AI Governance:
Law, Ethics and Policy

Reading Material
For
State-Level Conference on Speedy and Qualitative Disposal of Cyber Crime Cases: Issues, Challenges & Solutions

18th January, 2026

Prepared by :
Judicial Academy, Jharkhand

Judicial Academy Jharkhand

AI Governance: Law, Ethics and Policy

Reading Material

For

State-Level Conference on Speedy and Qualitative Disposal of Cyber Crime Cases: Issues, Challenges & Solutions

18th January, 2026

Prepared by :

Judicial Academy, Jharkhand

Near Dhurwa Dam, Dhurwa, Ranchi – 834004

Phone : 0651-2772001, 2772103, Fax : 0651-2772008

Email id : judicialacademyjharkhand@yahoo.co.in, Website : www.jajharkhand.in

DISCLAIMER

- This book is intended for Private Circulation Only.
- The information contained in this book is intended for information purposes only and should not be construed as legal advice on any subject matter.
- The cases and content provided are for educational purposes and illustrative understanding. For a comprehensive understanding, readers are encouraged to refer to the complete case laws and official sources.
- The government schemes, sections, and rules mentioned in this book are intended solely for professional understanding. For complete and authoritative information, readers are advised to refer to the relevant Bare Acts and official legal documents.

Justice
M.S. Sonak
The Chief Justice,
High Court of Jharkhand -cum-
Patron-in-Chief Judicial Academy,
Jharkhand

MESSAGE

Artificial Intelligence is steadily reshaping systems of governance and public administration. Its growing presence in institutional decision-making calls for careful reflection on legality, accountability, and constitutional propriety. In this context, the publication "*AI Governance: Law, Ethics and Policy*" is both relevant and timely.

The justice delivery system rests on human judgment, fairness, and responsibility. While technological tools may assist courts and allied institutions in improving efficiency and managing information, they must always remain subordinate to judicial discretion and constitutional values. This work appropriately recognises that Artificial Intelligence can aid the administration of justice, but cannot supplant human reasoning or dilute institutional accountability.

The Judicial Academy, Jharkhand deserves appreciation for undertaking this initiative as part of its continuing commitment to judicial education. The book offers a measured and insightful perspective on the opportunities and limits of Artificial Intelligence, particularly within the Indian constitutional framework.

I am confident that this publication will serve as a useful reading material for judges, legal professionals, and other stakeholders, and will contribute to a thoughtful and responsible engagement with technology in public institutions.

Warm regards,
Justice M.S. Sonak

Justice
Rongon Mukhopadhyay
Judge,
High Court of Jharkhand
cum-Judge-In-Charge,
Judicial Academy, Jharkhand

MESSAGE

The emergence of Artificial Intelligence as a pervasive force in public administration presents both opportunities and challenges for constitutional governance. As algorithmic systems increasingly inform decision-making processes across institutions, it becomes imperative to examine their compatibility with the principles of legality, fairness, accountability, and due process. The present publication, ***"AI Governance: Law, Ethics and Policy,"*** is a timely and considered response to these concerns.

Within the justice delivery system, the use of Artificial Intelligence requires particular circumspection. Courts and allied institutions are entrusted with the protection of personal liberty, fundamental rights, and human dignity. Any technological intervention in such domains must therefore be subjected to rigorous legal scrutiny and guided by established constitutional values. Artificial Intelligence may serve as an aid in managing information, enhancing efficiency, and supporting institutional functions; however, it cannot be permitted to dilute human responsibility or displace judicial discretion. This work adopts a balanced and principled approach by situating technological advancement within the framework of law and ethics.

The Judicial Academy, Jharkhand has undertaken this initiative in furtherance of its obligation to equip judicial officers and other stakeholders with contemporary knowledge relevant to the evolving landscape of law and governance. I appreciate the scholarly effort and institutional commitment that have gone into the preparation of this publication, which will serve as a valuable reference for judicial officers, law enforcement agencies, prosecutors, legal professionals, policymakers, academicians, and students.

It is my considered view that informed engagement with Artificial Intelligence, guided by constitutional discipline and ethical restraint, is essential to ensuring that technological progress strengthens, rather than undermines, public trust in institutions. I trust that this book will contribute meaningfully to such engagement.

I commend this publication to the readers.

Warm regards,
Justice Rongon Mukhopadhyay

Rajesh Sharan Singh
Director,
Judicial Academy, Jharkhand

PREFACE

Artificial Intelligence has, in recent years, transitioned from the realm of academic research and speculative discourse into the very centre of public administration and governance. Algorithm-driven systems today influence modes of communication, business processes, governmental decision-making, and increasingly, the administration of justice. This expanding role of Artificial Intelligence raises significant questions relating not merely to technology, but also to law, ethics, accountability, constitutional values, and human agency. The present book has been conceived and prepared in response to these evolving concerns.

The principal objective of this publication, titled ***"AI Governance: Law, Ethics and Policy"***, is to provide a clear, structured, and accessible understanding of Artificial Intelligence and its implications for key institutional domains, particularly the judiciary, police, prosecution, and the banking and financial sector, with specific reference to the Indian context. While discussions on Artificial Intelligence are often characterised by technical complexity, this work consciously adopts a non-technical and explanatory approach. It seeks to make the subject comprehensible and relevant for judicial officers, members of the Bar, police and prosecution officials, policymakers, academicians, students, and other stakeholders engaged in the justice delivery system.

The book commences with a foundational overview of Artificial Intelligence, tracing its conceptual evolution and explaining the functioning of contemporary AI systems. It thereafter examines global approaches to AI governance, including international principles, ethical frameworks, and regulatory responses adopted across jurisdictions. These comparative perspectives provide valuable insight into the emerging global consensus on the responsible development and deployment of Artificial Intelligence.

A substantial portion of the work is devoted to India's approach towards Artificial Intelligence. India's AI strategy reflects a calibrated balance between innovation and regulation. Through investments in digital public infrastructure, promotion of indigenous AI capabilities, and emphasis on capacity building, India seeks to harness the benefits of Artificial Intelligence for national development, while remaining anchored to constitutional values and socio-economic realities. This balanced approach assumes particular significance in a diverse and populous country, where technological interventions can have far-reaching consequences.

Special emphasis has been placed on the justice delivery system. Courts, police, and prosecutors function in areas where personal liberty, human dignity, and fundamental rights are directly

implicated. The application of Artificial Intelligence in these domains, therefore, warrants the highest degree of caution and oversight. Through an analysis of judicial pronouncements, policy initiatives, and practical illustrations, this book underscores a consistent and principled position emerging from Indian institutions—**technology must supplement and assist the administration of justice, and cannot substitute human judgment.**

The examination of Artificial Intelligence in the banking and financial sector further demonstrates how technological tools can enhance efficiency, accuracy, and security, while simultaneously posing new regulatory and ethical challenges. These sectoral experiences collectively reinforce the central theme of this work: Artificial Intelligence must operate within a framework of transparency, accountability, responsibility, and meaningful human control.

The **Judicial Academy, Jharkhand** takes considerable pride in presenting this publication as part of its mandate to promote continuous judicial education and capacity building in emerging areas of law and governance. The Academy places on record its profound gratitude to **His Lordship Hon'ble Mr. Justice Rongon Mukhopadhyay**, Judge, High Court of Jharkhand and Judge-in-Charge, Judicial Academy, Jharkhand, for his constant encouragement, visionary leadership, and invaluable guidance. His Lordship's sustained support and scholarly engagement have been instrumental in the successful culmination of this work.

This book is the outcome of the collective and dedicated efforts of **Sri Satyakam Priyadarshi**, Additional Director-I cum Senior Faculty Member **Sri Laxmi Kant**, Additional Director-II cum Senior Faculty Member, Judicial Academy, Jharkhand; **Sri Amikar Parwar**, Administrative Officer; and the Research Scholars **Ms. Jyotsna Singh, Ms. Sarita Akhuli, and Mr. Uday Narayan**. Their meticulous research, analytical rigour, and unwavering commitment have significantly contributed to the quality and depth of this publication.

The Academy also acknowledges the guidance and scholarly insights of **Prof. (Dr.) Avinash Dadhich**, Founding Director, Dhirubhai Ambani University – School of Law, whose academic mentorship has enriched the conceptual framework of this book and strengthened its alignment with constitutional principles, ethics, and the rule of law.

This work does not advocate either uncritical adoption or outright rejection of Artificial Intelligence. Rather, it promotes an informed, cautious, and ethically grounded engagement with technology. It is earnestly hoped that this publication will contribute to a reasoned and responsible discourse on AI governance and serve as a practical and thought-provoking resource for all stakeholders involved in the administration of justice and governance in the digital age.

If this book succeeds in fostering thoughtful and responsible engagement with Artificial Intelligence, it will have fulfilled its intended purpose. I hope and wish this reading material is going to be beneficial for the readers.

Rajesh Sharan Singh
Director,
Judicial Academy, Jharkhand

Contents

Chapter 1: Artificial Intelligence: Evolution, Conceptual Understanding, and Contemporary Forms

"Technology will integrate police, forensics, jails, and courts, and will speed up their work as well. We are moving towards a justice system that will be fully future-ready."

-Prime Minister, Shri Narendra Modi

Artificial Intelligence (AI) has emerged as one of the most transformative developments of the modern era, influencing nearly every sphere of human activity, from healthcare and education to governance, finance, and law. What was once a speculative idea confined to science fiction has now become a practical and powerful technological reality. Despite its widespread use, AI is often misunderstood, either viewed as an all-powerful autonomous system or reduced to simple automation tools. A clear understanding of its meaning, historical evolution, present structure, and different forms is therefore essential.

1.1. Understanding Artificial Intelligence

Artificial Intelligence broadly refers to the capability of machines or computer systems to perform tasks that normally require human intelligence. These tasks include reasoning, learning from experience, problem-solving, understanding natural language, recognizing patterns, and making decisions.

One of the earliest and most influential definitions was proposed in 1956 by John McCarthy, widely recognized as the father of Artificial Intelligence due to his astounding contribution in the field of Computer Science and AI. In his article named ***What is Artificial Intelligence?*****1**, McCarthy defines AI as following: *"It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable."* Where intelligence is *"the computational part of the ability to achieve goals in the world. Varying kinds and degrees of intelligence occur in people, many animals and some machines."* Later definitions refined this idea by emphasizing rational behaviour and learning ability.

According to the Organisation for Economic Co-operation and Development (OECD), AI is a machine-based system that can make predictions, recommendations, or decisions influencing real or virtual environments based on data and algorithms. The OECD definition of an AI system contained in the ***OECD AI Principles*** (OECD, 2019**2**); (OECD, 2019**3**) built on the conceptual view of AI detailed in ***Artificial Intelligence: A Modern Approach*** (Russell and

1 WHAT IS ARTIFICIAL INTELLIGENCE? John McCarthy, Computer Science Department Stanford University, Stanford, CA 94305, jmc@cs.stanford.edu, Revised November 12, 2007: https://www-formal.stanford.edu/jmc/whatisai/node1.html

2 OECD (2019), Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449.

3 OECD (2019), "Scoping the OECD AI principles: Deliberations of the Expert Group on Artificial Intelligence at the OECD (AIGO)", OECD Digital Economy Papers, No. 291, OECD Publishing, Paris, https://doi.org/10.1787/d62f618a-en.

Norvig, 2009**4**). It read: *"An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy".*

The definition was further updated by the OECD publishing titled ***Explanatory Memorandum On the Updated OECD Definition of an AI System***[5], on March 2024 as following:

"An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment."

What distinguishes AI from traditional computer programs is its capacity to learn and adapt. Conventional software follows predefined instructions, whereas AI systems can improve their performance by analysing data, identifying patterns, and modifying their responses accordingly. *Adaptiveness*, as reflected in the revised definition of an AI system, generally refers to AI systems based on machine learning that retain the capacity to evolve even after their initial development. Such systems adjust their functioning through continuous interaction with data and inputs, either before or after deployment. Common examples include speech recognition tools that gradually attune themselves to a user's voice or music recommendation platforms that refine suggestions based on listening preferences. AI systems may undergo training on a one-time, periodic, or ongoing basis and operate by identifying patterns and relationships within data. Through this learning process, certain AI systems may acquire the ability to generate new forms of inference that were not explicitly anticipated by their designers.

1.2. Historical Evolution of Artificial Intelligence

1.2.1. Early Conceptual Roots: From Ancient Thought to Modern Science

The idea of artificial intelligence long predates modern computing. Ancient Greek civilizations imagined artificial life through myths and mechanical inventions, reflected in early *automaton*, self-acting devices designed to imitate living beings. Notable examples include the mechanical dove attributed to *Archytas* in the 4th century BCE and **Leonardo da Vinci's** Renaissance-era *mechanical knight*.[6] While these early efforts demonstrated human curiosity about artificial cognition, substantive progress became possible only in the twentieth century with the advent of electronic computing. Growing philosophical and scientific inquiry into the nature of intelligence during this period laid the groundwork for formal AI research.

1.2.2. Birth of Artificial Intelligence (1950s)

Artificial intelligence emerged as a formal scientific discipline in the 1950s. In 1950, Alan Turing'**s** seminal paper *"Computing Machinery and Intelligence"* posed the question "Can machines

4 Russell, S. and P. Norvig (2009), Artificial Intelligence: A Modern Approach, 3rd edition, Pearson, London, http://aima.cs.berkeley.edu/.

5 OECD publishing, EXPLANATORY MEMORANDUM ON THE UPDATED OECD DEFINITION OF AN AI SYSTEM, OECD ARTIFICIAL INTELLIGENCE PAPERS, March 2024, No. 8: https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/03/explanatory-memorandum-on-the-updated-oecd-definition-of-an-ai-system_3c815e51/623da898-en.pdf

6 What is the history of artificial intelligence (AI)?, Tableau from Salesforce, https://www.tableau.com/data-insights/ai/history

think?" and introduced the Turing Test as a benchmark for machine intelligence.**7** The term "Artificial Intelligence" was coined in 1956 at the *Dartmouth Summer Research Project*, organized by John McCarthy, M. L. Minsky, N. Rochester, and C. E. Shannon,**8** widely regarded as the field's founding moment. Early researchers expressed optimism that human-level intelligence could be achieved within decades, focusing primarily on symbolic reasoning and rule-based systems such as the Logic Theorist, which demonstrated that machines could replicate limited forms of human problem-solving.

1.2.3. Expansion and Early Optimism (1960s–1970s)

The 1960s and early 1970s marked a period of rapid expansion in artificial intelligence research, dominated by symbolic or "good old-fashioned" AI, which relied on explicit rules and structured knowledge representations. This era also saw the development of specialized programming languages such as LISP to support AI research.

Significant milestones included Rosenblatt's Mark I Perceptron (1957), an early neural network designed for pattern recognition, which advanced machine learning and brain-inspired computation.[9] However, Minsky and Papert's *Perceptrons* (1969) exposed the limitations of early neural networks, particularly single-layer models, temporarily curtailing further work in this area. In natural language processing, Weizenbaum's ELIZA (1965) demonstrated that machines could simulate human conversation through pattern matching, though without genuine understanding.

This period also witnessed the growth of expert systems aimed at replicating human decision-making through rule-based logic, though they proved rigid and ill-suited to ambiguity. Artificial intelligence entered public consciousness through films such as *2001: A Space Odyssey* (1968) and popular characters like C-3PO and R2-D2 in *Star Wars* (1977). Early consumer applications, including the Speak & Spell educational toy (1978), illustrated initial attempts to integrate AI into everyday contexts. Despite widespread optimism, these systems performed reliably only in controlled settings and were constrained by limited computing power, insufficient data, and a lack of adaptability, contextual reasoning, and common-sense understanding.

1.2.4. The AI Winters: Decline and Reassessment (1970s–1990s)

Unrealistic expectations and technical limitations led to periods of reduced funding and skepticism, known as AI winters. The first occurred in the mid-1970s, as expert systems proved costly and inflexible. Japan's ambitious Fifth Generation Computer Project in the 1980s further highlighted the challenges of achieving general intelligence[10]. Despite setbacks, progress continued in niche areas such as robotics and autonomous systems. In 1984, **NAVLab** developed one of the first autonomous land vehicles, demonstrating the potential of AI in robotics and

7 What is artificial intelligence (AI)?, Cole Stryker, Staff Editor, AI Models, IBM Think, Eda Kavlakoglu, Business Development + Partnerships, IBM Research, https://www.ibm.com/think/topics/artificial-intelligence

8 A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, 1956, https://home.dartmouth.edu/about/artificial-intelligence-ai-coined-dartmouth

9 Artificial Intelligence, Explained, Jennifer Monahan, Carniegie Mellon University, Heinz College, https://www.heinz.cmu.edu/media/2023/July/artificial-intelligence-explained

10 Id.

real-world navigation despite broader setbacks in the field. A second AI winter in the late 1980s and early 1990s followed the collapse of the expert systems market and specialized AI hardware.

1.2.5. Revival Through Machine Learning and Data (1990s–2000s)

The 1990s marked a shift from rule-based AI to data-driven machine learning approaches. Techniques such as neural networks, decision trees, and support vector machines gained prominence, supported by advances in computing power and increased data availability. The rise of data science as an interdisciplinary field further accelerated progress. A landmark achievement of this era was IBM's Deep Blue defeating world chess champion Garry Kasparov in 1997, symbolizing AI's renewed potential.[11] By the early 2000s, AI found widespread application in search engines, recommendation systems, speech recognition, and robotics.

1.2.6. The Deep Learning Revolution (2000s–2010s)

The 2010s marked a transformative phase with the rise of deep learning, based on multi-layered neural networks capable of processing complex and unstructured data. This revolution was driven by large datasets, improved GPUs, and advanced training algorithms. Breakthroughs such as IBM Watson's victory on *Jeopardy!* in 2011, Baidu's deep neural network achieved human-level accuracy in image recognition using convolutional neural networks in 2015, *AlphaGo's* defeat of the world Go champion Lee Sedol in 2016, and major advances in image and speech recognition demonstrated AI's expanding capabilities. Significant investment by major technology companies accelerated both research and deployment.

1.2.7. The Age of Generative and Multimodal AI (2020s–Present)

The 2020s ushered in a new phase of artificial intelligence with the emergence of Generative Pre-trained Transformers and other large language models trained on vast datasets to produce coherent, context-aware, and human-like language. Systems such as GPT-3 marked a major breakthrough in natural language processing, demonstrating the ability to generate essays, translations, summaries, and creative content by learning linguistic and semantic patterns directly from data.

Generative AI rapidly evolved beyond basic text production to function as writing assistants, translators, creative tools, coding aids, and conversational systems. Recent advances have further enabled multimodal models capable of integrating text, images, and speech, thereby supporting richer and more natural human–machine interaction.

The deployment of tools such as ChatGPT, Bard, and Bing Copilot has significantly enhanced productivity across sectors including education, healthcare, law, research, and business, positioning generative AI as a central driver of digital transformation. At the same time, the field has moved toward greater efficiency through optimized training methods and smaller, high-performing models. The widespread adoption of generative AI has also intensified ethical and regulatory concerns relating to privacy, bias, misinformation, intellectual property, and transparency, prompting the development of governance frameworks for responsible AI use.

11 What is artificial intelligence (AI)?, Cole Stryker, Staff Editor, AI Models, IBM Think, Eda Kavlakoglu, Business Development + Partnerships, IBM Research, https://www.ibm.com/think/topics/artificial-intelligence

1.3. Fundamental Concepts of Artificial Intelligence

A clear understanding of the foundational concepts underlying artificial intelligence is essential to appreciate how AI systems function. These concepts constitute the building blocks of AI and provide the conceptual framework for designing intelligent algorithms and computational models.

1.3.1 Machine Learning (ML):

Most contemporary advances in AI are driven by machine learning, which enables systems to learn from data and make informed decisions without being explicitly programmed. Deep learning, a specialized subfield of ML, has been particularly transformative in pattern recognition, large-scale data processing, and high-accuracy prediction.

1.3.2 Neural Networks:

Neural networks are computational architectures inspired by the structure and functioning of the human brain. Composed of interconnected artificial neurons, they process data through mathematical operations to generate outputs. Neural networks are especially effective in tasks such as image recognition, natural language processing, and speech recognition, and they form the core of deep learning systems.

1.3.3 Deep Learning:

Deep learning is a branch of machine learning that employs multi-layered neural networks to identify complex patterns within large datasets. It enables machines to model hierarchical representations of data and to approximate aspects of human perception and decision-making.

1.3.4 Natural Language Processing (NLP):

Natural language processing focuses on enabling machines to understand, interpret, and generate human language. With the development of advanced models such as GPT-4, AI systems can now perform tasks including text summarization, report drafting, translation, and conversational interaction, significantly enhancing human–machine communication.

1.3.5 Computer Vision:

Computer vision refers to AI's capacity to interpret and analyze visual information. It has driven major advances in facial recognition, autonomous vehicles, and medical diagnostics, transforming the use of visual data across sectors—from disease detection in medical imaging to applications in retail and security.

1.3.6 Expert Systems:

Expert systems are AI applications designed to replicate the decision-making abilities of human specialists. They employ knowledge bases and inference engines to solve complex problems using rules and heuristics derived from human expertise, with applications historically spanning medical diagnosis, financial planning, and risk assessment.

1.4. Purpose and Use of Artificial Intelligence

The primary purpose of artificial intelligence is to augment human capabilities and support more informed and efficient decision-making. From a technical perspective, AI systems are designed to analyze data, recognize patterns, and assist humans in addressing complex problems with greater accuracy and speed.[12] From a broader philosophical standpoint, artificial intelligence holds the potential to reduce dependence on repetitive and labor-intensive work, enhance human well-being, and support the management of increasingly complex social, economic, and institutional systems in ways that serve collective human interests.

Although such transformative outcomes remain largely aspirational, contemporary AI applications are primarily directed toward improving operational efficiency, automating resource-intensive processes, and enabling data-driven decision-making across industries. Organizations increasingly rely on AI to generate insights into user behaviour and to deliver personalized recommendations. Predictive search engines, streaming platforms, and social media applications exemplify this trend, using historical data to anticipate user preferences, optimize content delivery, and streamline user interaction.

At an operational level, AI systems are widely employed for optimized information retrieval, rule-based and logic-driven automation, large-scale pattern detection, and probabilistic modeling to forecast future outcomes. Collectively, these functions illustrate AI's central role in transforming how data is processed, decisions are formulated, and services are delivered.

1.5. Advantages of Artificial Intelligence

Artificial intelligence has become an integral component of modern technological ecosystems, enhancing both everyday conveniences and critical infrastructures. Its key advantages include the reduction of human error, continuous 24/7 operational capability, and the automation of repetitive and time-intensive tasks. AI-powered digital assistants and intelligent systems enable faster and more consistent decision-making, grounded in rational, data-driven analysis.

AI has demonstrated particular value in fields such as healthcare, where it supports diagnostics and treatment planning, and in security systems that enhance surveillance, fraud detection, and risk mitigation. Additionally, artificial intelligence contributes to improved communication systems, personalized services, and efficient information management, thereby strengthening productivity and reliability across sectors.

12 Bureau of Police Research and Development. (2022). *AI in the service of law enforcement: An introduction*. National Crime Research & Innovation Centre (NCR&IC), Bureau of Police Research & Development, Ministry of Home Affairs, Government of India. https://bprd.nic.in/uploads/pdf/AI%20in%20the%20service%20of%20Law%20Enforcement-%20a%20n%20Introduction.pdf

1.6. Recent Developments in Artificial Intelligence and Their Implications for the Legal System

Recent advances in artificial intelligence are redefining how legal professionals conduct research, draft documents, and conceptualize justice. Traditional predictive tools have given way to powerful **Generative AI models**, such as large language models (LLMs), which generate coherent text and legal analysis by learning from extensive datasets. These models have demonstrated the ability to assist in legal research, document drafting, and case analysis, although concerns about accuracy and ethical obligations persist. Research indicates that while generative AI can improve efficiency, it may also "hallucinate" information producing plausible but incorrect outputs that can undermine legal judgement if not verified by users.[13]

Generative AI's influence on law extends beyond efficiency gains. It raises profound questions regarding professional duties and ethical standards. For instance, the use of generative AI in legal practice implicates core lawyer obligations such as competence and confidentiality, especially when client data are processed by third-party AI systems. Cambridge University Press[14] emphasize that lawyers must exercise rigorous oversight and ensure outputs derive from authoritative sources to avoid professional misconduct.[15]

Moreover, scholarly analyses highlight the integration of generative AI within law schools and legal education, suggesting that these technologies will require new competencies in legal reasoning and academic integrity frameworks.

Another frontier in AI research involves **agentic AI**, which differs from generative models by autonomously planning and executing sequences of actions without direct prompting. Such systems, characterized by autonomy, proactivity, and adaptability, pose unique challenges for legal governance as they transition from advisory roles to autonomous decision-making entities.[16] These transformations provoke pressing legal and ethical questions about accountability, responsibility, and the legal status of autonomous actions, particularly when multiple stakeholders (developers, deployers, users) interact with agentic systems[17].

An emerging area of theoretical and technical interest is the intersection of **quantum computing and AI**. Preliminary work suggests that quantum-enhanced AI could accelerate optimization and complex legal data analysis, potentially reshaping high-stakes tasks such as large-scale

13 Schwarcz, D., Manning, S., Barry, P. J., Cleveland, D. R., Prescott, J. J., & Rich, B. (2025). AI-powered lawyering: AI reasoning models, retrieval augmented generation, and the future of legal practice. Minnesota Legal Studies Research Paper No. 25-16. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5162111

14 Terzidou, K. (2025). *Generative AI systems in legal practice offering quality legal services while upholding legal ethics*. *International Journal of Law in Context, Vol. 21* Issue 3, https://www.cambridge.org/core/journals/international-journal-of-law-in-context/article/generative-ai-systems-in-legal-practice-offering-quality-legal-services-while-upholding-legal-ethics/34011A84AA58A2BAB556A406A4653A8D

15 Coles, T. (2025, October 31). *Legal research using generative AI*. Thomson Reuters Legal Solutions. https://legalsolutions.thomsonreuters.co.uk/blog/2025/10/31/legal-research-using-generative-ai/

16 Hosseini, S. (2025). *The role of agentic AI in shaping a smart future*. *Journal of Intelligent & Robotic Systems, Volume 26, July 2025, 100399*. https://www.sciencedirect.com/science/article/pii/S2590005625000268

17 Truby, J. (2025). *Global governance of transboundary risks from agentic AI: Enhancing international standards for auditing AI management systems to comply with state duties.* In P. Hacker (Ed.), *Oxford Intersections: AI in Society*. Oxford University Press. https://doi.org/10.1093/9780198945215.003.0155

predictive modeling or cryptographic evaluation.[18] Although still nascent, such developments may prompt legal systems to revisit foundational doctrines governing digital evidence, data security, and computational transparency.

Despite these transformative potentials, generative and agentic AI applications have already revealed practical limits and regulatory gaps in real legal environments. Courts in multiple jurisdictions have reported instances where lawyers submitted AI-generated filings containing fictitious cases or inaccurate legal citations, a phenomenon attributed to hallucinations inherent in current generative models. These cases have prompted judicial admonitions and disciplinary inquiries, illustrating the tension between innovation and professional accountability in AI-assisted practice.[19]

18 Sultanow, E., Tehrani, M. G., Dutta, S., Buchanan, W. J., & Khan, M. S. (2025). *Quantum agents: Integrating quantum computing with agentic artificial intelligence* (arXiv:2506.01536v1). arXiv. https://arxiv.org/abs/2506.01536v1

19 Associated Press. (2025, January 22). UK courts warn lawyers after fake AI-generated legal cases submitted to judges. AP News. https://apnews.com/article/uk-courts-fake-ai-cases-46013a78d78dc869bdfd6b42579411cb

CHAPTER 2: GLOBAL SCENARIO

2.1 OECD AI Principles: Guardrails to Responsible AI Adoption[20]

Amid the rapid expansion of artificial intelligence (AI) across the digital landscape, the Organisation for Economic Co-operation and Development (OECD) has formulated a comprehensive set of AI principles to guide the responsible, trustworthy, and effective adoption of AI technologies. These principles provide a global framework for shaping AI governance and are intended to influence how governments, institutions, and industries deploy AI in a manner that is ethical, transparent, and aligned with public interest.

Established in 1961, the OECD is an international organisation comprising 38 member countries, committed to promoting evidence-based policymaking and international cooperation to enhance economic and social well-being worldwide. It serves as a collaborative platform where governments exchange experiences, develop common standards, and work collectively to address shared global challenges.

2.1.1 The Five OECD Principles on Artificial Intelligence[21]

The formulation of the OECD AI Principles began in 2018, when the **Organisation for Economic Co-operation and Development (OECD)** constituted a multidisciplinary expert group on artificial intelligence. This group brought together representatives from member states, industry leaders, academic experts, civil society, and other stakeholders. Their collaborative efforts resulted in a set of principles that were subsequently updated in 2024 and now serve as a global benchmark for responsible AI governance.

The OECD AI framework is designed to help governments, businesses, and institutions harness the benefits of AI while minimising its risks. It aims to ensure that AI systems are developed and deployed in a manner that is ethical, transparent, trustworthy, and aligned with broader societal objectives. The framework is structured around **five core principles**, complemented by recommendations for national policies and international cooperation.

i. Inclusive Growth, Sustainable Development, and Well-being

AI systems should contribute positively to society by promoting inclusive economic growth, sustainable development, and human well-being. This principle stresses that the benefits of AI must extend to all sections of society, rather than being concentrated among a few.

The focus is on using AI to reduce inequalities, bridge digital divides, and support sustainable practices. AI-driven solutions are encouraged in critical areas such as climate change mitigation, healthcare delivery, education, and social welfare. When AI is applied with these objectives in mind, it can become a powerful tool for addressing global challenges and advancing collective prosperity.

20 *EU AI Act: first regulation on artificial intelligence*, Article of European Parliament, 19th Feb 2025, https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

21 OECD (2019), "Scoping the OECD AI principles: Deliberations of the Expert Group on Artificial Intelligence at the OECD (AIGO)", OECD Digital Economy Papers, No. 291, OECD Publishing, Paris, https://doi.org/10.1787/d62f618a-en.

ii. Human-Centred Values and Fairness

AI technologies must respect human rights, democratic values, and human dignity. This principle underscores that AI systems should be designed and used in ways that promote fairness, equality, and non-discrimination.

To achieve this, organisations must actively prevent bias in AI systems by ensuring diverse and representative training data, designing algorithms that minimise discriminatory outcomes, and conducting regular fairness audits. AI applications should not reinforce existing prejudices or create new forms of exclusion. By embedding human-centred values at every stage of AI development, trust and social acceptance of AI can be strengthened.

iii. Transparency and Explainability

AI systems should operate in a transparent and understandable manner. Users and stakeholders must be able to comprehend how AI systems function and how decisions or recommendations are generated.

Explainability involves making the logic, data sources, and reasoning behind AI outputs accessible and intelligible, particularly where AI decisions affect rights, opportunities, or legal outcomes. Clear documentation, user-friendly explanations, and mechanisms for questioning or reviewing AI decisions are essential. Transparency not only enhances trust but also enables accountability and informed use of AI technologies.

iv. Robustness, Security, and Safety

AI systems should be reliable, secure, and resilient throughout their entire lifecycle. This principle requires that AI technologies are designed to function safely under both normal and unexpected conditions.

Organisations must protect AI systems from cyber threats, data manipulation, and misuse, while ensuring consistent performance. Regular testing, stress assessments, and monitoring are necessary to detect vulnerabilities, biases, or failures. Safeguards and fail-safe mechanisms should be incorporated to prevent harm and ensure that AI systems remain dependable even in challenging environments.

v. Accountability

Those who design, deploy, and operate AI systems must be accountable for their functioning and impacts. Accountability ensures that AI-related decisions and outcomes can be traced, reviewed, and corrected when necessary.

This principle calls for clear governance frameworks that define roles and responsibilities, establish oversight mechanisms, and ensure compliance with legal and ethical standards. Organisations should maintain audit trails, enable independent review, and provide effective grievance redressal mechanisms. When errors or harms occur, prompt corrective action must be taken, reinforcing the principle that responsibility for AI systems ultimately rests with humans.

OECD Recommendations for Trustworthy AI Adoption

The degustation AI menu of the OECD closes with the following list of recommendations:

- Invest in AI Research and Development: Encourage innovation while addressing ethical and technical challenges.
- Foster a Digital Ecosystem for AI: Promote policies that support data access, infrastructure, and skills development.
- Shape an Enabling Policy Environment for AI: Develop legal and regulatory frameworks that facilitate AI adoption while protecting public interests.
- Build Human Capacity and Prepare for Labor Market Transformation: Equip the workforce with the necessary skills to thrive in an AI-driven economy.
- International Cooperation for Trustworthy AI: Collaborate globally to address cross-border AI issues and establish common standards.

2.2 EU AI ACT[22]

2.2.1 Four-point summary[23]

1.	**The AI Act classifies AI according to its risk:** • Unacceptable risk is prohibited (e.g. social scoring systems and manipulative AI). • Most of the text addresses high-risk AI systems, which are regulated. • A smaller section handles limited risk AI systems, subject to lighter transparency obligations: developers and deployers must ensure that end-users are aware that they are interacting with AI (chatbots and deepfakes). • Minimal risk is unregulated (including the majority of AI applications currently available on the EU single market, such as AI enabled video games and spam filters – at least in 2021; this is changing with generative AI

22 https://artificialintelligenceact.eu/high-level-summary/

23 EU AI Act: first regulation on artificial intelligence, Article of European Parliament, 19th Feb 2025, https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

2.	**The AI Act classifies AI according to its risk:** • Unacceptable risk is prohibited (e.g. social scoring systems and manipulative AI). • Most of the text addresses high-risk AI systems, which are regulated. • A smaller section handles limited risk AI systems, subject to lighter transparency obligations: developers and deployers must ensure that end-users are aware that they are interacting with AI (chatbots and deepfakes). • Minimal risk is unregulated (including the majority of AI applications currently available on the EU single market, such as AI enabled video games and spam filters – at least in 2021; this is changing with generative AI
3.	**The majority of obligations fall on providers (developers) of high-risk AI systems.** • Those that intend to place on the market or put into service high-risk AI systems in the EU, regardless of whether they are based in the EU or a third country. • And also third country providers where the high risk AI system's output is used in the EU.
4.	**Deployers are natural or legal persons that deploy an AI system in a professional capacity, not affected end-users.** • Deployers of high-risk AI systems have some obligations, though less than providers (developers). • This applies to deployers located in the EU, and third country users where the AI system's output is used in the EU.

General purpose AI (GPAI):

- All GPAI model providers must provide technical documentation, instructions for use, comply with the Copyright Directive, and publish a summary about the content used for training.
- Free and open licence GPAI model providers only need to comply with copyright and publish the training data summary, unless they present a systemic risk.
- All providers of GPAI models that present a systemic risk – open or closed – must also conduct model evaluations, adversarial testing, track and report serious incidents and ensure cybersecurity protections.

2.2.2 Prohibited AI systems (Chapter II, Art. 5)

AI systems:

- deploying subliminal, manipulative, or deceptive techniques to distort behaviour and impair informed decision-making, causing significant harm.

- exploiting vulnerabilities related to age, disability, or socio-economic circumstances to distort behaviour, causing significant harm.
- social scoring, i.e., evaluating or classifying individuals or groups based on social behaviour or personal traits, causing detrimental or unfavourable treatment of those people.
- assessing the risk of an individual committing criminal offenses solely based on profiling or personality traits, except when used to augment human assessments based on objective, verifiable facts directly linked to criminal activity.
- compiling facial recognition databases by untargeted scraping of facial images from the internet or CCTV footage.
- inferring emotions in workplaces or educational institutions, except for medical or safety reasons.
- biometric categorisation systems inferring sensitive attributes (race, political opinions, trade union membership, religious or philosophical beliefs, sex life, or sexual orientation), except labelling or filtering of lawfully acquired biometric datasets or when law enforcement categorises biometric data.
- 'real-time' remote biometric identification (RBI) in publicly accessible spaces for law enforcement, except when:
- targeted searching for missing persons, abduction victims, and people who have been human trafficked or sexually exploited;
- preventing specific, substantial and imminent threat to life or physical safety, or foreseeable terrorist attack; or
- identifying suspects in serious crimes (e.g., murder, rape, armed robbery, narcotic and illegal weapons trafficking, organised crime, and environmental crime, etc.).

Notes on remote biometric identification:

- Using AI-enabled real-time RBI is only allowed when not using the tool would cause harm, particularly regarding the seriousness, probability and scale of such harm, and must account for affected persons' rights and freedoms.
- Before deployment, police must complete a fundamental rights impact assessment and register the system in the EU database, though, in duly justified cases of urgency, deployment can commence without registration, provided that it is registered later without undue delay.
- Before deployment, they also must obtain authorisation from a judicial authority or independent administrative authority[24], though, in duly justified cases of urgency, deployment 1 can commence without authorisation, provided that authorisation is requested within 24 hours. If authorisation is rejected, deployment must cease immediately, deleting all data, results, and outputs.

24 Independent administrative authorities may be subject to greater political influence than judicial authorities (Hacker, 1 2024)

2.2.3 High risk AI systems (Chapter III)

2.2.3.1 Classification rules for high-risk AI systems (Art. 6)

High risk AI systems are those:

- ✓ used as a safety component or a product covered by EU laws in Annex I AND required to undergo a third-party conformity assessment under those Annex I laws; OR
- ✓ those under Annex III use cases (below), except if:
- the AI system performs a narrow procedural task;
- improves the result of a previously completed human activity;
- detects decision-making patterns or deviations from prior decision-making patterns and is not meant to replace or influence the previously completed human assessment without proper human review; or
- performs a preparatory task to an assessment relevant for the purpose of the use cases listed in Annex III.
    - ✓ The Commission can add or modify the above conditions through delegated acts if there is concrete evidence that an AI system falling under Annex III does not pose a significant risk to health, safety and fundamental rights. They can also delete any of the conditions if there is concrete evidence that this is needed to protect people.
    - ✓ AI systems are always considered high-risk if it profiles individuals, i.e. automated processing of personal data to assess various aspects of a person's life, such as work performance, economic situation, health, preferences, interests, reliability, behaviour, location or movement.
    - ✓ Providers that believe their AI system, which fails under Annex III, is not high-risk, must document such an assessment before placing it on the market or putting it into service.
    - ✓ 18 months after entry into force, the Commission will provide guidance on determining if an AI system is high risk, with list of practical examples of high-risk and non-high risk use cases.

2.2.3.2 Requirements for providers of high-risk AI systems (Art. 8-17)

High risk AI providers must:

- Establish a risk management system throughout the high risk AI system's lifecycle;
- Conduct data governance, ensuring that training, validation and testing datasets are relevant, sufficiently representative and, to the best extent possible, free of errors and complete according to the intended purpose.
- Draw up technical documentation to demonstrate compliance and provide authorities with the information to assess that compliance.

- Design their high risk AI system for record-keeping to enable it to automatically record events relevant for identifying national level risks and substantial modifications throughout the system's lifecycle.
- Provide instructions for use to downstream deployers to enable the latter's compliance.
- Design their high risk AI system to allow deployers to implement human oversight.
- Design their high risk AI system to achieve appropriate levels of accuracy, robustness, and cybersecurity.
- Establish a quality management system to ensure compliance.

2.2.4 General purpose AI (GPAI) (Chapter V)

GPAI model means an AI model, including when trained with a large amount of data using selfsupervision at scale, that displays significant generality and is capable to competently perform a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications. This does not cover AI models that are used before release on the market for research, development and prototyping activities.

GPAI system means an AI system which is based on a general purpose AI model, that has the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems.

GPAI systems may be used as high risk AI systems or integrated into them. GPAI system providers should cooperate with such high risk AI system providers to enable the latter's compliance.

2.2.4.1 All providers of GPAI models must (Art. 53):

- Draw up technical documentation, including training and testing process and evaluation results.
- Draw up information and documentation to supply to downstream providers that intend to integrate the GPAI model into their own AI system in order that the latter understands capabilities and limitations and is enabled to comply.
- Establish a policy to respect the Copyright Directive.
- Publish a sufficiently detailed summary about the content used for training the GPAI model.
- Free and open licence GPAI models – whose parameters, including weights, model architecture and model usage are publicly available, allowing for access, usage, modification and distribution of the model – only have to comply with the latter two obligations above, unless the free and open licence GPAI model is systemic.
- GPAI models are considered systemic when the cumulative amount of compute used for its training is greater than 10^25 floating point operations per second (FLOPS) (Art. 51). Providers must notify the Commission if their model meets this criterion within 2 weeks (Art. 52). The provider may present arguments that, despite meeting the criteria, their model does not present systemic risks. The Commission may decide on its own, or via

a qualified alert from the scientific panel of independent experts, that a model has high impact capabilities, rendering it systemic.

- In addition to the four obligations above, providers of GPAI models with systemic risk must also (Art. 55):
- Perform model evaluations, including conducting and documenting adversarial testing to identify and mitigate systemic risk.
- Assess and mitigate possible systemic risks, including their sources.
- Track, document and report serious incidents and possible corrective measures to the AI Office and relevant national competent authorities without undue delay.
- Ensure an adequate level of cybersecurity protection.
- All GPAI model providers may demonstrate compliance with their obligations if they voluntarily adhere to codes of practice until European harmonised standards are published, compliance with which will lead to a presumption of conformity (Art. 56). Providers that don't adhere to codes of practice must demonstrate alternative adequate means of compliance for Commission approval.

2.2.4.2 Codes of practice (Art. 56)

- Will account for international approaches.
- Will cover but not necessarily limited to the above obligations, particularly the relevant information to include in technical documentation for authorities and downstream providers, identification of the type and nature of systemic risks and their sources, and the modalities of risk management accounting for specific challenges in addressing risks due to the way they may emerge and materialise throughout the value chain.
- AI Office may invite GPAI model providers, relevant national competent authorities to participate in drawing up the codes, while civil society, industry, academia, downstream providers and independent experts may support the process.

2.2.5 Governance (Chapter VI)

- The AI Office will be established, sitting within the Commission, to monitor the effective implementation and compliance of GPAI model providers (Art. 64).
- Downstream providers can lodge a complaint regarding the upstream providers infringement to the AI Office (Art. 89).
- The AI Office may conduct evaluations of the GPAI model to (Art. 92):
  - ❖ assess compliance where the information gathered under its powers to request information is insufficient.
  - ❖ Investigate systemic risks, particularly following a qualified report from the scientific panel of independent experts (Art. 90).

- Timelines
  - ❖ After entry into force, the AI Act will apply by the following deadlines:
  - ❖ 6 months for prohibited AI systems.
  - ❖ 12 months for GPAI.
  - ❖ 24 months for high risk AI systems under Annex III.
  - ❖ 36 months for high risk AI systems under Annex I.
  - ❖ Codes of practice must be ready 9 months after entry into force.

2.3 OECD AI Principles as a Pathway to EU AI Act Compliance

The OECD AI Principles and the EU AI Act are closely aligned, so much so that the European legislation has, in essence, adopted the OECD's definition of what an AI system is. Hence, the OECD principles can become a valuable framework for organizations working towards EU AI Act compliance.[25]

OECD AI Principle	What the Principle Means (Simple Explanation)	Corresponding Focus under the EU AI Act	How It Helps with EU AI Act Compliance
Inclusive Growth, Sustainable Development & Well-being	AI should benefit society as a whole and support sustainable development	Ensuring AI systems are safe, lawful, and promote public interest, especially in high-risk areas	Encourages responsible deployment of AI systems that meet safety and societal impact requirements for high-risk AI
Human-Centered Values & Fairness	AI must respect human dignity, equality, and fundamental rights	Strong focus on non-discrimination, fairness, and protection of fundamental rights	Helps organizations address bias, prevent discriminatory outcomes, and ensure ethical AI use
Transparency & Explainability	AI decisions should be understandable and traceable	Mandatory transparency obligations, especially for high-risk AI systems	Supports compliance with documentation, user information, and explainability requirements
Robustness, Security & Safety	AI systems should be reliable, secure, and resilient against misuse	Risk management, cyber security, accuracy, and system reliability obligations	Assists in meeting technical standards related to safety, robustness, and resilience

25 https://code4thought.eu/2024/09/09/oecd-ai-principles-guardrails-to-responsible-ai-adoption/

OECD AI Principle	What the Principle Means (Simple Explanation)	Corresponding Focus under the EU AI Act	How It Helps with EU AI Act Compliance
Accountability	Clear responsibility should exist for AI design, deployment, and outcomes	Defined obligations for providers, deployers, and operators of AI systems	Helps establish governance frameworks, assign responsibility, and ensure legal accountability

2.4 The Framework Convention on Artificial Intelligence

The Council of Europe Framework Convention on Artificial Intelligence and human rights, democracy and the rule of law is the **first-ever international legally binding treaty** in this field. Opened for signature on 5 September 2024, it aims to ensure that activities within the lifecycle of artificial intelligence systems are fully consistent with human rights, democracy and the rule of law, while being conducive to technological progress and innovation.

The Framework Convention complements existing international standards on human rights, democracy and the rule of law, and **aims to fill any legal gaps that may result from rapid technological advances**. In order to stand the test of time, the Framework Convention does not regulate technology and is essentially **technology-neutral**.

How was the Framework Convention elaborated?

Work was initiated in 2019, when the ad hoc Committee on Artificial Intelligence (CAHAI) was tasked with examining the feasibility of such an instrument. Following its mandate, it was succeeded in 2022 by the Committee on Artificial Intelligence (CAI) which drafted and negotiated the text.

The Framework Convention was drafted by the 46 member states of the Council of Europe, with the participation of all observer states: Canada, Japan, Mexico, the Holy See and the United States of America, as well as the European Union, and a significant number of non-member states: Australia, Argentina, Costa Rica, Israel, Peru and Uruguay.

In line with the Council of Europe's practice of multi-stakeholder engagement, 68 international representatives from civil society, academia and industry, as well as several other international organisations were also actively involved in the development of the Framework Convention.

What does the Framework Convention require states to do?

Fundamental principles

Activities within the lifecycle of AI systems must comply with the following fundamental principles:

- Human dignity and individual autonomy

- Equality and non-discrimination
- Respect for privacy and personal data protection
- Transparency and oversight
- Accountability and responsibility
- Reliability
- Safe innovation

Remedies, procedural rights and safeguards

- Document the relevant information regarding AI systems and their usage and to make it available to affected persons;
- The information must be sufficient to enable people concerned to challenge the decision(s) made through the use of the system or based substantially on it, and to challenge the use of the system itself;
- Effective possibility to lodge a complaint to competent authorities;
- Provide effective procedural guarantees, safeguards and rights to affected persons in connection with the application of an artificial intelligence system where an artificial intelligence system significantly impacts upon the enjoyment of human rights and fundamental freedoms;
- Provision of notice that one is interacting with an artificial intelligence system and not with a human being.

Risk and impact management requirements

- Carry out risk and impact assessments in respect of actual and potential impacts on human rights, democracy and the rule of law, in an iterative manner;
- Establishment of sufficient prevention and mitigation measures as a result of the implementation of these assessments;
- Possibility for the authorities to introduce ban or moratoria on certain application of AI systems ("red lines").

Who is covered by the Framework Convention?

- The Framework Convention covers the use of AI systems by **public authorities** – including **private actors** acting on their behalf – and **private actors**.
- The Convention offers Parties two modalities to comply with its principles and obligations when regulating the private sector: Parties may opt to be directly obliged by the relevant Convention provisions or, as an alternative, take other measures to comply with the treaty's provisions while fully respecting their international obligations regarding human rights, democracy and the rule of law.
- Parties to the Framework Convention are not required to apply the provisions of the treaty to activities related to the protection of their national security interests but must ensure that such activities respect international law and democratic institutions and processes.

The Framework Convention does not apply to national defence matters nor to research and development activities, except when the testing of AI systems may have the potential to interfere with human rights, democracy, or the rule of law.

How is the implementation of the Framework Convention monitored?

The Framework Convention establishes a follow-up mechanism, the Conference of the Parties, composed of official representatives of the Parties to the Convention to determine the extent to which its provisions are being implemented. Their findings and recommendations help to ensure States' compliance with the Framework Convention and guarantee its long-term effectiveness. The Conference of the Parties shall also facilitate co-operation with relevant stakeholders, including through public hearings concerning pertinent aspects of the implementation of the Framework Convention.

Chapter 3. India's AI Revolution: A Roadmap to Viksit Bharat[26]

India is witnessing a significant transformation in the field of artificial intelligence, with the Indian Government playing an active role in shaping a robust and inclusive AI ecosystem. For the first time, public policy is focused on ensuring that essential resources such as computing capacity, GPUs, and research infrastructure are made widely available at affordable costs. This marks a departure from earlier phases, where access to AI technologies was limited to a small, privileged segment or dominated by large global technology corporations.

Through forward-looking and inclusive policies, the Indian Government is enabling students, startups, and innovators to access world-class AI infrastructure, thereby creating a more level and competitive environment. Flagship initiatives such as the IndiaAI Mission and the establishment of multiple Centres of Excellence for AI are reinforcing the national AI ecosystem and promoting innovation, self-reliance, and technological sovereignty. These initiatives align with the long-term vision of *Viksit Bharat @2047*, under which India aims to emerge as a global leader in AI by harnessing advanced technologies for economic development, improved governance, and social progress.

3.1 AI Compute and Semiconductor Infrastructure

India is rapidly strengthening its AI computing and semiconductor base to support its expanding digital economy. Following the approval of the IndiaAI Mission in 2024, the government committed ₹10,300 crore over a five-year period to enhance national AI capabilities. A central component of this mission is the creation of a shared high-end computing facility equipped with 18,693 GPUs, positioning India among countries with the largest AI compute capacities worldwide. This infrastructure is nearly nine times larger than that used by the open-source AI model DeepSeek and approximately two-thirds of the computing scale used by ChatGPT.

Key developments include the phased rollout of GPU capacity, with 10,000 GPUs already operational and the remainder to be deployed shortly. This infrastructure will support the development of indigenous AI solutions suited to Indian languages and local requirements. India has also introduced an open GPU marketplace, enabling startups, researchers, and students to access high-performance computing resources. Unlike systems dominated by large corporations, this approach democratizes access and encourages innovation by smaller players.

To ensure supply resilience, the government has empanelled ten companies to provide GPUs, creating a diversified and reliable supply chain. In parallel, India has set an objective to develop domestic GPU capabilities within the next three to five years, reducing dependence on imported technologies. A subsidised common compute facility is also being launched, allowing access to GPU resources at approximately ₹100 per hour, significantly lower than prevailing international costs. Alongside AI compute, India is advancing semiconductor manufacturing,

26 Ministry of Electronics & Information Technology. (2025, March 6). India's AI revolution: A roadmap to Viksit Bharat [Press release]. Press Information Bureau, Government of India. https://www.pib.gov.in/PressReleasePage.aspx?PRID=2108810&reg=3&lang=2

with five semiconductor fabrication units currently under construction, strengthening both AI innovation and the broader electronics ecosystem.

3.2 Advancing AI Through Open Data and Centres of Excellence

Recognising that data is fundamental to AI development, the Indian Government has introduced the IndiaAI Dataset Platform to provide access to high-quality, non-personal, anonymised datasets. This platform is intended to host one of the largest repositories of such data, empowering startups and researchers to build accurate and innovative AI applications. By facilitating access to diverse datasets, the platform is expected to reduce bias and enhance AI performance across sectors such as agriculture, climate analysis, transportation, and urban management.

In addition, the government has established three AI Centres of Excellence in Healthcare, Agriculture, and Sustainable Cities in New Delhi. The Union Budget 2025 announced a fourth Centre of Excellence focused on AI in education, supported by an allocation of ₹500 crore. Further plans include the creation of five National Centres of Excellence for Skilling, aimed at preparing the workforce for AI-driven industries. These centres, developed in collaboration with global partners, support the broader vision of "Make for India, Make for the World" in both manufacturing and AI innovation.

3.3 India's AI Models and Language Technologies

The Indian Government is actively supporting the development of indigenous foundational AI models, including Large Language Models (LLMs) and domain-specific solutions tailored to national needs. Under the IndiaAI framework, calls for proposals have been issued to encourage the creation of both large and small language models.

Digital India Bhashini serves as an AI-enabled language translation platform that facilitates access to digital services in Indian languages through text and voice interfaces. BharatGen, launched in 2024, represents the world's first government-funded multimodal LLM initiative, aimed at improving public service delivery and citizen engagement through advancements in language, speech, and computer vision. Other notable developments include Sarvam-1, a 2-billion-parameter language model optimised for major Indian languages, Chitralekha, an open-source video transcreation platform, and Everest 1.0, a multilingual AI system supporting dozens of Indian languages with plans for further expansion.

3.4 AI Integration with Digital Public Infrastructure

India's Digital Public Infrastructure (DPI) model has transformed digital innovation by combining public investment with private sector-driven applications. Foundational platforms such as Aadhaar, UPI, and DigiLocker form the backbone, while private entities develop specialized services on top. AI is now being integrated into this framework, enhancing efficiency in governance and financial services.

The effectiveness of AI-enabled DPI was demonstrated during Mahakumbh 2025, where AI tools were used for real-time crowd monitoring, multilingual assistance, and coordination between public agencies. These applications set a global example of inclusive, technology-driven management of large-scale events.

3.5 AI Talent and Workforce Development

India's growing AI ecosystem is supported by a rapidly expanding talent base. The country continues to attract global R&D investment, with Global Capability Centres being established at a rapid pace. To sustain this momentum, the government is modernising higher education curricula in line with the National Education Policy 2020, incorporating AI, 5G, and semiconductor design to ensure graduates are industry-ready.

Through initiatives such as IndiaAI Future Skills, AI education is being expanded across undergraduate, postgraduate, and doctoral programs. Fellowships support AI research at leading institutions, while Data and AI Labs are being established in Tier-2 and Tier-3 cities to broaden access. International assessments indicate that India ranks among the top globally in AI skill penetration and talent growth, including strong participation by women in AI-related fields.

3.6 AI Adoption and Industry Growth

India's generative AI ecosystem has expanded rapidly, transitioning from experimental applications to scalable, production-ready solutions. A large majority of Indian businesses now view AI as a strategic priority and plan to increase investments in AI technologies. Startup funding in the GenAI sector has grown substantially, supported by a strong incubator and accelerator network. AI adoption is also transforming workplaces and empowering small and medium enterprises by improving efficiency, personalization, and revenue generation.

Market projections indicate sustained high growth in India's AI economy, with AI creating new employment opportunities even as it automates routine tasks.

3.7 A Pragmatic Approach to AI Regulation

India has adopted a balanced and pragmatic approach to AI regulation, seeking to promote innovation while ensuring accountability. Rather than relying exclusively on prescriptive legislation, the Indian Government is investing in technical safeguards and research-led solutions to address challenges such as deepfakes, privacy risks, and cybersecurity threats. By funding leading universities and institutions to develop AI governance tools, India is pursuing a techno-legal model that encourages innovation, prevents monopolistic control, and proactively addresses ethical concerns. This approach aims to ensure that AI remains a driver of inclusive growth and societal benefit.

"India's future workforce must not only be digitally aware but AI confident."

- Shri Jayant Chaudhary,
Minister of State (Independent Charge)
for Skill Development & Entrepreneurship and
Minister of State for Education, Government of India

Presidential Leadership in AI Capacity Building: The 'Skill the Nation Challenge' Announced by the President Droupadi Murmu and the SOAR Initiative Endorsed by Jayant Chaudhary

India's evolving AI ecosystem is increasingly supported by high-level institutional and leadership-driven initiatives aimed at fostering nationwide AI awareness and future-ready skills. On 1st January 2026, the President of India announced the "*#SkillTheNation Challenge*," a national call to action encouraging citizens, policymakers, educators, professionals, and youth to participate in structured AI learning through the SOAR (Skilling for AI Readiness) programme hosted on the Skill India Digital Hub. The initiative positions AI literacy as a foundational national capability and reflects a policy commitment to building an inclusive, technologically empowered society.

Launched in July 2025, SOAR constitutes the Ministry of Skill Development and Entrepreneurship's flagship programme for AI capacity building, offering accessible micro-credential courses tailored for students, educators, working professionals, and lifelong learners. Designed to promote ethical, responsible, and application-oriented understanding of AI, the programme has, within six months, enrolled over 1.5 lakh learners nationwide, indicating growing public engagement with AI education.

The initiative also reflects a leadership-led model of skills adoption. The Minister of State for Skill Development and Entrepreneurship and Education publicly completed the foundational SOAR AI module and nominated senior public officials and institutional leaders to undertake the course, signalling a cascading approach to AI readiness. This leadership participation has extended to Parliament, where multiple Members of Parliament have completed the SOAR module, underscoring institutional commitment to AI capacity building within democratic governance structures. Complementing these efforts, school-level participation through nationally administered institutions highlights the government's attempt to embed AI awareness across all educational tiers.

Chapter 4: Justice Delivery In A Digitally Transforming Legal System

4.1 Human Rights And Judicial Integrity In The Age Of Artificial Intelligence: Unesco's Global Standards For Courts[27][28]

4.1.1 Introduction: Technology at the Threshold of Justice

Artificial Intelligence (AI) is steadily entering government systems across the world, including courts and tribunals. In simple terms, AI refers to computer-based tools that assist in organising information, analysing data, translating languages, preparing summaries, and managing large volumes of documents. Within judicial systems, AI is being explored primarily to address chronic delays, manage heavy caseloads, and improve administrative efficiency.

However, the growing presence of AI in justice delivery raises a fundamental concern: should technology merely assist justice, or can it begin to shape it? Recognising this tension, UNESCO has released the first-ever global ethical framework governing the use of AI in courts and tribunals. The *Guidelines for the Use of AI Systems in Courts and Tribunals (2025)* seek to ensure that technological efficiency never undermines judicial integrity, human rights, or the rule of law.

This landmark initiative draws a clear line between administrative assistance and judicial decision-making, reaffirming that justice must remain a fundamentally human function.

4.1.2 The Global Context: Efficiency Versus Ethical Risk

Judicial systems worldwide are under unprecedented strain. Several countries report millions of pending cases, creating immense pressure to adopt technological solutions. AI tools have demonstrated remarkable potential in this context. For example, AI-assisted systems in Argentina have reportedly increased case processing efficiency by nearly 300%, while in countries such as India and Egypt, automated transcription and translation tools are helping courts overcome language barriers in real time.

Yet, these gains come with serious risks. Generative AI systems, including large language models, across jurisdictions have encountered instances where legal filings contained non-existent case laws generated by AI tools. The use of opaque or "black box" algorithms in areas such as risk assessment or bail decisions raises deeper concerns relating to transparency, accountability, and the right to a fair trial.

Against this backdrop, UNESCO's intervention marks a decisive shift from enthusiasm for efficiency to a rights-based, ethics-first approach.

27 GUIDELINES FOR THE USE OF AI SYSTEMS IN COURTS AND TRIBUNALS by UNESCO; https://unesdoc.unesco.org/ark:/48223/pf0000396582;

28 United Nations Educational, Scientific and Cultural Organization (UNESCO), Global Toolkit on AI and the Rule of Law for the Judiciary (Paris, 2023)

4.1.3 Development of the UNESCO Guidelines

The UNESCO Guidelines were developed through an extensive and inclusive consultation process involving over 36,000 judicial actors from more than 160 countries. Judges, lawyers, technologists, academics, and civil society organisations contributed to shaping the framework.

At the heart of the Guidelines lies a central premise:

AI must function as an assistive tool, not as a substitute for human judgement.

The document explicitly states that AI systems cannot replace qualified legal reasoning, judicial discretion, or tailored legal analysis. Their role is limited to support and facilitation.

4.1.4 The Fifteen Universal Principles for AI in the Judiciary

The Guidelines articulate fifteen universal principles that operate as a checklist for courts and policymakers considering the use of AI technologies.

i. **Protection of Human Rights**

AI systems must respect, protect, and promote human rights, including special safeguards for women, children, minorities, and persons with disabilities. Efficiency cannot override fundamental freedoms.

ii. **Non-Discrimination and Equality**

AI must not reproduce or reinforce social biases. Courts must ensure representative training data and protect equality of arms so that unequal access to technology does not distort fairness.

iii. **Procedural Fairness**

No individual should be judged solely by a machine. AI use must never compromise the right to a fair trial.

iv. **Privacy and Data Protection**

Judicial data is highly sensitive. Robust safeguards are mandatory to prevent misuse, surveillance, or data breaches.

v. **Liberty and Security**

No deprivation of liberty may rest on opaque AI decisions. Where AI informs bail or parole decisions, the process must be transparent and contestable.

vi. **Proportionality**

AI tools must be used only where necessary and appropriate. Intrusive technologies must not be deployed if less invasive alternatives exist.

vii. **Feasibility and Public Benefit**

Courts must assess whether AI genuinely solves a problem. The Guidelines caution against "technological solutionism."

viii. **Safety**

AI systems must undergo rigorous testing to prevent unintended harm.

ix. **Information Security**

Judicial AI systems must be resilient against cyber threats that could compromise court records or evidence.

x. **Accuracy and Reliability**

AI outputs must be technically sound. Systems should function reliably across varied inputs and contexts.

xi. **Explainability**

Judges must understand and be able to explain how AI-assisted outputs are generated. Unexplainable "black box" tools are unsuitable for substantive judicial use.

xii. **Auditability**

AI systems must be open to independent audits of their design, data, and outputs.

xiii. **Transparency and Open Justice**

Courts must inform the public when and how AI is used in judicial processes.

xiv. **Human Oversight and Decision-Making**

Judges must remain fully responsible for decisions. Judicial mandates cannot be delegated to algorithms.

xv. **Multi-Stakeholder Governance**

AI governance must involve collaboration among technologists, legal experts, civil society, and affected communities.

4.1.5 Operational Guidance: Using AI with Caution

The Guidelines go beyond abstract principles and provide practical guidance, particularly for generative AI tools such as ChatGPT or CoPilot. They warn that commercial, general-purpose AI systems are not reliable sources of legal authority.

Key operational safeguards include:

- No confidential data should ever be entered into public AI platforms.
- All AI-generated content must be independently verified, including citations and legal reasoning.
- Disclosure of AI use is essential to preserve transparency and public trust.

4.1.6 The Indian Judicial Perspective[29]

In India, technological adoption in the judiciary has been cautious and incremental. Initiatives such as the e-Courts Project, virtual hearings, e-filing, automated cause lists, and AI-based translation platforms aim to improve access to justice and administrative efficiency.

At the same time, the Supreme Court of India has consistently affirmed that justice cannot be delivered by machines alone. Judicial decision-making is a human function rooted in reasoning, fairness, and conscience, protected under Articles 14 and 21 of the Constitution of India.

AI, therefore, is viewed strictly as a supporting tool. Excessive dependence on technology risks undermining judicial independence, transparency, and public confidence. Biased datasets, unexplained outputs, or inaccurate information can directly harm litigants' rights.

4.1.7 Safeguards, Training, and Institutional Responsibility

The UNESCO Guidelines emphasise the need for institutional preparedness. Courts must formulate clear internal policies defining permissible and prohibited uses of AI. Risk assessments should precede deployment, and periodic reviews must follow.

Training is equally critical. Judges and court staff must understand both the capabilities and limitations of AI tools. Individual judicial officers bear responsibility to apply independent legal reasoning, avoid blind reliance on technology, and protect confidentiality at all times.

Generative AI requires special caution. Its use must be strictly limited, clearly identified, and never extended to deciding cases or creating evidence.

4.1.8 Conclusion: Keeping Justice Human

Artificial Intelligence presents real opportunities to enhance efficiency and access to justice. However, it also carries profound risks if left unchecked. The UNESCO Guidelines make one principle unmistakably clear: technology must serve justice, not govern it.

In India and across the world, constitutional values, human rights, and judicial accountability demand that final authority remains with human judges. AI may assist, organise, and support—but it cannot replace judgment, discretion, or fairness.

As UNESCO aptly reminds us, justice is ultimately a human endeavour. While machines can process information, only the human mind can truly process justice.

29 https://justai.in/unesco-launches-guidelines-for-the-use-of-ai-use-in-the-judiciary-5-12-25/

4.2 Artificial Intelligence and the Indian Judiciary: Policy Guidance, Ethical Safeguards, and the Road Ahead[30]

A Consolidated Analysis of the Supreme Court of India's White Paper on AI and the Judiciary

4.2.1 Introduction: Justice in the Age of Intelligent Technology

The judiciary is one of the most vital constitutional institutions, entrusted with protecting rights, enforcing laws, and sustaining public confidence in the rule of law. As societies become increasingly digital, courts across the world—including in India—are facing unprecedented pressures in the form of growing caseloads, procedural complexity, and heightened expectations for efficiency and accessibility.

Against this backdrop, the Supreme Court of India, through its Centre for Research and Planning (CRP), released a landmark **White Paper on Artificial Intelligence and the Judiciary (2025)**. This document represents India's most comprehensive and principled articulation on how Artificial Intelligence (AI) may be integrated into judicial functioning—without compromising constitutional values.

The White Paper adopts a clear position: **AI may assist the judiciary, but it can never replace judges, judicial reasoning, or human responsibility**. Decisions affecting liberty, rights, and justice must always remain in human hands, guided by fairness, dignity, and the rule of law.

4.2.2 Understanding the White Paper: Purpose and Scope

The White Paper functions as a **roadmap for judges, lawyers, policymakers, court staff, and the public**, explaining how AI can be responsibly used in courts while guarding against its dangers. It combines:

- A conceptual explanation of AI and Generative AI
- A survey of global judicial practices
- A detailed account of India's AI initiatives
- An honest assessment of risks
- A structured ethical and institutional framework

Readers gain clarity on how AI systems work, why tools like Generative AI may "hallucinate" or fabricate information, and why blind reliance on such tools can threaten justice.

30 White Paper on Artificial Intelligence and Judiciary, Centre for Research and Planning, Supreme Court of India (November 2025)

4.2.3 From Paper Files to Intelligent Courts: India's Digital Evolution

To appreciate the present role of AI, it is essential to understand the judiciary's technological journey. For decades, Indian courts relied on paper-based systems—manual filing, physical records, and handwritten processes. This led to delays, inefficiencies, and frequent loss or mismanagement of files.

Phase I (2007 onwards): Digitisation	Phase II: Data-Driven Courts	Phase III (Current Phase): Intelligent Integration
• Launch of the **e-Courts Mission Mode Project** • Introduction of computers, basic software, and staff training • Digital cause lists and electronic case status access	• Introduction of **CIS 3.0** • Creation of the **National Judicial Data Grid (NJDG)** for real-time case statistics • E-Filing, online court fees, and mobile court services	• End-to-end digitisation • Interlinking courts with police and other justice agencies • Introduction of AI, machine learning, OCR, and language tools • Focus on **supporting judicial work—not automating judging**

Recognising these systemic limitations, the judiciary adopted Information and Communication Technology (ICT) reforms:[31]

4.2.4 The Global Landscape: Learning from International Practice

The White Paper surveys AI adoption in more than a dozen jurisdictions, including Brazil, China, Spain, the UAE, Singapore, and the United States. Common trends emerge:

- **Widespread Use**: AI is deployed for case management, transcription, legal research, and translation.
- **Ethical Emphasis**: Global frameworks stress transparency, accountability, and human oversight.
- **Shared Risks**: Overreliance, opacity, bias, and accountability gaps are universal concerns.

International organisations such as UNESCO and the OECD have issued AI ethics principles that strongly influence India's approach—especially the insistence on **human control and judicial accountability**.

4.2.5 India's Homegrown AI Initiatives in the Judiciary[32]

The White Paper highlights carefully designed Indian AI tools that operate strictly within supportive boundaries:

31 https://www.pib.gov.in/PressReleasePage.aspx?PRID=2040232&reg=3&lang=2

32 Digital Transformation of Justice: Integrating AI in India's Judiciary and Law Enforcement, Press Information Bureau

- **SUPACE (Supreme Court Portal for Assistance in Court Efficiency)**

  The **Supreme Court Portal for Assistance in Court Efficiency (SUPACE)** is an AI-driven system designed to assist judges by efficiently summarising and organising case files. It processes case records to extract relevant facts and information, providing structured inputs to support judicial understanding without influencing decision-making.

  SUPACE functions through four integrated components: **file preview**, which converts PDF case files into searchable text; a **text and voice-enabled chatbot** that provides quick case overviews and suggests follow-up questions; a **logic gate** that extracts key information such as case synopsis, evidence, chronology, and applicable case law; and a **notebook**, an integrated drafting interface that allows judges to compile summaries and notes using AI-extracted data and voice dictation.

- **SUVAS (Supreme Court Vidhik Anuvaad Software)**

  SUVAS is an AI-based translation tool launched in November 2019 to translate Supreme Court documents and judgments from English into regional languages such as Hindi, Punjabi, Gujarati, Tamil, Marathi, Bengali, and Kannada. Using machine-assisted translation, SUVAS has translated over **31,184 Supreme Court judgments into 16 Indian languages**, significantly improving public access to judicial decisions. Each translated judgment carries a disclaimer absolving the Supreme Court of responsibility for translation errors. During its development, challenges arose due to the absence of a standardised legal vocabulary in regional languages, leading to inaccuracies. To address this, the Bar Council of India has undertaken efforts to develop a uniform legal terminology framework for use across Indian courts.

- **TERES & AI Transcription Tools**

  AI transcription refers to the use of artificial intelligence systems to automatically convert spoken words into written text. Instead of relying on manual note-taking or time-consuming human transcription of audio recordings, AI-powered tools listen to speech in real time or from recorded audio and accurately transform it into text. In the judicial context, AI transcription helps create precise records of court proceedings, arguments, and testimonies, improving efficiency, accuracy, and accessibility while allowing judges and court staff to focus on substantive legal work.[33]

- **LegRAA (Legal Research Analysis Assistant) and AI-assisted Research Tools**

  LEGRAA is an AI-powered legal research and document analysis tool developed under the guidance of the eCommittee of the Supreme Court of India. Designed to assist judges in navigating large volumes of legal material, LEGRAA supports efficient legal research by analysing judgments, statutes, and case documents to provide structured and relevant insights. As part of the Indian judiciary's pilot initiatives on AI-assisted tools, LEGRAA aims to enhance judicial efficiency and informed decision-making while ensuring that final legal reasoning and conclusions remain entirely within human judicial control.[34]

(can be accessed at: https://www.pib.gov.in/PressReleasePage.aspx?PRID=2106239&reg=3&lang=2)

33 https://transcribe.com/blog/ai-transcription

34 https://www.pib.gov.in/PressReleasePage.aspx?PRID=2205770&reg=3&lang=1

- **AI-enabled e-Filing Systems**

  The **Supreme Court of India** has recently introduced a **pilot AI-assisted e-filing system** aimed at streamlining procedural compliance and improving the efficiency of case filings. The initiative is designed to identify and flag procedural defects at the filing stage, reduce avoidable delays, and make the filing process more user-friendly for litigants and advocates. By enhancing accuracy and ease of access, the system reinforces the constitutional objective of ensuring timely and meaningful access to justice.[35]

- **Nyaya Shruti app**

  It has been launched in 2024 under the Inter-operable Criminal Justice System (ICJS), to facilitate virtual appearances and testimonies of accused persons, witnesses, police officials, prosecutors, scientific experts, prisoners etc. through video conferencing, saving both time and resources while expediting case resolutions.[36]

- **e-Sakshya**

  The judiciary has introduced **digital recording of evidence through the e-Sakshya platform** to improve the accuracy, reliability, and transparency of judicial records. In addition, the **e-Summons platform** has been implemented to enable faster, more secure, and efficient service of court notices and summons. Together, these digital initiatives strengthen procedural efficiency, reduce delays, and enhance the overall effectiveness of court communication and record management.[37]

High Courts and district judiciaries are also adopting AI for summarisation, virtual court management, and administrative assistance—always under human supervision.[38]

4.2.6 Risks and Challenges: Why Caution Is Essential

The White Paper devotes significant attention to the dangers of AI misuse:

a. Hallucinations and Fake Citations

AI may generate non-existent cases, laws, or quotations. Courts in India and abroad have already faced incidents involving fabricated authorities.

b. Algorithmic Bias

AI systems trained on historical data may reproduce social or institutional biases, particularly dangerous in bail, sentencing, or risk assessment contexts.

c. Privacy and Confidentiality

Judicial data is highly sensitive. Feeding such information into public or third-party AI tools risks irreversible data leaks.

35 https://recordoflaw.in/supreme-court-introduces-ai-enabled-e-filing-system-a-new-era-for-digital-justice/

36 https://www.pib.gov.in/PressReleasePage.aspx?PRID=2205770&reg=3&lang=1

37 https://www.pib.gov.in/PressReleasePage.aspx?PRID=2205770&reg=3&lang=1

38 https://www.pib.gov.in/PressReleasePage.aspx?PRID=2100323&reg=3&lang=2

d. Opacity and the "Black Box" Problem

Many AI systems cannot explain how outputs are generated—contrary to the judicial requirement of reasoned decisions.

e. Deepfakes and Digital Evidence Manipulation

AI-generated audio, video, and images threaten evidentiary integrity and the truth-finding function of courts.

4.2.7 Core Ethical Principles Governing AI Use

To counter these risks, the White Paper establishes clear ethical foundations:

- **Human Primacy ("Human in the Loop")**: Judges remain fully responsible for decisions.
- **Accuracy and Verification**: All AI outputs must be independently checked.
- **Confidentiality**: Sensitive data must never be entered into unsecured systems.
- **Fairness and Non-Discrimination**: Continuous monitoring for bias.
- **Transparency and Accountability**: AI use must be auditable and explainable.

4.2.8 Institutional Safeguards and Practical Guidelines

The White Paper recommends concrete institutional measures:

- Creation of **AI Ethics Committees** within courts
- Development of **secure, in-house AI systems**
- Mandatory **training for judges, lawyers, and court staff**
- Disclosure and verification requirements for AI-generated content
- Clear role-based responsibilities for judges, advocates, and law clerks

AI use should remain confined to **administrative and supportive tasks** such as record management, translation, transcription, scheduling, and policy analysis. **Judicial discretion, reasoning, and authoritative legal interpretation must never be delegated to machines.**

4.2.9 Conclusion: Innovation with Integrity

The Supreme Court's White Paper represents a balanced and forward-looking vision. It acknowledges that Artificial Intelligence can significantly reduce delays, manage caseloads, and improve access to justice—but only if deployed with restraint, transparency, and ethical discipline.

Guided by constitutional values and reinforced by global ethical frameworks, the Indian judiciary has drawn a clear boundary: **technology must serve justice, not control it**. AI is a tool of assistance, while human judgment, accountability, and public trust remain the heart of the justice delivery system.

4.3 CASES

The use of Artificial Intelligence (AI) tools has increasingly come under judicial consideration before Indian courts. In certain instances, courts have even consulted AI platforms to gain broader contextual insights, though without allowing such tools to influence the judicial determination of rights or liabilities.

A notable example is the decision in *Jaswinder Singh v. State of Punjab*[39], decided on 27 March 2023 by the **Punjab and Haryana High Court**. In this case, while considering a petition for bail involving allegations of cruelty, the Court sought feedback from ChatGPT to understand global bail jurisprudence in cases where assault is accompanied by cruelty.

The Court expressly clarified that the reference to ChatGPT was made solely to obtain a broader and comparative perspective on bail principles and was not intended to influence the merits of the case. To this end, the Court posed the following question to the AI platform:[40]

> *"What is the jurisprudence on bail when the assailants assaulted with cruelty?"*
>
> *In response, ChatGPT explained that bail jurisprudence in such cases depends on the specific facts of each case and the legal framework of the jurisdiction concerned. Generally, where the alleged offence involves violence and cruelty—such as murder, aggravated assault, or torture—the accused may be considered a potential threat to public safety or a flight risk. Consequently, courts may be reluctant to grant bail or may impose stringent conditions, including a higher bail amount. Factors such as the gravity of the offence, the accused's criminal antecedents, and the strength of the evidence are relevant considerations in deciding bail applications.*
>
> *At the same time, the response emphasized that the presumption of innocence remains a foundational principle of criminal justice. Accordingly, even in cases involving cruelty, bail may still be granted if the court is satisfied that the accused does not pose a risk to the community or of absconding.*

After recording this response, the High Court categorically stated that any reference to ChatGPT, and the observations derived therefrom, did not constitute an expression of opinion on the merits of the case. It was further directed that the trial court should not rely upon or advert to these observations. The reference was made only to present a broader picture of bail jurisprudence where cruelty is a relevant factor.

Ultimately, the bail petition was dismissed. However, considering the length of the petitioner's custody, the Court held that the ends of justice would be served by expediting the trial. Accordingly, the trial court was requested to take up the matter on priority and endeavor to conclude the trial by 31 July 2023. This direction was made subject to the condition that the petitioner would not seek any adjournment; any such request would automatically result in recall of the order for expeditious trial under Sections 362 and 482 of the Code of Criminal Procedure. All pending applications were disposed of accordingly.

39 https://www.barandbench.com/columns/artificial-intelligence-in-context-of-legal-profession-and-indian-judicial-system

40 https://drive.google.com/file/d/1s8sNhO7yzW0OHxZ_GPmJezYvFCzoACY-/view

Similarly, in the case of *Md Zakir Hussain v. State of Manipur*, WP(C) No. 70 of 2023 (23 May 2024) the Manipur High Court used ChatGPT for research - but did not delegate the decision-making - in a service matter when the state government failed to furnish the judge with essential information on the service rules of Village Defence Force personnel.

In *Christian Louboutin SAS v. The Shoe Boutique,* 22 August 2023, the plaintiffs relied on a ChatGPT response affirming that they were known for their iconic, red-soled shoe to support an argument that the defendant had breached trademark restrictions. The Delhi High Court found in favour of the plaintiff but rejected the ChatGPT response, observing that:

> *"The responses from ChatGPT cannot be the basis of adjudication of legal or factual issues in a court of law. The response of a Large Language Model (LLM) based chatbots such as ChatGPT depends upon a host of factors including the nature and structure of the query put by the user, the training data etc. Further, there are possibilities of incorrect responses, fictional case laws, imaginative data etc. generated by AI chatbots. Accuracy and reliability of AI generated data is still in the grey area. At the present stage of technological development, AI cannot substitute either the human intelligence or the humane element in the adjudicatory process. At best the tool could be utilised for a preliminary understanding or for preliminary research and nothingmore."*
>
> ***Prathiba M. Sing, Judge, Delhi High Court, 2023***

4.4 Artificial Intelligence in the District Judiciary: Policy Framework and Safeguards under the Kerala High Court AI Policy[41]

4.4.1 Introduction: AI and the Justice Delivery System

The increasing use of Artificial Intelligence (AI) in governance and public institutions has inevitably extended to the justice delivery system. Contemporary AI tools are capable of assisting courts in tasks such as organising case records, summarising lengthy documents, translating legal texts, managing schedules, and supporting administrative workflows. These developments promise greater efficiency and convenience in judicial functioning.

However, the use of AI within courts raises concerns that are fundamentally different from those in other sectors. Unregulated or careless deployment of AI in judicial processes can threaten judicial independence, compromise confidentiality, introduce bias, and erode public trust in the justice system. Recognising both the opportunities and the risks associated with AI, the High Court of Kerala has adopted a comprehensive policy regulating the use of Artificial Intelligence tools in the District Judiciary.

4.4.2 Rationale and Philosophy of the Kerala AI Policy

The Kerala AI Policy reflects a cautious, principled, and constitutionally grounded approach. It proceeds on the clear understanding that **AI is only an assistive technology and can never substitute judicial reasoning, discretion, or responsibility**. The policy aligns closely with

41 Kerala_HC_AI_Guidelines.pdf accessed on https://images.assettype.com/theleaflet/2025-07-22/mt4bw6n7/Kerala_HC_AI_Guidelines.pdf

global ethical frameworks, particularly the principles articulated in the UNESCO Guidelines on AI and the Rule of Law.

At its core, the policy recognises that courts do not merely process information. Judicial decision-making involves the application of law to facts through human judgment, ethical reasoning, and constitutional values. Any technological intervention in this domain must therefore be strictly controlled to ensure that efficiency does not come at the cost of justice.

4.4.3 Distinct Nature of Judicial Functions and the Need for Regulation

Unlike commercial or administrative sectors, the judicial process operates within a framework of accountability, transparency, and reasoned decision-making. The policy acknowledges that indiscriminate use of AI—particularly cloud-based and generative AI tools—can:

- Expose sensitive judicial and personal data
- Compromise privacy and confidentiality
- Introduce factual or legal inaccuracies
- Reinforce systemic or data-driven bias
- Undermine confidence in judicial outcomes

Accordingly, the policy seeks to strike a careful balance: **permitting limited technological assistance while preventing misuse, overdependence, or substitution of human judgment**.

4.4.4 Scope and Applicability of the Policy

The scope of the Kerala AI Policy is deliberately broad to ensure uniform ethical standards across the judicial system. It applies to:

- Judicial officers of the District Judiciary
- Court staff
- Law clerks and interns

The policy covers **all forms of AI tools**, whether generative or analytical, and applies irrespective of whether such tools are accessed through personal devices or official court infrastructure. By adopting such wide applicability, the policy ensures consistent accountability and ethical compliance at every level of the judicial hierarchy.

4.4.5 Guiding Principles Governing AI Use[42]

At the heart of the policy lie foundational principles that mirror global standards for ethical AI use in judicial systems. These include:

- **Transparency**
- **Fairness and non-discrimination**
- **Accountability**

42 https://ssrana.in/articles/kerala-high-courts-new-ai-guidelines-set-national-standard-for-judicial-integrity/

- **Confidentiality and data protection**
- **Human oversight and control**

The policy explicitly states that AI systems are **neither neutral nor infallible**. Judicial officers remain fully responsible for all work produced in their name, regardless of whether AI tools were used at any stage of the process.

4.4.6 Data Protection and Verification Safeguards

A central safeguard under the policy is the strict regulation of data usage. Judicial officers and staff are expressly prohibited from uploading:

- Case records
- Personal or sensitive data
- Confidential or privileged information
- Court documents

to unapproved or cloud-based AI platforms. The policy cautions that such platforms may store, reuse, or disclose inputs beyond institutional control, posing serious risks to privacy and data security.

Even where AI tools are formally approved, the policy mandates **mandatory human verification**. AI-generated summaries, translations, legal citations, and references cannot be relied upon unless independently checked by judges or authorised personnel.

4.4.7 Permissible Use and Absolute Prohibitions

The policy draws a clear and uncompromising distinction between **assistance** and **substitution**:

Permissible Uses

AI tools may be used only for:

- Administrative and organisational support
- Court management and scheduling
- Non-adjudicatory assistance
- Routine and preparatory tasks

Prohibited Uses

AI tools shall **not** be used for:

- Judicial reasoning or legal analysis
- Drafting judgments or orders
- Arriving at findings of fact or law
- Granting reliefs or exercising discretion

The responsibility for judicial decisions is **personal to the judge** and cannot, under any circumstances, be delegated to artificial intelligence.

4.4.8 Oversight, Training, and Accountability Mechanisms

To ensure effective and ethical implementation, the policy establishes robust oversight mechanisms. Records of AI usage must be maintained, detailing the nature of assistance provided and the verification undertaken.

Judicial officers and court staff are required to undergo structured training programmes on the ethical, legal, and technical aspects of AI use. These programmes are to be conducted through the Judicial Academy or under the supervision of the High Court, fostering informed and cautious engagement with technology.

The policy also mandates prompt reporting of errors or irregularities arising from AI-generated outputs. Importantly, violations of the policy may invite **disciplinary action under applicable service rules**, reinforcing that misuse of AI in judicial work is a matter of serious institutional concern.

4.4.9 Conclusion: Innovation Anchored in Judicial Values

The AI policy of the High Court of Kerala represents a thoughtful and forward-looking response to the challenges posed by emerging technologies in the judicial domain. By allowing limited and carefully regulated use of AI while firmly preserving human oversight, accountability, and judicial independence, the policy aligns closely with international ethical standards.

It underscores a fundamental principle: innovation in the judiciary must be guided not by convenience alone, but by constitutional values, public trust, and the enduring truth that justice is ultimately a human responsibility.

4.5 Artificial Intelligence Through the Lens of the Supreme Court

4.5.1 Misuse of Artificial Intelligence in Court Filings: Supreme Court Raises a Red Flag[43]

In a rare and serious incident, the **Supreme Court of India** has taken note of the misuse of artificial intelligence in legal pleadings. The issue arose when a rejoinder filed before the Court was found to contain hundreds of fake judicial precedents, allegedly generated using AI tools. Terming the matter too serious to ignore, the Court decided to continue hearing the case on merits.

The controversy came before a Bench of Hon'ble Mr. **Justice Dipankar Datta** and Hon'ble Mr. **Justice A G Masih** during arguments in a dispute between ***Omkara Assets Reconstruction Private Limited*** *and* ***Gstaad Hotels Private Limited***, promoted by **Deepak Raheja**. Senior advocate **Neeraj Kishan Kaul** pointed out that the rejoinder relied on case laws that did not exist. He explained that even where case names were real, the legal principles mentioned were completely fabricated, suggesting deliberate use of AI to mislead the Court.

He warned that courts hear many matters every day, and if AI-generated false information is relied upon, it could seriously harm the justice system. On the other side, senior advocate **C A Sundaram** expressed embarrassment and placed on record an unconditional apology from the advocate-on-record, seeking permission to withdraw the filing.

The incident has raised serious concerns within the legal community and highlights the need for careful and responsible use of AI in legal work. While technology can assist lawyers, the

43 https://economictimes.indiatimes.com/news/new-updates/ai-fraud-in-court-supreme-court-detects-hundreds-of-fabricated-judgements-in-high-profile-corporate-battle/articleshow/125866172.cms?from=mdr

Supreme Court's response makes it clear that accuracy, honesty, and human responsibility in court filings cannot be compromised.

4.5.2 Hon'ble Mr. Justice Vikram Nath on AI and the Judiciary: Technology Must Assist, Not Replace Justice[44]

Hon'ble Mr. Justice **Vikram Nath** of the **Supreme Court of India** has cautioned against excessive dependence on artificial intelligence in the justice system, stressing that while technology can support courts, it can never replace human judgement. Speaking at a joint event organised by the **Supreme Court Bar Association** and the **Orissa High Court Bar Association**, he observed that AI may help inform judicial decision-making, but the true essence of justice can only be delivered by human intelligence guided by constitutional values, empathy, and lived experience.

Justice Nath acknowledged that AI has already made valuable contributions to Indian courts, particularly through the e-courts project, which has introduced digitisation of records, e-filing, virtual hearings, translation of judgments, and AI-assisted transcription, including for Constitution Bench matters. However, he warned that technology must be used with caution, noting that a judge is not an algorithm and justice is not a mechanical output. According to him, machines cannot understand human suffering, remorse, or social context, and therefore justice cannot be reduced to mere computation.

He expressed serious concern over algorithmic bias, explaining that AI systems trained on biased data may reproduce or even amplify existing social inequalities related to caste, gender, or economic status. He also highlighted the lack of transparency in many AI systems, describing them as "black boxes" that produce answers without revealing reasoning. In law, he stressed, reasoning is essential, and a decision without reasons cannot be called a judgment. Any AI used in the legal system, he said, must therefore be transparent and explainable.

Justice Nath also raised important questions of accountability, asking who would be responsible if an AI tool produces a wrong or defamatory output—the programmer, the platform, the user, or the AI itself. He noted that India's current legal framework is not fully prepared to address such issues, echoing earlier judicial concerns about AI tools generating fake citations and incorrect facts. He further warned about privacy risks, pointing out that AI requires vast amounts of data and that unregulated use of personal data raises serious constitutional concerns, especially in light of the fundamental right to privacy.

Calling for responsible regulation, Justice Nath urged the development of clear legal frameworks governing AI in the justice system. He suggested mandatory transparency requirements, regular audits to check bias and accuracy, clear rules on liability for AI-related errors, and strong data protection safeguards. He repeatedly emphasised that AI should function only as an aid—helping judges, lawyers, and litigants by improving access to justice, reducing delays, and assisting research—while the final act of adjudication must always remain a human responsibility.

He also warned about the growing misuse of AI through deepfakes and misinformation, referring to recent cases involving manipulated videos and the protection of personality rights of public figures such as **Rashmika Mandanna**, **Anil Kapoor**, and **Jackie Shroff**. He clarified that these issues go beyond celebrities and raise serious concerns for ordinary citizens as well.

44 https://www.medianama.com/2025/09/223-bias-privacy-justice-vikram-nath-indian-courts-ai-caution/

In his concluding remarks, Justice Nath advised young lawyers to understand AI-related manipulation techniques and rely on forensic evidence and recent judicial precedents while handling such cases. He ended on a lighter note, remarking that while AI may speed up legal processes, it cannot replace the human instinct and experience essential to courtroom practice. His message was clear and reassuring: technology should be treated as a junior assistant in the justice system, never as the judge itself.

4.5.3 Hon'ble Mr. Justice B R Gavai warns against blind reliance on AI in the Judiciary[45]

Hon'ble Mr. Justice **B R Gavai** of the **Supreme Court of India** has raised serious concerns about the growing use of generative artificial intelligence in the Indian judicial system. Speaking at the **Supreme Court of Kenya** in Nairobi, he cautioned that while technology has improved access to courts, it has also created new ethical risks that cannot be ignored.

Justice Gavai pointed out that AI tools such as **ChatGPT** have, on several occasions, generated fake case citations and incorrect legal facts. He warned that relying on AI for legal research without proper verification can seriously mislead courts. This concern is linked to the problem of "AI hallucinations," where systems produce information that appears correct but is actually false.

He referred to recent incidents in India, including an order of the **Income Tax Appellate Tribunal**, Bengaluru Bench, which cited non-existent judgments of the Supreme Court and the **Madras High Court**. Similar issues have also surfaced in trial courts and High Courts, where judges unknowingly relied on AI-generated wrong precedents.

Legal researcher **Dona Mathew** explained that the pressure of heavy workloads makes AI an attractive tool for judges and lawyers. However, she stressed that human judgment and careful verification are essential, especially because mistakes in criminal cases can affect personal liberty and fundamental rights. She warned that unregulated AI use could have serious consequences for justice delivery.

Justice Gavai echoed this concern, stating that justice is not just about efficiency but also about empathy, ethics, and understanding social context—qualities that machines do not possess. He emphasised that AI should assist the legal system, not replace human decision-making.

The message is clear: while AI can help with speed and research, courts must use it cautiously, with strong safeguards and constant human oversight. Justice, at its core, must remain a human process guided by constitutional values, not automated outputs.

45 https://www.medianama.com/2025/03/223-justice-gavai-flags-ai-risks-when-chatgpt-gets-legal-facts-wrong/

4.5.4 Hon'ble Mr. Justice Surya Kant on Artificial Intelligence and the Human Core of Justice[46]

> *"We can resist technology and risk stagnation, or we can shape and guide it, embedding our legal and ethical values within its design"*

In examining Artificial Intelligence through the lens of the Supreme Court, the views of Hon'ble Mr. **Justice Surya Kant** offer a clear and balanced constitutional perspective. Justice Kant has unequivocally stated that AI can neither replace the lawyer nor the judge, as justice will always remain a profoundly human enterprise. While AI tools may assist in legal research, drafting, and identifying inconsistencies, they lack the capacity to understand human emotions, moral responsibility, and the nuanced realities that shape judicial decision-making.

Speaking at the Bar Association of Sri Lanka's annual law conference on *"Technology in the Aid of the Legal Profession – A Global Perspective,"* Justice Kant cautioned against the overuse of AI in the legal domain. He emphasised that AI systems are not infallible and may produce inaccuracies, hallucinations, or biased outcomes rooted in their training data. Consequently, human oversight is indispensable, and the lawyer or judge must always remain the final authority in assessing and validating AI-generated outputs.

At the same time, Justice Kant acknowledged the positive role of technology in improving court administration. Innovations such as e-filing, digital registries, online hearings, and case management systems have enhanced efficiency, transparency, and access to justice. However, he stressed that the legal profession must adapt to technological change without losing sight of its ethical foundations and constitutional values.

Justice Kant has also urged the judiciary to evolve in response to emerging challenges such as digital exclusion, climate vulnerability, displacement, and migration, warning that stagnation could render the justice system ineffective. Emphasising the importance of legal aid, he described justice as a "living promise" and highlighted the role of young lawyers and law students in strengthening access to justice through India's expanding legal-aid framework.

Together, these reflections reinforce the Supreme Court's position that Artificial Intelligence is a valuable aid to the justice system, but never its substitute. Technology may enhance the path to justice, but it is human judgment, empathy, and constitutional conscience that must ultimately lead the way.

46 https://lawbeat.in/top-stories/ai-cannot-replace-lawyer-or-judge-human-oversight-non-negotiable-justice-surya-kant-1535932

Chapter 5 : Police and Artificial Intelligence[47]

Law enforcement agencies across the world are entrusted with the responsibility of maintaining public safety, preventing crime, investigating offences, and ensuring rule of law. These responsibilities have become increasingly complex due to urbanization, technological advancement, organized crime, cybercrime, terrorism, and transnational offences. To address these challenges, police agencies are increasingly relying on advanced technologies, particularly Artificial Intelligence (AI).

Artificial Intelligence refers to computational systems capable of performing the tasks that typically requires use of human intelligence. These tasks include in recognizing patterns, learning from the data's, making predictions, identifying objects or individuals, and analyzing large volumes of information at high speed. In the context of policing, AI is not meant to replace police officers, but to enhance their operational capacity, efficiency, and accuracy.

In recent years, AI-based policing technologies have moved from experimental use to operational deployment. Technologies such as facial recognition, predictive policing, video analytics, automated license plate readers, and AI-based evidence management systems are now being adopted by police agencies of varying sizes. These systems assist officers in investigation, surveillance, crime detection, prevention, decision-making, and judicial processes.

Predictive policing is only one visible outcome of this transformation. Several other policing practices—ranging from crime scene investigation to post-trial supervision—are also undergoing significant adjustments due to AI integration. This document explains **how police agencies use AI**, the **scope of such use**, and the **breadth of AI applications in policing,** without assessing benefits or harms

5.1 Use of Artificial Intelligence by Police: Investigative Applications and Crime Deterrence Strategies

5.1.1 Identification

AI systems can be used to identify individuals or verify their identity.

- **Face recognition** is a computer vision technology in which it analyzes faces in an image. It can be used for the things such as face identification (the identification of an individual based on a specific comparison with a pool of known individuals) and face verification (verifying that a given face corresponds to a specific person — for example, verifying that a person's face matches the photo on their identification card).
- **Iris recognition** identifies individuals by their iris patterns. A specialized camera is used to take an image of the boundaries and textures of the iris of their eye, then maps the iris image using over 200 distinct features.

47 Bureau of Police Research & Development. (2022). *AI in the service of law enforcement – an introduction*. Ministry of Home Affairs, Government of India. https://bprd.nic.in/uploads/pdf/AI%20in%20the%20service%20of%20Law%20Enforcement-%20a%20n%20Introduction.pdf

- **Automated fingerprint identification** has been in use in departments for decades; now, AI systems can be used to enable better matching even when a fingerprint is distorted or incomplete. AI also is being used to develop new systems which can take a person's fingerprint without physical contact.
- **Palm-print identification,** like fingerprint identification, is accomplished through analysis of ridges and valleys on the skin's surface. Some claim that this technique has advantages over face recognition technology — for example, palms have more details to tell one person from another, and it is harder to scan a person's palm without their consent.
- **Ear biometrics** can be used in identifying individuals who are difficult to identify through face recognition technology — for example, due to the individual wearing a mask.
- **Gait recognition** analyzes how people walk in order to identify them. It has advantages over other biometric systems, as it enables the identification of persons from a distance. Notably, accuracy is a serious issue due to the variability of environments and human bodies.
- **Voice recognition systems** are used to determine the identity of a person based on audio of their voice. These systems use specialized models known as acoustic models to process and analyze audio files.
- **DNA analysis** has long been used by law enforcement to identify suspects. Now, AI is being used to improve this process and make it more efficient and faster. New AI-powered forms of DNA analysis are now being developed, such as forensic DNA phenotyping, which attempts to predict externally visible characteristics such as eye, hair, and skin color, as well as the geographic origins of a person's ancestors.

5.1.2 Tracking

Policing agencies use AI systems to track the locations or movements of individuals.

- **Tracking algorithms** can detect objects and/or individuals in video files and track them across cameras based on the appearance, velocity, and motion of the thing being tracked. This feature could be used, for example, to search stored video footage from a particular neighborhood and identify all the times that a given individual was recorded.
- **Vehicle-surveillance systems,** also known as automated license plate readers, detect information about passing vehicles, such as a vehicle's color, make, and license plate number. This data can be stored, along with the location and time of capture, thus enabling police to ascertain the locations of vehicles over time. Some agencies now are using drone-based vehicle-surveillance systems.
- **Robotic birds:** Birds are everywhere, and most people do not give them much attention. This fact has brought about an interest in using bird forms to carry out surveillance. Robotic birds capable of autonomy and staying in the air for more than a few minutes. Many sophisticated bird robots are evolving depending upon the functionality they are required to deliver

- **Smart glasses:** Another way the surveillance might be monitoring its citizens is through "smart" glasses equipped with facial recognition software to scan faces and match them to persons of interest in seconds. These augmented reality (AR) eyewear provide law enforcement with a quick way to patrola large number of people

5.1.3 Detection

Policing agencies use AI to detect crime, anomalies, or suspicious events.

- **Anomaly detection** seeks to identify events or data points that are anomalous — that is, that deviate from what is expected. This technology is widely used by the private sector— for example, by financial institutions to detect fraudulent transactions or by network administrators to detect cyberattacks.
- Some vendors have developed systems designed to alert policing agencies to events such as shoplifting, fights, loitering, dangerous driving, and casing a location. At least one vendor is leveraging vehicle surveillance system data to try to identify driving patterns that may be associated with drug trafficking activity or other unlawful conduct.
- **Gunshot detection systems** use a network of outdoor acoustic sensors to detect and locate gunfire and alert police. Policing agencies use gunshot detection systems to reduce response times, in the hope of locating a shooter, getting help to victims, or finding evidence such as shell casings. New systems use two-source detection — sound and flash — to confirm gunshots.
- **Weapons detection systems** are used to identify the presence of weapons. Computer vision-based systems analyze images to detect objects that appear to be weapons. Other vendors use sensors and analytics to detect concealed weapons and identify their location — an alternative to traditional metal detectors.
- **Drug detection systems** are being used to detect drugs on-site using mobile spectrometers, as opposed to sending samples to a lab.
- **Object Detection:** Like face detection Al models are even being trained to detect objects like bags and weapons. Since the model has been trained on different types of images so the shape, color, size and type of the object does not matter as long as the model has been trained to identify it.

5.1.4 Prediction

Policing agencies use AI to try to predict the location and time of future crime, as well as those who may perpetrate or be the victims of it.

- **Place-based predictive policing systems** use historical crime data to identify areas prone to crime, and at what times. Systems also can analyze geographic features that increase the risk of crime, known as risk-terrain analysis.
- **Person-based predictive policing systems** seek to identify individuals who are at risk of committing crimes or becoming a victim. This can be based on data such as one's risk factors for violence or becoming a victim, and/or their frequenting high-crime locations.

5.1.5 Recognizing Emotions

Policing agencies are experimenting with AI systems to analyze an individual's sentiments or emotions.

- **Lie detection systems** claim to track eye movements and analyze micro-expressions to determine whether an individual is engaged in deception. Some systems are designed specifically for law enforcement use.
- **Sentiment analysis** is a natural language processing technique designed to classify individuals' sentiment as positive, negative, or neutral. Affective computing, which goes beyond sentiment analysis, seeks to understand and interpret specific emotions based on facial expressions, voice intonations, text, and physiological signals. Sentiment analysis/ affective computing might be used, for example, to flag problematic police interactions captured on bodyworn cameras for supervisor review.

5.1.6 Identifying Associations

Policing agencies use AI systems to help detect associations among individuals.

- **Convoy analysis** is a feature for Vehicle Surveillance Systems, or License Plate Readers, that identifies vehicles that travel together, and thus presumably are associated with one other. They allow officers to enter a license plate number and search for related vehicles.
- **Social network analysis tools** suggest how individuals are connected in society, visualized through graphs. For example, AI tools have been used to identify alleged associates based on social media data. Machine learning algorithms are used to identify patterns, trends, and anomalies in social networks.

5.1.7 Evidence Management and Analytics

Policing agencies use AI to help agencies find potentially relevant evidence in large datasets.

- **Automated metadata tagging** can automatically tag and label digital evidence, helping investigators to find relevant evidence in the future. Some body-worn camera systems use AI to tag and label videos with relevant contextual information, helping police locate specific events within large video databases.
- **Evidence matching tools** automatically search an agency's databases to find evidence that might be related to an incident under investigation.
- **CSAM detection tools** detect and flag the existence of child sexual abuse material (CSAM) on devices, helping police to locate such materials and identify victims more quickly.
- **Transcription tools** can be used to transcribe audio automatically from video and audio files. This enables agencies to search for keywords across potentially thousands of videos.

5.2 AI in police work

Law enforcement agencies are entrusted with the responsibility of maintaining public safety while addressing a wide range of operational challenges. To meet these demands, police forces increasingly rely on technology to assist in various aspects of their work. In recent years, artificial intelligence has emerged as a significant component of modern policing across the globe. As AI-based policing tools become more deeply embedded in law enforcement operations, traditional approaches to crime prevention and crime prediction are undergoing substantial transformation.

Predictive policing represents only one outcome of this technological evolution. Several other policing functions are also being reshaped as AI is integrated into daily operations, all in the interest of enhancing public safety. Police agencies worldwide have already begun to harness the potential of AI in several important areas.

5.2.1 Facial Recognition

Facial recognition technology has become an important tool for police departments. Law enforcement agencies use facial recognition to identify absconding criminals and locate missing persons through image analysis. Footage captured by street cameras is often of poor quality, making manual examination difficult and time-consuming. Many police departments also lack sufficient personnel or technical specialists to analyze the vast quantity of image data generated during investigations.

AI-based facial recognition systems offer greater accuracy than human review and significantly reduce the time required for image analysis. These systems can identify facial features using parameters that go beyond ordinary human perception. Some advanced AI technologies are capable of detecting a single individual within a large crowd, such as a packed stadium, a capability that has already been used to apprehend criminals at large public events.

5.2.2 Cameras and Video Analytics

In most urban areas, surveillance cameras are widely installed on streets and in commercial establishments. Law enforcement agencies routinely depend on this footage to reconstruct crimes and identify offenders. AI enhances the usefulness of such footage by enabling not only facial recognition but also the identification of objects and complex activities, such as traffic accidents.

Object recognition plays a crucial role during large public events like festivals and marathons, where police officers cannot physically monitor all areas at the same time. AI systems can automatically alert law enforcement if an individual is carrying a weapon or exhibiting unusual behavior that may pose a threat. Beyond security monitoring, object recognition also allows AI systems to identify vehicles based on predefined characteristics. For example, the system can retrieve records of every blue minivan that passed through a particular intersection within a specific timeframe. This capability is especially useful in cases involving stolen vehicles or suspects attempting to evade capture.

Police agencies are also increasingly deploying drone-based cameras, which enable wider area coverage and faster search-and-rescue operations. These drones are typically equipped with AI-powered facial and object recognition features.

5.2.3 Predictive Policing

AI-driven predictive policing refers to the ability to forecast where crimes are likely to occur, identify potential offenders and victims, and anticipate the types of crimes that may be committed. Although predictive policing remains controversial and is not yet widely implemented, police departments and technology companies are actively testing such systems.

Predictive algorithms analyze historical crime data to map crime hotspots, guiding police agencies in deploying patrols and surveillance resources more effectively. AI systems can also assess patterns to estimate who may be at risk of committing crimes or reoffending after release from prison. At the same time, the use of such information raises important debates regarding its ethical and legal implications.

One of the most promising applications of predictive policing lies in identifying potential future victims. For example, ongoing research focuses on preventing elder abuse by analyzing environmental factors that contribute to such crimes and using predictive models to anticipate likely forms of abuse. While elder abuse is only one application, similar approaches could be extended to a wide range of violent crimes.

5.2.4 Robots in Policing

Although the complete replacement of police officers by robots is not imminent, law enforcement agencies are increasingly deploying robotic systems to perform tasks ranging from routine duties to highly dangerous operations. Some countries are experimenting with robotic police officers capable of transmitting data to headquarters for human review. Dubai, for instance, is testing street robots equipped with touchscreens for crime reporting and multilingual communication.

Robots are also used in high-risk situations where officer safety is a concern. They can enter hazardous environments, identify threats, and even detonate explosive devices, thereby reducing the risk to human life.

5.2.5 Non-Violent Crimes

AI is particularly effective at detecting irregular patterns, making it well suited for identifying non-violent crimes such as fraud and money laundering. Financial institutions have already adopted AI as a core component of their security infrastructure, and law enforcement agencies increasingly collaborate with these institutions to detect financial crimes.

Through image analysis, AI systems can also identify counterfeit goods and currency with high accuracy, often detecting details that may escape human observation.

5.2.6 Pre-Trial Release and Parole

AI systems are used within the criminal justice process during the pre-trial phase and in determining parole conditions. These systems assess factors such as the risk of flight and the likelihood of reoffending by analyzing large datasets that include historical crime records and personal information about the accused.

For example, the United States criminal justice system employs COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) as a risk assessment tool to assist

courts in parole-related decisions. Such systems aim to facilitate faster and more efficient judicial decision-making, with proponents arguing that AI reduces human error.

5.2.7 Future of AI in Law Enforcement

Although AI is still relatively new to policing, its impact is already evident in areas such as surveillance, crime prevention, and crime investigation. Enhanced imaging technologies and advanced object and facial recognition systems reduce the burden of labor-intensive tasks, allowing officers to focus on more complex responsibilities. AI also has the potential to solve crimes that might otherwise remain unresolved and to identify offenders who might otherwise evade detection.

Predictive policing, in particular, is expected to have far-reaching implications for crime control and victim identification. While its promise is significant, further refinement is necessary to address existing limitations and concerns.

5.2.8 AI from A Human Rights Perspective

The rapid development and deployment of AI have far outpaced the evolution of corresponding legal frameworks. As a result, AI technologies are often implemented without adequate analysis of their human rights implications. The complexity and opacity of AI systems can obscure decision-making processes, lending an appearance of authority to outcomes that are difficult to question. Limited digital literacy further exacerbates concerns regarding accountability and rights protection.

AI has placed unprecedented surveillance capabilities in the hands of states, raising fears of total surveillance regimes. Over-reliance on these technologies without proper safeguards has prompted human rights and technology organizations worldwide to demand legal reforms. These concerns transcend national boundaries, as technologies that violate human rights in one country are likely to produce similar effects elsewhere.

The implications of AI use in policing are particularly serious because police possess coercive powers, including arrest, detention, and the use of force. Consequently, unchecked AI deployment in law enforcement can undermine human rights, especially in authoritarian contexts.

A major concern is algorithmic bias. Although developers often claim that AI systems are objective because they rely on data, numerous studies have shown that AI can perpetuate and amplify existing social biases. Since AI systems learn from historical data, biased data leads to biased outcomes. This phenomenon is especially evident in predictive policing systems that rely on past crime data.

For instance, systems such as PredPol and HART analyze historical crime records to predict future offenses. Critics argue that these records are already biased due to discriminatory policing practices, particularly against minority communities. In the United States, African-Americans have historically been disproportionately targeted, arrested, and incarcerated, resulting in biased datasets that feed AI systems. This creates a harmful feedback loop that stigmatizes certain individuals and neighbourhoods, further limiting their economic and social opportunities.

as expression, association, and peaceful assembly. Individuals often rely on a reasonable degree of anonymity when exercising these rights. The fear of being identified, tracked, or penalized may deter people from expressing dissenting opinions, participating in protests, or joining lawful assemblies.

Perhaps the most profound concern surrounding FRT is the erosion of privacy. The right to privacy is a core component of human dignity and encompasses both respect for private life and protection of personal data. Although "private life" cannot be exhaustively defined, it includes various dimensions of an individual's social identity. Continuous surveillance through FRT, and the resulting loss of anonymity, often leads to broader infringements on freedoms of expression and association. The use of FRT may also enable intrusive searches and even arbitrary arrests, further encroaching upon the right to privacy.

FRT relies on biometric processing of facial images, many of which are captured in public spaces and subsequently stored in centralized databases for future identification. The long-term retention and reuse of such sensitive biometric information intrude upon both the right to privacy and the right to personal data protection. Since AI systems operate by collecting and analyzing vast datasets, FRT effectively creates large repositories of biometric data without the informed consent of individuals. In the absence of robust legal safeguards, such data is vulnerable to misuse and abuse.

In India, facial recognition systems have already been adopted by several state authorities. For instance, the Punjab Police has implemented an AI-based FRT known as PAIS. At the national level, the Government of India has initiated the Automated Facial Recognition System (AFRS), with the National Crime Records Bureau (NCRB) designated as the implementing authority. The NCRB has invited bids from private companies to develop this system, which critics claim could become the largest facial recognition database in the world.

Beyond privacy concerns, critics argue that AFRS risks transforming India into a surveillance state. Notably, no legislation enacted by Parliament currently authorizes the implementation of AFRS. Although the NCRB relies on a Cabinet Note issued in 2009 to justify the project, such a note does not carry the legal authority of a parliamentary statute. Furthermore, India presently lacks a comprehensive data protection law, rendering the deployment of such an invasive technology particularly hazardous.

The proposed Personal Data Protection Bill has itself attracted substantial criticism. It permits the government to exempt its agencies from compliance with the law and allows the executive to determine the safeguards applicable to its own data practices. These provisions arguably grant expansive new surveillance powers to national security agencies. Given that India already permits surveillance under laws such as the Indian Telegraph Act and the Information Technology Act, the introduction of an intrusive system like AFRS is likely to significantly intensify state surveillance and pose a serious threat to the constitutional right to privacy.

5.2.9 Safeguarding Human Rights In AI Deployment

Artificial intelligence is a transformative technology with immense potential to contribute to both economic advancement and social development. While it offers significant opportunities to enhance human welfare, it is essential that AI systems are designed and deployed in a manner

that respects and protects human rights. A narrow or purely efficiency-driven approach to AI would be counterproductive. Instead, a proactive and balanced strategy is required—one that maximizes the benefits of AI while simultaneously preventing its misuse and safeguarding fundamental rights.

Discussions around the development of ethical AI have already commenced at the global level. The European Commission has released guidelines aimed at fostering ethical and trustworthy AI. These guidelines rest on three foundational pillars: first, AI systems must comply with existing laws; second, they must adhere to ethical principles; and third, they must be technically and socially robust. Although these guidelines do not have binding legal force, they represent a significant and progressive step toward responsible AI governance.

To ensure effective protection of human rights in the AI ecosystem, several measures can be adopted. Every country should put in place a legal framework that mandates human rights impact assessments before AI systems are developed, procured, or deployed. Alongside this, it is crucial to ensure that users possess adequate AI literacy, enabling them to understand, engage with, and critically assess the functioning of these systems.

Human oversight must remain central to the deployment of AI. Decision-making authority should not rest entirely with machines; instead, AI systems should operate under continuous human supervision. Monitoring and intervention by human operators at all stages of an AI system's lifecycle will help ensure accountability, regulation, and respect for human rights.

Robust data protection laws are equally essential. Such legislation should be capable of foreseeing, reducing, and remedying human rights risks arising from AI use. Since AI systems rely heavily on personal data, individuals must be recognized as having ownership over their data, along with the right to give informed consent for its use. Legislatures should clearly and narrowly define the legitimate purposes for which personal data may be accessed.

Transparency is another critical requirement. The public should be adequately informed about the deployment and use of AI systems. Moreover, the outcomes produced by such systems must be explainable, allowing affected individuals to understand how decisions were made and how they were validated.

Individuals who are adversely affected by AI-driven decisions must have access to effective remedies. This necessitates the establishment of independent authorities empowered to investigate complaints, review decisions, and adjudicate disputes arising from the use of AI technologies.

Preventing discrimination caused by embedded bias within AI systems is of paramount importance. Ensuring diversity in datasets and adopting a zero-tolerance approach toward biased systems is essential. Regular due diligence processes and periodic human rights impact assessments should be institutionalized to identify and address discriminatory outcomes.

The implementation of the UN Guiding Principles on Business and Human Rights is also crucial. These principles require businesses to prevent, address, and remedy human rights violations linked to their activities. Applying these standards to AI development and deployment would impose clear responsibilities on the private sector to respect human rights and avoid harmful practices, thereby supporting the creation of ethical AI systems.

AI-based risk assessment tools like COMPAS have also been criticized for discriminatory outcomes. Investigations by ProPublica revealed that COMPAS misclassified African-Americans as high risk at twice the rate of Caucasians. Since the criminal justice system is one of the most powerful institutions affecting individual rights, biased AI systems threaten the right to equality guaranteed under international human rights law and national constitutions.

Biased AI systems also challenge the presumption of innocence by labeling individuals as "high risk" based on historical data. Furthermore, proprietary algorithms are often protected under intellectual property laws, preventing accused persons from examining or challenging AI-generated decisions. This "black-box" nature undermines transparency and fairness, infringing upon the right to a fair trial.

Concerns regarding arbitrary arrest and detention have also emerged. Human Rights Watch has reported that predictive policing in China has enabled arbitrary detentions in Xinjiang. In the United States, courts have expressed reservations about over-reliance on AI risk assessments, emphasizing that such tools should not be determinative in sentencing decisions.

5.2.10 Surveillance And Facial Recognition Technology[48]

Facial Recognition Technology (FRT) is emerging as one of the most powerful applications of artificial intelligence. Across the globe, governments are increasingly installing CCTV networks to enable the use of FRT within their territories. Law enforcement agencies have adopted this technology for a variety of purposes, including monitoring borders to track migrants, supervising passengers at airports, and observing public spaces within cities. The primary objective of FRT is to assist police authorities in identifying individuals by matching their facial features with digital images stored in databases.

However, the extensive deployment of FRT for mass surveillance—most notably by the People's Republic of China—has generated serious debate regarding its implications for human rights. Concerns have been raised particularly about the alleged targeting and profiling of certain ethnic minorities. These apprehensions are not without basis, as FRT provides law enforcement agencies with a powerful mechanism to continuously track, monitor, and profile individuals or groups. Such issues were formally highlighted in 2019 by a Special Rapporteur to the United Nations Human Rights Council.

One of the primary issues associated with FRT is its lack of accuracy. Multiple studies have demonstrated that facial recognition systems frequently misidentify individuals. A federal study conducted in the United States revealed significant racial bias in these systems, largely attributable to non-diverse and unrepresentative training datasets. This inherent bias increases the likelihood of discriminatory outcomes, thereby reinforcing existing inequalities and resulting in violations of fundamental human rights.

Another major concern relates to discriminatory profiling. Beyond mere surveillance, FRT can be used to single out, identify, and subsequently target specific communities. In authoritarian settings, this capability can become an instrument for systematic oppression of particular groups. Even in democratic societies, the pervasive use of FRT may undermine freedoms such

48 The Times of India. (2020, March 12). *NCRB authorised to use facial recognition to track criminals, MHA informs Rajya Sabha*. https://timesofindia.indiatimes.com/india/ncrb-authorised-to-use-facial-recognition-to-track-criminals-mha-informs-rajya-sabha/articleshow/74481284.cms

Finally, promoting AI literacy is indispensable. Deploying AI without sufficient understanding among users and institutions increases the risk of rights violations. Comprehensive efforts must therefore be undertaken to enhance AI literacy across all sectors that rely on such technologies.

There is an urgent need to identify potential harms and to mobilize legal and institutional resources to address existing gaps in the AI regulatory framework. In the absence of due process and adequate safeguards, AI systems risk undermining the human rights regime that has been painstakingly constructed in the aftermath of the world wars. As AI introduces new and complex challenges, governments worldwide must take immediate and proactive steps to prevent such erosion and to ensure that AI is harnessed effectively for the collective benefit of humanity.

5.3 AI and DPDPA Guidelines

India entered a significant phase of digital governance in November 2025 with the notification of the Digital Personal Data Protection Act (DPDPA) Rules, 2025, bringing nearly 800 million internet users under a formal privacy law. Although the DPDPA was enacted in August 2023, its enforceability depended on these rules, which now set 13 May 2027 as the date of applicability. The rules detail operational mechanisms under the Act, including timelines, compliance obligations, and sector-specific requirements.[49]

Key features of the DPDPA Rules include enhanced protections for children's data through verifiable parental consent, with limited exceptions for healthcare and child protection. The rules also introduce the concept of a *consent manager*, a trusted third party enabling individuals to manage their data consents, and outline eligibility criteria, regulatory oversight, and responsibilities of the Data Protection Board. Additional provisions cover data retention schedules, breach notification timelines, and mandatory audits for significant data fiduciaries. The government has indicated that it may consider shortening the compliance timeline from 18 months to 12 months after further industry consultation.

Alongside data protection reforms, the Ministry of Electronics and Information Technology released the **India Artificial Intelligence Governance Guidelines** under the IndAI Mission. While not legally binding, these guidelines provide a foundational framework for safe, inclusive, and responsible AI adoption. The framework is built on seven guiding principles, or *sutras*, including trust, human-centricity, accountability, fairness, and safety, supported by six strategic pillars, a phased action plan, and practical guidance for stakeholders.

Complementary developments include a proposed amendment to the IT Rules, 2021 mandating abelling of AI-generated content, a comprehensive review of cyber laws relating to women by the National Commission for Women, and significant competition law enforcement against Meta and WhatsApp for data-sharing practices. Collectively, these initiatives signal India's evolving approach to data protection, AI governance, digital trust, and accountability in the digital ecosystem.

49 Nadkarni, S. (2025, November). *Notes from the Asia-Pacific region: India releases DPDPA rules, AI governance guidelines*. International Association of Privacy Professionals (IAPP). https://iapp.org

5.4 Case Study

- **Humanoid teacher makes learning fun in Kerala school**: Nova is an AI-powered humanoid teacher of Government LPGS, Punalur. The humanoid is mainly designed to improve the communication skills of children as they are encouraged to ask questions in all languages.[50]
- During the **Maha Kumbh in Prayagraj**, artificial intelligence transformed crowd management by ensuring safety, efficiency, and highly accurate headcounts. Around 550 AI-enabled cameras monitored the mela grounds, ghats, and entry–exit points, providing real-time data on devotee movement. Advanced AI-based crowd density algorithms and mathematical models achieved nearly 95% accuracy in estimating attendance, replacing earlier approximation methods.

  The system was operated from the Integrated Control & Command Centre, where police officials and technical experts analysed crowd patterns and issued alerts whenever thresholds were crossed. AI used object detection and machine learning to identify individuals in video frames, distinguish between moving and stationary groups, and improve accuracy over time. Along with facial recognition and automatic number plate recognition systems, the technology also assisted in vehicle monitoring, cleanliness surveillance, and fire alerts, making the Maha Kumbh 2025 safer and more efficiently managed than in the past.[51]
- The **Oxford Institute of Technology and Justice**[52] is using artificial intelligence to expand access to justice by making free legal support more accessible to vulnerable communities. Working with the Clooney Foundation for Justice (CFJ) and receiving technical assistance from Microsoft's AI for Good Lab, the Institute is developing scalable, technology-driven tools to help survivors of injustice claim their rights and seek redress.

  For **journalists**, CFJ has already supported the release of many unjustly detained reporters, but access to legal help has remained fragmented. To address this, the Institute is creating AI-based tools that provide journalists with clear legal information and connect those at risk to qualified lawyers offering pro bono representation. A key initiative is the **Journalists' Legal Assistant**, developed in partnership with the Committee to Protect Journalists (CPJ). This AI-driven chatbot serves as a centralised platform where journalists can seek legal guidance and be matched with vetted lawyers, with CPJ experts reviewing requests to ensure safety and efficiency.

  For **women and girls**, CFJ's *Waging Justice for Women* programme has delivered free legal aid to victims of child marriage, gender-based violence, and discrimination. To scale this impact, the Institute is developing AI tools that simplify and accelerate legal processes. In Malawi, where access to lawyers is extremely limited, a pilot **Pro Bono Assistant for Women** helps survivors narrate their experiences in their own language, automatically

50 The Hindu. (2024, November 23). *Humanoid teacher makes learning fun in Kerala school*. https://www.thehindu.com

51 Dixit, K., & Chakraborty, P. (2025, January 16). *How AI is helping in crowd mgmt, headcount accuracy*. *The Times of India*. https://timesofindia.indiatimes.com/city/lucknow/how-ai-is-helping-in-crowd-mgmt-headcount-accuracy/articleshow/117276798.cms

52 Oxford Institute of Technology and Justice. (n.d.). *Harnessing AI to expand access to justice*. Tech & Justice, Blavatnik School of Government, University of Oxford. https://www.techandjustice.bsg.ox.ac.uk/

generating legally compliant draft affidavits for protection orders, significantly reducing lawyers' workload and response time.

Additionally, the **Women's Legal Assistant** project uses an AI-powered WhatsApp chatbot to connect first responders and survivors in Malawi with free legal advice and representation from the Women Lawyers Association of Malawi. This initiative addresses severe shortages in legal resources and enables real-time access to legal support for women and girls facing violence, child marriage, and abuse.

Chapter 6: Prosecutors And Artificial Intelligence

6.1 Introduction: The Emerging Need For AI In Indian Prosecution

The prosecution system in India occupies a pivotal position in the criminal justice process, acting as the institutional bridge between investigation and adjudication. Public Prosecutors and Assistant Public Prosecutors are constitutionally entrusted with the duty not merely to secure convictions, but to assist courts in arriving at the truth in accordance with law. However, the contemporary realities of criminal litigation reveal that prosecutors are functioning under extreme systemic pressure. Ever-increasing pendency, voluminous police records, multiple investigating agencies, frequent transfers, shortage of trained support staff, and uneven digitisation have cumulatively weakened prosecutorial effectiveness. Despite significant technological investments in policing and judiciary, the prosecution has largely remained a manual, reactive, and document-heavy institution.

Artificial Intelligence, when viewed not as a substitute for human judgment but as an assistive and augmentative tool, offers a transformative opportunity for Indian prosecution services. Global experiences demonstrate that AI can reduce administrative burden, enhance consistency, improve evidence handling, and allow prosecutors to focus on substantive legal reasoning. In the Indian context, where fairness, transparency, and constitutional safeguards are paramount, AI must be adopted through a carefully designed, ethically governed, and legally compliant framework that strengthens the rule of law rather than undermining it.

6.2 Leveraging Artificial Intelligence in Prosecution: Accelerating Case Preparation and Timely Resolution

The adoption of artificial intelligence (AI) in prosecutorial work significantly enhances the speed and efficiency of case proceedings. AI-enabled systems are designed to undertake labour-intensive legal tasks that traditionally consume considerable time and human effort. By automating processes such as evidence review, data analysis, and document management, AI allows prosecutors to devote greater attention to strategic decision-making, legal reasoning, and effective case presentation, thereby contributing to faster case disposal and improved justice delivery.

The following illustrates how AI can transform prosecutorial functions and streamline criminal proceedings:

1. Automated Analysis of Evidence

Evidence examination is among the most demanding aspects of case preparation, particularly when it involves extensive video recordings, audio files, transcripts, or voluminous documents. Traditionally, prosecutors must manually review each item, a process that is both time-consuming and prone to oversight. AI-based tools can rapidly process vast quantities of data, identifying relevant patterns, detecting faces in videos, classifying activities, and highlighting inconsistencies that may otherwise remain unnoticed.

For instance, AI systems can scan hours of video footage to pinpoint critical segments or analyze lengthy transcripts to extract significant statements within minutes. This not only results in substantial time savings but also enhances accuracy and consistency in evidence assessment.

2. AI-Assisted Redaction of Sensitive Information

Before evidence is disclosed or presented in court, sensitive details often require redaction to protect privacy and comply with legal requirements. Manual redaction of personal identifiers from documents, videos, or audio recordings is a slow and meticulous task. AI-driven redaction tools can automatically detect and obscure sensitive elements such as faces, license plates, personal identifiers, medical information, or protected voices with a high degree of precision.

By automating this process, prosecution teams can ensure compliance with applicable data protection standards while significantly reducing the time required to prepare evidence for disclosure or trial.

3. Natural Language Processing for Transcription and Review

The transcription and examination of interviews, witness statements, and courtroom recordings traditionally demands extensive manual effort. AI systems equipped with Natural Language Processing (NLP) capabilities can automatically convert audio recordings into accurate, searchable transcripts in a short span of time. These systems can also identify and extract relevant portions of testimony, enabling prosecutors to focus on substantive legal analysis rather than repetitive listening.

Additionally, AI-powered language translation tools facilitate the interpretation of multilingual audio evidence, ensuring that critical information is not missed due to linguistic barriers—an especially valuable feature in cases involving diverse populations or cross-border elements.

4 Intelligent Data Management and Case Prioritisation

Criminal prosecutions often involve managing large volumes of unstructured information, including police reports, medical records, forensic findings, and witness accounts. AI systems can efficiently organize this data, identify connections, and assist prosecutors in quickly locating crucial evidence. By analysing case attributes and urgency, AI can also support informed prioritisation of cases, ensuring that serious or time-sensitive matters receive prompt attention.

Such intelligent workflow management reduces administrative bottlenecks and helps prevent delays or oversight in case handling.

5. Improved Inter-Agency Collaboration

AI-enabled case management platforms facilitate seamless coordination among law enforcement agencies, prosecutors, and court administration. Evidence and case materials can be securely uploaded, processed, indexed, and shared in real time, ensuring that all stakeholders have timely access to accurate and up-to-date information.

For example, evidence collected at a crime scene can be digitally uploaded by investigators, automatically analysed by AI tools, and made available to prosecutors almost immediately. This eliminates the need for manual file review and enables quick retrieval of relevant materials

such as video clips or witness statements. Court officials can also access organized case records, thereby enhancing procedural preparedness and overall efficiency of court proceedings.

6.3 Structural Challenges in The Indian Prosecution System

Indian prosecutors operate within a uniquely complex institutional ecosystem. They receive cases from diverse investigating agencies including State Police, CBI, ED, NIA, and specialised units, each following different formats, technological maturity levels, and investigative cultures. Charge-sheets often run into thousands of pages, particularly in economic offences, NDPS cases, cybercrime, organised crime, and terror-related prosecutions. Manual scrutiny of such records not only consumes time but increases the likelihood of human error, omissions, and inconsistencies that ultimately benefit the accused and erode public confidence.

Further, prosecutors are expected to manage bail matters, remand proceedings, witness coordination, trial scheduling, victim interaction, appeals, and compliance with evolving jurisprudence, often without institutional research support. Delays in sanction orders, lapses in statutory timelines under the BNSS, non-production of witnesses, and missing forensic reports are frequent causes of acquittals. In this backdrop, AI emerges as a structural necessity rather than a technological luxury.

6.4 AI-Enabled Office Case Management for Prosecutors

AI-driven office case management systems can fundamentally alter how prosecution offices function. By creating an intelligent digital case lifecycle, AI can track criminal cases from registration of FIR through investigation, cognisance, trial, and appeal. Such systems can continuously monitor procedural milestones, flag statutory deadlines, identify missing documents, and alert prosecutors about impending risks such as default bail, limitation expiry, or non-compliance with court directions.

In the Indian framework, AI-powered case management can be integrated with existing digital infrastructures such as e-Courts, ICJS, CCTNS, and state prosecution portals. This integration would allow prosecutors to access real-time case status, custody details, hearing schedules, and judicial orders from a unified interface. Over time, AI systems can learn from patterns of delays and acquittals, enabling supervisory prosecutors to identify systemic weaknesses and implement corrective measures at the institutional level.

6.5 AI-Assisted Scrutiny of Police Reports and Charge-Sheets

One of the most critical responsibilities of prosecutors is the scrutiny of police reports before cognisance and trial. AI can assist prosecutors by automatically analysing charge-sheets to identify missing witnesses, incomplete forensic evidence, contradictory statements, and non-compliance with mandatory legal requirements. Such assistance is particularly valuable in sensitive prosecutions under POCSO, NDPS, SC/ST (Prevention of Atrocities) Act, and economic offences where procedural lapses often have fatal consequences for the prosecution.

By training AI systems on statutory requirements, prosecution manuals, and judicial precedents, prosecutors can be alerted to defects at the pre-trial stage itself, enabling timely rectification through supplementary investigation. This proactive intervention can significantly improve

conviction rates while simultaneously upholding the rights of the accused to a fair and legally sound prosecution.

6.6 AI-Based Data Collection and Evidence Management

Modern criminal cases increasingly rely on digital evidence such as CCTV footage, call detail records, mobile data, financial transactions, and electronic communications. Prosecutors are often required to manually sift through vast quantities of unstructured data, a task that is not only inefficient but prone to oversight. AI-powered evidence ingestion systems can automatically collect, organise, and index digital evidence from multiple sources, converting it into searchable and intelligible formats.

AI-assisted transcription and analysis of audio and video evidence can help prosecutors quickly identify relevant portions of confessions, witness statements, or surveillance footage. Pattern recognition capabilities can assist in linking accused persons across cases, detecting organised crime networks, and identifying modus operandi in repeat offences. Importantly, such systems can also assist in identifying exculpatory evidence, thereby strengthening the ethical foundation of prosecution and reducing wrongful convictions.

6.7 Operational Efficiency and Prosecutorial Decision Support

AI can significantly enhance operational efficiency within prosecution offices by assisting in legal research, drafting, and decision support. AI legal assistants trained on Indian statutes, Supreme Court and High Court judgments, and state-specific circulars can generate case summaries, draft written submissions, prepare appeal opinions, and suggest relevant precedents within minutes. While final responsibility must always rest with the prosecutor, such assistance can drastically reduce preparation time and improve the quality of advocacy.

In bail and sentencing matters, AI systems can provide data-driven insights by analysing historical case outcomes, judicial trends, and comparable fact situations. This can help prosecutors make more consistent and proportionate submissions while avoiding unconscious bias. AI-assisted scheduling and witness management tools can further reduce adjournments by optimising court calendars, tracking witness availability, and ensuring timely service of summons.

6.8 Constitutional and Legal Safeguards for AI use in Prosecution

The use of AI in prosecution must be firmly anchored in constitutional principles, particularly Articles 14, 20, and 21 of the Constitution of India. AI systems must operate transparently, avoid arbitrariness, and ensure that accused persons retain the right to challenge prosecutorial material generated or assisted by AI. Under no circumstances should AI be allowed to autonomously determine guilt, recommend punishment, or override human discretion.

AI outputs must be explainable and auditable, enabling courts and defence counsel to understand how conclusions were reached. Data protection, privacy, and confidentiality of investigation records must be rigorously safeguarded, especially in light of emerging data protection jurisprudence. Human oversight must remain central, with prosecutors retaining full accountability for every decision taken.

6.9 Barriers to AI Adoption in Indian Prosecution

Despite its promise, AI adoption in Indian prosecution faces significant challenges. Data quality remains uneven, with many records still existing in paper form or poorly digitised formats. Resource constraints, particularly at the district level, limit the capacity to invest in advanced technology. There is also apprehension among legal professionals regarding over-reliance on technology, loss of discretion, and potential misuse.

Institutional resistance, lack of AI literacy, and absence of standard operating procedures further complicate adoption. Without a clear governance framework, there is a risk that AI could exacerbate inequalities rather than reduce them. These concerns underline the necessity of a phased, regulated, and capacity-driven approach.

6.10 Policy and Governance Framework for India

India requires a dedicated AI governance framework for prosecution services that balances innovation with accountability. Pilot projects should be initiated in selected districts and special courts to test AI applications in case management and evidence handling. State-level AI oversight committees comprising prosecutors, judges, technologists, and ethicists should evaluate tools before deployment and conduct periodic audits.

Capacity building must be prioritised through structured training programmes at judicial academies and prosecution institutes, fostering interdisciplinary collaboration with technical institutions. Indigenous AI development tailored to Indian laws, languages, and judicial practices should be encouraged to ensure transparency and sovereignty over prosecutorial data.

Chapter 7. Artificial Intelligence and Banking Sector in India

Artificial intelligence has become a transformative technology in the banking sector, powering both **internal operations and customer-facing services**. Banks worldwide are integrating AI into front, middle, and back-office functions to improve efficiency, risk management, regulatory compliance, and customer experience. This integration is driven by the need to remain competitive in a rapidly digitising financial services industry. AI applications in banking include customer service automation, fraud detection, wealth management, credit scoring, and compliance monitoring, among others[53].

7.1. Key Applications of AI in Banking

7.1.1 Customer Service and Engagement:

AI-powered solutions such as virtual assistants and chatbots enable banks to provide 24/7 customer support, respond to routine inquiries, and personalise interactions based on customer data. These tools reduce the workload of human agents and help standardise service quality across diverse customer segments.

7.1.2 Fraud Detection and Risk Management:

AI algorithms can analyse transaction patterns in real time to identify anomalies indicative of fraud. Machine learning models continuously adapt to emerging threats, enhancing security frameworks and reducing financial losses.

7.1.3 Operational Efficiency and Decision-Making:

Banks utilize AI to automate repetitive tasks such as data entry, document verification, and loan processing, resulting in faster turnaround times and lower operational costs. Predictive analytics powered by AI enables institutions to make more informed decisions regarding credit risk, investment strategies, and market trends.

7.1.4 Regulatory Compliance and Wealth Management:

AI assists in monitoring compliance with regulatory requirements by flagging suspicious activities and helping interpret complex rules through automated systems. In wealth management, AI supports personalised investment recommendations by analysing client data and financial trends.

7.2 Legal and Regulatory Considerations

The increasing reliance on AI in banking raises **critical legal and regulatory issues**. Data privacy and protection are paramount given the vast amount of sensitive customer information processed by AI systems. Legal frameworks must evolve to address algorithmic transparency, accountability for automated decisions, and the ethical deployment of AI technologies in

53 Finio, M., O'Brien, K., & Downie, A. (2025). AI in banking. IBM Think. https://www.ibm.com/think/topics/ai-in-banking

financial services. Banks are also required to maintain robust governance structures and risk frameworks to ensure that AI tools comply with existing financial regulation and consumer protection laws.

7.3. Indian Banking Context

In India, AI adoption within the banking sector is growing rapidly. Leading Indian banks such as State Bank of India, HDFC Bank, ICICI Bank, and Axis Bank have implemented AI-based solutions for customer engagement, fraud detection, predictive credit scoring, and regulatory compliance.[54]

7.3.1 AI-Driven Customer Support:

Indian banks use AI chatbots and virtual assistants to resolve customer queries in multiple languages, improving accessibility and satisfaction while reducing pressure on call centre operations.[55]

7.3.2 Enhanced Fraud Prevention:

AI in Indian banks helps monitor transactions in real time to detect irregular behaviour and prevent fraudulent activities, contributing to stronger cybersecurity and financial integrity.

7.3.3 Adoption Trends and Regulatory Response:

A *Reserve Bank of India* study indicates that mentions of AI-related terms in bank annual reports have increased significantly in recent years, showing greater institutional focus on AI tools.[56] IndiaAI Regulatory bodies such as the RBI and India's financial sector regulators are increasingly engaging with guidelines on AI governance, cybersecurity, and data protection to balance innovation with systemic risk management.

7.4. Government of India Initiatives on AI-Enabled Cybersecurity in the Financial Sector

The Government of India has adopted a series of institutional and technological measures to strengthen cybersecurity in the financial sector, acknowledging the growing vulnerability of digital financial ecosystems to cyber fraud and organised financial crime. A significant development in this regard is the introduction by the Reserve Bank of India of an artificial intelligence-based tool known as *MuleHunter*,[57] designed to identify and disrupt "money mule" networks used to channel proceeds of digital fraud. This initiative forms part of a broader national framework for

54 Dr. Sonia. (2025). The role of artificial intelligence in the banking sector in India: Opportunities, applications, and challenges. International Journal of Innovations & Research Analysis (IJIRA). https://inspirajournals.com/uploads/Issues/995799919.pdf

55 FutureSkills Prime. (2025, February). Artificial intelligence in Indian banking: Transforming the future. FutureSkills Prime, Ministry of Electronics & Information Technology, Government of India and nasscom. https://www.futureskillsprime.in/blogs/artificial-intelligence-in-indian-banking-transforming-the-future/

56 Stanly, M. (2024, October 25). *RBI study explores how Indian banks leverage AI*. IndiaAI, Ministry of Electronics and Information Technology, Government of India. https://indiaai.gov.in/article/rbi-study-explores-how-indian-banks-leverage-ai

57 Ministry of Finance. (2025, March 18). *Various measures have been taken by the government to strengthen cyber security in the financial sector: Artificial Intelligence (AI) based tool 'MuleHunter' for identification of money mule has been launched by RBI* [Press release]. Press Information Bureau, Government of India. https://www.pib.gov.in/PressReleasePage.aspx?PRID=2112323&reg=3&lang=2

financial cyber-fraud management, including the Citizen Financial Cyber Fraud Reporting and Management System, which enables real-time reporting of fraud incidents and facilitates rapid intervention and recovery mechanisms.

These measures are further supported by the establishment of the Indian Cyber Crime Coordination Centre under the Ministry of Home Affairs and the operationalisation of the National Cyber Crime Reporting Portal, which integrates complaint registration with law-enforcement response. Parallel technological interventions by the National Payments Corporation of India, such as enhanced device-binding protocols, multi-factor authentication systems, and AI/ML-driven fraud monitoring tools, have reinforced the resilience of digital payment infrastructures. Collectively, these initiatives reflect a coordinated policy approach towards leveraging artificial intelligence and advanced analytics to safeguard financial systems, protect consumers, and strengthen the integrity of India's digital financial architecture.

7.5. Reserve Bank of India, FREE-AI Report

On 13 August 2025, the Reserve Bank of India (RBI) released the report titled *Framework for Responsible and Ethical Enablement of Artificial Intelligence* (FREE-AI Report), drawing upon the recommendations of the FREE-AI Committee. The Committee, constituted in December 2024, was tasked with identifying risks associated with the adoption of Artificial Intelligence (AI) in the financial sector and proposing an appropriate regulatory framework to guide its responsible and ethical deployment.[58]

7.5.1. Key Findings from the RBI Survey on AI

As part of this initiative, the RBI conducted surveys among regulated entities including banks and non-banking financial companies (NBFCs) as well as fintech firms to assess the current level of AI adoption and the challenges encountered. The survey revealed that 20.8 per cent of the responding entities are presently deploying AI systems, primarily in areas such as customer support, sales, credit underwriting, and cybersecurity. Notably, while current deployment remains limited, a substantial majority, approximately 67% of the surveyed entities expressed a strong interest in exploring and expanding the use of AI across a wider range of applications in the financial sector.[59]

7.5.2. Benefits and Opportunities of AI in the Financial Sector

7.5.2.1 AI as an Enabler of Financial Inclusion

In developing economies such as India, AI holds particular promise for advancing financial inclusion. A large segment of the population remains outside the formal financial system due to the absence of collateral, credit histories, or formal documentation. AI-enabled alternative credit assessment models address this gap by analysing non-traditional data sources, such as utility payments, mobile usage patterns, GST filings, and e-commerce behaviour, thereby enabling the inclusion of "thin-file" and "new-to-credit" borrowers.[60]

58 *FREE-AI Committee Report: RBI directions on implementation of AI in financial services*, Khaitan & Co., 28th August 2025

59 *Id*

60 *FREE-AI Committee Report, Framework for Responsible and Ethical Enablement of Artificial Intelligence*, Reserve Bank of India, August 2025

In addition, AI-powered conversational interfaces can provide context-aware financial guidance, grievance redressal, and behavioural nudges to low-income and rural populations. Voice-enabled banking solutions in regional languages further expand access for illiterate or semi-literate users, reducing informational and linguistic barriers to financial participation.

7.5.2.2 Integration of AI with Digital Public Infrastructure

India's digital public infrastructure (DPI), including Aadhaar, Unified Payments Interface (UPI), and Account Aggregator frameworks, provides a robust foundation for responsible AI integration. Policy discussions increasingly emphasise the convergence of AI with DPI to enable intelligent, adaptive, and inclusive service delivery. Applications such as AI-assisted KYC, conversational interfaces embedded in payment systems, and personalised financial services through data-sharing frameworks illustrate the potential for next-generation DPI[61].

The development of AI models as public digital goods can also lower entry barriers for smaller and regional financial institutions, mitigating concentration risks and fostering innovation across the ecosystem.

7.5.2.3 Sector-Specific and Indigenous AI Models

A key strategic question in the Indian context concerns the development of financial-sector-specific AI models. Large foundation models, trained on vast datasets and fine-tuned for general use, offer flexibility but may not adequately reflect India's linguistic, cultural, and operational diversity. Reliance on foreign-developed models for core financial functions also raises concerns regarding systemic dependence and data sovereignty.

Alternative approaches include the development of small language models tailored to specific tasks, as well as "trinity models" designed around specific language–task–domain (LTD) combinations[62]. Such models can support multilingual inclusion, regulatory alignment, and resource-efficient deployment, making them particularly suitable for India's diverse financial ecosystem.

7.5.2.4 Autonomous AI Systems and Emerging Synergies

The emergence of autonomous AI agents marks a shift from task automation to decision automation. These systems can decompose complex objectives, coordinate with other agents, and generate adaptive solutions, potentially reshaping financial intermediation. For example, AI agents acting on behalf of borrowers could negotiate with multiple lenders, compare loan terms, and execute transactions in real time.[63]

7.5.2.5 Synergies with other Emerging Technologies:

AI is also beginning to intersect with other emerging technologies, including quantum computing and privacy-enhancing technologies. While such synergies remain at an early stage, they point towards the development of next-generation AI systems capable of handling complex computations while preserving data privacy through techniques such as federated learning.[64]

61 *Id.*

62 *Id.*

63 *Id.*

64 *Id.*

7.5.3. Emerging Risks and Sectoral Challenges[65]

7.5.3.1 Systemic and Governance Risks

The integration of AI into financial systems introduces a wide range of risks that challenge traditional risk management frameworks. These include concerns related to data privacy, algorithmic bias, market manipulation, concentration risk, operational resilience, cybersecurity vulnerabilities, explainability, and consumer protection.[10] Given the interconnected nature of financial markets, AI-driven failures can propagate rapidly, potentially undermining market integrity and financial stability.

7.5.3.2 Model Risk and Explainability Challenges

AI model risk arises when algorithmic outputs deviate from expected outcomes, leading to financial loss or reputational damage. Biases embedded in training data or model design, combined with the opacity of "black box" systems, make it difficult to audit decisions or assign responsibility. Risks related to data quality, model calibration, implementation, and deployment can interact to produce cascading failures across institutions.

Generative AI models introduce additional challenges, including hallucinations and reduced explainability, complicating compliance, auditability, and accountability.

7.5.3.3 Operational and Third-Party Risks

While automation can reduce human error, it can also amplify failures across high-volume transactions. Data pipeline failures, model drift, and inadequate monitoring can degrade system performance over time. Dependence on third-party vendors, cloud providers, and subcontractors introduces additional risks related to service continuity, regulatory compliance, and concentration, often compounded by limited visibility into vendor controls.

7.5.3.4 Market Conduct, Competition, and Financial Stability

AI systems may unintentionally reinforce pro-cyclicality, herding behaviour, and market volatility, particularly when similar models are widely deployed across institutions. The potential for algorithmic collusion, both unintended and deliberate, raises concerns for competition law and market conduct regulation. Historical episodes such as the "Flash Crash" illustrate how automated systems can exacerbate market stress when inadequately tested for extreme scenarios.

7.5.3.5 Cybersecurity, Data Protection, and Consumer Harm

AI is a double-edged sword in cybersecurity, enabling both advanced attacks and enhanced defence mechanisms. Risks include data poisoning, adversarial inputs, prompt injection, model inversion, and deepfake-enabled fraud. At the same time, AI-driven security systems can enhance threat detection and response capabilities.

From a consumer protection perspective, algorithmic opacity, behavioural manipulation, and data over-collection raise ethical concerns relating to consent, autonomy, and fairness. AI-driven decision-making may exacerbate existing power asymmetries between financial institutions and consumers, particularly vulnerable groups.

65 *Id.*

7.5.3.6 The Risk of AI Inertia

Finally, the failure to adopt AI presents its own set of risks. Institutions that lag in AI adoption may become less competitive, less resilient to cyber threats, and less capable of advancing financial inclusion objectives. At a systemic level, delayed adoption could widen access gaps and undermine India's broader digital and financial development goals.

7.5.4. Proposed Amendment to Existing Laws[66]

The Report acknowledges that the current legal framework, including the Information Technology Act, 2000 and rules thereunder, are sufficient to address current risks. The Report further analyses existing RBI guidelines and suggests the following amendments in respect of AI related aspects:

RBI Regulation	**Amendments Proposed**
RBI Guidelines on Managing Risks and Code of Conduct in Outsourcing of Financial Services by Banks, 2006	Incorporation of AI-specific risks and AI-usage disclosure requirements.
RBI Cyber Security Framework in Banks, 2016	Inclusion of AI-specific threats (e.g., model poisoning, adversarial attacks) and incident protocols.
RBI (Digital Lending) Directions, 2025	Disclosure of AI-driven credit assessments; fairness audits to mitigate algorithmic biases.
RBI Master Circular on Customer Service in Banks, 2015	AI-usage disclosure requirements and establishment of processes for customers to contest AI driven decisions.
RBI (Fraud Risk Management in Commercial Banks (including Regional Rural Banks) and All India Financial Institutions) Directions, 2024	Implementation of AI-driven fraud detection, along with testing the accuracy and bias in these processes.
RBI (Information Technology Governance, Risk, Controls and Assurance Practices) Directions, 2023	Introduction of AI-specific access control measures for autonomous AI.
RBI (Outsourcing of Information Technology Services) Directions, 2023	Requirement of AI-usage disclosure by service providers and AI-specific risk assessments.

66 *Id.*

7.5.5 The Seven Sutras: Guiding Principles

The Committee emphasises that the future trajectory of AI adoption in the financial sector must be grounded in a principles-based regulatory framework. Accordingly, it has articulated seven *Sutras*, or foundational principles, intended to guide the design, deployment, and governance of AI systems within the financial ecosystem. These principles provide normative direction for ensuring that AI innovation remains responsible, ethical, and aligned with public interest objectives.

7.5.5.1 Sutra 1: Trust is the Foundation

Trust is non-negotiable in a sector entrusted with public funds. AI systems must reinforce institutional credibility, with trust embedded by design rather than treated merely as a compliance requirement.

7.5.5.2 Sutra 2: People First

AI should augment, not replace, human judgment. Ultimate decision-making authority must remain with humans, and individuals must be informed when interacting with AI systems. Human safety, dignity, and agency are central to sustainable AI deployment.

7.5.5.3 Sutra 3: Innovation over Restraint

The framework promotes responsible, purpose-driven innovation. AI should act as a catalyst for value creation and inclusion, with safeguards that enable progress rather than unduly constraining it.

7.5.5.4 Sutra 4: Fairness and Equity

AI systems must ensure non-discriminatory outcomes and should be leveraged to promote financial inclusion and equitable access to financial services.

7.5.5.5 Sutra 5: Accountability

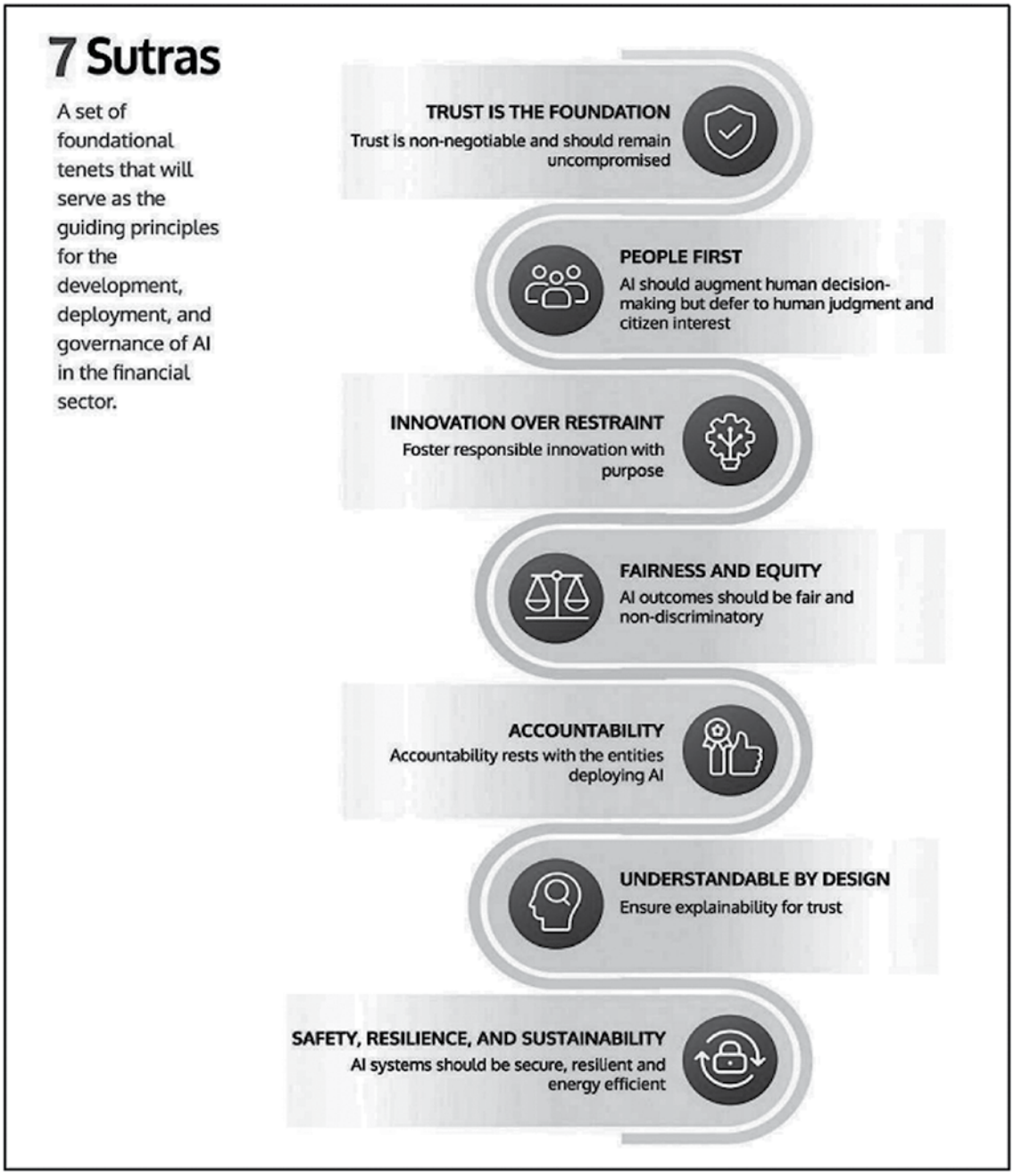
Clear accountability must rest with deploying institutions. Responsibility for AI-driven outcomes cannot be delegated to algorithms or vendors.

7.5.5.6 Sutra 6: Understandable by Design

Explainability must be embedded into AI systems. Institutions should be able to interpret and justify AI outputs, particularly in high-impact and customer-facing contexts.

7.5.5.7 Sutra 7: Safety, Resilience, and Sustainability

AI systems should be secure, resilient, and capable of safe operation under diverse conditions, with built-in safeguards and attention to long-term sustainability.



Taken together, the seven Sutras form an integrated and mutually reinforcing framework for the responsible innovation and adoption of AI. Reflecting the Sanskrit meaning of the term *sutra*, a "thread", these principles are intended to be woven throughout the entire lifecycle of AI systems. They constitute the normative foundation of the FREE-AI framework and apply to all institutions engaged in the development, deployment, or governance of AI within the Indian financial sector. Far from being abstract ideals, the Sutras are actionable principles that should be embedded within institutional policies, governance arrangements, operational processes, and risk management frameworks.

Chapter 8: Conclusion

The discourse on Artificial Intelligence and governance is neither static nor capable of definitive closure. As this book has demonstrated, Artificial Intelligence is not merely a technological development; it is a transformative force that intersects with law, ethics, public policy, and constitutional governance. Its increasing presence across institutions—particularly within the justice delivery system—requires sustained reflection rather than final answers.

Throughout this work, Artificial Intelligence has been examined as a tool with significant potential to enhance efficiency, consistency, and access to justice, while simultaneously posing serious challenges relating to transparency, accountability, bias, privacy, and human oversight. The discussions across judicial, policing, prosecutorial, and financial domains underscore a central insight: the impact of Artificial Intelligence is shaped not by technology alone, but by the legal and ethical frameworks within which it is deployed.

In the Indian context, this inquiry assumes particular importance. India's constitutional structure, social diversity, and scale of governance demand a cautious and context-sensitive approach to technological adoption. As highlighted in this book, innovation cannot be pursued in isolation from constitutional values such as equality before law, due process, dignity, and the protection of fundamental rights. Artificial Intelligence, therefore, must remain subject to the discipline of law, democratic accountability, and institutional responsibility.

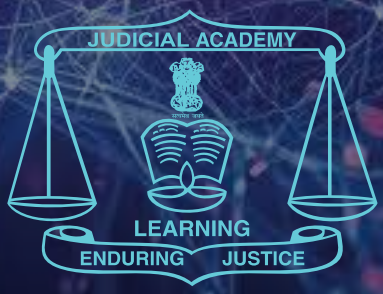
A recurring theme across the chapters is the indispensability of human judgment. Whether in adjudication, investigation, prosecution, or financial regulation, Artificial Intelligence may assist by organising information, identifying patterns, and supporting decision-making processes. However, it cannot replace the normative reasoning, ethical discernment, and contextual understanding that are intrinsic to human decision-makers. The legitimacy of public institutions ultimately rests on human accountability, not algorithmic outcomes.

This book does not seek to prescribe a singular model of AI governance, nor does it claim to exhaust the subject. On the contrary, it recognises that Artificial Intelligence is an evolving phenomenon, and that legal and policy responses must adapt over time. Questions concerning regulatory design, standards of explainability, allocation of liability, judicial review of algorithmic decisions, and the role of independent oversight bodies remain open and require continued engagement by courts, legislatures, administrators, and civil society.

The intention of this work is to contribute to an informed and principled conversation on Artificial Intelligence—one that moves beyond uncritical enthusiasm as well as unwarranted apprehension. By situating technology within the broader framework of constitutionalism, rule of law, and ethical governance, this book invites readers to reflect on how Artificial Intelligence can be harnessed in a manner that strengthens, rather than undermines, public trust in institutions.

The discussion on Artificial Intelligence and governance must, therefore, remain ongoing. As technology evolves, so too must our legal understanding, ethical sensitivity, and institutional preparedness. It is hoped that this book will serve not as a conclusion to the debate, but as a foundation for further inquiry, dialogue, and responsible action in shaping the future of AI governance in India.

❑❑❑



Prepared by :
Judicial Academy, Jharkhand
Near Dhurwa Dam, Dhurwa, Ranchi – 834004
Phone : 0651-2772001, 2772103, Fax : 0651-2772008
Email id : judicialacademyjharkhand@yahoo.co.in, Website : www.jajharkhand.in