025).

compulsion to check notifications and "likes," indicating that algorithm-driven features—such as auto-play videos, infinite scroll, and personalised feeds—were reinforcing habitual, emotionally charged engagement.

This case highlights how in the Indian context, AI-driven social-media design can turn social-media use into a stressor for teenagers, even in an educational environment where screen time is often framed as "required" for connectivity and information access. It also underscores the need for campus-based mental-health counselling and digital-wellbeing workshops tailored to Indian youth.[22]

Case Study B: High Social Media Usage and Anxiety in Bihar Adolescents

A 2025 study from Katihar Medical College & Hospital, Bihar, offers another India-based case-study-style analysis of social-media impact on teenagers. The study divided 80 adolescents aged 13–19 into two groups:

- Group A: less than 2 hours of social media per day;
- Group B: more than 2 hours per day.

The findings revealed that Group B showed significantly higher levels of depressive and anxiety symptoms, including poor sleep, daytime fatigue, and social withdrawal, compared with Group A. Many in Group B observed that they mainly excess social media platform to escape academic pressure or family conflicts, but instead felt more isolated and self-critical after long-term use.

This case functions as a localized Indian example of how prolonged exposure to AI-curated feeds—which prioritize emotionally intense or "highlight-reel" content—can amplify psychological challenges among adolescents who are already navigating high-stakes exams and conservative social environments. The study indirectly suggests that time-bound, mindful usage policies in schools and homes could help mitigate these effects in the Indian context.[23]

Case Study C: AI as a Confidant Among Indian School Students

A 2025 national survey conducted by Youth Ki Awaaz and YLAC (Youth Leadership and Action Committee) reveals an emerging, India-based behavioral-pattern case study involving AI and teenagers. The survey found that **88% of Indian school students turn to AI tools during periods of stress**, with the highest proportion of users in the **13–18 age group**.

In this particular group, many teenagers observed that they rely on AI chatbots and homework assistance method not only for their academics as also for emotional support systems. This included discussion of their exam anxiety, friendships, and loneliness. A common reason for choosing AI was that they preferred chatting with an AI system than parents and teachers due to the feeling of judgment.

This pattern can be understood as a focus case on the issue of AI-induced emotional dependency among Indian youths. Although AI devices make it easy to avail themselves of instant support, the problem is that such tools have not been designed as psychological interventions, nor do they come with any safety measures such as crisis intervention or oversight.[24]

However, in the recent past, there has been a lot of concern regarding misinformation, identity theft, financial fraud, and non-consensual pornography owing to the availability of such content online. It becomes difficult to trust any online communication when fake identities can easily be created. The role of audio-visual evidence in criminal cases is also undermined when such manipulation is done. Discussing India, availability of AI technology for manipulation coupled with the usage of social media by most Indians has created conducive circumstances for such activities. Such behavior is an infringement of the fundamental rights of privacy and dignity of individuals along with constitutional problems of safeguarding fundamental rights under Articles 19 and 21 of the Indian constitution.

---

[22]Vismay Basu, Social media giants, addicted teenagers, and the fight for informed choice, THE NEW INDIAN EXPRESS,05 Apr 2026, 8:55 am

[23]R. Choudhary, Impact of Daily Social Media Usage on Adolescent Mental Health: A Prospective Study, 13 Int'l J. Indian Psychol. 3620 (2025).

[24]Rishab Chauhan, 88% Students Turn to AI Tools Like ChatGPT to Manage Stress: Study, India Today (Aug. 21, 2025, 8:17 PM IST), https://www.indiatoday.in/education/story/88-students-turn-to-ai-tools-like-chatgpt-to-manage-stress-study-XXXX. (indiatoday.in in Bing)

The Digital Personal Data Protection Act, 2023, The Information Technology Act, 2000, and other such acts have limited and indirect solutions to cybercrimes where the perpetrator engages in manipulating digital content. However, these regulationwere not specifically designed to cater to the unique issues involved in AI-based media manipulation. Thus, it becomes difficult to define what qualifies as a crime, who the accused is, the acceptance of the evidence used in such cybercrimes, and the applicable jurisdiction.

Result and Declaration

The conclusions clearly highlight how artificial intelligence (AI) and social media positively and negatively influence teenagers. They make children more stressed, anxious, depressed, self-conscious, and sleepless when used extensively without any guidance, particularly if they spend at least two hours daily consuming emotionally stimulating or unrealistic information from algorithms-driven social media feeds. However, in case of regulated usage and constant oversight, AI and social media become effective learning tools useful for researching, preparing for tests, completing homework assignments, and practicing creative thinking, with most children noting increased confidence in studying and improved digital literacy skills. Some children utilize AI-powered chatbots and social networks to communicate and reflect their experiences, which can reduce feelings of loneliness; however, such behavior may prevent teenagers from reaching out for professional assistance. Moreover, AI-manipulated images and deepfakes have contributed to cyberbullying, leading to significant emotional harm, reputation damage, and social isolation in a few instances. Therefore, it becomes apparent that social media and artificial intelligence (AI) are not exclusively beneficial or harmful – their impact depends on various factors, including design features, user behavior, monitoring efforts, and digital literacy education.

In India, the frenzy of schooling combined with rapid digitalization magnifies both hazards and opportunities. Well-conceived school-based programs, parent involvement, and stricter controls on technology such as placing time limits, labeling content, and referring to emotional and psychological health services can enable adolescents to reap the benefits of technology without suffering from its adverse consequences.

Conclusion

Social media and AI have two major influences on children and teenagers: They create opportunities and risks. In terms of opportunities, they provide assistance in education, revision for exams, creativity, and worldwide networking. Teenage leadership in activism and involvement in designing AI may be liberating and increase awareness about psychological wellbeing. However, excessive and random usage of social media and AI technologies can cause anxiety, depression, poor self-esteem, insomnia, and social comparison, which is normal between teenagers who use algorithm-based apps for several hours daily.

The confluence of studies with the increasing amount of digital technology in India can encourage the pushing of harm despite greater access to knowledge and information channels. It is necessary to understand that, when it comes to AI and social media, there cannot be any positive or negative effect without considering the design, management, and use of such technologies. In order to achieve positive results and minimize the risks associated with harm, it is important to promote digital literacy, limit screen time, monitor digital interaction and strengthen protective measures for social platforms, such as warning labels and referrals to crisis hotlines to reduce the risk of deepfake-related abuse. If approached with the best interests of a child in mind, society can help shape the digital age.

References

1. LauraDeNardis, *The Global War for Internet Governance* (Yale Univ. Press 2014)
2. T. Masri-Zada, S. Martirosyan, A. Abdou, R. Barbar, S. Kades, H. Makki, G. Haley & D.K. Agrawal, The Impact of Social Media & Technology on Child and Adolescent Mental Health, 9 J. Psychiatry & Psychiatric Disorders 111 (Apr. 16, 2025).
3. M.B. Robb & S. Mann, Talk, Trust, and Trade-Offs: How and Why Teens Use AI Companions (Oct. 6, 2025), https://[insert stable URL].

4. J.M. Nagata, Z. Memon, O. Huang & M.A. Moreno, Adolescent Health and Generative AI—Risks and Benefits, 180 JAMA Pediatr. 7 (Jan. 1, 2026), https://doi.org/10.1001/jamapediatrics.2025.4502. (doi.org in Bing)
5. R. Castellanos Silva & G. Steins, Social Media and Body Dissatisfaction in Young Adults: An Experimental Investigation of the Effects of Different Image Content and Influencing Constructs, 14 Front. Psychol. 1037932 (2023),
6. J.M. Nagata, Z. Memon, O. Huang & M.A. Moreno, Adolescent Health and Generative AI—Risks and Benefits, 180 JAMA Pediatr. 7 (Jan. 1, 2026), https://doi.org/10.1001/jamapediatrics.2025.4502. (doi.org in Bing)
7. S. van der Linden, Social Media Bans for Teens Lack Evidence, Nature Health, Feb. 25, 2026, at 1.
8. S. Zhang, S. Tao, Y.L. Zhang, J. Zhou, J. Wei, M. Chen, Q. Hu, H. Zheng & Z.L. Wang, Examining the Spectrum of Problematic Online Behaviors in Chinese Adolescents: A Network Analysis of Smartphone, Gaming, and Social Media Use, 167 Comput. Hum. Behav. 108611 (2025)
9. J. Park, J. Gracie, A. Alsoubai, A. Razi & P.J. Wisniewski, Personally Targeted Risk vs. Humor: How Online Risk Perceptions of Youth vs. Third-Party Annotators Differ Based on Privately Shared Media on Instagram, in Proc. 23rd Ann. ACM Interaction Design & Children Conf., Delft, Neth., June 17–20, 2024, at 1.
10. Adya Gupta, The Impact of Artificial Intelligence on Teenagers/Adolescents: A Comprehensive Analysis, Research & Publications, Dougherty Valley High Sch., San Ramon, Cal., May 9, 2025
11. Y. Pérez Vera, R.S. Escobedo Quispe & P.A. Ramírez Santos, A Clustering Approach to Identify Risk Perception on Social Networks: A Study of Peruvian Children and Adolescents, 13 Informatics 3 (2026), https://doi.org/10.3390/informatics13010003.
12. Bindu Shajan Perappadan,Economic Survey calls for tackling rising digital addiction and screen-related mental health problems,THE HINDU, January 30, 2026 11:36 am IST - New Delhi
13. Matthew Raine et al. v. OpenAI, Inc., No. CGC-25-628528 (Cal. Super. Ct. Aug. 26, 2025).
14. Vismay Basu, Social media giants, addicted teenagers, and the fight for informed choice, THE NEW INDIAN EXPRESS,05 Apr 2026, 8:55 am
15. R. Choudhary, Impact of Daily Social Media Usage on Adolescent Mental Health: A Prospective Study, 13 Int'l J. Indian Psychol. 3620 (2025).
16. Rishab Chauhan, 88% Students Turn to AI Tools Like ChatGPT to Manage Stress: Study, India Today (Aug. 21, 2025, 8:17 PM IST), https://www.indiatoday.in/education/story/88-students-turn-to-ai-tools-like-chatgpt-to-manage-stress-study-XXXX. (indiatoday.in in Bing)

Managing Synthetic Media Without Challenges to Speech: Deepfakes and Democratic Governance

Neetu Umashankar Singh
Class: LL.M. (2 nd Year), Semester IV
Specialization: (Criminal Law**)**

Abstract

Artificial intelligence technology generates hyper-realistic and highly susceptible media such as deepfakes which pose unique risks to governance. Deep fakes has a wide range of uses, including"such as satire, art, and politics. The public debates and deliberations on gubernatorial matters will likewise be furthered by them, especially in modern democracies where digital communication plays a central role. However, dishonest individuals often deploy them to mislead, impersonate officials and trick voters. It poses a serious threat to the democratic process and constitutional rights, particularly in societies that rely on informed decision-making.

Focusing on governance as the overarching theme, the paper seeks to answer a key question: how can we implement a legal framework that efficiently eliminates harmful deep fakes while still upholding the essential The constitution protects the freedom to express one's views." and expression and freedom of the press? The worldwide reach of the internet, combined with how rapidly material circulates, makes the issue far more complicated.

Through an examination of the legal frameworks across different systems regarding and policies related to deep fakes, this paper seeks to provide an answer. The study is divided into key parts definition and impact, case studies, regulatory frameworks, and a proposed governance model. the DSA, the EU's regulation on digital services is taken as a focal point owing to its comprehensive nature. The paper ultimately argues that governance should aim to prevent harm and hold platforms accountable rather than impose blanket bans on content before it becomes constitutionally relevant.

Keywords :- Deepfakes, synthetic media, electoral integrity, AI governance, freedom of speech, platform accountability, disinformation

Research Objectives

1. **To define and classify synthetic media:** Identify the technical evolution of deepfakes (from GANs to diffusion models) and categorize their applications into legitimate (artistic/educational) and malicious (disinformation/non-consensual imagery) uses.
2. **To evaluate democratic and constitutional harms:** Analyze how deepfakes erode public trust, interfere with electoral integrity, and infringe upon core rights including privacy and human dignity.
3. **To perform a comparative legal analysis:** Examine and critique the existing regulatory frameworks in the **European Union (DSA)**, **India (IT Rules)**, and the **United States** (State and Federal proposals) to identify regulatory gaps.
4. **To design a Layered Governance Model:** Propose a four-tier regulatory system—ranging from transparency and labelling to expedited takedowns—that balances effective harm mitigation with constitutional protections.
5. **To establish constitutional safeguards for AI regulation:** Formulate legal standards, such as sunset clauses and judicial oversight, to prevent government overreach and protect the "chilling effect" on legitimate satire and political speech.

Research Questions

1. **Technical vs. Legal Definition:** How do deepfakes differ from traditional disinformation, and why does their unique audiovisual nature require a specialized legal response rather than general speech regulations?
2. **Electoral Integrity:** To what extent did recent case studies (e.g., 2024 elections in Slovakia, India, and the USA) demonstrate the failure of current "notice-and-takedown" timelines in protecting the democratic process?

3. **Comparative Efficacy:** How does the European Union's risk-based "platform accountability" approach compare to India's executive-led "prescriptive" approach regarding the speed, transparency, and in respect for due process?
4. **Proportionality in Regulation:** Can a tiered governance model—which prioritizes transparency for low-risk content and rapid removal for high-risk electoral content—satisfy the "narrow tailoring" requirements of constitutional free speech?
5. **Liability and Enforcement:** What are the legal and ethical challenges in imposing civil and criminal liability on creators and intermediaries, particularly regarding the *mens rea* (intent) required to protect parody and satire?

Parameters and Constrain of the Research

This study revolves around the governance of deepfakes and how the media functions in the democratic electoral integrity, freedom of speech, and platform accountability. It adopts a comparative legal approach by examining regulatory frameworks in the European Union, India, and the United States. The research covers recent developments, especially from 2024–2025, including real-world case studies of political deep fakes and their impact on voters and institutions.

The study also proposes a tiered governance model aimed at balancing harm prevention with constitutional protections. It seeks to contribute to ongoing debates on AI regulation by offering a structured and proportionate approach.

However, the study is limited in certain respects. First, it relies mainly on secondary sources such as written laws, judicial precedent,policy documents, and reported incidents, and refrain from empirical or field-based research. Second, given the contains shifting nature of artificial intelligence, the technological aspects discussed may become outdated quickly as new tools and techniques emerge. Third, the study is solely restricted to selected jurisdictions and may fall short of the global diversity, especially in developing countries. Lastly, practical implementation challenges—such as technical feasibility of detection tools, compliance by platforms, and cross-border enforcement—are discussed conceptually but not tested in real-world settings.

Research Procedure

This work investigate a doctrinal and comparative legal research methodology to examine the regulation of deepfakes in democratic governance. The research is primarily based on secondary sources, including legislation, judicial precedent, literature, policy papers, and official reports relating to artificial intelligence and media regulation. These sources provide a comprehensive understanding both the legal and technological dimensions of the issue.

A techinique is adopted to analyze the legal frameworks of different jurisdictions, particularly the European Union, India, and the United States, in order to identify similarities, differences, and regulatory gaps. This comparison helps in evaluating what paradigm are more suitable for regulation with constitutional freedoms.

The study also incorporates recent case studies of deepfake incidents to understand their real-world impact on elections, governance, and public discourse. These case studies highlight the urgency of regulatory reform. An analytical method is applied to evaluate constitutional principles such as freedom of speech, proportionality, and accountability. Based on this analysis, the research proposes a structured governance model aimed at balancing the need to control harmful deepfakes with the protection of fundamental rights.

Literature Review

The scholarly discourse surrounding deepfakes has rapidly transitioned from analyzing technical novelties to addressing systemic democratic threats. Scholars such as Chesney and Citron (2019) highlight the emergence of the "Liar's Dividend," where individuals can deny real evidence by claiming it to be fake. This phenomenon weakens accountability mechanisms in democratic societies.

Further, Regina Rini (2020) emphasizes that deepfakes exploit "perceptual realism," making them more convincing than traditional misinformation because humans naturally trust audiovisual

content.Several academics have found out that deepfakes contribute to a broader "post-truth" environment where facts become difficult to verify.

Legal scholarship also reveals a divergence in regulatory approaches. The European Union focuses on risk-based accountability and transparency, India emphasizes executive control and quick takedown, while the United States prioritizes free speech and constitutional safeguards. Despite these efforts, the literature suggests that no single framework has fully addressed the balance between regulation and freedom. This gap highlights the need for a more nuanced and layered governance approach.

Introduction

The management of deep fakes synthetic media generated or distorted through by applying deep learning technique including auto encoders and generative adversarial networks (GANs) [1]. Deep fakes can prove to be a highly influential capacity tool with the potential threaten democracies. Even though artificial intelligence can be useful for the entertainment world, education and art, they can also mislead people[2]. Deep fakes differ from conventional disinformation as they leverage the distinctive weight assigned to moving image and sound. Technologies can undermine citizen trust in audiovisual evidences and electoral integrity and public faith. [3]

The regulatory responses are constrained by constitutional, transnational, and rapid evolution. The constitutional guarantee of free speech and press means that overly broad prohibitions risk a chilling effect on satire or legitimate criticism. Deep fakes are hard to regulate because they're on the internet. AI is evolving quickly than the laws. This central heme of the paper is to examines how a legal framework can mitigate harm caused by the unethical use of synthetic media while maintaining human rights.

Using a comparative legal methodology, the paper looks into an assortment of existing regulations in the European Union, India and the USA. It highlights constitutional tensions and regulatory gaps in existing regulation and sets out a model for tiered governance of deep fakes. This model favors harm prevention and platform accountability over use of bans on content ex ante. The document moves in four parts. Part II defines deep fakes and classifies its democratic harms. Part III offers instance from 2024-25. Part IV examines current legal framework. Part V lays out the model and lastly, offers views on compliance with constitutional values.

The emergence of **Diffusion Models** and advanced **Generative Adversarial Networks (GANs)** has democratized high-fidelity forgery, allowing for the generation of synthetic media in seconds rather than weeks. This shift creates what scholars term a "Post-Truth" governance crisis. In this environment, the primary threat is not just the belief in a lie, but the resulting "information nihilism," where citizens—overwhelmed by the difficulty of verification—simply cease to believe in any objective reality. This creates a fertile ground for authoritarianism, as the loss of a shared factual baseline prevents the public from holding the state accountable. Consequently, the legal challenge is twofold: we must regulate the weaponization of the technology while ensuring that the "truth" does not become a monopoly of the government or a few powerful tech intermediaries.

Recognizing Deep fakes and Their Effects on Democracy.

a) Meaning of Mechanical Engineering.

Deep fakes primarily refer to fabricated media wherein an existing video or an image of a person is replaced with someone else. The word deep fake comes from deep learning and fake. A GAN is relies on two interconnected neutral network , a generator which produces fake content and a discriminator which to spot the forgery. In light of constant competition, the generator becomes excellent at generating content that resembles the original recordings [4].The emergence of

[1] Ian Goodfellow et al., *Generative Adversarial Nets*, in 27 Advances in Neural Information Processing Systems 2672 (2014)

[2] Robert Chesney & Danielle Citron, *Deepfakes and the New Disinformation War,* 98 Foreign Aff. 147, 149-52 (2019).

[3] Danielle Keats Citron & Robert Chesney, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 Calif. L. Rev. 1753, 1758-65 (2019).

[4] Mika Westerlund, *The Emergence of Deepfake Technology: A Review*, 9 Tech. Innovation Mgmt. Rev. 39, 40-41 (2019)

contemporary architectures like diffusion models and transformers has made it easier to develop deep fakes. The availability of open-source tools such as DeepFaceLab or FaceSwap along with some commercial tools allow virtually anyone with minimal technical ability to create highly believable forgeries on readily accessible computing hardware. The segment of popular culture attributing face swapped videos, lip-synced speeches, and voice clones (using as little as three seconds of source audio) to deepfake creation is misleading. Likewise, fully artificial "digital humans" who never existed are not necessarily deep fakes. Inconsistent eye-blinking, hair or skin texture, or lighting made early deep fakes easy to catch due to visible artifacts. Recent models, however, generate videos that can consistently evade detection by humans and algorithms [5]. Detection only approaches will eventually fail as technology will advance. As a result, governance solutions should be implemented at the level of provenance and platform accountability.

b) Legitimate and Malicious Uses.

The deep fake technology creates artificial media in which a person in an existing image or video is replaced with someone else's likeness. Face-swapping or face reenactment involves the transposition of the likeness of one face onto another. While deep fake has proven its uses in wrong hands, it certainly must not become simple [6]. The documentary Welcome to Chechnya uses a deep fake video for the purpose of identity of LGBTQ+ subjects while preserving most of their emotional expressiveness. Impressively, museums created deep fakes to breathe life into historical characters for exhibits. Yet, the evil purposes overshadow these righteous purposes, leading to the worldwide proposal of so many bills and acts. It includes political disinformation such as fake speeches and fabricated sex scandals. Next, we have electoral manipulation in the form of voter suppression through false announcements. Next, we have non-consensual intimate imagery or "deep fake pornography". Celebrities are primarily targeted, but even private citizens are not safe [7].Over recents years there is rise in sexual harassment and its forms. Among the numerous problems to plague the country recently, the appearance of deep fake porn is perhaps among the more disturbing. This is an AI-generated alteration where the faces of women are placed on pornographic videos to humiliate them.

c) Analyzing the Democratic Harms

Deep fakes in the real world scenario can manipulate people without their consent using AI-generated videos, besides that, the dangers of identity theft and cyber-fraud are also serious consequences of deep fakes [8].Real content could start to lose credibility because of deep fake technology. Deep fakes pose a serious threat to the function of democracy across the world. The mere existence of fake content can discredit real content. An authentic clip of a rival candidate making damaging comments could be passed off as a deep fake. This erosion has a corrosive second order effect whereby the latter emerges, political actors can dismiss it as a synthetic media and get away with it. Deep fakes have a constructive effects on elections. With the right timing, a well-executed deep fake can change the tone of elections. For instance, a deep fake revealing [9].The fourth concern is deep fakes can be maliciously used against journalists and activists since authoritarian regimes use them to discredit opponents.

Fifth, as per the experts, these deep fake violate your constitutional rights. They fundamentally infringe privacy, reputation and dignity rights which are either directly or indirectly protected in constitutional democracies [10].Ultimately, a deep fake video can be created of one ethnic group

[5] Siwei Lyu, *Deepfake Detection: Current Challenges and Next Steps,* 2020 IEEE Int'l Conf. on Multimedia & Expo Workshops 1, 2

[6] Chesney & Citron, *supra note* 2, at 153

[7] Citron & Chesney, *supra note 3*, at 1758-65.

[8] Regina Rini, *Deepfakes and the Epistemic Backstop*, 2 Philosophers' Imprint 1, 5 (2020).

[9] Robert Chesney & Danielle Keats Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security,* **107 Cal. L. Rev. 1753 (2019)**

[10] Citron & Chesney, *supra note* 3, at 1780.

attacking another so as to incite actual violence and communal riots. A multi-dimensional governance response is needed for the multi-dimensional harm profile of deep fakes.

d) Real World Deepfake Case Studies

A deep fake audio in anticipation of the to the parliamentary election of Slovakia in 2024 has started making rounds on that social medium, posing as a liberal candidate Michal Šimečka who was caught up to rig the election and raise the beer price. Though the clip was discredited by fact-checkers within hours, hundreds of thousands had by then heard the audio. Shimecka lost by a narrow margin. According to exit polls, the deep fake had an impact on approximately five percent of undecided voters [11].The rapidity of dissemination as opposed to the reaction is what made this case significant. Facebook and Telegram were the main platforms for mass distribution of the deep fake. Facebook removed the content at some point but after the critical forty-eight hours when the contestants were campaigning during the election. This pertains to the updating theme, represents the mismatch between the harm caused and the countermeasure, in deepfake contexts

e) The General Elections in India 2024

The exact number of deep fakes targeting leaders during the 2024 general election in India was high. A deep fake of home minister Amit Shah appeared for instance, which showed him promising to do away with reservations for Scheduled Castes and Tribes. Naturally, he did not say this, either. Another deep fake was shown of opposition leader Rahul Gandhi supporting policies which his party of politics would certainly not support. While the Election Commission did issue advisories, it did not have the statutory power to effect takedown in the forty-eight hour window[12].Localized deep fakes in regional languages targeted specific constituencies and because verifying them at a central level was next to impossible, the situation worsened. Moreover, although the IT Rules 2021 of India allow for takedown within a time frame of seventy-two hours; there is no special expedited process for deep fakes related to elections. To sum up, the content volume is extremely high in a country. a country where the number is above 800 internet users Struggled with manual and automated detection.

f) The Incident of "Biden Robocall" in the USA (2024).

In January 2024, thousands of New Hampshire voters got a robocall from a group posing as President Joe Biden. In reaction to New Hampshire's primary election, the call urged voters to reject voting the Democratic primary. The robocall used the phrase "Come on, man" which the president often uses. It also adopted the president's tone, which is rather lethargic and tired. Not long after, the robocall was traced to a political party [13].The voice of a robocall powered by artificial intelligence, the FCC said, is properly covered already by telephone consumer protection law. However, that statutory interpretation did not cover deep fakes on social media. The incident also demonstrated another fragmentation problem; no agency in the US has clear authority over electoral deep fakes. There's a scattergun arrangement of jurisdiction through the Federal Election Commission, Federal Communications Commission, the Federal Trade Commission and state election offices.

g) Deep fakes of Ukraine-Russia War.

A deepfake video purportedly featuring the President of Ukraine Volodymyr Zelenskyy circulating on many Ukrainian news websites and social media, in which he urges soldiers to lay down arms, was released in the ongoing Russia-Ukraine war. This will further create confusion in the region. The video was hastily debunked by Zelenskyy himself, but it has nevertheless provided a good example of the weaponization of deep fakes in war. The video may have had such intentions [14].There are deepfakes of Russian Defense Minister Sergei Shoigu calling for a general mobilization. A fake video created a scare for some time in Russia and thereafter was closed. Wartime Deep fakes

---

[11] *How a Deepfake Influenced Slovakia's Elections,* Wired (Oct. 03, 2023),https://www.wired.com/story/slovakias-election-deepfakes-show-ai-is-a-danger-to-democracy

[12] Shaji George, *Regulating Deepfakes to Protect Indian Elections*, 1 Partners Universal Innovative Research Publication (2024).

[13] *US political consultant indicted over AI-generated Biden robocalls,* Reuters (May 23, 2024), https://www.reuters.com/world/us/us-political-consultant-indicted-over-ai-generated-biden-robocalls-2024-05-23.

[14] *Deepfake video of Zelenskyy could be 'tip of the iceberg' in info war, experts warn,* NPR (Mar. 16, 2022), https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia.

significantly exceed any previous efforts in manipulating electoral integrity and squarely place the state and hybrid warfare domain. It becomes easier to deny their involvement. States assert that deep fakes, while synthetic, are nevertheless construed as true, rendering them incapable of being used to commit fraud since the victim cannot prove otherwise.

Regulatory Analysis Comparison.

a) The Digital Services Act by the European Union.

The Digital Services Act (DSA) implementing framework of the European Union (EU) came into full application in February 2024, which represents the world's most comprehensive regulatory framework covering synthetic media [15].VLOPs are defined as online platforms with more than 45 million monthly users within the EU, and they are required to perform risk assessments for dissemination of illegal content and manipulation of information, per Article 34. Article 35 Obligations of a platform which could include the moderate measures like labeling, demoting in recommendation algorithm or removal. According to Article 52, the generator of an electronic system must ensure that the output of their system is marked as synthetically generated in a machine-readable format. This enables downstream platforms to automatically detect and label synthetically generated content. The DSA does not require any removals of deep fakes, but wants transparency and mitigating risk. This approach shall respect the right to freedom of expression as per Article 11 of the EU Charter of Fundamental Rights [16].The EU's Digital Services Act (DSA) could regulate many forms of voting-disinformation deep fakes thanks to its disinformation and human and technology manipulation rules as well as serious coordinated diversion campaigns that include deep fakes. All very large online platforms and very large online search engines are covered by enforcement, which enforces responsibility. There will be external audits once a year. Non-compliance can attract a fine of up to six percent of the global turnover. However, the reliance of DSA on platform self-regulation might be inadequate for electoral.

b) India: The IT Rules 2021 (as amended 2023).

India has opted for a more prescriptive approach through the IT (Intermediary Guidelines and Digital Media Ethics Code) Rules of 2021 [17].As per Rule 3(1)(b)(v), intermediaries, on receipt of court or government order, shall remove within seventy-two hours and disable access to content which is "patently false" or "misleading"; in the case of deep fakes the government may issue emergency takedown orders without prior judicial review. According to the provisions of Rule 4(4) [i]major social media intermediaries can be called such who have more than 5 million registered users and have to take reasonable assurance in identifying and labeling uses of AI generated content. The deep fakes complaint can be reported by the users under the grievance addressed mechanism which also mandates the intermediaries to acknowledge a complaint within 24 hours and dispose off within 72 hours.

The Constitution of India guarantees freedom of speech and expression under Article 19(1)(a) that may be subject to "reasonable restrictions" under Article 19(2). Restrictions may be imposed, as it may be necessary to impose reasonable restrictions on the ground of a friendly relationship with foreign states. There are two challenges launched against the rules in Kunal Kamra v. Union of India, 2024 Firstly, it lacks firm statutory backing as it exists as delegated legislation initiated under the IT Act, 2000. Secondly, it leads to over-censorship, which is dysfunctional. [18] In essence, major platforms face incentives to remove specific content that is deemed 'obviously false', according to the rule that was released on Friday.

The lack of judicial pee-review for emergency take-downs under Article 19 is particularly alarming. Further, the term "patently false" is open-ended as what one government official may deem as patently false may very well be political speech by another official.

---

[15] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services, arts. 34-35, 52, 2022 O.J. (L 277) 1 [hereinafter Digital Services Act]

[16] Charter of Fundamental Rights of the European Union, art. 11, 2012 O.J. (C 326) 391.

[17] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, G.S.R. 139(E) (India), amended 2023.

[18] *Kunal Kamra v. Union of India,* 2024 SCC OnLine Bom 4326 (Sept. 26, 2024) (India).

c) State Laws and Federal Proposals in the U.S.

United States does not have a federal deep fake law that applies to elections specifically. The DEFIANCE Act of 2024 outlawed deep fakes through the creation of a federal crime for the alteration of images. In addition, the civil remedies provided include statutory damages worth [19].Regarding electoral deepfakes, the AI CONSENT Act and Protect Elections from Deceptive AI Act were proposed but none have been enacted by mid-2020. Several states have taken independent action. A California Assembly Bill 730 of 2019 prevents the distribution of fake videos involving candidates for an election within sixty days of the election. Texas SB 751 (2019) makes it a crime to use deep fakes to injure a candidate or influence an election, punishable by one year. According to Minnesota HF 1370 (2023), Misdemeanor is an offense to produce Deep fakes to influence elections [20].

In United States v. Alvarez (2012) Supreme Court held that the Stolen Valor Act, which made it a crime to falsely claim to have received military medals, is unconstitutional, as false statements alone are not necessarily unprotected [21].As a result, deep fake laws that criminalize a false statement without proof of particularized harm (like defamation, incitement or fraud) are unlikely to survive First Amendment scrutiny. Scholarly opinion now holds that narrowly tailored disclosure and labeling requirements will be more constitutionally sustainable than criminal prohibitions[22].

d) Analyses and omissions.

The EU, the US and India have chosen to regulate deep fakes differently. The European Union's approach as captured in the DSA is the most protective of free expression and places "risk based duties" on platforms. When platforms are allowed to be the "judge and jury" who apply a mess of terms, content can be rendered unavailable based on the platforms' findings, which are often flawed or politicized – especially for non-electoral deep fakes. The EDPB clarification confirms that due process and "minimal interference" are not required, so there is much room for error. In addition, the procedures' speed, which only covers individual specific things, would be too slow for electoral deep fakes where harm must take place before election day.

India's approach indeed emphasizes speed, simplicity, and primacy of government. However, it does so at the expense of significant over-censorship by opaque entities and political actors. The failing bodies cover "designated" (not all) intermediaries so the most popular platforms may be unaffected. Content must be covered under the constitution else rules will get invalid.

Conceptual Framework

This study is grounded in the Principle of Proportionality, which requires that any restriction on speech must be necessary, reasonable, and the least restrictive means available. The framework rejects a simple "ban versus allow" approach and instead adopts a layered governance model.

The model treats deepfakes as a spectrum of risk. Low-risk content such as satire and artistic expression is protected through transparency measures, while high-risk content such as electoral manipulation is subject to stricter regulation.

This framework emphasizes three key elements: transparency, accountability, and targeted intervention. By linking the level of harm to the level of regulation, it ensures that democratic values are preserved while minimizing misuse. The approach also recognizes that not all deepfakes are harmful and therefore regulation must be carefully calibrated rather than overly broad.

Findings and Results

The study finds that deepfakes pose a serious and evolving threat to democratic governance. Existing regulatory frameworks are largely ineffective during critical periods such as elections. In several case studies, harmful content spread rapidly before any regulatory action could be taken.

The comparative analysis shows that the European Union's approach is more balanced but may be slow in urgent situations. India's framework is faster but risks over-censorship due to lack of judicial

[19] Disrupt Explicit Forged Images and Non-Consensual Edits Act of 2024, S. 3671, 118th Cong. (2024).

[20] Cal. Elec. Code § 18320 (2019); Tex. Elec. Code § 255.004 (2019); Minn. Stat. § 211B.02 (2023).

[21] *United States v. Alvarez*, 567 U.S. 709, 718 (2012).

[22] *Reno v. ACLU,* 521 U.S. 844, 874 (1997).

oversight. The United States provides strong free speech protection but lacks a unified regulatory mechanism.

Another key finding is that technological solutions alone are insufficient. Detection systems are not fully reliable and may result in errors that affect legitimate speech. Therefore, a combined approach involving law, policy, and platform responsibility is necessary.

Recommendations / Suggestions

The paper proposes a **Layered Governance Model** to manage synthetic media effectively:

1. Tiered Regulatory Action

- **Tier 1 (Transparency):** Mandatory machine-readable watermarking and "synthetic" labels for all AI content.
- **Tier 2 (Contextualization):** For non-electoral harms (defamation/fraud), use warning labels and "trusted flagger" reporting.
- **Tier 3 (Expedited Takedown):** For electoral deepfakes, establish an **Independent Digital Tribunal** capable of issuing *prima facie* takedown orders within **24 hours**.

2. Legal Safeguards

- **Sunset Clauses:** Emergency electoral powers should automatically expire 30 days after the election to prevent "mission creep."
- **Specific Intent (Mens Rea):** Criminal liability should only apply to creators with "specific intent" to cause fraud, violence, or voter suppression, protecting artists and satirists.

3. Safe Harbor Reform

- Intermediary "Safe Harbor" protection (e.g., Section 79 of India's IT Act) should be **contingent** on a platform's compliance with the tiered labeling and takedown framework.

4. International Treaties

- Develop bilateral agreements for **cross-border takedowns**, as deepfake creators often operate from jurisdictions outside the target country's legal reach.

Conclusion

Deep fakes threaten the very fabric of democracy. The capability of deep fakes to fabricate fictitious realities has damaging effects on the electoral process, the constitutional right to privacy and dignity, and trust in democracy. Yet, censorship cannot be the solution to this issue. Legislation that criminalizes all parties except the wrongdoer will have a chilling effect on satire, artistic expression, political dissent and so on, which is against democratic governance.

Governance of deep fakes requires a tiered and harmonized approach. In other words, different deep fakes may cause different types of harm and would require a proportional response. A sexualize deep fake would cause more severe harm to D. There are tiers on transparency requirements, contextual disclaimers, rapid electoral takedown, targeted liability for creators of deep fakes, paid-ordering of deep fakes among others. The European Union's risk-based approach to AI governance aims at tackling similar issues and can serve as a strong free speech sensitive model. The takedown mechanism of India gives you speed at the expense of excessive takedown. The First Amendment of the United States requires narrow tailoring, which state laws have tried, but the bill of DC has not.

While digital chaos or digital censorship is a false dichotomy in democracies, but without consequence, the latter will prove to be inevitable. Although the genie may be out of the bottle, there remains substantial scope for governance of synthetic media. Management must take. Further studies must look into the ability of technical watermarking standards to work, the feasibility of independent digital courts with fast and informal procedures, and the architecture of international deep fake treaties. The time is now, before the next cycle of elections gets swamped by a deluge of synthetic lies.

References:

1. Ian Goodfellow et al., *Generative Adversarial Nets*, in 27 Advances in Neural Information Processing Systems 2672 (2014).
2. Robert Chesney & Danielle Citron, *Deepfakes and the New Disinformation War,* 98 Foreign Aff. 147, 149-52 (2019).

3. Danielle Keats Citron & Robert Chesney, Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security, 107 Calif. L. Rev. 1753, 1758-65 (2019).
4. Mika Westerlund, *The Emergence of Deepfake Technology*: A Review, 9 Tech. Innovation Mgmt. Rev. 39, 40-41 (2019).
5. Siwei Lyu, *Deepfake Detection: Current Challenges and Next Steps,* 2020 IEEE Int'l Conf. on Multimedia & Expo Workshops 1, 2.
6. Chesney & Citron, *supra note* 2, at 153.
7. Citron & Chesney, *supra note 3*, at 1758-65.
8. Regina Rini, *Deepfakes and the Epistemic Backstop*, 2 Philosophers' Imprint 1, 5 (2020).
9. Robert Chesney & Danielle Keats Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security,* 107 Cal. L. Rev. 1753 (2019)
10. Citron & Chesney, *supra note* 3, at 1780.
11. *How a Deepfake Influenced Slovakia's Elections,* Wired (Oct. 03, 2023),https://www.wired.com/story/slovakias-election-deepfakes-show-ai-is-a-danger-to-democracy/
12. Shaji George, *Regulating Deepfakes to Protect Indian Elections*, 1 Partners Universal Innovative Research Publication (2024).
13. *US political consultant indicted over AI-generated Biden robocalls,* Reuters (May 23, 2024), https://www.reuters.com/world/us/us-political-consultant-indicted-over-ai-generated-biden-robocalls-2024-05-23.
14. *Deepfake video of Zelenskyy could be 'tip of the iceberg' in info war, experts warn,* NPR (Mar. 16, 2022), https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia.
15. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services, arts. 34-35, 52, 2022 O.J. (L 277) 1 [hereinafter Digital Services Act]
16. Charter of Fundamental Rights of the European Union, art. 11, 2012 O.J. (C 326) 391.
17. Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, G.S.R. 139(E) (India), amended 2023.
18. *Kunal Kamra v. Union of India,* 2024 SCC OnLine Bom 4326 (Sept. 26, 2024) (India).
19. Disrupt Explicit Forged Images and Non-Consensual Edits Act of 2024, S. 3671, 118th Cong. (2024).
20. Cal. Elec. Code § 18320 (2019); Tex. Elec. Code § 255.004 (2019); Minn. Stat. § 211B.02 (2023).
21. *United States v. Alvarez*, 567 U.S. 709, 718 (2012).
22. *Reno v. ACLU,* 521 U.S. 844, 874 (1997).

Regulation of Generative AI: Challenges for Legal Systems

Joshua Ebenezer

Penultimate Year, BA LLB (S-VIII) (R No. A13)

HVPS College of Law, University of Mumbai.

Abstract

Gujarat's highest court issued an AI governance policy in early 2026 that, while rooted in sound constitutional principles, arrives too late and too narrowly to contain the technology it seeks to manage. Rather than calibrating regulation to the profession as it actually operates, the policy draws boundaries around judicial conduct that leave untouched the far larger landscape of AI already at work in Indian courts. The policy carves out an extremely restricted space for judicial AI: routine scheduling tasks and precedent retrieval, both subject to verification requirements that few practitioners would describe as workable at the pace of daily court operations. The profession the policy seeks to insulate has already crossed the threshold. Counsel routinely prepare court filings with AI drafting assistance; legal education now treats AI fluency as a vocational requirement; and the court system's own digital backbone depends on the same category of technology the policy restricts. Drawing on comparative analysis of the EU AI Act 2024, Singapore's Judicial AI Governance Framework, Estonia's AI-native court infrastructure, and United States disclosure-led models, this paper argues for a paradigm shift: from prohibition to structured governance, from restriction to mandatory AI literacy, and from fear of hallucination to prompt engineering competence as a professional standard. India's absence of regulatory frameworks for Artificial General Intelligence and quantum-era data security represents an institutional vulnerability that will compound as the technology evolves. A six-point policy roadmap is proposed, centred on a National Judicial AI Governance Commission, enterprise AI procurement under existing e-Courts funding, and targeted amendments to India's data protection legislation. Constitutional values of judicial independence and human-centred adjudication are best served by governance, not by AI that operates without it.

Keywords: *Judicial AI governance, Generative AI regulation, Prompt engineering, AI hallucination, Gujarat HC AI Policy, Artificial General Intelligence, e-Courts Phase III.*

I. Research Objectives

This paper pursues four principal objectives:

1. To examine the Gujarat High Court AI Policy 2026 and assess whether its prohibition-centred architecture is adequate to govern AI in the Indian judicial ecosystem as it actually operates.
2. To conduct a comparative analysis of regulatory frameworks adopted in the European Union, Singapore, Estonia, and the United States, with a view to identifying governance models suited to the Indian context.
3. To argue for mandatory prompt engineering literacy as a professional competence standard for judicial officers and legal practitioners.
4. To propose a six-point institutional and legislative policy roadmap suited to India's constitutional framework and technology trajectory.

II. Research Questions

1. Is a restriction-centred regulatory framework capable of functioning as meaningful AI governance in a legal system where AI is already pervasively integrated into professional practice?
2. What do the comparative experiences of the EU, Singapore, Estonia, and the United States suggest about effective AI governance in judicial institutions?
3. How should mandatory prompt engineering competence be structured as a professional standard for judicial officers?
4. What legislative and institutional reforms are required to prepare India's judicial system for the AGI, ASI, and quantum computing transition?

III. Scope and Limitations

This paper examines the regulatory architecture of AI governance in the Indian judicial system, with primary focus on the Gujarat High Court AI Policy 2026. The comparative analysis is confined to four jurisdictions: the European Union, Singapore, Estonia, and the United States. The paper does not undertake a technical evaluation of specific AI systems or language model architectures, nor does it assess the empirical performance of any particular judicial AI deployment. The policy roadmap proposed is institutional and legislative in character; implementation modalities and cost-benefit assessments lie outside the paper's scope. Claims about AGI and ASI timelines are necessarily speculative and are presented as governance planning considerations rather than technical predictions.

IV. Research Methodology

This paper employs a doctrinal and comparative legal methodology. Primary source analysis draws on the Gujarat HC AI Policy 2026, the EU AI Act (Regulation (EU) 2024/1689), Singapore's Judicial AI Governance Framework 2024, and relevant Indian legislation including the Digital Personal Data Protection Act 2023 and the e-Courts Phase III mission document. Constitutional provisions of the Constitution of India, in particular Articles 21, 225, and 227, are examined as primary sources.

The comparative component analyses regulatory instruments from four jurisdictions selected on the basis of their diversity of approach and relevance to the Indian governance debate. Secondary sources include publicly available government policy documents, judicial press releases, and technical standards publications from NIST. The paper's normative arguments are advanced through doctrinal reasoning, comparative synthesis, and a structured policy design exercise.

V. Literature Review

Academic engagement with AI governance in judicial contexts has grown substantially since 2022 but remains uneven in its geographic coverage. The majority of scholarship addresses the United States and European contexts, with the Indo-Pacific region receiving substantially less systematic attention.

Within the Indian context, the existing literature on judicial AI governance is sparse. Commentary has clustered around two moments: the Supreme Court's engagement with its own SUPACE platform, and the Gujarat HC policy of 2026. Neither has yet produced a comprehensive doctrinal analysis of the governance architecture question. The present paper seeks to address that gap.

The EU AI Act has generated a substantial body of commentary addressing its risk-classification architecture and its implications for high-stakes institutional AI deployment. Singapore's governance framework has been noted for its sandbox-led approach to judicial competence-building. The hallucination problem has attracted significant computer science and legal scholarship following the Mata v. Avianca incident in 2023, which demonstrated in a United States courtroom the professional liability consequences of unverified AI-assisted legal research.

The prompt engineering literature is primarily technical but has begun to attract legal education scholarship addressing the professional competence implications of language model fluency. The AGI and ASI governance literature is predominantly speculative but has generated significant institutional responses from the EU AI Office and equivalent bodies, which this paper treats as evidence of a regulatory consensus that governance architecture should be designed for the technology trajectory rather than the current moment alone.

VI. Introduction: A Policy Issued Against the Current

On a morning in April 2026, a legal intern at a subordinate court in Ahmedabad uses an AI assistant to research bail precedents in a narcotics case. Simultaneously, a trade consultant in Mumbai uses the same category of tool to draft a regulatory reply. A law student in Delhi runs a mock argument through an AI agent to sharpen her moot court preparation. A High Court judge, by contrast, is now bound by a freshly notified policy that treats those very tools as risks requiring strict institutional containment.

This is not a thought experiment. It is a precise description of the Indian legal profession in 2026. The Gujarat High Court's policy on judicial AI use reflects careful legal drafting, a serious engagement with constitutional principles, and a real awareness of the dangers that poorly governed AI poses to the administration of justice. Rigorous engagement with the policy requires asking whether its chosen instrument, prohibition, is adequate to the problem it seeks to solve.[1]

The question this paper poses is not whether the risks the policy identifies are real. They are. The question is whether, in 2026, a restriction-centred framework can function as meaningful governance of a technology that has already become integral to the professional ecosystem surrounding the courts it seeks to regulate. A policy aimed at what judges may not do with AI cannot govern what every advocate, clerk, paralegal, and litigant is already doing with it, just outside the courtroom threshold, every working day.

VII. Conceptual Framework

A. The Architecture of Restriction

The Gujarat HC policy constructs a two-lane regulatory framework: one very narrow lane of permitted AI engagement, and one very broad lane of absolute prohibition. The policy permits two categories of AI engagement: administrative scheduling through anonymised docket data, and retrieval-based research with mandatory cross-verification against primary legal databases before any reliance is placed on the output.[2]

The prohibited zone is drawn far more broadly. No AI system may participate in the assessment of facts, the evaluation of evidence, the exercise of discretion on bail or sentencing questions, or the construction of reasoned outcomes at any stage of a proceeding. Information about parties, witnesses, or pending matters cannot be entered into any publicly accessible AI platform.[3] Breach of the policy's terms is treated as professional misconduct with exposure under India's information technology, data protection, and criminal law frameworks.[4]

B. The Governance Gap: AI on Only One Side of the Bar

The legal profession's integration with AI tools is not a future prospect. It is a present reality. The judiciary's own digital infrastructure tells the story. Hundreds of crores of pages of judicial records have been digitised under the e-Courts programme.[5] Dedicated Supreme Court platforms have brought AI-assisted legal research and machine-translated judgments across nineteen Indian languages into routine judicial practice.[6][7] A publicly accessible data grid lets any citizen track court proceedings in real time.[8]

A policy that prohibits a judicial officer from using an AI research tool, while the advocate filing submissions before that same officer has prepared every document in the file using an identical tool, does not remove AI from the courtroom. It ensures that AI operates on only one side of the bar, the side that faces no restriction. The policy's own preamble registers awareness of the generational dimension: that judicial officers now entering service are digitally native professionals for whom AI is an ambient environment, not an external tool.[9]

C. The Global Mirror: Four Jurisdictions, One Direction

---

[1] High Court of Gujarat, Policy on the Use of Artificial Intelligence in Judicial and Court Administration cl. 1 (2026) [hereinafter Gujarat HC AI Policy].

[2] Gujarat HC AI Policy, supra note 1, cls. 7.1, 7.2.

[3] Gujarat HC AI Policy, supra note 1, cl. 8.

[4] Gujarat HC AI Policy, supra note 1, cl. 12.

[5] e-Courts Mission Mode Project, Phase III, Dep't of Justice, Ministry of Law and Justice, Gov't of India 3 (2023) [hereinafter e-Courts Phase III].

[6] Supreme Court of India, SUPACE: Supreme Court Portal for Assistance in Court Efficiency, https://main.sci.gov.in (last visited Apr. 2026).

[7] Supreme Court of India, SUVAS: Supreme Court Vidhik Anuvaad Software, https://main.sci.gov.in (last visited Apr. 2026).

[8] National Judicial Data Grid, Ministry of Law and Justice, Gov't of India, https://njdg.ecourts.gov.in (last visited Apr. 2026).

[9] Gujarat HC AI Policy, supra note 1, pmbl.

The European Union's AI Act, in force since 2024, rests on proportionate risk management rather than categorical exclusion.[10] AI deployed in the administration of justice is classified as high-risk, activating pre-deployment conformity assessments, mandatory human oversight, and ongoing post-deployment monitoring.[11]

Singapore's judicial AI governance framework, published in 2024, starts from a premise that inverts the Gujarat HC policy's logic: judicial independence is protected not by distancing judges from AI tools, but by ensuring every judicial officer who uses those tools has been rigorously trained to do so. The framework operates through structured sandbox pilots and mandatory disclosure protocols.[12][13]

Estonia has incorporated AI assistance directly into the operational architecture of its court system through the KIS case management platform, integrating AI-assisted scheduling and automated transcription with mandatory human certification as standard features. United States federal and state courts have settled on a disclosure-centred model: AI-generated content in court filings must be identified, and the responsible officer must certify independent verification. The consequences for submitting AI-generated fabrications as authentic authority have proved significant.[14]

Table 1: Comparative Regulatory Framework Overview

Regulatory Approach	**India (Gujarat HC 2026)**	**EU (AI Act 2024)**	**USA (Exec. Order 2023)**	**Singapore (AI Gov Framework)**
Stance on AI in Courts	Ultra-restrictive; admin and research only	Risk-based; high-risk AI regulated, not banned	Guidance-based; sector-specific principles	Collaborative; sandbox-led governance
GenAI Permitted?	No; public LLMs barred from case work	Yes, under safeguards and conformity assessment	Yes, with disclosed safeguards	Yes, with human oversight
Accountability Model	Personal judicial accountability; strict discipline	Provider and deployer liability; CE marking	Agency oversight; NIST framework	Industry-government co-regulation
Hallucination Response	Prohibition-led; verify all outputs	Transparency obligations; accuracy requirements	Agency discretion on enforcement	Technical safeguards required
AGI / ASI Readiness	Silent; policy may become obsolete	AI Office for frontier model oversight	Executive action limited;	Forward-looking; iterative framework

[10]Council Regulation 2024/1689, 2024 O.J. (L 1689) 1 (EU) [hereinafter EU AI Act].
[11]EU AI Act, supra note 10, arts. 6, 9-11.
[12]Singapore Judiciary, Artificial Intelligence Governance Framework for Courts pt. II (2024) [hereinafter Singapore AI Framework].
[13]Singapore AI Framework, supra note 12, pt. III.
[14]Mata v. Avianca, Inc., No. 22-cv-1461 (PKC), 2023 WL 4114965, at *1 (S.D.N.Y. June 22, 2023).

Regulatory Approach	**India (Gujarat HC 2026)**	**EU (AI Act 2024)**	**USA (Exec. Order 2023)**	**Singapore (AI Gov Framework)**
			Congress lagging	

Source: Author's compilation from primary regulatory instruments cited herein.

VIII. Findings / Results

A. The Prompt Engineering Imperative

The single most effective regulatory intervention available to India's judiciary in 2026 is the establishment of mandatory AI literacy, with prompt engineering at its centre, as a condition of professional competence for every judicial officer and court administrator in the country.

Prompt engineering is the disciplined practice of formulating instructions to AI systems in ways that constrain their outputs toward accuracy, source-anchoring, and explicit acknowledgement of uncertainty. A judicial officer with prompt engineering competence can direct an AI research tool to retrieve only verified case law from named authoritative databases, flag any citation it cannot confidently verify, and present outputs in a format that supports systematic independent verification. An officer without that competence asks the tool a question and accepts the answer. That is the difference between AI as a governed instrument and AI as a liability.

The fear of hallucination that drives the policy's strictest provisions is understandable. The early 2026 incident before the Supreme Court demonstrated what unmanaged AI use can produce in a legal context.[15] But prohibition does not reduce that risk. It relocates it underground, where it operates without institutional visibility, training support, or accountability. The appropriate response is to equip every officer who will encounter AI-generated material with the skills to identify its failure modes and manage them professionally.

B. The Technology Horizon: AGI, ASI, and Quantum Computing

The Gujarat HC policy addresses the large language model era of AI. It says nothing about what comes next. The policy allows AI for research because research is retrieval; it prohibits AI from reasoning because reasoning is judgment. The systems that leading research organisations project will be operational within two to five years of 2025 will not respect that boundary. An AGI-capable legal assistant will reason about the applicability of precedents to a fact pattern, identify tensions in the jurisprudence, evaluate competing lines of authority, and propose resolutions. A regulatory framework built on the premise that AI is a retrieval tool will be structurally incoherent when applied to a reasoning system.

The European Union has created dedicated institutional machinery for oversight of frontier AI through the AI Office. India, meanwhile, has proposed an AI Safety Institute but still lacks a judicial governance architecture designed for highly capable frontier systems. The quantum computing transition represents a third governance horizon the policy does not acknowledge. NIST has finalised initial post-quantum cryptographic standards and urged organisations to begin migration planning now. Judicial data systems should be incorporated into national post-quantum security planning during the current development cycle rather than after vulnerable systems must be replaced at scale.

IX. Recommendations / Suggestions

The following six-point policy roadmap is proposed.

1. Establish a National Judicial AI Governance Commission.

A statutory commission anchored jointly in the Supreme Court's e-Committee and the Ministry of Law and Justice should develop and maintain a national framework for AI governance in judicial proceedings; commission annual independent bias audits of AI systems deployed across court

[15]Supreme Court of India, Press Release: Chief Justice Surya Kant's Statement on AI-Generated Petitions (Feb. 2026).

infrastructure; maintain active liaison with equivalent bodies in the EU and Singapore; and issue rolling guidance as the technology evolves through the AGI and ASI phases.

2. Replace Prohibition with Mandatory Disclosure.

Every court document, bench memo, or research note prepared with AI assistance should carry a standardised notation identifying the tool used, the nature of the assistance provided, and the verification steps taken. This approach builds a transparent institutional record and preserves full personal accountability without requiring the pretence that prohibition is enforceable against a generation of professionals for whom AI assistance is a default working mode.

3. Mandate AI Literacy as a Qualification Condition.

Completion of a structured Judicial AI Literacy Programme should become a condition of appointment and promotion for all judicial officers. The curriculum should cover: how language models generate text and produce false information; practical prompt formulation skills calibrated to legal research; structured verification habits for evaluating AI outputs; and confidentiality and data protection obligations applicable to AI use in judicial settings.[16]

4. Procure Enterprise AI Infrastructure.

The procurement of judiciary-specific AI infrastructure under formal data processing agreements that bar providers from training on judicial inputs would address the confidentiality concern that underlies the current blanket prohibition. The existing e-Courts Phase III budget[17] provides the funding vehicle for this procurement.

5. Amend the Digital Personal Data Protection Act 2023.

Three significant gaps require legislative attention: the absence of a clear processing framework for litigant data handled by AI systems in court infrastructure; unresolved liability questions when AI deployed in a judicial context produces a data breach; and the absence of any provision for the quantum computing transition that will render the current encryption baseline obsolete before the end of the decade.[18]

6. Extend Governance to the Bar and Legal Education.

The Bar Council of India should establish mandatory AI competency standards as part of continuing legal education requirements for practising advocates. Law schools should treat AI governance as a core compulsory subject within the BA LLB and LLB programmes. The next generation of Indian legal professionals must graduate equipped both to use AI effectively and to be held professionally accountable for how they use it.

X. Conclusion

Regulatory frameworks that seek to contain technologies already embedded in professional practice rarely achieve containment. What they tend to achieve is a formal boundary that defines the terms of a practice continuing informally, without the transparency, training, or accountability mechanisms that good governance would otherwise require. The Gujarat HC policy is a document of institutional seriousness deployed in precisely this situation.

The honest assessment is this: the window for keeping AI outside the Indian courtroom closed before the policy was notified. Every participant in the justice system except the judicial officer on the bench is already working with these tools. A framework that addresses only what judges may do, while leaving entirely unaddressed the AI-mediated submissions arriving on their desks every morning, does not govern the problem. It governs one corner of the problem while the rest continues unremarked and unmanaged.

India has the institutional ambition, the constitutional framework, the existing technology investment, and the legal talent to build a judicial AI governance architecture that leads the world rather than lags behind it. The e-Courts programme demonstrates institutional capacity for large-scale digital transformation. The Supreme Court's own AI platforms demonstrate a willingness to engage

[16] Gujarat HC AI Policy, supra note 1, cl. 3.6.

[17] e-Courts Phase III, supra note 5, at 14.

[18] Digital Personal Data Protection Act, No. 22 of 2023 (India).

with the technology. What has been absent is a governance philosophy equal to the scale of the challenge: one that treats AI not as a threat to be walled out but as a force to be directed, trained around, and held accountable.

The High Court of Gujarat has acted with constitutional fidelity and genuine institutional care. The Government of India must now act with the institutional ambition and technological foresight that the moment requires. The two imperatives are not competing. Properly understood, they are the same.

XI. References / Bibliography

Primary Sources

i. INDIA CONST. arts. 21, 225, 227.
ii. High Court of Gujarat, Policy on the Use of Artificial Intelligence in Judicial and Court Administration (2026).
iii. Digital Personal Data Protection Act, No. 22 of 2023 (India).
iv. Information Technology Act, No. 21 of 2000 (India).
v. Bharatiya Nyay Sanhita, No. 45 of 2023 (India).
vi. Council Regulation 2024/1689, 2024 O.J. (L 1689) 1 (EU) (AI Act).
vii. Exec. Order No. 14,110, 88 Fed. Reg. 75,191 (Oct. 30, 2023).
viii. Singapore Judiciary, Artificial Intelligence Governance Framework for Courts (2024).
ix. Mata v. Avianca, Inc., No. 22-cv-1461 (PKC), 2023 WL 4114965 (S.D.N.Y. June 22, 2023).

Secondary Sources

i. e-Courts Mission Mode Project, Phase III, Dep't of Justice, Ministry of Law and Justice, Gov't of India (2023).
ii. National Judicial Data Grid, Ministry of Law and Justice, Gov't of India, https://njdg.ecourts.gov.in (last visited Apr. 2026).
iii. Supreme Court of India, SUPACE: Supreme Court Portal for Assistance in Court Efficiency, https://main.sci.gov.in (last visited Apr. 2026).
iv. Supreme Court of India, SUVAS: Supreme Court Vidhik Anuvaad Software, https://main.sci.gov.in (last visited Apr. 2026).
v. Supreme Court of India, Press Release: Chief Justice Surya Kant's Statement on AI-Generated Petitions (Feb. 2026).
vi. Nat'l Inst. of Standards and Tech., Post-Quantum Cryptography Standardization Project, https://csrc.nist.gov/projects/post-quantum-cryptography (last visited Apr. 2026).

"Integrating Artificial Intelligence Tools in Legal Education: Opportunities, Ethical Challenges and the Future of Legal Learning"

Bhagubai Changu Thakur College of Law, New Panvel

Chaitanya Arunkumar Pulekar

Abstract

Context- Artificial Intelligence (AI) is quickly changing the field of education, including legal education. Tools such as AI legal research assistants, document analysis software, and language models are changing the way law students study, research, and develop legal writing skills. This research studies the introduction of AI tools into the legal education and how they affect teaching, learning, and professional skill development. With rapid digitalisation, the growth of online learning, and the increasing use of AI-based research tools, legal training is evolving across the world. In India, the need to modernise legal education and prepare students for technology-driven legal practice makes this topic especially important. This study uses a doctrinal, desk-based research method. These include journal articles, government and institutional reports, legal education policy documents as well as additional credible online sources. Thematic analysis was used to examine the opportunities, ethical concerns, and future impact of AI in legal education. The research finds that AI tools make legal research faster, improve access to legal information, support personalised learning, and helps to develop legal drafting and case analysis. However, several concerns also exist. These include risks to academic integrity, over-dependence on AI, data privacy issues, bias in AI outputs, and the absence of clear institutional guidelines. Also, having a digital divide of the student and not training teachers also makes it more difficult to adopt AI. Finally, the study highlights that AI would play a supportive rather than replacing role in teaching students. Law institutions should create ethical guidelines, improve digital literacy, and update the curriculum to ensure a balance between technological innovation, academic integrity, and professional responsibility.

Keywords-Artificial Intelligence; Legal Education; Digital Learning; Legal Research Tools; Ethics in AI

Research Objectives-

This study examines the opportunities, ethical challenges, and future implications of integrating AI tools into legal education.

The objectives are:

(1) To analyse the benefits of AI tools in enhancing legal learning,

(2) To identify ethical and academic risks associated with AI use, and

(3) To understand how legal education can prepare for change responsibly given technology.

Research Questions-

1. What are AI tools to improve legal education quality and accessibility?
2. What ethical and academic integrity concerns arise from AI use in legal learning?
3. What ethical and academic integrity problems does your use in legal learning account for?

Scope-

This doctrinal research investigates the teaching of Artificial Intelligence (AI) tools in legal education based on a review of existing literature, policies and academic discourse It focuses on the opportunities AI offers in improving legal research, writing, accessibility, and personalised learning, with special reference to the Indian context and the growing need to modernise legal education. The study also explores ethical and professional concerns such as academic integrity, data privacy, bias, and and the next steps in how AI will impact legal learning and practice.

Limitations-

The research is based solely on secondary sources and does not include any field work such as survey or interview. The research does not analyse the technical functioning of AI tools or evaluate

specific software in detail. Due to limited India-specific research, the paper refers to global literature, and since AI is rapidly evolving, some observations may need updates in the future.

Research Methodology-

This research follows a doctrinal research methodology that is based on the study and analysis of existing literature, legal materials, and policy documents regarding the utilization of Artificial Intelligence in legal education. The study relies completely on secondary sources of information. Data was collected from peer-reviewed journal articles, books, government and institutional reports, legal education policies, and trusted academic databases. The selected sources mainly focus on topics such as AI in education, legal technology, academic integrity, digital learning, and the ethical issues connected with AI use.

After collecting the materials, the information was examined and organised into key themes such as the AI uses in legal learning, the challenges and risks involved, ethical concerns, and the impact on legal education. Using an thematic analysis method, common themes and patterns were identified within the literature. It allows us to systematically study both the opportunities and challenges that AI tools bring, as well as provide a clear and structured framework to address the research questions.

Literature Review-

The legal profession is undergoing a transformation relevant to the discussion of forming partnerships with technology that is automating many routine legal tasks (McGinnis and Pearce, 2014). She maintains that changes within law schools are necessary to train students for a future with lawyers collaborating with smart machines. The report explains the importance of incorporating tech skills, legal analytics and digital research methods into the classroom discussion.[1]

Ashley (2017) explains that AI tools significantly improve the speed and efficiency of legal research.This research demonstrates that AI-driven applications can sift through millions of cases, find germane precedents, and help students learn to engage in legal reasoning. The study indicates that AI can be used in legal education to train students on research efficiently so they would actually focus more on critical thinking and problem-solving ideas rather than devoting countless hours to manual research.[2]

Selbst et al. (2019) outlines the potential moral dangers involved in using AI within educational and career pathways. For instance, they warn about bias in algorithms, lack of transparency and hypertension due to excessive automated systems reliance. When it comes to legal education, their research indicates students might become reliant upon AI content, which could impair originality, academic integrity and independent legal reasoning.[3]

Holmes, Bialik and Fadel (2019) discuss the role of AI in creating personalised learning experiences. Their research indicates that AI can personalize lesson planning based on the requirements and learning speed of particular students. In legal education, this allowed students to grapple with complex legal subjects, develop writing skills and receive immediate feedback. Nevertheless, the authors also insist that teachers need to closely monitor students to ensure that they use AI tools responsibly. [3]

Cath et al. (2018) examine the governance and regulation of artificial intelligence. T The study underlines the necessity of reframing curricular objectives for legal education, integrating knowledge into education curricula focused on digital literacy, ethics and technology law to prepare future lawyers with the necessary resources for a technologically integrated legal society. [4]

---

[1] Ashley, K. D. (2017). *Artificial Intelligence and Legal Analytics: New tools for law practice in the digital age*. Cambridge University Press.

[2] Selbst, A. D., Boyd, D., Friedler, S., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems.

[3] Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in Education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

[4] Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial intelligence and the "good society.

Research Problem Statement-

Legal education has always been based on classroom teaching textbooks, and manual legal research. But AI is changing the legal profession at a rate that has implemented tools to automate research, analyse documents and provide assistance in drafting written documents. But even against this backdrop of massive technology change, many law schools are still slow to update their teaching techniques and course offerings to include AI-based instruction. This gap poses a problem because it means that law graduates may go into practice without the requisite technological skills to successfully participate in modern legal practice.

Simultaneously, the expanding implementation of AI in the legal education context has dire moral and academic implications. The concerns about over reliance on AI tools, academic dishonesty and data privacy along with the risk of biased or inaccurate AI-generated content raise questions as to how best these technologies should be used responsibly. Moreover, there are no clear rules and policies to govern the use of AI in legal education.

Accordingly, the principal research question analysed in this paper is how to use AI-powered tools responsibly and meaningfully in legal education while ensuring ethical compliance, academic integrity and skills development. It aims to plot the possibilities and protections required for AI-based learning in law.

Introduction-

AI is impacting global professional education at a rapid rate. Legal education, once upon a time, primarily the product of textbooks and case law analysis taught in classrooms is going through an innovative technological change. Law students and legal professionals already use AI-powered tools like legal research platforms, document review software, generative AI writing assistants. The rising demand for technology-led legal services and the presence of the huge boom in legal tech start-ups in India point towards the modernisation of legal education so that students are well-prepped to meet future challenges of actual practice.

As social, economic and technological developments have occurred over the years, the legal profession has always adapted. With the recent rise of digital revolution, a new genre of work environment came into picture where this legal work was not constricted to physical libraries and manual research. Legal practitioners rely on huge databases, internet case sites and automated instruments that assist in any respect orders. Consequently, law firms are increasingly making it clear that they want new trainees to prepare for the modern-day realities of legal practice and this pressure continues to increase on law schools to adapt their teaching methods and curricula accordingly. It has thus made the integration of Artificial Intelligence in legal education not an option but a necessity.

Using AI, Legal education can be made more interactive, accessible and pragmatic Add hub for your language Intelligent research assistants help students to find relevant judgements in a matter of seconds, analyse and navigate through vast amounts of legal documents, or even receive feedback on their legal writings. This technology can facilitate students mastering complex legal theories more quickly while enabling teachers to devote greater time to areas such as critical thinking, ethics and problem-solving skills. AI also plays a vital role in personalised learning where each student has its own pace, making legal education more inclusive and flexible.

However, the continued and expanding use of AI in legal education also raises important ethical and professional concerns. Questions about academic integrity, over-dependence on technology, data privacy, and the reliability of AI-generated legal information must be carefully considered. The risk is that students might depend too much on automated tools without developing essential legal reasoning and analytical abilities. Moreover the applications of AI in legal education brings challenges related to bias, transparency, and accountability, which are particularly significant in a profession that values justice and fairness.

Despite the increasing availability of AI tools, their integration into legal education remains unstructured and inconsistent. In many cases, there are no direct guidance on whether the use of AI in assignments, research and examinations is ethical or not. Concerns regarding plagiarism, academic

integrity, data privacy, and algorithmic bias are becoming more prominent. Additionally, there is uncertainty about how AI affects critical thinking, legal reasoning, and professional ethics among law students. [5]

This paper explores the opportunities that may arise along with ethical and moral dilemmas by the addition of AI tools into legal education. It seeks to examine how AI can enhance teaching and learning while also identifying the risks to ensure responsible use. The paper also addresses the future of legal learning and how law schools can leverage technological innovations while staying true to traditional legal values. This study shows the need for a thoughtful and balanced approach to the digital transformation of legal education.

Analysis-

1. AI Opportunities In Legal Education

AI has opened great prospects for raising the quality and speed of legal education. One of its biggest benefits is the speed and accuracy it provides for legal research. One advantage is the speed and accuracy it brings to legal research. Traditional legal research often requires students to spend long hours searching through case law, statutes, and commentaries. AI-powered research tools can quickly analyse large databases and identify relevant cases, saving time and focusing more on understanding and analysing legal issues.

AI tools also assist in the legal drafting skills by suggesting better structure, grammar, and legal terminology, which enhances the overall quality of written work such as assignments, moot court memorials, and legal opinions. AI tools help provide personalised learning experiences by adjusting explanations according to individual learning needs.

Many students also use AI study assistants to explain complex legal concepts in simple language, generate notes, prepare for examinations, and create practice questions. These technologies make legal research faster, more affordable, and available at any time, which especially benefits students with limited access to physical libraries.

Legal research tools help students quickly locate relevant case law, provide summaries of judgments, and suggest proper citations which helps students execute better legal research. AI case-summarisation tools also help simplify complex and long judgements through quick identification of core facts and legal principles within seconds. [6]

AI is also worldwide because it gives the availability of legal systems across countries which allows students to study international law or compare different legal systems in a world stage. This is especially helpful to students who may lack access to comprehensive law libraries or costly legal databases. Overall, AI helps make legal education more accessible, efficient, and skill oriented.

2. Ethical and Academic Integrity Challenges

Despite its benefits, the application use of AI in legal education raises serious ethical and academic concerns. One of the controversial issues is the risk of plagiarism and misuse of AI tools in academic work. Students may rely on AI to generate assignments or research material and no legitimately attribute which challenges traditional ideas of originality and academic honesty.

Another concern is the risk of over-dependence on technology. There is also the risk of becoming too dependent on technology. Reading judgments is essential to develop independent legal reasoning skills, and if students shift too much burden on AI-generated summaries or answers rather than reading judgments themselves, it may become detrimental. This can weaken their critical thinking and analytical abilities.

Because AI systems rely on existing data which may contain social or historical biases, they can sometimes generate biased or incorrect information. This creates a risk of students unknowingly using

[5] Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in Education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

[6] Selbst, A. D., Boyd, D., Friedler, S., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems.

inaccurate or biased legal content. This poses a risk that students will inadvertently use unreliable or biased legal written material.

Another significant issue is academic integrity. AI tools could also be misused to write assignments, take assessments or create content that crosses a line into plagiarism if not disclosed properly. Moreover, the aspect of AI in doing legal research is not completely reliable.

Law also requires human judgement, moral reasoning, and ethical interpretation, which technology cannot fully replace. Excessive reliance on AI may reduce creativity and original thinking among students. In addition, unequal access to AI tools has caused a widening gap in experience with advanced technology. [7]

3. Preparing Future Lawyers

With the increasing prevalence of AI in law, it is critical that legal profession makes it essential for law schools to prepare students for technology-driven legal practice. Future lawyers will be expected to work alongside AI tools in areas such as legal research, document review, and contract analysis. As a result of this development, educating law students to be AI literate has emerged as a significant learning goal in legal education. Law schools need to update their curriculum to include subjects related to legal technology, digital research skills, and and AI ethics. Integrating traditional legal education with technology can make sure that graduates are more ready to practice law in a modern environment. By training students as such, lawyers who can use technology effectively while toting professional responsibility and high ethical standards will be hatched. Lawyers working along with AI rather than without it is the future of legal education. With technology hitting the mainstream legal practice learning will shift from rote memorisation to analytical and solving problems. AI-assisted tools will complement research done manually, arming students with precisely the types of skills aligned with what practitioners require.

Findings-

1. AI Opportunities In Legal Education

Artificial Intelligence has significantly improved the efficiency and accessibility of legal education. AI-powered research tools enable students to locate relevant case laws, statutes, and legal materials within a short time, reducing the effort required for manual research. As a result, students can devote more time to analysing legal issues and developing arguments rather than searching for information. The personalized adjustability of the AI writing assistants boosts clarity, structure, and language accuracy for a more standardized quality drafting.

Another, related point, is that AI tools have the capacity to provide students with worldwide access to legal resources so they can analyse comparative as well as international developments amongst legal systems. Access to this information is often helpful for students who may not have a large access to law libraries. Overall, the findings suggest that AI contributes to more efficient learning, skill development, and wider access to legal knowledge.

2. Ethical and Academic Integrity Challenges

The research identifies several ethical and academic concerns linked to the growing use of AI tools in legal education. An important revelation is the increasing risk of academic dishonesty by some students, who may adopt AI-generated material for their class tasks without disclosing it. This raises questions about originality, authorship, and academic honesty. The study also finds evidence of over-dependence on AI tools, which may reduce students' engagement with primary legal sources such as judgments and statutes. This could stop the development of critical thinking and independent reasoning skills, which are essential for legal professionals. Another key finding is the presence of bias and inaccuracies in AI-generated outputs. Since AI systems rely on existing data, they may sometimes produce incomplete, outdated, or biased information. These risks underscore the importance of responsible and supervised use of AI in legal education. [8]

Should AI Be Banned or Regulated?

[7] UNESCO. (2021). *AI and education: Guidance for policy-makers*. UNESCO Publishing.

[8] Susskind, R. (2019). *Tomorrow's Lawyers: An introduction to your future* (2nd ed.). Oxford University Press.

Artificial Intelligence should be regulated rather than banned in legal education. Completely prohibiting AI is not practical because technology has become an unavoidable part of modern learning and the legal profession already relies on AI tools in many areas. Instead of banning it, institutions should focus on creating clear rules and guidelines to ensure responsible and ethical use.

3. Preparing Future Lawyers

The use of AI in the legal profession makes technological literacy an essential skill for future lawyers. Legal practice is increasingly using AI tools for tasks such as legal research, document review, and case management. Therefore, law students must be trained to use these tools effectively and ethically. The research also finds a gap between the current legal education curriculum and the technological skills required in modern legal practice. Many law schools have not yet fully incorporated AI-related subjects into their programs. In addition, the research uncovers a disconnect between current legal education curriculum and the necessary technological skills needed to practice law today. The implications indicate that curriculum reform, digital literacy, and an ethical awareness are needed to prepare lawyers to live in a technology-driven legal environment.

An ideal approach to legal education is to blend AI tools with traditional learning practices. AI can assist students by speeding up research, summarising cases, and improving drafting, but the responsibility for analysis, interpretation, and reasoning should remain with humans. This balanced method allows students to benefit from technology while still developing strong critical thinking and legal skills. [9]

Recommendations-

Develop clear institutional guidelines for AI use – Law schools should create formal policies that explain how students and teachers can use AI tools responsibly. These guidelines should clarify what counts as acceptable assistance, how AI use should be disclosed in assignments, and what actions may be considered academic misconduct. Clear rules will help reduce confusion and misuse.

1) **Promote responsible use through disclosure policies**-Students should be encouraged to use AI as a support tool while openly acknowledging its use in academic work. A transparent disclosure policy can protect academic integrity while allowing students to benefit from technological advancements.
2) **Introduce AI literacy and legal technology courses**- Legal education should include subjects that teach students about AI tools, their benefits, limitations, and ethical risks. Courses on legal technology, digital research skills, and AI ethics will help students use technology responsibly and prepare them for modern legal practice.
3) **Provide training for teachers and faculty**- Educators need proper training to understand AI tools and guide students effectively. Workshops and development programs for faculty could assist teachers integrate AI into teaching while maintaining academic standards.
4) **Encourage critical thinking alongside AI use**- Law schools should design assignments and assessments that focus on analysis, reasoning, and problem-solving rather than simple information collection. This will ensure that AI supports learning without replacing independent thinking.
5) **Encourage research on AI and legal education**- Universities should promote further research on the long-term impact of AI on legal learning and professional practice. Ongoing research will help institutions develop better policies and teaching strategies in the future. [10]
6) **Bridge the digital divide**-. All students should have equitable access to AI tools and digital resources, supported by the institutions. Providing institutional subscriptions, computer labs, and technical support can reduce inequality in access to technology.

[9] Pasquale, F. (2015). *The Black Box Society: The secret algorithms that control money and information*. Harvard University Press.

[10] OECD. (2021). *OECD digital education outlook 2021: Pushing the frontiers with AI, blockchain and robots*. OECD Publishing.

7) **Regularly update curriculum and policies**- Since AI technology is rapidly evolving, legal education institutions should regularly review and update their curriculum, policies, and teaching methods. Continuous evaluation will help maintain a balance between innovation and academic integrity.

Conclusion-

Artificial Intelligence is gradually transforming legal education by improving research efficiency, enhancing legal writing skills, and increasing access to legal knowledge. There is a possibility that AI tools can make legal learning more interactive, flexible, and skill-oriented. Similarly, the research draws attention to key issues such as academic integrity, overreliance on technology, risks associated with data privacy and risk of bias in content generated by AI. This was equated to challenges faced in other industries, demonstrating that adopting AI technologies into legal education should be informed by robust ethical frameworks and demonstrated institutional policies.

First, AI should be perceived as an assistant and not a substitute for human learning reasoning. Legal education is in dire need of reform to integrate digital literacy, responsible use of AI, and experimentation with legal technology. The practice of integrating traditional teaching with technological innovation (rather than replacing the former with the latter) can put legal institutions in a position where they are able to train future lawyers who are not only skilled in technology—located around information ethics—but even more accountable to ethical standards. In conclusion, a balanced and regulated way will undoubtedly provide that AI would help to reinforce legal education while protecting the foundation values of the legal vocation.

References-

1. Ashley, K. D. (2017). *Artificial Intelligence and Legal Analytics: New tools for law practice in the digital age*. Cambridge University Press.
2. Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial intelligence and the "good society": The US, EU, and UK approach. *Science and Engineering Ethics, 24*(2), 505–528.
3. Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in Education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.
4. Luckin, R., et al. (2016). *Intelligence Unleashed*. Pearson. UNESCO. (2021). *AI and Education Guidance*. American Bar Association. (2023). *Future of Legal Education Report*.World Economic Forum. (2020). *Future of Jobs Report*.
5. McGinnis, J. O., & Pearce, R. G. (2014). The great disruption: How machine intelligence will transform the role of lawyers in the delivery of legal services. *Fordham Law Review, 82*(6), 3041–3066.
6. Pasquale, F. (2015). *The Black Box Society: The secret algorithms that control money and information*. Harvard University Press.
7. Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.
8. Selbst, A. D., Boyd, D., Friedler, S., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 59–68.
9. Susskind, R. (2019). *Tomorrow's Lawyers: An introduction to your future* (2nd ed.). Oxford University Press.
10. UNESCO. (2021). *AI and education: Guidance for policy-makers*. UNESCO Publishing.
11. \OECD. (2021). *OECD digital education outlook 2021: Pushing the frontiers with AI, blockchain and robots*. OECD Publishing.

Theoretical Analysis of AI-Induced Cognitive Offloading and its Impact on MHL of Adolescents

Author**:**

Sakshi Sachdev

Counselling Psychologist

PhD Research Scholar, Chhatrapati Shivaji Maharaj University (CSMU)

Co-Author:

Dr. Avani M Mistry

Head, Department of Law

Chhatrapati Shivaji Maharaj University

Abstract

Artificial Intelligence (AI) and digital platforms have become deeply embedded in modern existence, fundamentally reshaping the landscapes of human cognition, behavioral patterns, and emotional well-being. This theoretical article investigates the profound psychological impacts of AI-driven platforms on late adolescence—a critical developmental period marked by neuroplasticity and the refinement of executive functions. The central focus of this inquiry is "Cognitive Offloading": the systemic externalization of critical reflection, synthesis, and evaluative thinking to algorithmic systems. While these platforms enhance efficiency, they simultaneously curate "choice architectures" that create feedback loops, reinforcing existing user behaviors and significantly limiting individual autonomy. Drawing upon the ***Extended Mind Thesis*** *and* ***Cognitive Load Theory****, this paper argues that the pervasive reliance on AI for both academic tasks and emotional processing facilitates a "hollowing" of internal cognitive architecture. By delegating higher-order mental processes to black-box algorithms, adolescents bypass the "desirable difficulties" essential for cognitive growth and psychological resilience. This process is hypothesized to diminish* ***Mental Health Literacy (MHL)*** *by replacing active, effortful introspection with algorithmically-driven content that prioritizes engagement over accuracy. Such offloading reduces the capacity for self-regulation and diminishes the ability to recognize or articulate complex emotional states, potentially exacerbating symptoms of anxiety and depression. Through a critical synthesis of psychological theory and AI ethics, this article highlights the risks of "cognition-by-proxy" and argues for the urgent implementation of human-centered design and ethical frameworks. By prioritizing transparency and cognitive agency, stakeholders can mitigate the psychological risks inherent in behavioral design and algorithmic bias, ensuring that the development of technology does not compromise the psychological development of adolescents.*

Keywords: Cognitive offloading, Mental health literacy, Adolescent psychology, Artificial intelligence, Algorithmic bias, Digital platforms

I. Introduction:

The Algorithmic Mediation of the Adolescent Mind

The pervasive integration of Artificial Intelligence (AI) into everyday life has transformed it from a technological tool into an external cognitive resource that increasingly shapes human thinking, memory, and decision-making[1]. For the "iGen" and subsequent cohorts, digital platforms are not just tools but environmental constants that shape how information is processed, synthesized, and stored.[2] In late adolescence—a developmental window characterized by significant neuroplasticity, the pruning of synaptic connections, and the refinement of executive functions in the prefrontal cortex—

[1] Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

[2] Livingstone, Sonia (2017) Book review: iGen: why today's super-connected kids are growing up less rebellious, more tolerant, less happy – and completely unprepared for adulthood. Journal of Children and Media, 12 (1). pp. 118-123. ISSN 1748-2798 DOI: 10.1080/17482798.2017.1417091

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds and Machines, 30*(4), 681–694.

Sam Gilbert, Cognitive Offloading (2019)

the reliance on AI-driven systems introduces a novel psychological phenomenon: systemic **"Cognitive Offloading.[3]"**

Cognitive offloading refers to the externalization of mental tasks to physical or digital artifacts to reduce the cognitive demand on the biological individual. [4]While historically human cognition has always leaned on external scaffolds—such as the transition from oral traditions to written notebooks—the advent of adaptive, predictive, and generative AI marks a qualitative and ontological shift.[5] Unlike static tools, modern AI performs "integrative reasoning"—tasks of synthesis, evaluation, and creative generation that were previously the exclusive domain of the biological mind.[6] This paper seeks to establish that this frictionless externalization leads to a "hollowing" of the internal cognitive architecture, which in turn diminishes **Mental Health Literacy (MHL)**.

The central thesis of this analysis is that as adolescents increasingly delegate their critical thinking and emotional processing to algorithmic systems, their capacity for self-regulation and psychological insight atrophies. By exploring the intersection of developmental psychology and AI ethics, we aim to uncover how the reduction of cognitive effort in the digital sphere translates into a deficit in mental health knowledge and introspective depth.

II. Theoretical Frameworks: The Extended Mind and Cognitive Load

A. The Extended Mind Thesis (EMT) and the Parity Principle

Understanding how AI affects teenage psychology requires first considering the mind as a vast system[7]. Proposed by Clark and Chalmers (1998)[8], the **Extended Mind Thesis** posits that the mind does not exclusively reside within the skull but extends into the environment via "coupled systems." When an adolescent uses an AI chatbot to draft an essay or a mood-tracking algorithm to interpret their emotional state, that digital resource becomes a functional part of their cognitive machinery.[9]

However, the EMT assumes a "parity principle"—that the external resource should be as reliable and transparent as an internal mental process. In the age of "black-box" algorithms, this parity is compromised.[10] The adolescent "extends" their mind into a system that is not neutral but driven by commercial incentives and algorithmic biases. This "active externalism" means that the adolescent's sense of self and their cognitive identity are increasingly determined by external, opaque processes. [11]This raises critical questions about the quality of the "extended" cognition: if the external tool is biased or optimized for engagement rather than truth, the internal psychological state of the adolescent follows suit[12].

B. Cognitive Load Theory (CLT) and the Path of Least Resistance

Cognitive Load Theory (CLT) asserts that working memory possesses a limited processing capacity, making effective learning dependent on managing cognitive demands and minimizing unnecessary mental load[13]. Traditionally, offloading "low-level" tasks (like basic arithmetic) was seen as a rational trade-off to free up "germane load" for higher-order learning.[14] However, AI-induced

---

[3] Sarah-Jayne Blakemore & Suparna Choudhury, "Development of the Adolescent Brain: Implications for Executive Function and Social Cognition," Journal of Child Psychology and Psychiatry, 47(3–4), 296–312 (2006).
Casey, B. J., Jones, R. M., & Hare, T. A. (2008).

[4]Evan F. Risko & Sam J. Gilbert, "Cognitive Offloading," Trends in Cognitive Sciences, 20(9), 676–688 (2016).

[5]Andy Clark, Supersizing the Mind: Embodiment, Action, and Cognitive Extension (Oxford University Press, 2008).

[6] Luciano Floridi & Massimo Chiriatti, "GPT-3: Its Nature, Scope, Limits, and Consequences," Minds and Machines, 30(4), 681–694 (2020).

[7] Watney, Murdoch. (2024). Privacy and Personal Information Protection by Social Media Companies in an AI era. European Conference on Social Media. 11. 266-272. 10.34190/ecsm.11.1.2089.

[8] Andy Clark & David J. Chalmers, "The Extended Mind," Analysis, 58(1), 7–19 (1998)..

[9] Andy Clark & David J. Chalmers, "The Extended Mind," Analysis, 58(1), 7–19 (1998).

[10] Shoshana Zuboff, The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power (PublicAffairs, 2019).

[11] Jenna Burrell, "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms," Big Data & Society, 3(1), 1–12 (2016).

[12] Eli Pariser, The Filter Bubble: What the Internet Is Hiding from You (Penguin Press, 2011).

[13] Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.

[14] John Sweller, "Cognitive Load During Problem Solving: Effects on Learning," Cognitive Science, 12(2), 257–285 (1988).

offloading often targets the higher-order tasks themselves—analysis, critical reflection, and emotional synthesis.

Adolescents, whose prefrontal cortex is still maturing, are naturally inclined toward the "path of least cognitive resistance."[15] When a digital platform provides an immediate, curated answer or an emotional "label" via an algorithm, the adolescent avoids the **"desirable difficulty"** necessary for cognitive growth.[16] This systemic avoidance leads to a decline in cognitive stamina.

Over time, the internal "mental muscles" required for complex psychological introspection begin to weaken, as the external algorithm performs the heavy lifting of meaning-making.[17] This shift represents a transition from "cognition-with-tools" to "cognition-by-proxy," where the individual becomes a passive recipient of their own cognitive outputs.

III. The "Hollowed Mind": Mechanisms of Internal Cognitive Architecture Erosion

In the traditional developmental trajectory of late adolescence, the brain undergoes a process of "synaptic pruning" and myelination, which streamlines the neural pathways required for complex reasoning and emotional self-regulation. This period is essential for building robust **Internal Mental Models**—the cognitive frameworks individuals use to understand the world and their own place within it.[18] However, the rise of AI-induced cognitive offloading threatens to disrupt this biological scaffolding, leading to what this paper defines as a "hollowing" of the internal cognitive architecture.

A. Erosion of the "Desirable Difficulty" and Schema Acquisition

According to cognitive psychology, learning and retention are improved by "desirable difficulties"—tasks that call for laborious thinking.[19] When an adolescent grapples with a complex text or attempts to synthesize disparate ideas for a PhD assignment, the mental strain signals the brain to encode that information into long-term memory, building a complex schema[20].

When AI systems—such as Large Language Models (LLMs)—provide immediate summaries or generate synthesized arguments, the adolescent bypasses the "encoding" stage of learning. [21]The external system provides the **product** of cognition without the user undergoing the **process**. Consequently, the internal schema remains superficial.[22] For a counseling psychologist, this is particularly concerning: if the cognitive architecture for critical reflection is never built, the individual lacks the internal resources to perform complex problem-solving or emotional nuance in real-world scenarios where an AI is not present.

B. Metacognitive Decay and the Illusion of Explanatory Depth[23]

One of the primary casualties of excessive offloading is **Metacognition**—the ability to monitor and regulate one's own thinking processes.[24] Teenagers are more likely to encounter the "Illusion of Explanatory Depth," a cognitive bias in which people think they comprehend a concept more thoroughly than they actually do because they have access to an outside source that clarifies it [25].

John Sweller, Paul Ayres & Slava Kalyuga, Cognitive Load Theory (Springer, 2011).

[15] B. J. Casey, Rebecca M. Jones & Todd A. Hare, "The Adolescent Brain," *Annals of the New York Academy of Sciences*, 1124, 111–126 (2008).

[16] Robert A. Bjork & Elizabeth L. Bjork, "Making Things Hard on Yourself, But in a Good Way: Creating Desirable Difficulties to Enhance Learning," in *Psychology and the Real World* (Worth Publishers, 2011).

[17] Evan F. Risko & Sam J. Gilbert, "Cognitive Offloading," *Trends in Cognitive Sciences*, 20(9), 676–688 (2016).

[18] Sarah-Jayne Blakemore & Uta Frith, *The Learning Brain: Lessons for Education* (Blackwell Publishing, 2005); B. J. Casey, Rebecca M. Jones & Todd A. Hare, "The Adolescent Brain," *Annals of the New York Academy of Sciences*, 1124, 111–126 (2008).

[19] Robert A. Bjork & Elizabeth L. Bjork, "Making Things Hard on Yourself, But in a Good Way: Creating Desirable Difficulties to Enhance Learning," in *Psychology and the Real World* (Worth Publishers, 2011).

[20] John Sweller, "Cognitive Load During Problem Solving: Effects on Learning," *Cognitive Science*, 12(2), 257–285 (1988).

[21] Evan F. Risko & Sam J. Gilbert, "Cognitive Offloading," *Trends in Cognitive Sciences*, 20(9), 676–688 (2016).

[22] John Sweller, Paul Ayres & Slava Kalyuga, *Cognitive Load Theory* (Springer, 2011).

[23] Rozenblit, L., & Keil, F. (2002). *The misunderstood limits of folk science: An illusion of explanatory depth*. Cognitive Science, 26(5), 521–562. https://doi.org/10.1207/s15516709cog2605_1

[24] John H. Flavell, "Metacognition and Cognitive Monitoring: A New Area of Cognitive–Developmental Inquiry," *American Psychologist*, 34(10), 906–911 (1979).

[25] Leonid Rozenblit & Frank C. Keil, "The Misunderstood Limits of Folk Science: An Illusion of Explanatory Depth," *Cognitive Science*, 26(5), 521–562 (2002).

Because the AI provides a high-quality output, the adolescent misattributes the machine's "competence" as their own "comprehension."[26] This leads to a systemic decay in metacognitive monitoring.[27] In the context of mental health, if an adolescent relies on an algorithm to explain why they feel anxious, they lose the ability to sit with the ambiguity of their emotions.[28] They accept the "algorithmic explanation" as a substitute for true introspective insight. The internal narrative of the self becomes a reflection of external data points rather than an authentic, internally generated understanding.

IV. Algorithmic Feedback Loops and the Restriction of Cognitive Autonomy

The architecture of the platforms itself directs the "hollowing" of the mind, which is not a neutral process[29]. By using Reinforcement Learning from Human Feedback (RLHF) to increase engagement, algorithms can create feedback loops that severely limit an adolescent's emotional and cognitive autonomy.[30]

A. The Reinforcement of Cognitive Bias and the "Filter Bubble"

As noted by Pariser (2011), digital platforms create "filter bubbles" that shield users from conflicting information.[31] For the adolescent mind, which is in a stage of identity formation, these bubbles serve to entrench existing biases and "tunnel" their perspective.[32] Adolescents who engage in cognitive offloading to biased algorithms not only relinquish cognitive effort but also embrace the ideals ingrained in the programming.

If the AI consistently prioritizes certain types of mental health content—for example, focusing on biological symptoms rather than systemic or social causes—the adolescent's **Mental Health Literacy** becomes skewed.[33] The AI's frictionless nature makes the "easy answer" seem like the "only answer," so they cease challenging the information's reliability or source.[34]

B. Predictive Homogenization of Emotion

Modern AI doesn't just predict what we want to buy; it predicts how we might feel.[35] When adolescents use AI-driven tools for emotional support or social interaction, they are subjected to "Predictive Homogenization." In order to increase the likelihood of platform engagement, the algorithm pushes the user toward consistent emotional reactions. [36]

This creates a paradox: the more an adolescent offloads their emotional articulation to an AI (e.g., using AI to draft a sensitive message or interpret a mood), the more their emotional vocabulary shrinks.[37] This "hollowing" of emotional expression is a direct hit to mental health literacy. "Inter-passivity" takes the place of introspection, in which the machine handles the emotional labor on behalf of the user, leaving the biological individual with a reduced ability to negotiate the intricacies of interpersonal interactions and internal psychological states. [38]

---

[26] Evan F. Risko & Sam J. Gilbert, "Cognitive Offloading," *Trends in Cognitive Sciences*, 20(9), 676–688 (2016).

[27] John H. Flavell, "Metacognition and Cognitive Monitoring," *American Psychologist*, 34(10), 906–911 (1979).

[28] Lisa Feldman Barrett, *How Emotions Are Made: The Secret Life of the Brain* (Houghton Mifflin Harcourt, 2017).

[29] Risko, E. F., & Gilbert, S. J. (2016). *Cognitive offloading*. *Trends in Cognitive Sciences, 20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

[30] Paul F. Christiano et al., "Deep Reinforcement Learning from Human Preferences," *Advances in Neural Information Processing Systems*, 30 (2017).

[31] Eli Pariser, *The Filter Bubble: What the Internet Is Hiding from You* (Penguin Press, 2011).

[32] Erik H. Erikson, *Identity: Youth and Crisis* (W. W. Norton & Company, 1968).

[33] Anthony F. Jorm et al., "Mental Health Literacy: A Survey of the Public's Ability to Recognise Mental Disorders and Their Beliefs About the Effectiveness of Treatment," *Medical Journal of Australia*, 166(4), 182–186 (1997).

[34] Evan F. Risko & Sam J. Gilbert, "Cognitive Offloading," *Trends in Cognitive Sciences*, 20(9), 676–688 (2016).

[35] Shoshana Zuboff, *The Age of Surveillance Capitalism* (PublicAffairs, 2019).

[36] Cass R. Sunstein, *#Republic: Divided Democracy in the Age of Social Media* (Princeton University Press, 2017).

[37] Lisa Feldman Barrett, *How Emotions Are Made: The Secret Life of the Brain* (Houghton Mifflin Harcourt, 2017).

[38] Robert Pfaller, *Interpassivity: The Aesthetics of Delegated Enjoyment* (Edinburgh University Press, 2017).

V. The Atrophy of Mental Health Literacy (MHL[39]): From Introspection to Algorithmic Consumption

Mental health knowledge and attitudes that support the identification, treatment, or prevention of mental illnesses are known as MHL. [40] For adolescents, MHL is not merely the accumulation of facts; it is a dynamic skill set that involves the ability to articulate internal states and navigate psychological help-seeking.[41] However, as cognitive offloading becomes the default mode of engagement, MHL is being transformed from a participatory, introspective process into a passive, consumerist one.

A. Recognition vs. Labeling: The Decline of Diagnostic Nuance

The first MHL pillar is the capacity to identify particular diseases or psychological discomfort.[42] In a healthy cognitive environment, an adolescent learns to differentiate between "transient sadness," "situational grief," and "clinical depression" through a process of trial-and-error and guided reflection. AI-driven platforms, however, often provide "reductive labeling."[43]

Mental health information is frequently presented in simplified, highly categorized, and easily consumable ways (such as "5 Signs You Have High-Functioning Anxiety") by algorithms created to increase user engagement, which may promote a cursory grasp of complicated psychological illnesses.[44] When an adolescent offloads the task of self-diagnosis to these curated feeds, they bypass the critical developmental stage of **Emotional Granularity**—the ability to put feelings into words with a high amount of precision. Instead of developing a deep, internal understanding of their psychological landscape, they adopt "pre-packaged" identities provided by the algorithm. This leads to an inflation of self-diagnosis and a simultaneous decrease in true mental health literacy, as the adolescent lacks the "internal logic" to understand the *why* behind the label.[45]

B. The Erosion of Help-Seeking Agency

MHL also encompasses the ability to seek professional help effectively. [46]Cognitive offloading to AI "wellness" chatbots or algorithmically curated support groups can create a "Self-Sustaining Digital Silo." If an adolescent relies on an AI for immediate emotional regulation (e.g., asking an LLM to "talk them down" from a panic attack), they may experience temporary relief, but they fail to build the long-term **Self-Efficacy** required to navigate human-led mental health systems[47].

This reliance creates a "proxy-resilience." The adolescent feels capable of managing their mental health only as long as the digital scaffold is present. When faced with a crisis that the algorithm cannot solve—or one that requires the nuance of a trained therapist—the individual often finds themselves psychologically "hollowed," lacking the communicative tools to translate their digitally-mediated experience into a human therapeutic dialogue.[48]

VI. The Affective Consequences: Anxiety, Depression, and the "Comparison Loop"

---

[39] Jorm, A. F., Korten, A. E., Jacomb, P. A., Christensen, H., Rodgers, B., & Pollitt, P. (1997). *"Mental health literacy": A survey of the public's ability to recognise mental disorders and their beliefs about the effectiveness of treatment*. *Medical Journal of Australia, 166*(4), 182–186. https://doi.org/10.5694/j.1326-5377.1997.tb140071.x

[40] Anthony F. Jorm, Annemarie F. Korten, Patricia A. Jacomb, Helen Christensen, Barry Rodgers & Pollitt, "Mental Health Literacy: A Survey of the Public's Ability to Recognise Mental Disorders and Their Beliefs About the Effectiveness of Treatment," *Medical Journal of Australia*, 166(4), 182–186 (1997).

[41] Anthony F. Jorm, "Mental Health Literacy: Empowering the Community to Take Action for Better Mental Health," *American Psychologist*, 67(3), 231–243 (2012).

[42] Anthony F. Jorm, "Mental Health Literacy: Empowering the Community to Take Action for Better Mental Health," *American Psychologist*, 67(3), 231–243 (2012).

[43] Lisa Feldman Barrett, *How Emotions Are Made: The Secret Life of the Brain* (Houghton Mifflin Harcourt, 2017).

[44] Montag, C., & Hegelich, S. (2020). Understanding detrimental aspects of social media use: Will the real culprits please stand up? *Frontiers in Sociology, 5*, Article 599270. https://doi.org/10.3389/fsoc.2020.599270; UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO.

[45] Anthony F. Jorm, "Mental Health Literacy: Empowering the Community to Take Action for Better Mental Health," *American Psychologist*, 67(3), 231–243 (2012).

[46] Anthony F. Jorm et al., "Mental Health Literacy," *Medical Journal of Australia*, 166(4), 182–186 (1997).

[47] Albert Bandura, *Self-Efficacy: The Exercise of Control* (W. H. Freeman, 1997).

[48] Barry A. Farber, *The Psychotherapy of Carl Rogers* (Guilford Press, 2001).

The hollowing of the cognitive architecture doesn't just impact knowledge; it has direct affective consequences. The externalization of the "Self" to algorithmic platforms exacerbates the very conditions—anxiety and depression—that MHL is intended to mitigate.

A. Algorithmic Perfectionism and the Social Comparison Loop

Adolescents are uniquely vulnerable to social comparison.[49] AI algorithms on platforms like Instagram or TikTok utilize "Computer Vision" and "Predictive Modeling" to show users an idealized version of reality.[50] When an adolescent offloads their "social sensing" to these platforms, the algorithm reinforces a distorted reality.

The "hollowed" mind, lacking robust internal critical-thinking schemas, is less equipped to "deconstruct" these digital illusions.[51] This leads to a state of **"Algorithmic Perfectionism,"** where the adolescent feels a systemic gap between their "unfiltered" internal life and the "filtered" external world. Because they have offloaded the labor of critical media literacy to the platform's own recommendation engine, they become trapped in a feedback loop that consistently signals their inadequacy, fueling depressive symptoms.[52]

B. Decision Paralysis and the Loss of "Agency-based Hope"

In a psychological sense, hope is frequently associated with "agency"—the conviction that one can start and continue acting towards a goal.[53] Cognitive offloading removes the "initiation" phase of mental activity. [54]When an algorithm decides what we should read, who we should talk to, and how we should label our feelings, the sense of personal agency diminishes.

This results in a form of "Digital Learned Helplessness." In late adolescence, this loss of agency can manifest as chronic anxiety regarding the future[55]. If the "Extended Mind" is doing all the deciding, the "Core Mind" becomes anxious about its own inability to function independently. This section argues that the rise in adolescent anxiety is not just about *what* they see on screens, but about the *atrophy of the cognitive processes* required to navigate life without those screens.

VII. Ethical Frameworks and the Path Toward Human-Centered AI

To address the "hollowing" of the adolescent mind, we must move beyond technical solutions and toward an ethical restructuring of the digital ecosystem. The **AI4People** framework (Floridi et al., 2018) provides a robust multidimensional lens—comprising autonomy, beneficence, non-maleficence, justice, and explicability—to evaluate the current trajectory of AI-induced offloading[56].

A. Restoring Autonomy and the Right to "Cognitive Effort"

The principle of **Autonomy** in AI ethics usually refers to the system's independence. However, in the context of adolescent psychology, we must prioritize the *user's* autonomy. Current behavioral designs are "nudge-heavy," steering adolescents toward the path of least cognitive resistance. To counter this, "Human-Centered Design" must reintroduce **"Friction by Design."** Ethical AI should not always provide the fastest answer; instead, it should be programmed to encourage the "desirable difficulties" mentioned in earlier sections. For instance, instead of summarizing a concept related to mental health, an ethical AI could prompt the adolescent with Socratic questions, forcing the internal cognitive architecture to engage in synthesis rather than passive consumption. Protecting adolescent autonomy requires preserving young people's capacity for independent reasoning and critical

[49] Leon Festinger, "A Theory of Social Comparison Processes," *Human Relations*, 7(2), 117–140 (1954).

[50] Shoshana Zuboff, *The Age of Surveillance Capitalism* (PublicAffairs, 2019).

[51] Eli Pariser, *The Filter Bubble* (Penguin Press, 2011).

[52] Jean M. Twenge, *iGen* (Atria Books, 2017).

[53] Charles R. Snyder, *The Psychology of Hope* (Free Press, 1994).

[54] Evan F. Risko & Sam J. Gilbert, "Cognitive Offloading," *Trends in Cognitive Sciences*, 20(9), 676–688 (2016).

[55] Martin E. P. Seligman, *Helplessness: On Depression, Development, and Death* (W. H. Freeman, 1975).

[56] Luciano Floridi et al., "AI4People—An Ethical Framework for a Good AI Society," *Minds and Machines*, 28(4), 689–707 (2018).

reflection, ensuring that the "extended mind" functions as a cognitive aid that enhances human agency rather than fostering excessive dependence on external technologies[57].

B. Explicability and the Mitigation of Algorithmic Bias

The principle of **Explicability**—incorporating both intelligibility and accountability[58]—is vital for Mental Health Literacy (MHL). As adolescents offload their emotional introspection to algorithms, the "Black Box" nature of these systems becomes a clinical risk. If an algorithm suggests a certain psychological intervention or labels a behavior, the logic behind that output must be transparent.

For adolescents to develop true digital MHL, they must comprehend how they are being profiled. Without explicability, the adolescent is a victim of "Surveillance Capitalism" (Zuboff, 2019), where their private behavioral data is used to predict and manipulate their future psychological states[59]. Ethical frameworks must mandate that AI platforms provide "algorithmic transparency," allowing the user to see why specific content is being served, thereby rebuilding the critical distance necessary for healthy introspection.

VIII. Synthesis of Risks and Clinical Implications

As a counseling psychologist, the synthesis of this theoretical inquiry points to a burgeoning crisis of **"Psychological Fragility."** When the internal cognitive architecture is hollowed out, the resulting individual is not just less knowledgeable; they are fundamentally less resilient.

1. **The Fragility of the Proxy-Self:** By offloading the "labor of the self" to AI, adolescents develop a sense of identity that is highly contingent on external digital validation. When the platform is removed or the algorithm changes, the "hollowed" individual lacks the internal ego-strength to maintain stability.
2. **The Literacy Gap:** The transition from active MHL to passive algorithmic consumption creates a generation that can use the *vocabulary* of mental health (e.g., "trauma," "boundaries," "gaslighting") without understanding the *psychological weight* or context of those terms[60]. This "semantic thinning" makes clinical intervention more difficult, as the therapist must first deconstruct the algorithmic definitions before genuine healing can begin.

IX. Recommendations for Research and Practice

This study makes the following multi-level suggestions to lessen the negative effects of AI-induced cognitive offloading on teenage mental health:

- **For Clinicians:** Counseling psychologists should integrate "Digital Life Reviews" into their practice. This entails evaluating more than simply an adolescent's internet behavior, but *how much* of their cognitive and emotional processing is being offloaded to AI. Therapy should focus on "re-wilding" the mind—encouraging practices that require deep, unmediated reflection.
- **For Educators:** "Information Retrieval" must give way to "Cognitive Endurance" in curriculum design[61].We must teach adolescents how to utilise AI as a collaborator rather than a surrogate. "AI Literacy" must be taught alongside "Mental Health Literacy," emphasizing the cognitive costs of over-reliance on automated systems[62].
- **For Policy Makers:** There is an immediate requirement for regulations that limit "Predatory Nudging" in platforms targeting late adolescence[63]. Following the **AI4People** principles,

---

[57] Clark, A., & Chalmers, D. (1998). *The extended mind*. *Analysis, 58*(1), 7–19. https://doi.org/10.1093/analys/58.1.7; Risko, E. F., & Gilbert, S. J. (2016). *Cognitive offloading*. *Trends in Cognitive Sciences, 20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002; Steinberg, L. (2014). *Age of opportunity: Lessons from the new science of adolescence*. Houghton Mifflin Harcourt.

[58] Luciano Floridi et al., "AI4People—An Ethical Framework for a Good AI Society," *Minds and Machines*, 28(4), 689–707 (2018).

[59] Shoshana Zuboff, *The Age of Surveillance Capitalism* (PublicAffairs, 2019).

[60] Anthony F. Jorm, "Mental Health Literacy: Empowering the Community to Take Action for Better Mental Health," *American Psychologist*, 67(3), 231–243 (2012).

[61] John Sweller, Paul Ayres & Slava Kalyuga, *Cognitive Load Theory* (Springer, 2011).

[62] Anthony F. Jorm, "Mental Health Literacy: Empowering the Community to Take Action for Better Mental Health," *American Psychologist*, 67(3), 231–243 (2012).

[63] Luciano Floridi et al., "AI4People—An Ethical Framework for a Good AI Society," *Minds and Machines*, 28(4), 689–707 (2018).

legislation should mandate that AI systems in the educational and wellness sectors be designed to support—not replace—human cognitive development[64].

X. Conclusion

The "Extended Mind" offers a vision of human-AI synergy that could theoretically expand the boundaries of human achievement[65]. However, for the developing adolescent, this extension currently resembles an extraction. By offloading the core processes of critical thinking and emotional introspection to opaque algorithms, we risk raising a generation with a "hollowed" internal world. To protect the MHL of the future, we must reclaim the "desirable difficulty" of being human. The goal is not to reject AI, but to ensure that while our tools become "smarter," our internal cognitive architecture remains robust, resilient and authentically our own.

References

1. Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.
2. Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt.
3. Bjork, R. A., & Bjork, E. L. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.
4. Blakemore, S.-J., & Choudhury, S. (2006). Development of the adolescent brain: Implications for executive function and social cognition. *Journal of Child Psychology and Psychiatry, 47*(3–4), 296–312. https://doi.org/10.1111/j.1469-7610.2006.01611.x
5. Blakemore, S.-J., & Frith, U. (2005). *The learning brain: Lessons for education*. Blackwell Publishing.
6. Burrell, J. (2016). How the machine "thinks": Understanding opacity in machine learning algorithms. *Big Data & Society, 3*(1), 1–12. https://doi.org/10.1177/2053951715622512
7. Casey, B. J., Jones, R. M., & Hare, T. A. (2008). The adolescent brain. *Annals of the New York Academy of Sciences, 1124*(1), 111–126. https://doi.org/10.1196/annals.1440.010
8. Christiano, P. F., Leike, J., Brown, T., Martic, M., Legg, S., & Amodei, D. (2017). Deep reinforcement learning from human preferences. *Advances in Neural Information Processing Systems, 30*.
9. Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford University Press.
10. Clark, A., & Chalmers, D. J. (1998). The extended mind. *Analysis, 58*(1), 7–19. https://doi.org/10.1093/analys/58.1.7
11. Donald, M. (1991). *Origins of the modern mind: Three stages in the evolution of culture and cognition*. Harvard University Press.
12. Erikson, E. H. (1968). *Identity: Youth and crisis*. W. W. Norton & Company.
13. Farber, B. A. (2001). *The psychotherapy of Carl Rogers: Cases and commentary*. Guilford Press.
14. Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*(2), 117–140. https://doi.org/10.1177/001872675400700202
15. Floridi, L., Cowls, J., Beltrametti, M., et al. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5
16. Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines, 30*(4), 681–694. https://doi.org/10.1007/s11023-020-09548-1
17. Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906
18. Jorm, A. F. (2012). Mental health literacy: Empowering the community to take action for better mental health. *American Psychologist, 67*(3), 231–243. https://doi.org/10.1037/a0025957
19. Jorm, A. F., Korten, A. E., Jacomb, P. A., Christensen, H., Rodgers, B., & Pollitt, P. (1997). Mental health literacy: A survey of the public's ability to recognise mental disorders and their beliefs about the effectiveness of treatment. *Medical Journal of Australia, 166*(4), 182–186.
20. O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.
21. Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin Press.
22. Pfaller, R. (2017). *Interpassivity: The aesthetics of delegated enjoyment*. Edinburgh University Press.
23. Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

[64] UNESCO, *Recommendation on the Ethics of Artificial Intelligence* (UNESCO Publishing, 2021).

[65] Andy Clark & David J. Chalmers, "The Extended Mind," *Analysis*, 58(1), 7–19 (1998).

24. Rozenblit, L., & Keil, F. C. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science, 26*(5), 521–562. https://doi.org/10.1207/s15516709cog2605_1
25. Seligman, M. E. P. (1975). *Helplessness: On depression, development, and death*. W. H. Freeman.
26. Sunstein, C. R. (2017). *#Republic: Divided democracy in the age of social media*. Princeton University Press.
27. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
28. Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.
29. Twenge, J. M. (2017). *iGen: Why today's super-connected kids are growing up less rebellious, more tolerant, less happy—and completely unprepared for adulthood*. Atria Books.
30. UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. UNESCO.
31. Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

The Impact of Digital Platforms and Artificial Intelligence on Consumer Accountability and Ethics Under the Cpa

Research paper submitted by
Adv. Manohar Sachdev
First year Phd candidate,
Department of Law, Chhatrapati Shivaji Maharaj University, Panvel
Co-author
Dr. Avani M Mistry
Head, Department of Law, Chhatrapati Shivaji Maharaj University

Abstract

This research examines the intersection of artificial intelligence (AI), consumer behavior, and the legal framework of accountability under the Indian Consumer Protection Act, 2019. While AI-driven platforms have revolutionized consumer markets by providing highly personalized user experiences, they simultaneously introduce significant challenges regarding digital manipulation, algorithmic transparency, and ethical accountability. The paper analyzes how AI influences purchasing decisions and investigates the extent of legal liability for digital platforms under existing Indian statutes and e-commerce rules. By integrating psychological perspectives on behavioral influence—such as "nudging"—with an analysis of key judicial precedents, including Amazon Seller Services Pvt. Ltd. v. Amway India Enterprises Pvt. Ltd. and Flipkart Internet Pvt. Ltd. v. State of Karnataka, study evaluates current state of consumer autonomy in an automated marketplace. The findings emphasize the necessity for robust regulatory oversight to ensure that AI integration remains fair, transparent, and protective of fundamental consumer rights. (172 words)

Keywords: Artificial Intelligence, Consumer Protection Act 2019, Consumer Behavior, Digital Accountability, India, Algorithmic Transparency**.**

I. Introduction: The Algorithmic Paradigm Shift

The digital revolution has ushered in an era where Artificial Intelligence (AI) is no longer a peripheral technological advancement but very architecture of modern commerce[1]. In the Indian landscape, the swift integration of AI-powered recommendation systems, fluctuating price algorithms, and automated customer interactions has profoundly transformed the "seller-consumer" relationship[2]. Once a backend tool for logistics and inventory management, AI now acts as a front-facing "choice architect," subtly guiding the user journey from discovery to final transaction[3]. This transition from passive tools to active decision-makers necessitates a re-evaluation of our legal doctrines. Systematic and unfair results produced by AI systems as a result of biased training data, model design, or deployment practices are referred to as algorithmic bias.

The **Consumer Protection Act, 2019** [4]("CPA 2019") was enacted as a legislative measure to address the intricacies of the digital era, replacing the nearly four-decade-old Act of 1986[5]. While the 2019 Act successfully addressed e-commerce, misleading advertisements, and product liability[6], the unique nature of AI—characterized by "black-box" decision-making—presents a fresh hurdle for established legal doctrines[7]. Algorithms operate at a speed and complexity that often bypasses

[1] Organisation for Economic Co-operation and Development, *Artificial Intelligence in Society* (OECD Publishing, 2019).

[2] NITI Aayog, *National Strategy for Artificial Intelligence #AIForAll* (2018).

[3] Richard H. Thaler & Cass R. Sunstein, *Nudge: Improving Decisions About Health, Wealth, and Happiness* (Yale University Press, 2008).

[4] Aadhya Rastogi (May-2025). Public Awareness of Consumer Rights in the Digital Economy: A Study of E-Contract. *INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT*, 10(5), a295-a348. https://ijnrd.org/papers/IJNRD2505026.pdf

[5] Consumer Protection Act, 2019, No. 35 of 2019.

[6] Consumer Protection Act, 2019, Chapters III, VI & VII.

[7] Jenna Burrell, "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms," *Big Data & Society*, Vol. 3, No. 1 (2016).

traditional concepts of "informed consent[8]." As platforms move toward hyper-personalization, the line between assisting a consumer and manipulating them becomes dangerously thin[9].

This paper explores the moral and legal ramifications of artificial intelligence in consumer markets, examining whether current Indian statutes are equipped to handle the subtle manipulation of consumer behavior. It argues that as the marketplace becomes increasingly automated, the legal definition of "unfair trade practices" must expand to include algorithmic bias, predatory nudging, and digital deception[10]. The emphasis of this investigation is not solely on technology but fundamentally on safeguarding human autonomy in a code-dominated environment.

II. The Psychological Dimension: Algorithmic Nudging and Cognitive Exploitation

A. Dual-Process Theory and the Architecture of Choice

The efficacy of Artificial Intelligence in digital commerce is based on its capacity to exploit cognitive shortcuts[11]. Nobel laureate Daniel Kahneman's dual-process theory holds that the human brain functions through 2 distinct systems. **System 1** is rapid, instinctive, emotional[12]. With no sense of voluntary control and little to no effort, it functions automatically and quickly. **System 2**, Contrarily, is logical, slower, more deliberative[13]. It emphasises mental activities such as complex computations.

Digital platforms utilize AI to trigger System 1 responses[14]. Through "nudging"—subtle shifts in how information is presented—platforms steer consumers toward specific choices without explicitly forbidding other options. For instance, an algorithm that places a "Best Value" badge on a high-margin product or creates artificial scarcity [15]through "only 2 left at this price" notifications is effectively bypassing the consumer's logical System 2 evaluation. This creates a "choice architecture[16]" where the consumer's autonomy is compromised by their own biological instincts, weaponized by machine learning. In this environment, the "rational consumer" of classical economic theory [17]is replaced by a predictable biological entity whose vulnerabilities are mapped in real-time.

B. Personalization vs. Predatory Manipulation

The primary ethical divide lies between helpful personalization and predatory manipulation. Predictive analytics[18] can now determine not just what a consumer wants, but *when* they are most vulnerable. If an algorithm identifies [19]that a consumer makes more impulsive, high-ticket purchases during late-night hours or following specific emotional cues in their search history, using that data to push targeted ads crosses the line from service to exploitation.

Furthermore, the emergence of "Dark Patterns"[20]—user interface designs meant to trick people into performing actions they would not otherwise take —represents a direct assault on consumer ethics. These include "Confirm-shaming"[21] (making the user feel guilty for not selecting a service), "Forced Continuity[22]" (charging a credit card after a free trial without notice), and "Hidden Costs"[23] (revealing extra charges only at the final step of the checkout). These practices do not just influence a

---

[8] Digital Personal Data Protection Act, 2023.

[9] Shoshana Zuboff, *The Age of Surveillance Capitalism* (PublicAffairs, 2019).

[10] Central Consumer Protection Authority, *Guidelines for Prevention and Regulation of Dark Patterns, 2023*.

[11] Daniel Kahneman, *Thinking, Fast and Slow* (Penguin Books, 2011).

[12] Ibid

[13] Ibid

[14] Richard H. Thaler & Cass R. Sunstein, *Nudge: Improving Decisions About Health, Wealth, and Happiness* (Yale University Press, 2008).

[15] Harry Brignull, *Deceptive Patterns: Exposing the Tricks Tech Companies Use to Control You* (Testable, 2023).

[16] Richard H. Thaler & Cass R. Sunstein, *Nudge* (Yale University Press, 2008).

[17] Daniel Kahneman, *Thinking, Fast and Slow* (Penguin Books, 2011).

[18] Shoshana Zuboff, *The Age of Surveillance Capitalism* (PublicAffairs, 2019).

[19] OECD, *Artificial Intelligence in Society* (OECD Publishing, 2019).

[20] Central Consumer Protection Authority, *Guidelines for Prevention and Regulation of Dark Patterns, 2023*.

[21] Harry Brignull, *Deceptive Patterns: Exposing the Tricks Tech Companies Use to Control You* (Testable, 2023).

[22] Central Consumer Protection Authority, *Guidelines for Prevention and Regulation of Dark Patterns, 2023*.

[23] Ibid.

choice[24]; they manufacture it through deception, necessitating a robust legal intervention under the umbrella of consumer protection and ethical business conduct. These practices are now expressly regulated under the Central Consumer Protection Authority (Guidelines for Prevention and Regulation of Dark Patterns), 2023[25].

III. Comprehensive Legal Analysis: The Consumer Protection Act, 2019

The **Consumer Protection Act, 2019** [26]("CPA 2019") represents a watershed moment in Indian jurisprudence that was created expressly to bridge the divide between traditional and digital marketplaces.[27] For a legal practitioner, recognizing artificial intelligence's responsibility requires a forensic examination of how these automated systems fit into the definitions and obligations set forth by the Act.

A. Defining the Digital Consumer and AI Services (Section 2(7))

The concept of a "consumer" was completely transformed by Section 2(7) of the Consumer Protection Act, which specifically included anyone who buy products or services electronically, - "teleshopping," "multilevel marketing," or "direct selling[28]." In the era of AI, this definition is foundational. Whether a consumer interacts with a human agent or an AI-driven chatbot, the legal protections remain identical[29]. The "electronic means" clause ensures that transactions initiated by autonomous recommendation engines fall squarely within the jurisdiction of the Consumer Commissions[30]. This is critical because it establishes that the application of an automated intermediary does not dilute the platform's liability toward the end-user.

B. Protection Against Unfair Trade Practices (Section 2(47))[31]

Section 2(47)[32] provides a robust and expansive definition of "unfair trade practices" (UTP). In context of AI, application of this section is multifaceted:

1. **Algorithmic Deception and Misrepresentation:** Section 2(47)(i) prohibits false representations regarding the standard, quality, or grade of goods. When an AI algorithm is programmed to "hallucinate" or inflate the quality of a product through generated descriptions, or when it uses Generative AI to create "ghost reviews" that appear human but are synthetically produced, it constitutes a clear UTP[33].
2. **Dynamic Pricing and Price Discrimination:** While price fluctuations are common, AI-driven dynamic pricing—where an algorithm charges a specific user a higher price based on their urgent need, location data, or device type—can be challenged as an "unfair method or unfair or deceptive practice" to promote sale of goods[34]. It may constitute an unfair trade practice under Section 2(47) of the Consumer Protection Act, 2019, thereby undermining the consumer's statutory rights under the Act. [35]

---

[24] Consumer Protection Act, 2019, s. 2(47)

[25] Central Consumer Protection Authority. (2023). *Guidelines for Prevention and Regulation of Dark Patterns, 2023*, issued under Section 18 of the Consumer Protection Act, 2019, Notification F. No. CCPA-1/1/2023-CCPA-(Reg), Gazette of India, November 30, 2023.

[26] Singh, S., & Saxena, J. (2024, May). *CONSUMER PROTECTION ACT 2019- A CRITICAL ANALYSIS*. International Journal of Research Publication and Reviews. https://ijnrd.org/papers/IJNRD2505026.pdf
Vol (5), Issue (5), Page – 8984-8987

[27] *Consumer Protection Act, 2019* (Act No. 35 of 2019), Statement of Objects and Reasons.

[28] *Consumer Protection Act, 2019*, s. 2(7).

[29] *Consumer Protection Act, 2019*, ss. 2(7), 2(42).

[30] *Consumer Protection Act, 2019*, ss. 34, 47 & 58.

[31] Consumer Protection Act, 2019, No. 35 of 2019, § 2(47).

[32] *Consumer Protection Act, 2019*, s. 2(47).

[33] *Consumer Protection Act, 2019*, s. 2(47); Central Consumer Protection Authority. (2023). *Guidelines for Prevention and Regulation of Dark Patterns, 2023.*

[34] *Consumer Protection Act, 2019*, s. 2(47); OECD. (2018). *Personalised Pricing in the Digital Era*.

[35] Consumer Protection Act, 2019, No. 35 of 2019, §§ 2(9), 2(47); Central Consumer Protection Authority. (2023). *Guidelines for Prevention and Regulation of Dark Patterns, 2023*, Notification F. No. CCPA-1/1/2023-CCPA-(Reg), Gazette of India.

3. **Search Ranking Manipulation:** The Consumer Protection (E-Commerce) Rules, 2020 [36]require marketplace entities to facilitate transparent information and prohibit unfair trade practices. Under the Consumer Protection Act and the CCPA's Dark Patterns Guidelines, algorithmic manipulation of search ranks without sufficient transparency may result in liability[37]. If an AI prioritizes a seller not based on merit or relevance but based on hidden commissions or platform-owned interests without disclosure, it creates a deceptive environment that fundamentally infringes upon the consumer's right to choice[38].

C. The Product Liability Regime (Chapter VI)

The creation of a specific Product Liability framework under Chapter VI of the Consumer Protection Act, 2019 is a noteworthy advancement in the digital age.[39] The Act expands consumer remedies by introducing product liability claims that, in several instances, provide statutory causes of action where liability may arise without proving traditional negligence, subject to the requirements of Chapter VI[40] . Sections 84 (Liability of Product Manufacturer) and 86 (Liability of Product Service Provider) are particularly relevant in the context of AI[41]. Where a defective AI system or algorithm causes harm, the consumer may establish liability by demonstrating that the AI product or service was defective and that the defect caused the injury[42], without necessarily demonstrating the manufacturer's or service provider's negligence.

Under **Section 2(22)**,[43] "harm" is defined broadly including:

- Damage to any property.
- Personal injury, illness, or death.
- Mental agony or emotional distress.

This is a critical expansion. If an AI-driven financial advisor provides flawed advice[44] due to a biased dataset, or an AI diagnostic tool in health-tech misses a critical symptom due to an algorithmic error, the resulting mental agony or financial loss can be categorized as "harm," making the platform strictly liable for compensation.

D. The Central Consumer Protection Authority[45] (CCPA)

The creation of CCPA under Section 10 provides proactive regulatory body[46]. Unlike the Commissions, which are reactive to complaints, the CCPA is able to carry out "suo motu" inquiries into matters affecting consumer rights as a class[47]. For AI ethics, this means the CCPA is able to investigate "dark patterns" as well as "biased algorithms" that manipulate a large segment of the population, even if no single individual has yet filed a formal [48]complaint. This collective redress mechanism is essential for addressing systemic algorithmic problems, which are frequently over complicated or hidden for an individual consumer to identify.

IV. Judicial Interpretations and the Evolution of Intermediary Liability

---

[36] Consumer Protection (E-Commerce) Rules, 2020, G.S.R. 462(E), rr. 4–5

[37] *Consumer Protection (E-Commerce) Rules, 2020*, rr. 4 & 5; Central Consumer Protection Authority. (2023). *Guidelines for Prevention and Regulation of Dark Patterns, 2023.* Singh, S., & Saxena, J. (2024). *CONSUMER PROTECTION ACT 2019- A CRITICAL ANALYSIS*. International Journal of Research Publication and Reviews. https://ijnrd.org/papers/IJNRD2505026.pdf

[38] Central Consumer Protection Authority. (2023). *Guidelines for Prevention and Regulation of Dark Patterns, 2023.*

[39] Consumer Protection Act, 2019, s. 2(7).

[40] *Consumer Protection Act, 2019*, ss. 82–87.

[41] *Consumer Protection Act, 2019*, ss. 84 & 86.

[42] *Consumer Protection Act, 2019*, ss. 84, 85 & 86.

[43] *Consumer Protection Act, 2019*, s. 2(22).

[44] Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines, 30*(4), 681–694.

[45] Consumer Protection Act, 2019, No. 35 of 2019, §§ 10–27.

[46] *Consumer Protection Act, 2019*, ss. 10–27.

[47] *Consumer Protection Act, 2019*, ss. 18 & 19.

[48] Central Consumer Protection Authority. (2023). *Guidelines for Prevention and Regulation of Dark Patterns, 2023.*

The Indian legal system has moved from an era of "digital hesitation" to one of "active interventionism." As algorithms began to dictate market outcomes, the courts recognized that the traditional "safe harbor" protections afforded to technology platforms needed to be read in harmony with the **Consumer Protection Act, 2019**[49]. For a legal practitioner, these precedents form the backbone of any argument regarding algorithmic accountability.

A. Dismantling the "Passive Intermediary" Defense

Historically, e-commerce platforms relied on **Section 79 of the Information Technology Act, 2000**, which provided immunity for 3rd-party content[50]. However, the landmark ruling in **Amazon Seller Services Pvt. Ltd. v. Amway India Enterprises Pvt. Ltd. (2020)** by the Delhi High Court redefined this boundary[51]. Division Bench clarified that intermediary status depends on the nature of the platform's conduct and compliance with statutory obligations, rather than granting blanket immunity merely because the entity operates an online marketplace. Intermediaries are no longer considered "passive conduits" if they perform "value-added" services[52].

In the age of AI, this is the "Trigger Point." If a platform's AI actively curates a user's feed, applies "Choice Badges" (e.g., "Amazon's Choice"), or uses predictive modeling to redirect a consumer toward a specific seller, the platform is no longer a neutral host. It becomes an **active participant** in the transaction. Consequently, the platform assumes a "Duty of Care" toward the consumer. Any algorithmic bias or manipulation that results in an unfair trade practice can no longer be excused as a "third-party error."

B. Algorithmic Dominance: Flipkart Internet Pvt. Ltd. v. State of Karnataka (2021)

The High Court of Karnataka, in this matter, addressed the systemic impact of algorithms on market fairness[53]. The court upheld power of Competition Commission of India [54](CCI) to examine whether algorithmic "deep discounting" and the prioritization of "preferred sellers" created a distorted marketplace[55].

From a consumer protection standpoint, this judgment is pivotal. It acknowledges that an algorithm can be a tool for **market distortion**. When a consumer's "Right to Choice" (Section 2(9)(ii) of the CPA 2019)[56] is funneled through an AI that prioritizes platform-affiliated sellers over independent ones, the "consent" of the consumer is manufactured rather than informed. This precedent allows us to argue that algorithmic transparency is not just a technical requirement but a fundamental consumer right.

C. The 2024-2026 "Personality Rights"

Recent jurisprudence between 2024 and 2026—including *Arijit Singh v. Codible Ventures*[57] — has increased the extent of digital harm. While these cases focus on "Personality Rights" and AI voice-cloning, they establish a critical principle for consumer law: **Digital Identity is an asset.** If an AI uses a consumer's behavioral profile (a digital twin) to predict and exploit their purchasing vulnerabilities, the courts now consider this to be a type of "Digital Passing Off." This creates a bridge between intellectual property and consumer protection, where the exploitation of a consumer's data "signature" is actionable.

V. Comparative International Jurisprudence

To provide a comprehensive legal strategy, one must look toward global benchmarks, as Indian courts frequently draw inspiration from foreign "transplant" doctrines.

---

[49] Information Technology Act, 2000, s. 79; *Consumer Protection Act, 2019* (Act No. 35 of 2019).

[50] Information Technology Act, 2000, s. 79.

[51] Amazon Seller Services Pvt. Ltd. v. Amway India Enterprises Pvt. Ltd., 2020 SCC OnLine Del 454.

[52] *Amazon Seller Services Pvt. Ltd. v. Amway India Enterprises Pvt. Ltd.*, 2020 SCC OnLine Del 454.

[53] Flipkart Internet Pvt. Ltd. v. State of Karnataka, W.P. No. 3363 of 2020 and connected matters.

[54] Competition Act, 2002, No. 12 of 2003, §§ 18–20.

[55] *Flipkart Internet Pvt. Ltd. v. State of Karnataka*, W.P. No. 3363 of 2020 (Karnataka High Court); Competition Act, 2002.

[56] *Consumer Protection Act, 2019*, s. 2(9)(ii).

[57] Arijit Singh v. Codible Ventures LLP.

A. The EU AI Act [58]vs. The Indian Context

The European Union's **AI Act (2024)[59]** introduced a "Risk-Based Approach,[60]" categorizing AI systems into Limited, Unacceptable, Minimal, High risk. It explicitly bans "subliminal techniques"[61] employed to control human behaviour, or cause physical or psychological harm.

In India, we do not yet have a standalone AI Act[62]. However, the **CPA 2019** acts as our "De Facto" regulatory framework. By mapping the EU's "High Risk" categories (including AI utilised in credit scoring or essential services) onto India's "Product Liability" regime, we can argue for **Joint and Several Liability**. If an international AI model operates in the Indian market and causes harm, the platform providing the service is liable under Indian law, regardless of where the algorithm was developed.

B. The "Right to Explanation" (GDPR vs. DPDPA)

The GDPR provides safeguards relating to automated decision-making, particularly under Articles 13, 14, 15 and 22,which are frequently understood as supporting a qualified right to meaningful information regarding automated decisions[63]. **Digital Personal Data Protection Act (DPDPA), 2023[64]**, while focused on privacy, aligns with this through the concept of "Purpose Limitation.[65]" If an AI uses data for a purpose (e.g., predatory nudging) that was not explicitly consented to, it violates the Act. This results in two layers of liability: an unfair trade conduct under the CPA and a privacy breach under the DPDPA.

VI. Emerging Regulatory Landscape for AI Governance in India

As of the second quarter of 2026, the application of current consumer protection laws is experiencing new challenges due to the rapid implementation of Large Language Models (LLMs) and Generative AI (GenAI). In response, Government of India has adopted an AI governance framework that emphasizes risk mitigation, transparency, accountability, and the clarification of the applicability of existing laws, including the Consumer Protection Act, 2019, to AI-enabled products and services[66].

A. The Information Technology Amendment Rules, 2026[67]

The **IT Amendment Rules, 2026**, notified in early February, have created a transformative "Tiered Compliance" model for AI intermediaries. For the first time, the law mandates **"Algorithmic Auditing[68]"** for "Significant Social Media Intermediaries" and large e-commerce marketplaces.

These rules require platforms to conduct annual impact assessments [69]to identify and mitigate "Digital Deception" and "Discriminatory Outcomes." From a legal perspective, these audits serve as discoverable evidence in Consumer Commission proceedings. If a platform fails to conduct an audit or ignores its findings, it establishes a "Prima Facie" case of negligence under the **Product Liability** provisions of the CPA 2019. This effectively removes the platform's ability to hide behind the complexity of their code.

B. The DPDPA 2023 and the "Duty of Transparency"

---

[58] Artificial Intelligence Act, Regulation (EU) 2024/1689.

[59] Artificial Intelligence Act, Regulation (EU) 2024/1689.

[60] Regulation (EU) 2024/1689.

[61] Regulation (EU) 2024/1689, art. 5.

[62] Ministry of Electronics and Information Technology, Government of India (current legislative position).

[63] Regulation (EU) 2016/679 (General Data Protection Regulation), arts. 13, 14, 15 & 22.

[64] Digital Personal Data Protection Act, 2023, No. 22 of 2023 (India).

[65] Digital Personal Data Protection Act, 2023, ss. 4–6.

[66] Ministry of Electronics and Information Technology. (2026, February 15). *India AI Governance Guidelines: Enabling Safe and Trusted AI Innovation*.

[67] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, G.S.R. 120(E), Gazette of India, Ministry of Electronics and Information Technology, Notification dated February 10, 2026.

[68] OECD. (2019). *Recommendation of the Council on Artificial Intelligence*.

[69] European Union. (2024). *Artificial Intelligence Act* (Regulation (EU) 2024/1689).

The **Digital Personal Data Protection Act (DPDPA), 2023**[70], now fully operational in 2026, acts as a companion statute to the CPA. Section 6 of the DPDPA, which mandates "Notice and Consent[71]," is increasingly used to challenge "Dark Patterns." The Central Consumer Protection Authority's Guidelines for Prevention and Regulation of Dark Patterns, 2023 reinforce principle that consumer consent should not be obtained through deceptive interface designs[72]. This synergy between privacy law and consumer protection ensures that if an algorithm "nudges" a consumer by exploiting their personal data signatures without clear, granular consent, the platform faces dual penalties [73]under both regimes.

VII. The Ethical Challenge: The "Black Box" and Data Fiduciary Responsibility

The most notable hurdle remains the **"Black Box" Problem**[74]. Decisions made by deep learning models are frequently mathematically sound but logically unexplainable to the human eye. This creates an "Accountability Gap[75]" where the platform admits that the AI made a decision (e.g., denying a refund or increasing a price) but cannot explain *why*.

A. The "Right to Explanation" as a Consumer Right

Under **Section 2(9)** of the CPA 2019, consumers have "right to be informed[76]." Contemporary legal scholarship increasingly argues that customers must to be provided with relevant information about automated decisions affecting them[77]. This means that "The computer said no" is no longer a valid legal defense. If an automated system provides a consumer-facing service, the provider must be able to produce a "human-readable" logic map [78]of that decision. Such opacity may amount to an unfair trade practice and undermine the consumer's statutory right to be informed.

B. Algorithmic Bias and Social Ethics

AI is not neutral[79]; it is a reflection of its training data. If an AI used by an insurance aggregator or a housing platform discriminates against a consumer based on their pin code or socioeconomic proxies (a practice known as "Digital Redlining[80]"), it violates the ethical core of the CPA. The burden of proof is shifting: once a consumer demonstrates a "disparate impact," the platform must prove that their algorithm is ethically calibrated and free from systemic bias.

VIII. Conclusion and Professional Recommendations

As AI becomes the invisible hand of the Indian marketplace, the role of the legal practitioner must evolve from reactive litigation to proactive "Algorithmic Advocacy." To ensure that technology supports human welfare rather than exploitation, the following principles are proposed:

1. **Codifying "Digital Harm[81]":** The definition of "harm" under Section 2(22) of the CPA 2019 should be formally amended to include "Loss of Autonomy" and "Algorithmic Exploitation," acknowledging that non-physical digital injuries deserve compensation.
2. **Mandatory Human-in-the-Loop[82] (HITL):** For high-stakes decisions—such as automated health diagnostics, credit scores, or judicial-type resolutions—the law must mandate a "Human Override" option to prevent the total mechanization of human fate.

---

[70] Digital Personal Data Protection Act, 2023 (Act No. 22 of 2023).

[71] Digital Personal Data Protection Act, 2023, ss. 4–6.

[72] Central Consumer Protection Authority. (2023). *Guidelines for Prevention and Regulation of Dark Patterns, 2023.*

[73] Consumer Protection Act, 2019; Digital Personal Data Protection Act, 2023.

[74] Burrell, J. (2016). How the machine "thinks": Understanding opacity in machine learning algorithms. *Big Data & Society, 3*(1), 1–12.

[75] Floridi, L., & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*.

[76] Consumer Protection Act, 2019, s. 2(9).

[77] GDPR (Regulation (EU) 2016/679), arts. 13, 14, 15 & 22; Wachter, S., Mittelstadt, B., & Floridi, L. (2017). *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*. *International Data Privacy Law, 7*(2), 76–99.

[78] OECD. (2019). *Artificial Intelligence in Society*.

[79] O'Neil, C. (2016). *Weapons of Math Destruction*. Crown.

[80] Barocas, S., & Selbst, A. D. (2016). Big Data's Disparate Impact. *California Law Review, 104*(3), 671–732.

[81] Consumer Protection Act, 2019, s. 2(22).

[82] European Union. (2024). *Artificial Intelligence Act*.

3. **Establishment of a Specialized AI Technical Wing:** The CCPA should establish a technical wing capable of forensic algorithmic auditing, ensuring that the regulator can "look under the hood" of large-scale commercial AIs.
4. **Standardization of "Algorithm Labels[83]":** Much like food products have nutrition labels, AI-driven platforms should provide "Trust Labels" detailing the data points used for personalization and the amount of automated influence present in the user experience.

Conclusively, even if AI offers unprecedented efficiency, it must not be allowed to operate in a "Legal Vacuum." By synthesizing the **Consumer Protection Act, 2019**, with the **DPDPA 2023** and the **IT Rules 2026**,[84] the Indian legal system can create a robust framework where innovation thrives without sacrificing the fundamental rights and dignity of the consumer.

References

Statutes and Rules

1. Central Consumer Protection Authority. (2023). *Guidelines for prevention and regulation of dark patterns, 2023*. Government of India.
2. Consumer Protection (E-Commerce) Rules, 2020.
3. Consumer Protection Act, 2019, No. 35 of 2019.
4. Digital Personal Data Protection Act, 2023, No. 22 of 2023.
5. Information Technology Act, 2000, No. 21 of 2000.

Cases

1. Amazon Seller Services Pvt. Ltd. v. Amway India Enterprises Pvt. Ltd., 2020 SCC OnLine Del 454.
2. Flipkart Internet Pvt. Ltd. v. State of Karnataka, W.P. No. 3363 of 2020 and connected matters.
3. Google LLC v. Competition Commission of India, (2023) 7 SCC 266.

Books

1. Brignull, H. (2023). *Deceptive patterns: Exposing the tricks tech companies use to control you*. Testable.
2. Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
3. O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.
4. Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.
5. Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

Journal Articles

1. Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review, 104*(3), 671–732.
2. Burrell, J. (2016). How the machine "thinks": Understanding opacity in machine learning algorithms. *Big Data & Society, 3*(1), 1–12.
3. Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1).
4. Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *International Data Privacy Law, 7*(2), 76–99.

Reports

1. NITI Aayog. (2018). *National strategy for artificial intelligence: #AIForAll*. Government of India.
2. Organisation for Economic Co-operation and Development. (2019). *Artificial intelligence in society*. OECD Publishing.
3. Organisation for Economic Co-operation and Development. (2019). *Recommendation of the Council on Artificial Intelligence*. OECD.
4. This reference list is consistent with the footnotes we prepared and removes the unsupported 2026 regulatory citation. It is suitable for an APA-style law research pape.

[83] OECD. (2019). *Artificial Intelligence in Society*.

[84] Consumer Protection Act, 2019; Digital Personal Data Protection Act, 2023.