IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

ACCOUNTABILITY AND TRANSPARENCY IN ALGORITHMIC DECISION-MAKING: EXAMINING THE USE OF AI AND MACHINE LEARNING IN INDIA'S PUBLIC SECTOR

Authors: Raj Shah, Dr. Uma Shinde
Designation - Student, JSPM University, Pune; Professor, JSPM University, Pune
Department of Law,
JSPM University, Pune, Maharashtra, India.

Abstract:
India is increasingly using artificial intelligence (AI) and Machine Learning (ML) in the public sector to improve efficiency, governance and public service delivery. These technologies are being used in areas such as welfare schemes, predictive policing, taxation systems, healthcare, and smart cities projects. While ai can make administration faster and more data driven, it also creates serious concerns regarding accountability and transparency. This paper examines the various issues arising from the use of ai based decision making systems in the Indian public sector. One major concern is that many ai systems operate as "black boxes, where it is difficult to pinpoint and understand how decisions were made. This opaqueness and lack of transparency along with other issues can affect important constitutional principles such as equality, fairness, right to information, privacy, etc under article 14 and 21 of the constitution. This study also highlights lack of proper laws and regulatory mechanisms governing AI in India. This paper argues that India needs a strong regulatory framework to ensure responsible use of ai in governance. It suggests measures such as the use of Explainable AI(XAI), independent auditing of algorithms and stronger human oversight in decision making. These safeguards are necessary to ensure that technological advancement does not weaken constitutional rights, public accountability, and citizens trust in government institutions.

1. INTRODUCTION:

Public administration in India is rapidly changing due to digital technology and increasing use of artificial intelligence (AI) and machine learning (ML). under initiatives such as digital India, the government has started using ai systems in areas like welfare distribution, healthcare, taxation, policing, smart cities, public infrastructure management, etc. Schemes like NITI Aayog's national strategy for artificial intelligence also promotes the use of AI in governance and public service. These technologies are deployed to increase efficiency, make administration faster and reduce human errors and corruption.

However, the growing use of ai in government decision making also creates serious concerns regarding transparency and accountability. Traditionally, administrative decisions made by government officials and humans were expected to follow principles of fairness, reasonableness, accountability, transparency, etc. Citizens usually have the right to know why a decision was made and can challenge unfair decisions before courts or authorities. But AI based systems especially advanced machine learning models, often work as "black boxes" meaning that even developers may not fully understand how the system reaches a particular decision. As a result, when governments rely on such systems it becomes difficult for citizens to know the reasons behind decisions that affect their benefits and rights.

This creates important constitutional concerns in India. Article 14 of the constitution guarantees equality before law and protects citizens against arbitrary state actions. Similarly, article 21 protects the right to life

and personal liberty which includes fairness and due process if an AI system wrongly denies welfare benefits, flags a person as suspicious or makes biased decisions without proper explanations citizens find it difficult to challenge such actions legally. The lack of clear human responsibility and accountability in algorithmic decision making is also a measure weakness.

Another major issue is the absence of a strong legal framework regulating AI in India. Although the Digital Personal Data Protection (DPDP) Act, 2023 provides safeguards for personal data and privacy, it does not specifically regulate automated decision-making systems or establish liability for errors caused by AI. There are currently no comprehensive laws regarding AI and accountability of algorithmic decision making in India. Many developed countries have already established or are in the process of establishing AI and ML laws.

This paper analyses the use of AI and Algorithmic decision making in India public sector and analyses the challenges it creates for accountability and transparency. It also studies the implications of such technologies and highlights the need for stronger safeguards. The paper argues that India must adopt measures such as explainable ai (XAI), independent auditing mechanisms, and stronger human oversight to ensure that technological advancements remains consistent with constitutional values, public trust, and democratic governance.

2. RESEARCH PROBLEM:

Despite growing awareness of the harms caused by unaccountable AI systems, no comprehensive and practicably operational accountability framework has been established in India that addresses legal, ethical and operational dimensions of the problem.

3. RESEARCH OBJECTIVES:

i) To track and understand the deployment of AI and algorithmic decision-making systems across key public sectors in India.

ii)To identify and analyse transparency and accountability failures arising from public sector AI deployment in India through documented case studies.

iii) To examine solutions for ensuring transparency and accountability in AI systems.

iv) To recommend a suitable governance framework that is responsible and adequate.

4. HYPOTHESIS:

India’s current legal framework is inadequate to ensure AI accountability, necessitating a structured and enforceable model inspired by global standards.

5. RESEARCH QUESTIONS:

I. Which public sectors in India have deployed the AI and algorithmic decision-making systems?
II. What are the impact of the AI and algorithmic decision-making systems that are implemented?
III. What institutional gaps in India's current framework allow unaccountable AI?
IV. What technical and legal solutions can effectively address transparency and accountability in public sector AI?

6. SIGNIFICANCE OF RESEARCH:

This research contributes to the growing field of AI governance and algorithmic accountability by examining existing scholarly literature on transparency, accountability, and decision making in AI systems, along with the solutions and regulatory approaches proposed by various scholars. It analyses the deployment of AI and machine learning systems across different public sector domains in India and highlights the increasing need for a comprehensive legal and institutional framework governing algorithmic accountability.

The research further emphasises that such a framework must align with constitutional principles such as transparency, fairness, due process, equality, and accountability. By examining the implementation of AI systems across the country, the research also aims to identify the legal and governance challenges emerging from their use in public administration.

The research is significant for policymakers, legal practitioners, researchers, and technology developers, as it provides practical recommendations for strengthening transparency, fairness, human oversight, and accountability in AI driven governance systems.

7. RESEARCH METHODOLOGY:

The study follows a qualitative and doctrinal research design. It contains theoretical elements. The research relies entirely on secondary sources and does not involve interaction with human participants. It studies real world case studies.

8. LIMITATIONS:

Due to limited time available the research has been limited by the extent of research that could be done during the time span available. The technical aspects of AI system design are addressed at a conceptual level; this is not a computer science paper and does not involve the development or testing of algorithms.

9. LITERATURE REVIEW:

i) O. E. Olorunniwo, Mayowa Alamge, Olatunji Isreal in their research paper titled Accountability in AI Decision-Making talk about important decisions that AI is today making areas like healthcare, finance, and criminal justice. The author talks about how vital it is to make sure these systems can be held responsible. The research breaks ai accountability into three main areas : legal, ethical and operational accountability. It then talks about the black box problem and how it poses as a challenge into identifying how ai makes its choices. The paper conducts case reviews of ai being implemented into various sectors like automotive, human resource, healthcare sectors, insurance underwriting, etc. It recommends implementing accountability by design which means implementing fairness and safety checks into the ai right from the start. They also focus on creating Explainable AI(XAI), which ensures that humans can easily understand how the ai came to its final decision. Finally, it talks about the need for cooperation between countries to implement a uniform and comprehensive framework so that there are no fragmented laws scattered across globe as AI will be used globally.[1]

ii) Research paper titled Data in New Delhi's predictive policing system by Vidushi and Shivangi is a study that analyses Delhi's CMAPS system. It argues that predictive policing reinforces existing biases rather than eliminating them. Based on a 2-year field study 2017-2019 the authors show that the system relies on flawed data shaped by human decisions, institutional practices and social inequalities. It highlights issues such as outdated mapping tools, incomplete address databases, and the target pressure that is present on call operators which lead to inaccurate crime recording. Migrant settlements and settlements consisting of marginalised groups are often misrepresented and targeted. The study identifies three biases: Historical Bias (reflecting societal discrimination), representation Bias (over policing of poorer areas), and measurement Bias (inaccurate location data). Then there is Lack of transparency which is partly due to it allegedly falling under RTI exemptions, which worsens accountability concerns. This even leads to constitutional concerns under article 14 and 15 due to discriminatory impacts on marginalised groups. The authors conclude that without institutional reform, technology only reproduce and intensify existing inequalities.[2]

iii) Orozco Brenda, in her paper discusses the growing importance of ai in public administration due to its efficiency and ability to improve governance. However, she also highlights significant concerns regarding transparency, accountability, bias, privacy and legality. The paper examines that opaque ai algorithms often make decisions that are difficult to understand or explain which can threaten fairness, due process and citizens' rights. Excessive dependence on automated systems may also weaken human judgment and democratic values in governance.

The paper further explains that transparency in AI governance refers to the extent to which AI decision-making processes are understandable and accessible to human stakeholders. It includes both technical transparency and procedural transparency. Technical transparency relates to understanding how AI systems process data and make decisions. However, many machine learning models lack explainability, making it difficult even for experts to trace their reasoning. Procedural transparency focuses on how AI systems are implemented within public institutions, including the rules governing their use, decision-making standards, and available remedies for affected individuals.

The paper also introduces the accountability triangle, which consists of legal, political and administrative accountability. Legal accountability is maintained through judicial and legislative oversight, political accountability through democratic participation and public scrutiny and administrative accountability

[1] O.E. Olorunniwo, Mayowa Alomge & Olatunji Isreal, *Accountability in AI Decision-Making* (2025).
[2] Vidushi Marda & Shivangi Narayan, *Data in New Delhi's Predictive Policing System* (2020).

through internal control and ethical practices. The paper concludes that effective AI governance requires accountability mechanisms across all three dimensions, although practical implementation remains challenging.[3]

iv) The paper written by Dr. Priyadarshi examines the growing challenge of assigning legal liability and accountability in AI driven decision making. The author argues that traditional legal frameworks are not sufficiently equipped to deal with AI systems because of their complexity, autonomy and lack of transparency. A main concern is the black box nature of AI models which makes it difficult to determine how decisions are reached and who should be held responsible when harm occurs. The paper identifies several obstacles to effective regulation including vague and ambiguous legal definitions of AI-related concepts, algorithmic bias and discrimination, jurisdictional conflicts across borders, and the difficulty of identifying responsibility among developers, operators, and users. These challenges create uncertainty in legal proceedings and weaken accountability mechanisms. To address these issues, the paper advocates a combination of technical, legal, and ethical solutions. It emphasizes the importance of explainable AI and Transparency-Enhancing Technologies to improve the interpretability and auditability of AI systems. The author also mentions about regulatory sandboxes which allow AI technologies to be tested under regulatory supervision before full deployment. Additionally, the paper highlights the need for ethical guidelines, stakeholder participation and interdisciplinary collaboration among lawyers, ethicists and developers. The paper concludes that effective AI governance requires clearer legal standards, stronger transparency requirements, international regulatory cooperation, and adaptive accountability frameworks capable of keeping pace with rapid technological change.[4]

v) Ben Chester Cheong in his paper examines the legal and ethical challenges of implementing transparency and accountability in artificial intelligence systems. The author highlights four main frameworks in his research. It includes Technical Approaches like Explainable AI and LIME method of interpretation to tackle black box. It then involved Legal and Regulatory framework examination like data protection and privacy by analysing provisions like GDPR and American Laws. The third focus involved ethical and societal consideration wherein the author points out that AI systems should be developed with Ethical considerations and trust should be built among users regarding the systems, and lastly the author talks about Interdisciplinary and collaborative approaches wherein he recommends collaboration between different disciplines like law, ethics, sociology, computer science, etc.. to address complex challenges.

vi) Al Jazeera conducts research to examine a central Algorithmic system called Parivar Pehchan Patra database which was established to determine the eligibility of welfare claimants. The Haryana state government has multiple welfare schemes, one among them provides montly pension to people over the age of 60 if they have income below 3,00,000 rs per annum. But the government stopped these pensions of thousands in span of three years because the algorithmic system had been wrongfully declaring them dead or was doing wrongful predictions and had other errors. This happened with several other welfare schemes too where it would wrongfully deny the eligibility of citizens.[5]

vii) Telangana's government deployed an algorithmic system called Samagra Vedika to determine eligibility for its food security scheme which provides subsidized grains to the poor. Between 2014 and 2019 over 1.86 million food security cards were cancelled and 1,42,086 fresh applications were rejected without notice.

The system uses data from up to 30 government databases to build citizen profiles and flags anyone it deems ineligible. However, due to faulty data and mismatched records, many legitimate beneficiaries were wrongly excluded. A Supreme Court ordered reverification found that at least 7.5% of rejected cards had been wrongfully cancelled.

A key problem is that officials trusted the algorithm over ground reality. Once excluded, the burden fell entirely on the removed beneficiaries to prove their eligibility to government agencies. And even when they did so, officials often favoured the algorithm decision. The systems source code remains opaque and

[3] Brenda Guadalupe Orozco, *Artificial Intelligence in Public Sector Decision-Making: Accountability, Transparency and Ethical Governance* (2025).

[4] Priyadarshi Nagda, *Legal Liability and Accountability in AI Decision Making: Challenges and Solutions* (2025).

[5] Kumar Sambhav Shrivastava, Tapasya & Divij Joshi, *In India, an Algorithm Declares Them Dead; They Have to Prove They're Alive*, Al Jazeera (Jan. 25, 2024), https://www.aljazeera.com/economy/2024/1/25/in-india-an-algorithm-declares-them-dead-they-have-to-prove-theyre. (last visited May 31, 2026)

has never been independently audited. Critics argue this lack of accountability, combined with institutional pressure to find fraud, systematically pushes poor and vulnerable families out of the welfare net.[6]

viii) An investigation by the Wire and Pulitzer Centre found that Delhi Police used AI powered facial recognition technology (FRT) to identify and arrest individuals in connection with the 2020 Northeast Delhi riots, sometimes relying solely on this technology. Accused persons spent years in pretrial detention Suffering alleged custodial abuse and financial ruin. The Delhi Police admitted to using FRT in over 750 riot related cases and the home minister confirmed that over 1900 perpetrators were identified through it, yet more than 80% of cases heard so far have ended in acquittals or discharges. The technology operates without any legal framework, privacy impact assessment, or meaningful oversight, and experts warn it disproportionately targets Muslim majority and over Policed neighbourhoods in Delhi. No legal opinion was sought before deploying it, and police training on FRT remains minimal.[7]

ix) The Dutch childcare benefits scandal was one of the most controversial scandals that took place in 2021. A self-learning algorithm used by the Dutch Tax and Customs administration in childcare benefit claims functioned wrongly. Between 2005 and 2019, approximately 26,000 parents were wrongly accused of making fraudulent benefit claims resulting in demands to repay their allowances in full. Racial profiling had built into the algorithm's design disproportionately flagging people from ethnic minorities and those with non-Dutch nationality as high risk. A discriminatory cycle was reinforced by a self-learning mechanism with no meaningful human oversight. Families lost their homes, jobs and savings and at least 3500 children were taken into foster care. The severity of the scandal led to the resignation of the entire Rutte cabinet in 2021. A parliamentary inquiry concluded it violated fundamental principles of the rule of law. [8]

x) Nair and Gundaboina look at how India's Tax Department is using AI to tackle its long-standing problem of low tax collection and widespread evasion. Their focus is Project Insight, which is launched in 2017, which pulls together data from banks, property records, GST and credit cards, etc to identify taxpayers who may be hiding income. It also sends SMS and email reminders nudging people to correct their returns voluntarily under their Nudge Plan. The results have been promising. Over 1 crore revised returns and 11,000 crore rs in extra taxes collected. But the authors also pen a caution note. The system can wrongly flag innocent people, especially those with complicated financial situations. It can also carry forward existing prejudices from past data. Worse, there is currently no independent body to review complaints, no error reporting and no way for ordinary taxpayers to meaningfully challenge the system's decisions.[9]

xi) A recent report on Hyderabad's AI-based traffic management system highlights both the benefits and challenges of using automated technology for law enforcement. The report notes that thousands of incorrect e-challans were issued in 2025 due to errors in the automatic number plate recognition system. These errors occurred when traffic cameras incorrectly matched vehicle numbers with detected violations or misread unclear number plates. The report also identifies the problem of duplicate chalans where the same offence is recorded by multiple cameras. Although manual verification is conducted before issuing chalans, officials state that the large number of daily challans makes thorough checking and verification difficult. The findings suggest that while AI and digital surveillance can improve traffic regulation and reduce human effort, technical inaccuracies can affect fairness and public trust. The report concludes that

[6] Kumar Sambhav Shrivastava, Tapasya & Divij Joshi, *How an Algorithm Denied Food to Thousands of Poor in India's Telangana*, Al Jazeera (Jan. 24, 2024), https://www.aljazeera.com/economy/2024/1/24/how-an-algorithm-denied-food-to-thousands-of-poor-in-indias-telangana . (last visited May 31, 2026)

[7] Astha Savyasachi, *As AI Took Over Policing in Delhi, Who Bore the Brunt?*, The Wire (July 2, 2025), https://pulitzercenter.org/stories/ai-took-over-policing-delhi-who-bore-brunt. (last visited May 31, 2026)

[8] Amnesty International, *Dutch Childcare Benefit Scandal an Urgent Wake-Up Call to Ban Racist Algorithms*, Amnesty Int'l (Oct. 25, 2021), https://www.amnesty.org/en/latest/news/2021/10/xenophobic-machines-dutch-child-benefit-scandal/. (last visited May 31, 2026)

[9] Sthanu R. Nair & Ranjith Gundaboina, *AI-Powered Tax Governance in India and Its Challenges*, The Hindu (Feb. 18, 2026), https://www.thehindu.com/opinion/op-ed/ai-powered-tax-governance-in-india-and-its-challenges/article70762627.ece. (last visited May 31, 2026)

better system accuracy, stronger verification procedures and improved grievance mechanisms are necessary to ensure accountability and reliability in AI-driven traffic enforcement.[10]

xii) A report discusses the Reserve Bank of India's plan to develop Digital Payments Intelligence Platform to improve security in the rapidly growing digital payments sector. The platform will use artificial intelligence to analyse data from multiple sources and identify potential fraudulent transactions before they are completed. According to RBI officials, the system will generate real-time risk alerts enabling banks and customers to make informed decisions about whether to proceed with a transaction. The initiative is being developed by the RBI Innovation Hub and is expected to strengthen fraud prevention mechanisms in digital payments. The proposed platform reflects a growing reliance on AI-driven monitoring systems to enhance financial security, improve fraud detection and maintain trust in India's digital payment ecosystem.[11]

International frameworks:

i) The EU AI ACT published in the official journal of the European Union on 12 July 2024 and entering into force on 1 august 2024, represents the world's first comprehensive legal framework for AI. It adopts a risk based approach, categorising ai systems into unacceptable-risk (prohibited), high-risk, limited-risk, and minimal-risk tiers. High-risk AI systems including those used in education, employment, biometric identification, critical infrastructure, law enforcement, migration, and administration of justice are subject to extensive obligations including transparency, human oversight, accuracy, data governance, and post-market monitoring requirements.[12]

ii) The General Data Protection Regulation (GDPR) remains the primary data protection instrument. Enforced in 2018, is the European Union's framework for data protection and privacy. It regulates how organizations collect, process, and store personal data of individuals within the European Union. GDPR is built on principles such as lawfulness, transparency, purpose limitation, and data minimization. It grants individuals rights like access, rectification, erasure (right to be forgotten), and data portability. Organizations must obtain clear consent and ensure strong data security measures. Non-compliance can lead to heavy fines, making GDPR one of the strictest data protection laws globally.[13]

10. DEPLOYMENT OF AI SYSTEMS IN INDIAN PUBLIC SECTOR:

I) Samagra Vedika :

Samagra Vedika is an algorithmic data profiling system used by the state government of Telangana to determine welfare eligibility. While the system was introduced to catch welfare fraud and streamline the process, it has wrongfully cut off thousands of legitimately impoverished residents from accessing subsidised food security programs.

Between 2014 and 2019, the Telangana government cancelled over 1.86 million existing food security cards and rejected more than 142000 new applications. While the state claimed this weeded out fraudulent subsidy claimants, investigations revealed thousands were wrongfully excluded due to bad data and faulty algorithmic decisions. This was introduced in 2016 (initially by the police to track criminals), Samagra Vedika pulls data from around 30 different government databases to build a 360-degree digital profile of residents. Under state food rules, family are disqualified if they meet certain criteria, such as owning a four-wheeler vehicle, family owning a property, owning a business or fuel pump, or having formal employment. Because names and addresses vary wildly across databases the algorithm relies on machine learning and entity resolution to match identities. While officials claim the error margin is under 5%, a supreme court ordered reverification of cancelled cards revealed a much higher error rate of at least 7.5% of the processed cases. They were found to be wrongfully rejected. Experts note that the push to lower welfare fraud (false negatives), inherently increases the rate of false positives, thereby unjustly stripping valid families of aid.

[10] *25k Wrong Challans: Tech Glitch Drives Hyd Motorists Up the Wall*, Times of India (Nov. 9, 2025), https://timesofindia.indiatimes.com/city/hyderabad/25k-wrong-challans-tech-glitch-drives-hyd-motorists-up-the-wall/articleshow/125191689.cms. (last visited May 31, 2026)

[11] *RBI Developing AI-Driven Tool to Flag Fraudulent Transactions*, New Indian Express (Oct. 1, 2025), https://www.newindianexpress.com/business/2025/Oct/01/rbi-developing-ai-driven-tool-to-flag-fraudulent-transactions. (last visited May 31, 2026)

[12] Artificial Intelligence Act, Regulation (EU) 2024/1689, 2024 O.J. (L 1689).

[13] General Data Protection Regulation, Regulation (EU) 2016/679, 2016 O.J. (L 119) 1.

Bismillah Bee is a 67-year-old widow whose benefit were cancelled because the algorithm mistakenly tagged her late husband who was a rickshaw puller as a car owner by confusing his name with someone else. Consequently, her family was stripped of access to subsidized rice during the height of the covid 19 pandemic and her husband's terminal cancer battle. Similarly, another women Maher Bee was denied aid because the system tagged her family with owning a car despite her husband being paralysed in one leg and unable to drive.

The algorithm is entirely opaque. The state IT department denied right to information requests to share the systems source code, citing proprietary right of the private developer, Posidex technologies.[14]

II) Parivar Pehchan Patra (PPP) :

In Haryana, the state government introduced an algorithmic database called the Parivar Pehchan Patra to verify eligibility for welfare schemes like old age and widow pensions. The system wrongly declared hundreds of thousands of people dead cutting off their pensions. A 102-year-old man named Dhuli Chand had to literally stage a mock wedding procession to convince officials he was alive. According to the government data, pensions of over 2,77,000 elderly citizens and 52,000 widows were stopped because the algorithm marked them as dead.

The important problem was complete lack of transparency and accountability. The government refused to respond to Right To Information (RTI) applications asking about the algorithms design and logic. When affected citizens tried to get corrections made, they were sent from office to office in endless loops, one family was told to get an income certificate to fix their PPP record, but getting that certificate required a valid PPP record in the first place.

The new age system without proper framework led to an opaque AI decision making, no transparency, no grievance mechanism that actually works, and vulnerable people bearing the consequences of algorithmic errors with no easy recourse.[15]

III) Predictive Policing and ADMS :

Predictive policing and artificial intelligence-based tools for law enforcement represents another domain where algorithmic decision-making carries serious legal and ethical consequences. Across India, several state police forces are in the race to adopt AI systems that performs functions such as analyse historical crime data to forecast where crimes are likely to occur and who is most likely to do them. Therefore, many state police departments have started using AI Based Automated decision-Making Systems (ADMS). The scale at which India is using these systems is substantial. These include the Facial Recognition System (FRS) Crime Mapping, Analytics & Predictive System (CMAPS) in Delhi[16], YAKSHA[17], Trinetra[18] and CrimeGPT in Uttar Pradesh[19], Punjab AI System (PAIS) in Punjab, Automatic Number Plate Recognition System (ANPR) in Madhya Pradesh[20], Artificial-Intelligence based Human Efface Detection (ABHED) system in Rajasthan[21], Telangana State Police – COP (TSCOP) and SHEILD in Telangana.[22],

---

[14] Kumar Sambhav Shrivastava, Tapasya & Divij Joshi, *How an Algorithm Denied Food to Thousands of Poor in India's Telangana*, Al Jazeera (Jan. 24, 2024), https://www.aljazeera.com/economy/2024/1/24/how-an-algorithm-denied-food-to-thousands-of-poor-in-indias-telangana. (last visited May 31, 2026)

[15] Kumar Sambhav Shrivastava, Tapasya & Divij Joshi, *In India, an Algorithm Declares Them Dead; They Have to Prove They're Alive*, Al Jazeera (Jan. 25, 2024), https://www.aljazeera.com/economy/2024/1/25/in-india-an-algorithm-declares-them-dead-they-have-to-prove-theyre. (last visited May 31, 2026)

[16] *Delhi Police Implements Crime Mapping, Analytics & Predictive System (CMAPS)*, BW Police World, https://www.bwpoliceworld.com/article/delhi-police-implements-crime-mapping-analytics-predictive-system-cmaps-134283 (last visited May 31, 2026).

[17] *UP Government Launches AI-Based App YAKSH to Strengthen Policing in State*, NDTV, https://www.ndtv.com/india-news/up-government-launches-ai-based-app-yaksh-to-strengthen-policing-in-state-10040969 (last visited May 31, 2026).

[18] *UP Police Revolutionizes Policing with Trinetra App 2.0, Enhancing Crime Prevention and Investigation*, Times of India (Feb. 2024), https://timesofindia.indiatimes.com/city/lucknow/up-police-revolutionizes-policing-with-trinetra-app-20-enhancing-crime-prevention-and-investigation/articleshow/108537077.cms (last visited May 31, 2026)

[19] *Meet CrimeGPT: The UP Police Department's AI Tool for Fighting Crime*, IndiaAI, https://indiaai.gov.in/article/meet-crime-gpt-the-up-police-department-s-ai-tool-for-fighting-crime (last visited May 31, 2026).

[20] *Soon Indore Will Be Under Cam Eye*, Hindustan Times, https://www.hindustantimes.com/indore/soon-indore-will-be-under-cam-eye/story-3HwFq5SKi8SBbaIkfsbgJN.html (last visited May 31, 2026).

[21] *Criminal Investigation: AI-Based Human Efface Detection System in Rajasthan*, YourStory (June 2017), https://yourstory.com/2017/06/criminal-investigation. (last visited May 31, 2026)

[22] *TSCOP App to Use Facial Recognition*, New Indian Express (Aug. 3, 2018), https://www.newindianexpress.com/cities/hyderabad/2018/Aug/03/tscop-app-to-use-facial-recognition-1852485.html. (last visited May 31, 2026)

Maharashtra has developed MARVEL to increase its policing efficiency and intelligence capabilities[23], Delhi and Gujarat use prophecy platforms developed by Innefu labs which combines facial recognition, open source intelligence and data fusion to profile criminal networks in real time, and Goa police use Deep Trace system[24]. More states are working towards deploying their independent ML and AI based intelligence platforms. All these systems are like ticking time bombs waiting to explode without proper provisions to govern them.

Investigation done by Vidushi Marda, Shivangi Narayan on The Delhi's CMAPS system shows that how inaccurate collection of data and feeding of historic data heightens Bias and discrimination of marginalised and disadvantaged communities and that without proper auditing, AI ethics and bias officers, redressal systems the platform can do more damage than good.[25]

Real world consequences have already begun. An investigation by the wire and Pulitzer centre uncovered instances in Delhi where individuals were arrested solely on the basis of facial recognition matches, with no corroborating evidence or credible witness accounts.[26]

The mass deployment of law enforcement AI in Delhi was at a all time high during the investigation of the communal violence that erupted on February 24,2020 in areas like Chand Bagh. The violence broke out during protests against the Citizenship Amendment Act and resulted in numerous casualties. The union Home Minister informed parliament that the delhi police had identified 1,922 accused using advanced facial recognition technology (FRT). This software was developed by private defense firms.[27]

The important finding of the investigation is that the Delhi police routinely arrested individuals based on almost entirely low quality facial recognition matches. The software flagged people from video footage even where there was no solid corroborating evidence, physical recovery, or credible public witness testimony to place them at scene of crime.

The investigation follows suspects like Ali, a resident arrested who was forced to undergo more than 4 and half years of pretrial incarceration. Because of severe backlogs in courts, he sat in prison without a trial even starting, until he was finally granted bail years later. The investigation also mentions other suspects one of whom also spent more than 2 years in prison before he was granted bail. These suspects lost their livelihoods, saw their families going into debt to pay fees and faced depression.[28]

Media reports state that from the people accused for violence more than 80 % cases heard till now were being acquitted or dismissed showing how unreliable FRT systems can be.[29]

IV) Project Insight – Taxation

India has struggled with low tax to GDP ratio and high levels of tax evasion, losing approximately 4.3% of tax revenues annually. To address this, the income tax department launched Project Insight in 2017, fully operational by 2019. The system uses Ai and advanced data analytics which aggregates financial data from banking transactions, property deals, credit card payments, GST records, and high value purchases to build a 360 degree financial profile of every taxpayer. When inconsistencies are found the NUDGE strategy sends SMS and email reminders nudging taxpayers to pay the appropriate tax amount voluntarily. The results have been good with over one crore revised returns filed from 2020-22 generating over 2.11 lakh crore rupees in additional taxes. Refund processing time dropped 93 days to 17.

But the legal and ethical concerns are to be looked at here. The system cannot distinguish genuine tax evasion from legitimate financial complexity sometimes and variable income professionals, joint family structures or simple clerical errors can trigger false positives, placing the burden of proof on ordinary

[23] *Maharashtra Cabinet Approves ₹23 Cr MARVEL AI Project for Police Force*, Times of India, https://timesofindia.indiatimes.com/city/mumbai/maharashtra-cabinet-approves-23cr-marvel-ai-project-for-police-force/articleshow/108554366.cms (last visited May 31, 2026).

[24] *Goa Police's AI-Powered Deep Trace to Probe Deep, Dark Crimes*, Times of India, https://timesofindia.indiatimes.com/city/goa/goa-polices-ai-powered-deep-trace-to-probe-deep-dark-crimes/articleshow/120833963.cms (last visited May 31, 2026).

[25] Vidushi Marda & Shivangi Narayan, *Data in New Delhi's Predictive Policing System* (2020).

[26] Astha Savyasachi, *As AI Took Over Policing in Delhi, Who Bore the Brunt?*, The Wire (July 2, 2025), https://pulitzercenter.org/stories/ai-took-over-policing-delhi-who-bore-brunt (last visited May 31, 2026).

[27] Press Information Bureau, *Home Minister's Statement on Delhi Riots*, PIB (Mar. 11, 2020), https://www.pib.gov.in/PressReleasePage.aspx?PRID=1606194&reg=3&lang=2. (last visited May 31, 2026)

[28] Astha Savyasachi, *As AI Took Over Policing in Delhi, Who Bore the Brunt?*, The Wire (July 2, 2025), https://pulitzercenter.org/stories/ai-took-over-policing-delhi-who-bore-brunt. (last visited May 31, 2026)

[29] Amnesty International, *Dutch Childcare Benefit Scandal an Urgent Wake-Up Call to Ban Racist Algorithms*, Amnesty Int'l (Oct. 25, 2021), https://www.amnesty.org/en/latest/news/2021/10/xenophobic-machines-dutch-child-benefit-scandal/. (last visited May 31, 2026)

taxpayers to disprove an algorithmic suspicion. Ai models trained on historical enforcement data risk socio economic and geographic biases, identifying fraud more frequently in certain communities not because they evade more but because they were historically scrutinised more.

The Dutch childcare benefits scandal where a similar algorithmic system disproportionately flagged ethnic minority families stands as a cautionary parallel[30]. More importantly India lacks an AI ombudsperson, mandatory algorithmic impact assessments, public reporting of false positive rates, or external audits of risk scoring models. Without these safeguards, project insight risks becomes less of a tool to catch Tax fraud and more as an opaque surveillance platform, one where taxpayers are flagged by an algorithm they cannot understand, for reasons they are never fully told.[31]

V) Traffic management systems

Cities like Ahmedabad, Bengaluru, Pune, Mumbai, Varanasi, Kolkata, etc deploy AI algorithms for Red light violation detection and Automatic Number Plate Recognition. These systems process visual data from surveillance cameras, match plate numbers with the ministry of Road transport's database and generate algorithmic e-challans without human intervention. Municipalities use computer vision algorithms to run behavioural studies on traffic density. The algorithms adjust the timing of traffic lights in real time based on vehicle volume rather than fixed timers. But these AI systems are not properly calibrated, many times it generates false positives leading to issuance of thousands of false challans where there was no traffic violations. [32]

For example, in Hyderabad Data reveals more than 50,000 erroneous e-challans were issued by advanced surveillance systems. Motorists were slapped with wrong fines for offences like overseeding, helmetless driving, and parking violations that they never committed. The core issue stems from a technical desynchronization in the Automatic Number Plate Recognition camera network. The network relies on two separate automated components: one system detects the traffic violation with a separate computer module reading the number plate. The glitch occurs when these two fail to sync correctly, pairing a genuine traffic offense committed by one vehicle with the license plate of completely innocent vehicle passing by.

Senior officials admitted that during manual verification officers usually only check if the registration number exists in the transport database, they do not verify whether the vehicle pictured in the camera frame actually matches the vehicle model or colour being fined. These AI systems are capable of generating thousands of challans everyday but manually verifying all of them proves to be very difficult due to the scale in big cities.[33]

VI) Digital Payments Intelligence Platform (DPIP)

The RBI is developing an AI powered platform called DPIP to detect and prevent digital payment fraud before transactions are completed. This is a shift from reactive to proactive fraud prevention. It is being prototyped at the RBI innovation hub.

DPIP functions as a shared intelligence layer across all payment participants. It analyses real time transaction signals, including mule accounts, telecom metadata, device ID, IP address, geographic origin and transaction history to generate a risk score for each transaction. If a transaction is flagged users or banks can intervene before the payment goes through. Importantly, the system uses federated learning meaning fraud patterns are analysed collectively without exposing individual user data.

Each UPI transaction will carry a real time risk score; suspicious handles will be flagged with in app warnings like High-Risk handle or unusual transfer detected and verified merchants will receive a verified merchant badge. Fraudulent or mule accounts will be automatically frozen across multiple banks.[34]

---

[30] *AI Governance Lessons from the Dutch Childcare Benefits Scandal*, AICC Responsible AI Hub, https://www.aicc.co/responsible-ai-hub/keep-up-to-date-with-responsible-ai/blog-posts/responsible-ai-lessons-from-the-dutch-childcare-benefits-scandal (last visited May 31, 2026).

[31] Sthanu R. Nair & Ranjith Gundaboina, *AI-Powered Tax Governance in India and Its Challenges*, The Hindu (Feb. 18, 2026), https://www.thehindu.com/opinion/op-ed/ai-powered-tax-governance-in-india-and-its-challenges/article70762627.ece. (last visited May 31, 2026)

[32] *AI in Indian Traffic Management: Transforming Urban Mobility Challenges*, IndiaAI, https://indiaai.gov.in/article/ai-in-indian-traffic-management-transforming-urban-mobility-challenges (last visited May 31, 2026).

[33] *25k Wrong Challans: Tech Glitch Drives Hyd Motorists Up the Wall*, Times of India (Nov. 9, 2025), https://timesofindia.indiatimes.com/city/hyderabad/25k-wrong-challans-tech-glitch-drives-hyd-motorists-up-the-wall/articleshow/125191689.cms. (last visited May 31, 2026)

[34] *DPIP: RBI's AI Tool to Detect Fraud in Real Time*, Economic Times, https://economictimes.indiatimes.com/news/economy/policy/dpip-rbis-ai-tool-to-detect-fraud-in-real-time/articleshow/124536025.cms (last visited May 31, 2026).

Alongside this, rbi has also introduced risk-based authentication guidelines effective April 1, 2026, allowing layered security checks beyond standard two factor authentication. Private and public banks together reported 21,000+ fraud cases involving crores of rupees in 2025 alone. And this system intends to significantly reduce such cases.[35]

But for successfully implementing this algorithmic system it must properly integrate data from banks, telecoms, and payment platforms. Then it must manage false positives without causing problems to other users and define legal liability when the system misses or wrongly flags a transaction and while staying ahead of evolving fraud tactics.

The RBI is trying a proactive approach and this leads to it creating high risks for users and account holders. The system will after a point be able to freeze accounts and block transactions. Without proper Human in the loop, ombudsman and grievance redressal, high accuracy and Data Bias check officers this system can lead to significant problems to the citizens.

VII) Other Areas:

Apart from these above-mentioned areas the government is also trying to implement AI systems in healthcare, education, etc. As the time goes on new systems will be implemented across almost all the sectors.

11. BRIDGING THE ACCOUNTABILITY GAP: LEGAL AND TECHNOLOGICAL SOLUTIONS FOR TRANSPARENT AI GOVERNANCE.

After looking at various sources the research has found below solutions that can be considered and applied. This is not an exhaustive list and as technology improves new challenges and solutions will keep on developing.

i.Explainable AI (XAI) as a Solution for Transparency and Accountability:

Explainable Artificial Intelligence has emerged as an important solution to address the lack of transparency and accountability in modern AI systems.[36] Many advanced AI models, especially deep learning systems, operate as black boxes, meaning their internal decision making processes are difficult for humans to understand.[37] This is known as opacity of these AI models and when it is difficult for an ai model to give exact reasons for why it arrived at a decision than it is said that an ai model is opaque. This creates serious concerns when AI is used in high stakes decision making such as healthcare, insurance, education, policing, recruitment, and welfare administration.

XAI aims to make systems more interpretable and understandable to users, regulators, and affected individuals. It gives explanations for how and why an AI system arrived at a particular decision or prediction. One major technique developed under XAI is LIME (local interpretable model-agnostic explanations) which explains individual AI predictions using simpler interpretable models. This is a sort of de-encryption technique developed to de-encrypt how the model reached its decision. Such methods help identify the factors influencing AI decisions and make complex systems easier to analyse.[38]

The importance of XAI can be seen in Houston Federation of Teachers vs Houston Independent School District, where teachers challenged as AI based evaluation system because they could or contest its results. The case demonstrated that opaque AI systems can undermine fairness and due process.[39]

XAI improves accountability by enabling individuals to question automated decisions and by assisting organisations and regulators in auditing AI systems for errors or bias. As the European Union's GDPR "right to explanation"[40] and the EU AI Act, which require transparency for high-risk AI systems[41] have been enacted XAI becomes extremely accountable.

---

[35] *RBI Developing AI-Driven Tool to Flag Fraudulent Transactions*, New Indian Express (Oct. 1, 2025), https://www.newindianexpress.com/business/2025/Oct/01/rbi-developing-ai-driven-tool-to-flag-fraudulent-transactions. (last visited May 31, 2026)

[36] Ben Chester Cheong, *Transparency and Accountability in AI Systems: Safeguarding Wellbeing in the Age of Algorithmic Decision-Making* (2024).

[37] S.R.A. Parisineni & M. Pal, *Enhancing Trust and Interpretability of Complex Machine Learning Models Using Local Interpretable Model Agnostic SHAP Explanations*, 23 Int'l J. Data Sci. & Analytics 458 (2023).

[38] Marco Tulio Ribeiro, Sameer Singh & Carlos Guestrin, *"Why Should I Trust You?" Explaining the Predictions of Any Classifier*, in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 1135 (2016).

[39] *Houston Fed'n of Teachers v. Houston Indep. Sch. Dist.*, 251 F. Supp. 3d 1168 (S.D. Tex. 2017).

[40] General Data Protection Regulation, Regulation (EU) 2016/679, art. 22, 2016 O.J. (L 119) 1.

[41] Artificial Intelligence Act, Regulation (EU) 2024/1689, 2024 O.J. (L 1689).

However, XAI faces challenges. More accurate AI models are often less interpretable, creating a trade off between performance and explainability.[42] This means that sometimes when XAI is deployed and AI is made interpretable it acts slower and duller than what it would have if XAI was not deployed. Additionally, technical explanations may still be difficult for ordinary users to understand. Therefore, while XAI is a vital step toward transparent and accountable AI, it must be combined with human oversight, legal safeguards, and proper auditing mechanisms to ensure responsible AI governance.

ii. Third Party Auditing and Algorithmic Impact Assessment:

Algorithmic auditing and impact Assessments are important tools for improving accountability and fairness in AI systems. As AI is increasingly influencing decisions in different sectors, there is a growing need to ensure that these systems do not produce biased, discriminatory or harmful outcomes.[43]

Third-Party Auditing:

Kroll et al. (2017) propose algorithmic audits as a means to detect discrimination and other biases in AI systems. These audits involve scrutinizing the inputs, processes, and outputs of AI systems to identify and mitigate biases that may not be apparent during the initial development stages (Kroll et al., 2017)[44]

One efficient way of conducting these audits is through third part auditing. Third party auditing involves independent external bodies examining an AI systems design, training data, decision making logic, and real world outputs to access whether it is functioning fairly, accurately and legally. These auditors operate without any stake in the systems commercial success, making their findings more credible and objective.

If independent audits of systems are conducted at regular intervals then it can lead to avoiding large scale harm caused by high stakes decision making systems. Third party audits typically examine three things whether the training data is representative meaning it accommodates all the parties involved equally and free from historical bias, whether the models outputs are consistent and accurate across different demographic groups and whether the system includes adequate safeguards against misuse. These Audit findings should than be made publicly available to ensure accountability and systems that fails audits should be suspended until corrected.

Therefore, audits involve examining the inputs, processes and outputs of AI systems to identify problems such as bias, unfair treatment or inaccurate decision that may not be visible during the design or development stage. And by analysing how these AI systems function in practice, audits can improve transparency and help organisations correct unfair outcomes which they themselves cannot see sometimes.

Algorithmic Impact Assessments:

Reisman et al. (2018) advocate for algorithmic impact assessments, which involve evaluating the societal implications of AI systems before deployment. These assessments help in understanding potential risks and benefits, thereby informing better design and governance practices.[45]

An AIA typically requires the deploying authority to identify who will be affected by the system and how, assess the risk of discrimination or exclusion of different population groups, evaluate whether the intended benefits justify the potential harms, and propose corrective measures for identified risks. Crucially, it shifts the responsibility of proving safety and fairness onto the deployer rather than leaving the affected citizens to prove harm after the fact.

AIA also has another important function. By requiring public disclosure of assessment findings, they allow general society, journalists, and affected communities to scrutinize and examine AI systems and hold authorities accountable before harm occurs rather than after the system is deployed. This is a important Democratic function.

Both these measures i.e. Party audits and AIA work best in combination. The AIA is conducted before deployment, ensuring the system is designed with safety in mind. The audit happens before and after deployment ensuring the system continues to perform as intended in real world conditions and without errors or bias. Together they create a strong solution.

[42] Sandra Wachter, Brent Mittelstadt & Luciano Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the GDPR*, 7 Int'l Data Privacy L. 76 (2017).

[43] Ben Chester Cheong, *Transparency and Accountability in AI Systems: Safeguarding Wellbeing in the Age of Algorithmic Decision-Making* (2024).

[44] Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. Pa. L. Rev. 633 (2017).

[45] Dillon Reisman et al., *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability*, AI Now Institute (2018), https://ainowinstitute.org/publication/algorithmic-impact-assessments-report-2.

iii.Privacy and Data Protection:

Data protection and privacy laws are crucial for enhancing transparency and accountability in AI systems. It is also a significant challenge to enforce.

A key issue is de-identification, the practice of removing personal identifiers from datasets to avoid legal obligations. While this may satisfy the technical requirements of data protection law, AI systems can often re-identify individuals or draw sensitive attributes from supposedly anonymised data. This means deidentification provides a false sense of legal compliance without actually protecting people's privacy. These systems are capable of unredacting the redacted aspects and then form patterns using the data. These systems use all the data they can get to see the patterns and so this may lead to sometimes the data being hacked or leaked or even if the company themselves are using the data than to form a response that uses their information to unconsciously target users.[46]

Similarly, most data protection frameworks do not cover inferences that AI systems make about individuals even when those inferences are deeply personal such as predicting helth conditions, financial behaviour or political views. Generally, these inferences are not covered by legal provisions as it is not formally regulated.[47]

Though new technical processes are upcoming that are trying to solve these issues but strong laws are needed expanding the definition of personal data to include inferences and derived data.

iv.Ambiguity in legal definitions:

A important problem with AI being used everywhere today is the legal terminology being used. Because it has developed recently, the AI terminology being used across is not clear and uniform. The absence of clear definitions can lead to different interpretations in legal settings which can lead to a complicated mesh of precedents being created as these phrases play a substantial role in establishing accountability. Therefore, because of ambiguous legal terminology, the law needs to properly define phrases. Many international forums and bodies are in the process of creating AI guidelines and frameworks and defining terms, and hence the world needs to come to an agreement or follow a single interpretation of the terminology being used. AI is not just being used in a single nation, it is everywhere. One country's company produces it, another stores it, another uses it. Without a common understanding, a lot of conflicts can arise.[48]

v.Jurisdictional issues:

Another problem closely related with legal terminology is jurisdictional issues. The use and deployment of AI on worldwide scale can result in conflicts.[49] When trade between countries started it resulted in many conflicts due to jurisdictional issues. That's when GATT was created followed by WTO to open up countries and have a dispute resolution system. And though AI comes under services and hence GATS and WTO but the impact it is having and the scale of it needs a separate framework. Defining liability in AI and algorithmic decisions has proved to be an difficult affair and therefore needs strong legal frameworks that can effectively regulate AI technology , guaranteeing their responsible and ethical use.

vi.Transparency-Enhancing Technologies (TET's):

Transparency enhancing technologies are technological mechanisms designed to increase transparency, accountability, and user awareness in AI and ML systems. Their primary purpose is make the operations involved, the functions and the impacts of such systems more visible and understandable to users. It is like getting a manual on how the particular system works and making it open for parties to read it. TET's help monitor how data is collected, processed, and used while also enabling oversight and auditing of automated systems. It creates records and explanations that help examine if AI systems are operating fairly and responsibly.

Examples of TET's include algorithmic audit logs, privacy dashboards, transparency reports, model cards, data tracking systems, ethical black boxes, and bias detection tools, etc.

---

[46] Luc Rocher, Julien M. Hendrickx & Yves-Alexandre De Montjoye, *Estimating the Success of Re-Identifications in Incomplete Datasets Using Generative Models*, 10 Nature Commc'ns 1 (2019); Kai Ppackhäuser et al., *Deep Learning-Based Patient Re-Identification Is Able to Exploit the Biometric Nature of Medical Chest X-Ray Data*, 12 Sci. Rep. 14851 (2022).

[47] Sandra Wachter, Brent Mittelstadt & Luciano Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the GDPR*, 7 Int'l Data Privacy L. 76 (2017).

[48] Brent D. Mittelstadt et al., *The Ethics of Algorithms: Mapping the Debate*, 3(2) Big Data & Soc'y (2016).

[49] *Id.*

TET's are related to XAI. XAI are used to make AI systems more comprehensible and interpretable. For ex if one needs to know where a decision made by AI or algorithm came from XAI helps in telling why that decision was made and on what basis. But TET's are different they don't just tell where the decision came from but they give you all the data, logs, records, etc giving system wide transparency. TETs play a broader role than XAI which mainly concentrate on individual AI decisions.

Therefore, both TET's and XAI play an important role in promoting accountability and transparency.[50]

vii. Enforcing regulations:

If anything has to be enforced than laws have to be made for it. If transparency and accountability have to be enforced than there must be provisions and regulations made to ensure that all companies developing these algorithms are obligated to provide details of their systems and their methodology.[51]

Another thing is establishing standards and certifications. Guidelines can be set that; systems need to adhere to particular set of standards before their implementation. This can be done through examinations and validation of AI systems that establishes conformity to standards. Therefore, if the system meets standards than it is allowed or otherwise not. This will also prove to develop trust among users.[52]

viii. Regulatory sandboxes:

As mentioned above sandboxes are a brilliant way of piloting an algorithmic system before it is deployed. Regulatory sandboxes involve setting up these special zones where AI and ML systems can be tested without companies having to fear about adhering to all stringent regulations like strict liability. Public sector has started using AI and algorithms everywhere and it can lead to serious risks. High risk sectors like policing, transport, recruitment have already implemented these systems and there also have been cases which shows how detrimental it can be. So, if any new technology that has the direct power to affect and in pact so many people is going to be used than it must be first Beta tested. This enables the detection and reduction of possible dangers prior to the deployment.[53]

These sandboxes have several advantages. They promote innovation, technological breakthroughs, empower them to rival bigger corporations, etc.

But the biggest reason Sandboxes can be really helpful is because it lets us put a barrier. Countries usually have a tendency to compete with each other and race each other. Now they are competing for AI, and even though all jurisdictions are still in the phase of developing and implementing laws for AI and Ml or automated decisions, but their deployment is another story. AI are being deployed left, right and centre and many have had to be shut down too for being biased, full of errors and propagating discrimination. And hence it must be made as a mandate to first test AI in a limited safe zone area and test its effectiveness and then rectify the errors and then roll it out to rest of the country. Fast is not always better especially without proper safeguards.

ix. Documentation:

Maintaining detailed documentation of AI systems, including their design, data sources, and decision-making processes, to facilitate transparency and accountability is also a step to be made compulsory. Documentation of all the important aspects helps to take care of errors later.[54]

x. Interdisciplinary collaboration:

Interdisciplinary collaborations include legal specialists, technologists/developers, ethicists working together to create an ecosystem where the algorithm is developed taking in all considerations. Legal professionals guide on legality and liability concerns, developers take care of the technical aspects and ethicists provide guidance on developing systems that adhere to ethical standards including privacy, sustainability, etc...[55]

This is an important step because normally developers and companies only think about technical parts and taking in profits because they are only experts in those fields but for a technological platform that has high

[50] Nick Bostrom & Eliezer Yudkowsky, *The Ethics of Artificial Intelligence*, in The Cambridge Handbook of Artificial Intelligence 316 (Nick Ramsey & Murray Shanahan eds., Cambridge Univ. Press 2014).

[51] Luciano Floridi & Josh Cowls, *A Unified Framework of Five Principles for AI in Society*, 1(1) Harvard Data Sci. Rev. (2019).

[52] Iyad Rahwan, *Society-in-the-Loop: Programming the Algorithmic Social Contract*, 20 Ethics & Info. Tech. 5 (2018).

[53] Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 Harv. J.L. & Tech. 353 (2016).

[54] Ben Chester Cheong, *Transparency and Accountability in AI Systems: Safeguarding Wellbeing in the Age of Algorithmic Decision-Making* (2024).

[55] *Id.*

impact it is important to look from all sides of the spectrum and interdisciplinary collaborations provide that. This will ensure that the systems are better equipped for the age of tomorrow.

xi.Ombudsman:

A grievance redressal body which will be ombudsman should be set up to take all complaints. An AI ombudsman is an independent grievance redressal authority that provides affected individuals a formal mechanism to challenge such decisions and seek review.

The main role of ombudsman is to investigate whether an AI system operated fairly, transparently and in compliance with legal and ethical standards.it examines whether there is\was any bias, violations, discriminatory practices, etc. it requires the companies to provide records, audit reports, explanations and other relevant information. Similar to how ombudsman are established for insurance, banking and other sectors even here the ombudsman will be accessible and provide fast paced relief to the complainant. Ombudsman if implemented correctly work brilliantly to improve public trust and solve ground problems instead of directly overburdening court system.

xii.Human in the loop:

This refers to the practice where a human being is involved in AI decision making particularly in high stakes decision. Under this, instead of an AI having power to directly take decisions, human will be involved to have the final say. When a human is involved in the loop it can look at errors made by the system and change the decision or rectify it. If humans are involved it will create an extra step where the AI cannot directly give the decisions.

Human in the loop does not mean that humans will just approve the decisions without them using their judgements it requires genuine human judgement at decision stage. Humans can look at the decision AI takes and can see and cross check if it is valid and verified. One way of ensuring these is placing some burden on the officer like making him accountable if the outcome causes harm. But this can be only applied in cases where the AI is not taking 1000s of decisions every day. Only the cases which have high priority and affects the individuals can be revived by such officers.

12. CONCLUSION

Indias public sector has embraced artificial intelligence with remarkable speed and scale across different areas such as welfare delivery, law enforcement, taxation, traffic management, and financial systems. This paper has demonstrated that while the advantages of these AI systems are many but the impacts are also significant. The opaqueness of the system and its inherent flaws have to be addressed.

There is a legal vacuum in India currently as it has no comprehensive legislation and framework to properly govern the deployment and establish various liabilities. The DPDP act 2023, though a step in the right direction, does not address many concerns. It talks about data privacy, but is silent on algorithmic accountability and liability.

This paper argues that addressing this gap requires action on multiple fronts simultaneously. Legally, India must enact a dedicated AI governance statute, expand data protection definitions, and establish an independent AI Ombudsperson. Technically, Explainable AI, Transparency-Enhancing Technologies, and regulatory sandboxes must be adopted as requirements. Institutionally, mandatory algorithmic impact assessments, independent third-party audits, and human-in-the-loop mandates must be built into every high-stakes AI deployment. And comparatively, India must draw carefully from global frameworks.

AI will not slow down. Its integration into governance is irreversible. The question is not whether India will use AI to govern it already does. The question is whether it will govern AI. Handled correctly, these systems can deepen inclusion, reduce corruption, and serve millions more effectively. Handled carelessly, they will institutionalize bias, opacity, and exclusion at a scale no human bureaucracy ever could. India's constitutional commitment to equality, dignity, and due process demands that it get this right.

13. ACKNOWLEDGEMENT

I would like to express my sincere gratitude to all those who supported and guided me during the completion of this research paper titled “Accountability and Transparency in Algorithmic Decision-Making: Examining the Use of AI and Machine Learning in India’s Public Sector.”

I extend my heartfelt thanks to Dr. Uma Shinde, Professor, Department of Law, JSPM University, for her valuable guidance, continuous support, constructive suggestions, and encouragement throughout the course of this research. Her insights and academic direction played an important role in shaping this study.

I am also grateful to JSPM University, Pune, for providing the academic environment and resources necessary to undertake this research work.

I would like to acknowledge the contribution of various authors, researchers, institutions, reports, judicial decisions, and scholarly publications whose work formed the foundation of this study and greatly assisted in developing a broader understanding of accountability, transparency, and AI governance.

Finally, I express my sincere appreciation to my family and well-wishers for their constant encouragement and support throughout the preparation of this paper.

14. REFERENCES:

1. Bostrom, Nick & Eliezer Yudkowsky. *The Ethics of Artificial Intelligence*. In *The Cambridge Handbook of Artificial Intelligence*, edited by Nick Ramsey & Murray Shanahan. Cambridge University Press, 2014.
2. Cheong, Ben Chester. *Transparency and Accountability in AI Systems: Safeguarding Wellbeing in the Age of Algorithmic Decision-Making*. 2024.
3. Floridi, Luciano & Josh Cowls. *A Unified Framework of Five Principles for AI in Society*. 1(1) Harvard Data Science Review (2019).
4. Kroll, Joshua A. et al. *Accountable Algorithms*. 165 University of Pennsylvania Law Review 633 (2017).
5. Marda, Vidushi & Shivangi Narayan. *Data in New Delhi's Predictive Policing System*. 2020.
6. Mittelstadt, Brent D., Patrick Allo, Mariarosaria Taddeo, Sandra Wachter & Luciano Floridi. *The Ethics of Algorithms: Mapping the Debate*. 3(2) Big Data & Society (2016).
7. Nagda, Priyadarshi. *Legal Liability and Accountability in AI Decision Making: Challenges and Solutions*. 2025.
8. Olorunniwo, O.E., Mayowa Alomge & Olatunji Isreal. *Accountability in AI Decision-Making*. 2025.
9. Orozco, Brenda Guadalupe. *Artificial Intelligence in Public Sector Decision-Making: Accountability, Transparency and Ethical Governance*. 2025.
10. Packhäuser, Kai et al. *Deep Learning-Based Patient Re-Identification Is Able to Exploit the Biometric Nature of Medical Chest X-Ray Data*. 12 Scientific Reports 14851 (2022).
11. Parisineni, S.R.A. & M. Pal. *Enhancing Trust and Interpretability of Complex Machine Learning Models Using Local Interpretable Model Agnostic SHAP Explanations*. 23 International Journal of Data Science and Analytics 458 (2023).
12. Rahwan, Iyad. *Society-in-the-Loop: Programming the Algorithmic Social Contract*. 20 Ethics and Information Technology 5 (2018).
13. Reisman, Dillon, Jason Schultz, Kate Crawford & Meredith Whittaker. *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability*. AI Now Institute (2018). https://ainowinstitute.org/publication/algorithmic-impact-assessments-report-2.
14. Ribeiro, Marco Tulio, Sameer Singh & Carlos Guestrin. *"Why Should I Trust You?" Explaining the Predictions of Any Classifier*. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 1135 (2016).
15. Rocher, Luc, Julien M. Hendrickx & Yves-Alexandre De Montjoye. *Estimating the Success of Re-Identifications in Incomplete Datasets Using Generative Models*. 10 Nature Communications 1 (2019).
16. Scherer, Matthew U. *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*. 29 Harvard Journal of Law & Technology 353 (2016).
17. Wachter, Sandra, Brent Mittelstadt & Luciano Floridi. *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the GDPR*. 7 International Data Privacy Law 76 (2017).
18. Artificial Intelligence Act, Regulation (EU) 2024/1689, 2024 O.J. (L 1689).
19. General Data Protection Regulation, Regulation (EU) 2016/679, 2016 O.J. (L 119) 1.
20. *Houston Fed'n of Teachers v. Houston Indep. Sch. Dist.*, 251 F. Supp. 3d 1168 (S.D. Tex. 2017).
21. Amnesty International. *Dutch Childcare Benefit Scandal an Urgent Wake-Up Call to Ban Racist Algorithms*. Amnesty Int'l (Oct. 25, 2021). https://www.amnesty.org/en/latest/news/2021/10/xenophobic-machines-dutch-child-benefit-scandal/.(last visited May 31, 2026).
22. *AI Governance Lessons from the Dutch Childcare Benefits Scandal*. AICC Responsible AI Hub. https://www.aicc.co/responsible-ai-hub/keep-up-to-date-with-responsible-ai/blog-posts/responsible-ai-lessons-from-the-dutch-childcare-benefits-scandal (last visited May 31, 2026).

23. *AI in Indian Traffic Management: Transforming Urban Mobility Challenges*. IndiaAI. https://indiaai.gov.in/article/ai-in-indian-traffic-management-transforming-urban-mobility-challenges (last visited May 31, 2026).
24. *Criminal Investigation: AI-Based Human Efface Detection System in Rajasthan*. YourStory (June 2017). https://yourstory.com/2017/06/criminal-investigation. (last visited May 31, 2026)
25. *Delhi Police Implements Crime Mapping, Analytics & Predictive System (CMAPS)*. BW Police World. https://www.bwpoliceworld.com/article/delhi-police-implements-crime-mapping-analytics-predictive-system-cmaps-134283 (last visited May 31, 2026).
26. *DPIP: RBI's AI Tool to Detect Fraud in Real Time*. Economic Times. https://economictimes.indiatimes.com/news/economy/policy/dpip-rbis-ai-tool-to-detect-fraud-in-real-time/articleshow/124536025.cms (last visited May 31, 2026).
27. *Goa Police's AI-Powered Deep Trace to Probe Deep, Dark Crimes*. Times of India. https://timesofindia.indiatimes.com/city/goa/goa-polices-ai-powered-deep-trace-to-probe-deep-dark-crimes/articleshow/120833963.cms (last visited May 31, 2026).
28. *Maharashtra Cabinet Approves ₹23 Cr MARVEL AI Project for Police Force*. Times of India. https://timesofindia.indiatimes.com/city/mumbai/maharashtra-cabinet-approves-23cr-marvel-ai-project-for-police-force/articleshow/108554366.cms (last visited May 31, 2026).
29. *Meet CrimeGPT: The UP Police Department's AI Tool for Fighting Crime*. IndiaAI. https://indiaai.gov.in/article/meet-crime-gpt-the-up-police-department-s-ai-tool-for-fighting-crime (last visited May 31, 2026).
30. Nair, Sthanu R. & Ranjith Gundaboina. *AI-Powered Tax Governance in India and Its Challenges*. The Hindu (Feb. 18, 2026). https://www.thehindu.com/opinion/op-ed/ai-powered-tax-governance-in-india-and-its-challenges/article70762627.ece. (last visited May 31, 2026).
31. Press Information Bureau. *Home Minister's Statement on Delhi Riots*. PIB (Mar.11,2020).https://www.pib.gov.in/PressReleasePage.aspx?PRID=1606194&reg=3&lang=2. (last visited May 31, 2026).
32. *RBI Developing AI-Driven Tool to Flag Fraudulent Transactions*. New Indian Express(Oct.1,2025). https://www.newindianexpress.com/business/2025/Oct/01/rbi-developing-ai-driven-tool-to-flag-fraudulent-transactions. (last visited May 31, 2026).
33. Savyasachi, Astha. *As AI Took Over Policing in Delhi, Who Bore the Brunt?* The Wire (July 2, 2025). https://pulitzercenter.org/stories/ai-took-over-policing-delhi-who-bore-brunt. (last visited May 31, 2026).
34. Shrivastava, Kumar Sambhav, Tapasya & Divij Joshi. *How an Algorithm Denied Food to Thousands of Poor in India's Telangana*. Al Jazeera (Jan. 24, 2024). https://www.aljazeera.com/economy/2024/1/24/how-an-algorithm-denied-food-to-thousands-of-poor-in-indias-telangana. (last visited May 31, 2026).
35. Shrivastava, Kumar Sambhav, Tapasya & Divij Joshi. *In India, an Algorithm Declares Them Dead; They Have to Prove They're Alive*. Al Jazeera (Jan. 25, 2024). https://www.aljazeera.com/economy/2024/1/25/in-india-an-algorithm-declares-them-dead-they-have-to-prove-theyre. (last visited May 31, 2026).
36. *Soon Indore Will Be Under Cam Eye*. Hindustan Times. https://www.hindustantimes.com/indore/soon-indore-will-be-under-cam-eye/story-3HwFq5SKi8SBbaIkfsbgJN.html (last visited May 31, 2026).
37. *TSCOP App to Use Facial Recognition*. New Indian Express (Aug. 3, 2018). https://www.newindianexpress.com/cities/hyderabad/2018/Aug/03/tscop-app-to-use-facial-recognition-1852485.html. (last visited May 31, 2026).
38. *25k Wrong Challans: Tech Glitch Drives Hyd Motorists Up the Wall*. Times of India (Nov. 9, 2025). https://timesofindia.indiatimes.com/city/hyderabad/25k-wrong-challans-tech-glitch-drives-hyd-motorists-up-the-wall/articleshow/125191689.cms. (last visited May 31, 2026).
39. *UP Government Launches AI-Based App YAKSH to Strengthen Policing in State*. NDTV. https://www.ndtv.com/india-news/up-government-launches-ai-based-app-yaksh-to-strengthen-policing-in-state-10040969 (last visited May 31, 2026).
40. *UP Police Revolutionizes Policing with Trinetra App 2.0, Enhancing Crime Prevention and Investigation*. Times of India (Feb. 2024). https://timesofindia.indiatimes.com/city/lucknow/up-police-revolutionizes-policing-with-trinetra-app-20-enhancing-crime-prevention-and-investigation/articleshow/108537077.cms. (last visited May 31, 2026).