DEMYSTIFYING THE BLACK BOX: AI, ALGORITHMIC AND REGULATORY EVOLUTION

Mr. Saumitra Sharma, Head of Law Department, United University, Prayagraj

Bhumika Agrawal, LL.M., United University, Prayagraj

ABSTRACT

With risk assessment, recidivism forecasting, and punishment prediction, the use of AI in court processes via predictive justice technologies is transforming the decision-making process for the effectiveness and uniformity of the legal system. Predictive justice is the use of AI-enabled technology to analyze vast amounts of data in order to forecast the results of court conflicts. However, since AI is a black box technology, its increasing use in supporting court rulings is also posing serious questions about accountability, transparency, and justice. Our incapacity to comprehend how deep learning algorithms make judgments is known as the black box dilemma. The judicial system is susceptible to systemic prejudices sustained by built-in AI algorithms due to the opacity of these technologies, which creates significant ethical and legal issues such biases, discrimination, and accountability gaps in different AIassisted court rulings. The loophole is that such algorithms are not being sufficiently regulated. This paper reveals all the challenges and problems concerning a black box nature of the technologies in the AI-based judicial system. Solutions to technology problems encouraged in this article include algorithm auditing standards to detect bias, explainable AI (XAI) to ensure transparency, and liability frameworks to hold accountable. To suggest the introduction of international best practices into our legal system, the study also evaluates several global systems. The article views technical innovation as ethically governed in a broader perspective and provides options based on a policy approach to align AI integration to the legal system.

Keywords: Artificial Intelligence, Legal System, Blackbox Dilemma, Integration, Accountability.

INTRODUCTION

In an AI-powered judicial system, the Black Box dilemma poses a serious threat to accountability, justice, and openness. AI systems may be made more transparent and less biased by using strategies like Explainable AI (XAI), algorithmic audits, and stringent laws. Strong ethical and legal underpinnings for AI use in the judicial system will be ensured by a well-defined set of legislation. In an effort to reduce case backlogs and advance justice, courts across a range of legal systems are using AI techniques to help judges forecast penalties and recidivism.

The following question is still relevant as artificial intelligence continues to influence how the legal system develops in the future: Is it feasible to have faith in a system that defies comprehension? In order to guarantee that AI enhances rather than diminishes the intrinsic justice and accountability of judicial adjudication, this study aims to provide policy-informed suggestions.

UNDERSTANDING THE AI AND BLACK BOX DILEMMA

There are many formal definitions of artificial intelligence systems that came from national legislatures and independent organizations. The European Union Artificial Intelligence ACT 2024 (EU AI Act 2024) defines an AI system as "a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and 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."[1]

The two main components of this description are infers and autonomy, which distinguish an AI system from conventional software in which a rigid algorithm predetermines the result (if a then b). For instance, a simple voice command system simply reacts to pre-programmed words, but a virtual assistant like Siri or Alexa deduces human intent and becomes better with time. By using a wide and technology-neutral approach, the definition guarantees that the Act will continue to be relevant as AI advances.[2]

[1] European Union Artificial Intelligence Act (EU AI Act), 2023,Art.3(1)
[2] The EU Artificial Intelligence Act: our 16 key takeaways. Retrieved from https://www.stibbe.com/ publications-and-insights/the-eu-artificial-intelligence-act-our-16-key-takeaways(visited on 6th February, 2025)

AI that only assesses potential outcomes is not the same as AI that learns from data. Examine the following fictitious computer software that is intended to screen a résumé for a position. The program's job is to evaluate the applicant's resume and make choices based on the applicant's credentials, background, and particular abilities. To select individuals, the conventional AI or rigorous algorithm will adhere to predetermined guidelines. For instance

(i) A candidate passes if they have at least three years of experience.

(ii) They pass if they have a master's degree.

(iii) A candidate passes if their CV contains certain keywords like Python or Project Management.

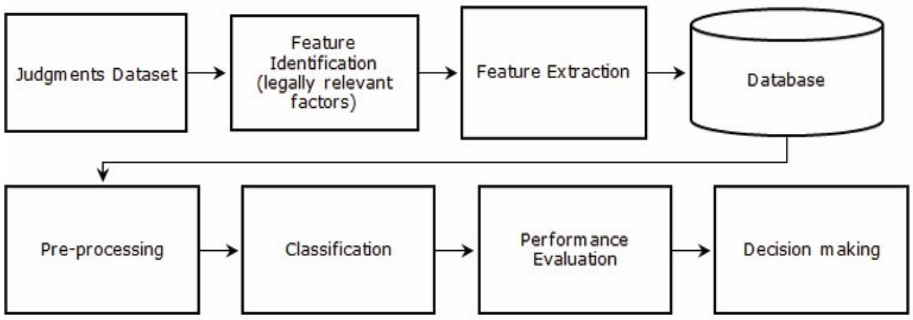
HOW DO MACHINE LEARNING ALGORITHMS WORK

Step 1: Collection of Information. The first step is the collection of the relevant information. The information in this case can include the criminal history of the accused, such as age, education level, criminal record and prior recidivism convictions. For instance: a) The age of defendant 'A' is 25, he is employed, has high school education, has two prior convictions, and has already committed another crime during the last two years. b) Defendant 'B': 40 years old, unemployed, college degree, has no prior convictions, and no recidivism. This data is used to train the machine learning system to identify trends related to recidivism.

Step 2: Data Pre-Processing: A raw data is usually not consistent or even complete and, therefore, it needs to be cleansed and converted into a useful form of data. As an illustration, missing values or categorical data, e.g. criminal record or education level, have to be converted to numerical values. This, besides making the data ready to be analyzed, introduces another level of complexity such that the final judgment is beyond trace.

Step 3: Model Training: The algorithm is trained on the pre-processed data to identify patterns and correlations. An example of this is that the model might find that individuals with low educational attainment and those who have been previously convicted of crimes have higher chances of returning to crime. However, the model's decision-making process is hard to understand since the training procedure involves millions of computations by analyzing the vast number of such previous situations.

Step 4: Model Evaluation: The performance of the trained model is next assessed using a different dataset. The model's accuracy for the new dataset is evaluated using several measures.



WHAT IS THE BLACK BOX DILEMMA

The Blackbox Dilemma is often described as our incapacity to completely comprehend an AI's decision-making process and our incapacity to forecast its choices or outputs. Consider a magic trick where the technique is concealed but the beginning and end are visible. This is the core of the Black Box issue. The degree to which humans can comprehend or audit AI systems decision-making processes depends on their lack of openness.

The AI system learns from data, as was covered in the preceding section. It finds links, correlations, and patterns in large datasets. It makes judgments or forecasts based on this knowledge. With the use of algorithms and input from the data it analyzes, an AI system modifies its settings. These deep learning model-based systems function at a level of complexity that makes it challenging to understand the reasoning behind their decisions.

The purpose of this program is to ensure that legal information is easily comprehensible by the non-native English speakers as India is a multilingual state. The translations of AI are prone to misunderstanding legal terms, altering the meaning of legal decisions, or even failure to work at all due to reliance on natural language processing (NLP) methods. Without independent verification of the legal documents that are translated, some ambiguity and incorrect interpretation of the law could be evident. Courts use CMS, an artificial intelligence platform, to handle pending cases, schedule hearings, and automate case submissions. It may shorten the time it takes to process cases and increase the efficiency of Indian courts. AIbased case prioritizing and scheduling automation may lead to inadvertent bias in case management and incomprehensible decision-making. The idea of equal access to justice may be jeopardized when AI systems prioritize certain kinds of cases that cannot be explained.

It uses AI algorithms to monitor enormous amounts of data, allowing authorities to track down people connected to terrorism or criminal activity. Because people may be watched and judged without the court's consent, such systems violate privacy. People are concerned about improper monitoring, data privacy, and violations of human rights due to the imprecise way AI identifies security threats.

The Crime and crime Tracking Network & System, or CCTNS, is a single database that uses artificial intelligence (AI) to monitor and preserve crime records, automate police reports, and promote interagency cooperation. CCTNS may reinforce preexisting prejudices in the legal system since it uses historical crime data to determine risk. AI algorithms may lead to erroneous arrests, biased policing, and a decline in public confidence in the police if they unfairly classify certain areas as high-risk based on faulty data.

The National Crime Records Bureau uses the Automatic Facial Recognition System to help police trace offenders using picture databases and CCTV cameras. The system often makes errors, particularly when it comes to recognizing members of the disadvantaged population. False arrests and privacy breaches have been caused by AI facial recognition errors. Furthermore, there is no established policy regarding the storage, access, or use of the data. Trust is necessary, but accountability is also necessary. It will be challenging to identify the bias and determine who should be held responsible if an AI system makes an unfair or biased decision. It is challenging to guarantee equality and prevent prejudice when there is a lack of openness. The Indian judicial system would be significantly impacted by the lack of accountability and legislation governing AI.

REVIEW OF LITERATURE

In order to improve accountability without jeopardizing the AI creators' intellectual property rights, the study also proposes the use of counterfactual explanations.

"The Black Box Society: The Secret Algorithms That Control Money and Information" by Frank Pasquale (2015) In this book, the author emphasizes how AI decision-making systems' opacity restricts openness in industries including government, healthcare, and finance. According to the author, businesses and governmental organizations may make significant choices without being held accountable because to the Black Box of AI. The research also highlights the risks of opaque algorithms influencing social and economic results, but it

provides very little insight into how these systems make judgments. The author cautions that if these AI systems are let to function unchecked in judicial systems, it may result in unjust and unchallengeable rulings that go against the rule of law.

"Harnessing the Power of Artificial Intelligence in the Indian Justice System: An Empirical Study" by Janees Rafiq (2024) This paper's author investigates AI's possibilities within the framework of Indian law. In order to shorten the total length of trials, the study assesses how AI can assist judges in making decisions, facilitate the efforts of lawyers, and address such problems as a case backlog and delays in the process. The author also analyzed the applicability of existing AI applications to enhancing judicial productivity and transparency in the Indian legal system. One of the other points highlighted in the report is that proper legal frameworks and ethical issues should be in place to govern the use of AI in court proceedings.

Sahil Girhepuje (2023) published the article Are Models Trained on Indian Legal Data Fair? In the article, the authors discuss the bias and fairness of AI models trained on Indian legal data.

EMERGENCE OF GOVERNING LEGISLATIONS AI: AN INTERNATIONAL PERSPECTIVE

"*The development of full artificial intelligence could spell the end of the human race.*"

–Stephen Hawking.

Otherwise, the prevalence of AI technologies in terms of efficiency and efficiency in decision-making by a diverse range of industries, such as the world judicial system, is fraught with severe dangers. The Black Box problem of AI systems raises a number of concerns, such as as bias, accountability, and ethical compliance. It has also influenced nations and international organizations to develop legal frameworks to regulate the development, implementation, and use of AI systems to resolve ethical, legal, and sociological concerns posed by these systems.

European Union Artificial Intelligence Act 2024 (EU AI ACT 2024): The European Union Artificial Intelligence Act is the first legislative document on artificial intelligence that regulates the risks of this technology posed by AI. The Act is aimed at ensuring the moral and humanistic evolution of artificial intelligence. The Act defines four categories of dangers of AI

systems that require the different degrees of regulation[3].

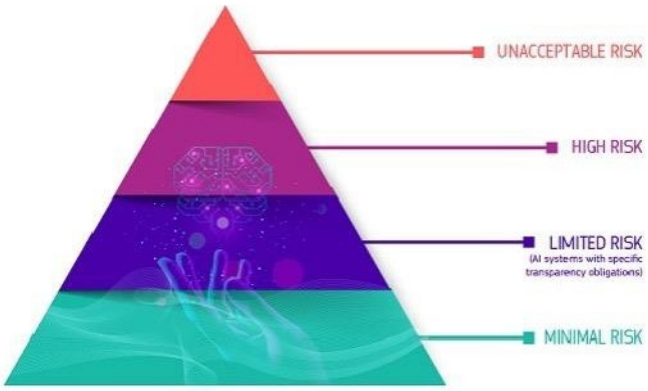


Figure 2: Levels of Risk

European Union Artificial Intelligence Act 2024 (EU AI ACT 2024): The European Union Artificial Intelligence Act is the first legislative document on artificial intelligence that regulates the risks of this technology posed by AI. The Act is aimed at ensuring the moral and humanistic evolution of artificial intelligence. The Act defines four categories of dangers of AI systems that require the different degrees of regulation. Self-driving cars with AI in the transportation sector, to the point, take decisions in a fraction of a second to protect the safety of its occupants. The medical sphere is one of the areas where AI can be used in surgery and diagnostics and a small mistake can have terrifying consequences. The educational system can apply AI to assess the performance of students and mark the tests. The AI system should not be biased and should not have any prejudice, and human control should ensure equitable outcomes. Unacceptable Risk: The European Union does not allow AI systems that threaten the rights, livelihoods, and safety of people. Under the AI Act, AI systems that perform the following activities are not allowed.

• Dangerous deceit and manipulation;

• Negative exploitation of weaknesses;

[3] Savio Jacob, "AI Regulations Around the World: A Comprehensive Guide To Governing Artificial Intelligence"(2024), Retrieved fromhttps://www.spiceworks.com/tech/artificial-intelligence/articles/ ai-regulations-around-the-world

• Social Evaluation

• Evaluation or forecasting of the danger of individual criminal offenses;

• The ability to identify emotions in both educational and professional settings.

UNITED STATES: ALGORITHMIC ACCOUNTABILITY ACT 2023 & AI BILL OF RIGHTS 2022

One of the proposed regulations is Algorithmic Accountability, which requires companies that apply AI to various sectors including as banking, healthcare, and legal systems to disclose their algorithms and ensure that their AI applications are unbiased and non-discriminatory. However, since this Act does not impose required transparency requirements and lacks stringent enforcement tools, compliance with it is mostly optional.[4]

The design, application, and implementation of AI systems are guided by these five fundamental principles:

• Robust and secure AI systems;

• Protections against algorithmic discrimination;

• Privacy of Data

• AI systems' explainability.

• Human supervision[5].

CHINA: STATE-CONTROLLED RULES FOR AI GOVERNANCE

China has taken the lead in developing some of the first legislation pertaining to artificial intelligence. China released a complete set of regulations pertaining to generative intelligence services, deepfakes, and algorithms in 2021. Additionally, it establishes a license system for

[4] National Institute of Standards and Technology (NIST),'AI Risk Management Framework'(2023) https://www.nist.gov/, accessed 11 February 2025.
[5] James Gong, "Ai Governance in China: Strategies, Initiatives, and Key Considerations", (Bird&Bird, 2024) Retrieved fromhttps://www.twobirds.com/en/insights/2024/china/ai-governance-in-china-strategiesinitiatives-and-key-considerations

the introduction of generative AI services. Before being implemented, AI systems in China must comply with national regulations and undergo a number of examinations and clearances. This centralized regulation guarantees that AI systems must adhere to the political and ideological principles of the state, which may restrict innovation and transparency in a number of areas, including the legal system.

OECD:

The first intergovernmental standards on AI, OECD AI principles set forward a number of criteria for the responsible development and use of AI technology[6]. The ethical use of these technologies, as well as the accountability and openness of AI systems, are the main concerns of the governance framework.

These guidelines support the use of creative, reliable AI systems that respect democratic norms and human rights. Among these values-based precepts are:

• Human-centered principles and inclusivity;

• Explainability, Fairness, and Transparency;

• Sturdiness and security; responsibility.[7]

It also offers suggestions for politicians to fund collaboration for reliable AI, and risk-based regulation of AI systems. Although these guidelines are not legally obligatory, they provide useful foundations that may influence global policy development.[8]

EMERGENCE OF LEGISLATION GOVERNING IN INDIA

India currently lacks a specific AI legislation, and the country's current legal system mostly addresses criminal law, cybersecurity, digital data protection, and electronic evidence. AIrelated topics are indirectly regulated by a number of laws, including those pertaining to

[6] OECD, 'OECD Principles on Artificial Intelligence'(OECD, 2019)https://www.oecd.org/going-digital/ai/principles/, accessed 12 February 2025.
[7] OECD, 'Recommendation of the Council on Artificial Intelligence' (2019) https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 accessed 12 February 2025.
[8] OECD,'Governance of Artificial Intelligence: Who is Responsible?' (2021) https://www.oecd.ai/en/governanceaccessed 12 February 2025.

data protection, digital forensics, surveillance, and electronic recordkeeping.

UNPRECEDENTED ISSUES AND CHALLENGES POSED BY THE GROWTH OF AI AND BLACK BOX

Concerns about accountability, justice, openness, and due process are raised by the incorporation of artificial intelligence into the judicial system. When used in the legal system, the Black Box aspect of AI technologies poses significant problems despite their efficiency and data-driven insights. Regulators find it difficult to understand how many AI applications, such as automated sentencing models, risk assessment tools, and predictive policing, operate. AI models eliminate public scrutiny while concentrating authority on individuals who create the algorithms.[9]. These hazards increase in the absence of standardized legal guidelines for AI-assisted legal judgments. Courts lack particular legislation to ensure AI tool transparency and equality in legal proceedings, despite the fact that they have begun to take into account AI-related concerns such biases in face recognition systems and inexplicable algorithmic choices. Because AI systems are unable to justify their activities, the Black Box poses a serious danger to basic rights like due process and the right to a fair trial in legal systems.

"We are conscious of the fact that artificial intelligence has a flip side. For instance, it would be very difficult and difficult for us to allow artificial intelligence to determine conviction in criminalcases. I don't think that we, as judges, would ever want to see the discretion we exercise, based on sound judicial principles, being overtaken by AI in sentencing policy."

This claim emphasizes the judiciary's hesitation to fully depend on AI since it lacks human reasoning, accountability, and openness. The important concerns raised by AI in the legal system, especially in the Indian context, will be examined in this section.

Lack of Accountability and Transparency: One fundamental issue with AI systems is that their algorithms are opaque. These are almost often sophisticated machine learning models that have been developed on large, challenging-to-understand data sets. Most AI-based decisions in criminal justice remain Black Box due to the absence of transparency to the way these algorithms arrive at the end-result. The artificial intelligence systems are made in such a way that defendants, judges, and other legal professionals cannot see or challenge the reasoning of

[9] Frank Pasquale,"The Black Box Society: The Secret Algorithms That Control Money and Information" (Harvard University Press 2015)

important judgments, including refusal of bails, sentence proposals, and parole decisions.

The worst part about these algorithms is that they do not make sense of their projected results in a manner that can be understood by a human being. In the traditional juridical decision-making process, judges normally consider various factors, which may include a set of facts, evidence, and legal provisions. Instead, AI systems operate based on computer algorithms that apply to large datasets without a general understanding of the logic of specific outputs. An example is that, when a predictive model is utilized to determine whether an individual would commit another crime, the system will give a risk score. But that score can be pegged on a variety of factors, some of which might be hard to elucidate even by the creator of the system. A deficiency in transparency in decision making kills public trust in the justice system which should deliver justice and equality to everyone. Individuals begin to question the validity of the legal system when they cannot understand the logic of some decisions or fight them successfully.

ALGORITHMIC BIAS AND DISCRIMINATION

AI systems are trained using a variety of algorithms and learn based on the data to distinguish patterns. Since the various groups of engineers have very different biases and assumptions to the development of algorithms and the datasets, they are in turn susceptible to both algorithms biases as well as, in most instances, databases biases. different legal AI systems use different algorithms and, in most cases, different datasets. Therefore, diverse legal AI systems are expected to lead to different outcomes. The systematic and unfair discrimination of the algorithms can further realign the existing injustices in society. These algorithms predict outcomes based on the historical data that might introduce bias between racial, gender, socioeconomic differences, leading to unintended bias and compromising the fairness of the justice system even more.

CONCLUSION

Though the introduction of AI to the judicial system presents efficiency, accountability, transparency, and prejudice are questioned. The reliance on general laws, such as the IT Act and DPDP Act and BNSS, which leave AI-based judicial tools unchecked, brings up due process and justice issues in India. It requires a robust legal structure as the options made with the help of AI are not explicable, which threatens fundamental rights. Close liability regimes,

audits of AI algorithms, and the creation of an AI regulatory body are needed to guarantee accountability and a solution to AI-related legal mistakes. AI should be upheld as an instrument, and judicial control should assure just and unbiased justice. In order to implement AI usages in accordance with ethical and constitutional standards of the legal framework in India, urgent changes in legislation are required.[10]

[10] Data Protection and Digital Privacy (DPDP)Act, 2023, (NO. 22 OF 2023).

REFERENCES

- European Union, Artificial Intelligence Act 2024 (EUAIAct2024).
- Frank Pasquale, The Black Box Society: The Secret Algorithms That Control Money and Information (Harvard University Press 2015).
- Government of India, Digital Personal Data Protection Act 2023.
- https://bigbrotherwatch.org.uk/
- https://internetfreedom.in/automated-facial-recognition-report/
- https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facialrecognition.html/
- https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
- https://www.spiceworks.com/tech/artificial-intelligence/articles/ai-regulations-around-theworld/
- https://www.technologyreview.com/2021/01/07/1016221/whats-wrong-with-ai-crimeprediction-everything/
- Justice KS Puttaswamy (Retd) v Union of India (2017)10 SCC 1.
- National Crime Record Bureau, Automated Facial Recognition System: Implementation Report (2021).
- Organisation for Economic Co-operation and Development (OECD), OECD AI Principles (2019).
- R(Bridges) v South Wales Police [2020] EWCACiv1058.
- Sandra Wachter, Brent Mittelstadt, and Chris Russell, 'Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR' (2018) 31 Harvard JL & Tech 841.
- State v Loomis,881NW2d 749(Wis2016).

- United Nations Educational ,Scientific and Cultural Organisation (UNESCO), Recommendation on the Ethics of Artificial Intelligence (2021).
- US Congress, Algorithmic Accountability Act 2023.
- White House Office of Science and Technology Policy ,AI Bill of Rights 2022.
- Yavar Bathaee, 'The Artificial Intelligence Black Box and the Failure of Intent and Causation' (2018) 31 Harvard JL & Tech 889.