AI ALGORITHMS IN SENTENCE DETERMINATION: JUSTICE, TRANSPARENCY AND ACCOUNTABILITY

Abstract

Technology has now become visible everywhere, engulfing all the institutions, including the judiciary, in many parts of the world. Humanity is embracing it in a manner that was not seen before. The march of artificial intelligence (AI) is one such scientific miracle that has come with excitement but unknown fear as well. The application of AI algorithms and machine learning tools will certainly assist the judges in conducting smooth legal research for their judgments and complex orders. This interesting setting is a combination of traditional and modern aspects, which prominently highlights the interrelation of law and technology. In this paper, we will explore different nuances of AI application in the administration of the criminal justice system in general and the sentencing process in particular, with specific reference to the United States, and will also discuss the Loomis (2016) judgment where for the first time COMPAS algorithm was used in sentence determination. We will also inquire as to how we can make the best use of AI algorithms in the sentencing process.

I Introduction

BIG DATA has undoubtedly emerged as the most essential feature of the modern world. Data streams orbit the globe at the speed of light, digital algorithms create virtual realities, and intelligent robots operate factories says Thomas Fuchs. He further says that artificial intelligence is beginning to learn and is already surpassing the first achievements of human intelligence. The subjective mind itself appears in the end only as a sum of algorithms as it is no longer bound to the earthly body.[1]

Once a matter of science fiction, it now permeates all aspects of our lives, including the criminal process. The surge in experiencing the increasing use of algorithms in judicial decision-making and the administration of justice around the world.

The scientific history is replete with examples that scientific inventions and discoveries have benefited mankind immensely, but initially, such scientific inventions and discoveries have also created suspicion, doubts, and unknown fear in different sections of the society, which fades away only with the passage of time as well as unfolding of benefits of all such discoveries and inventions.However, scientific history makes it clear that if the invention/discovery was really beneficial, it was accepted by the world with enthusiasm.The empirical data shows that artificial intelligence tools and machine learning (ML) has brought positive results and benefits for almost all institutions including judiciary in various parts of the world.

---

1 Fuchs, T., In Defense of Human Being: Foundational Questions of an Embodied Anthropology (Oxford University Press, U.K., 2021).

The march of the new era of artificial intelligence, machine learning, and deep learning is bringing unimaginable changes to the entire world. These exponential changes remind us that a scientifically advanced technology is often found to be indistinguishable from magic. The increasing application of AI and ML techniques is, indeed, one such magical change that the present world is now experiencing. This magical technology will certainly place judges in a more robust position to appreciate matters of fact and the law, equip them to make predictions about the possible consequences of their decision-making, and improve the quality of judicial orders and judgments. However, there are well-founded inhibitions and doubts about the presence of bias and absence of fairness in AI-based sentencing determinations as experienced in the United States in the recent past. Unfortunately, the US experience tells us that algorithms may reflect and sustain the existing discriminatory practices as already embedded in the available data. Undoubtedly, with the improvements and advancements of algorithms supported by more intense efforts, it can be demonstrated that there will be more clarity, precision, and coherence in how most of the decisions will be made on facts and law, leading to not only qualitative judgments but also greater disposals.

History tells us that the new scientific avatar of artificial intelligence is the outcome of the exceptional efforts made by four American computer scientists, namely J. McCarthy, M. Minsky, N.Rochester and C. Shannon who conducted serious research on the Dartmouth Summer Research Project on artificial intelligence, often remembered as the Dartmouth Conference, which began on June 18, 1956, and continued for eight weeks. The conference united some of the brightest minds in mathematics, cognitive psychology, and computer science at the time.[2]

In fact, these scientists, along with some four dozen people they had invited, embarked on an ambitious journey to create intelligent machines. McCarthy says that their proposal was aimed to find out "how to create machines that use language, from abstractions and concepts, and offer solutions to problems which until then was the domain of human beings." This conference to plan and create intelligent machines was unique in itself and could be compared to the Big Bang of physics and geography, as what we understand now about neural networks, machine learning, and deep learning has its origins in the 1956 conference at New Hampshire.[3] Interestingly, David Cunning and some of his colleagues predicted that the possibility of some inherent conflict between ML performance (*i.e.*, predictive accuracy) and its explanation of related issues. They said that often the highest performing methods such as deep learning are the least explainable while the most explainable methods called decision trees are the

2 Peter, S., "How AI was born at a Summer Camp in United States 68 years ago", *The Hindu in School*, Sunday, Sep. 8, 2024.

3 *Ibid.*

least accurate.[4] J. Ryberg offers the possible solution of dealing with the lack of explanation of more complicated machine learning systems by drawing on explainable artificial intelligence (x AI) which is a second algorithm that is created to explain *post hoc* the working of the system known as black box system.[5]

The shift in application of AI from computer science and mathematics to legal system and then to the judicial system is a remarkable shift in human history. We will now discuss the issue of how to apply AI algorithms at a micro-level in the judicial branch in general and the sentencing justice process in particular *i.e.*, sentencing a convict after the trial in the specific context of the United States.

II Prejudice inputs seeping into sentencing *dicta*

"Black Data" warrants of historical arrest and banks on conviction data sheet that espouses an over-representation of marginalized groups. The Indian criminal sentencing landscape shows a contradictory trajectory - scheduled castes and tribes remain abysmally under-represented on the bench, yet are disproportionately the subjects of criminal prosecutions. An algorithm having been trained on these oblique inputs will not only retroflex but also amplify the caste- and class-based sentencing parities, opaqueness, and communication abstraction?[6] The operation of black boxes is essentially catered to with proprietary code and unrevealed data sources in the proposed AI sentencing systems in India. Judges opting for such risk scores lack the privilege of the rudimentary logic, and defendants cannot coherently challenge algorithmic connotations at trial.[7]

Constitutional vexation and procedural impediments

Two major bedrocks of the constitutional scheme are eroded – article 14, which guarantees the right to equality and article 21 which espouses the right to life and personal liberty- by algorithmic sentencing which to a greater degree implicates these much-harassed rights. It essentially takes place through deprivation of a reasoned order to defendants so that they might contest on appeal.

---

4 Gunning, D., *et al*., " XAI- Explainable artificial intelligence Z", *Science Robotics*, 4(2006).

5 Ryberg, j., "Criminal Justice and Artificial Intelligence: How should we assess the Performance of Sentencing Algorithms?" *Philosophy and Technology* 37:9 (2024).

6 Mudit Singh and Pooja Shree, "Judiciary and Caste: A Study on the Caste System affecting Judicial System," 7(I) *International Journal of Law Management and Humanities*, 937–942 (2024).

7 Sakshi Tripathi, "Algorithmic Sentencing in India: The Next Frontier or a Constitutional Threat?" *Lawful Legal* (June 6, 2025), *available at:* https://lawfullegal.in/algorithmic-sentencing-in-india-the-next-frontier-or-a-constitutional-threat/.

Automation of severity thresholds without a reasonable and readable justification that is comprehensible to the human mind. Tasnimul Hassan presents such a sentencing approach with a caveat that in the absence of transparency and accountability safeguards, an erosion of the individualized justice that Indian jurisprudence hinges on is brought to bear.[8]

Emphasizing judicial discretion

India's sentencing framework underscores evaluating mitigating factors—compunction, social background, prospects for reform. Completely automated recommendations risk sidelining these qualitative judgments unless explicitly kept "advisory" by statute and confined to serving as one input among many.[9]

III Constitutional and procedural concerns

Right to equality (article 14) and nonarbitrariness

It is ironic to mention that the courts in many parts of the world are under pressure to diminish prison populations, which is increasingly turning to algorithmic risk-assessment tools—systems that assimilate a defendant's age, criminal history, and socio-economic indicators to generate a single "recidivism score" used for bail, pretrial release, and sentencing decisions.[10] These algorithmic tools are conditioned on historical arrest data, implying they learn statistical correlations rather than causal factors. Algorithmic sentencing tools portend "objective" risk scores, despite being trained on prejudiced historical data and often culminate in arbitrary—rather than individualized—outputs.[11] The Supreme Court has perpetually enunciated that every state action must harbor the twin pillars of article 14 *i.e.*, equality before the law and safeguard against arbitrary executive power. In *Bachan Singh* v. *State of Punjab*, the court annulled the mandatory death penalty provisions for being innocuous and bolstered the necessity for subtle, case-by-case judicial discernment.[12] The application of algorithmic sentencing in India would also violate the guidelines relating to individualized sentencing in death penalty cases issued by the Supreme Court in *Manoj* judgement.[13] Black-box algorithms,

8 Tasnimul Hassan Md, "The Perils and Promises of Artificial Intelligence in Criminal Sentencing," 19 (2) *Indian Journal of Law and Technology* (2024), *available at:* https://repository.nls.ac.in/ijlt/vol19/iss2/1/.(last visited on May 22, 2025).

9 Muskan Shokeen andVinit Sharma, "Artificial Intelligence and Criminal Justice System in India: A Critical Study" 5 (4) *International Journal of Law, Policy and Social Review* 156–162 (Dec. 2023).

10 Karen Hao, "AI Is Sending People to Jail—and Getting It Wrong," MIT Technology Review, 21 January 2019, *available at:* https://www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai/.

11 Tasnimul Hassan Md, "The Perils and Promises of Artificial Intelligence in Criminal Sentencing," 19(2) *Indian Journal of Law and Technology*, (2024), *available at:* https://repository.nls.ac.in/ijlt/vol19/iss2/1/.

12 *Bachan Singh* v. *State of Punjab* (1980) 2 SCC 684, para. 203.

13 *Manoj* v. *State of M.P.*, 2021 SCC OnLine SC 3219.

which neither come clean with their decision logic nor allow tailoring to individual circumstances, speaks of great friction with this constitutional mandate.

Right to life and personal liberty (article 21)

"*Procedure Established by Law*"

Article 21 guarantees that no person may be deprived of life or personal liberty except "according to procedure established by law." In *Maneka Gandhi* v. *Union of India*, the court emphasized the requirements of fairness, reasonableness and justice.[14] Algorithmic risk assessments—devoid of any statutory framework harboring data quality standards, transparency obligations, or audit mechanisms—cannot gratify the court's insistence that procedural rules themselves be just and reasonable. The use of algorithmic tools in the criminal justice system is bound to adversely affect the offender's right to reasoned orders and the *audi alteram partem* principle. Indian jurisprudence caters that every judicial order consists of intelligible reasons, not only to allow standardized appellate review but also to uphold the Audi alteram partem (hear-the-other-side) doctrine. In *Shyam Singh* v. *State of Rajasthan*, the court annulled the bare orders devoid of any rationale as a clear violation of article 21's due-process guarantee.[15] When a judge banks on an inscrutable algorithm, neither party can be sure of which facts or risk factors affected the sentence, effectively forestalling any challenge to the AI's conclusions.

Right to privacy and data protection (**article** 21)

The monumental judgment in *Justice K.S. Puttaswamy (Retd.)* v. *Union of India* discerns privacy inclusive of informational autonomy as a fundamental right under article 21.[16]AI-based sentencing systems assimilate a vast chunk of personal data (criminal histories, socioeconomic profiles, even social-media footprints) without proper consent or data-governance protection. This unregulated data-harvesting and profiling violates the court's insistence on proportionality and minimal infringement, due process, and the right to a fair hearing. A core component of due process is the ability of a defendant to know, test, and rebut the evidence against him. Opaque AI recommendations, sometimes treated by judges as authoritative risk scores, deny litigants the basic right to cross-examine or critique underlying data and model parameters. As Tasnimul Hassan warns that without transparency and accountability mechanisms put in place to check the machine bias, algorithmic sentencing threatens to "erode the individualized justice that Indian jurisprudence demands; consequently, they routinely judge economically backward and minority defendants as "high risk" despite no re-offence of such individuals, perpetuating biased stimulus and reinforcing

14 AIR 1978 SC 597, para 13.

15 (2007) 13 SCC 773, para. 17

16 (2017) 10 SCC 1, paras. 145–146.

cycle of surveillance and incarceration.[17] Moreover, because the algorithms operate as proprietary "black boxes," neither defendants nor judges can scrutinize which variables or weightings drove a particular score, making it impossible to challenge or audit erroneous outcomes despite mounting evidence that these systems exacerbate existing racial and economic disparities.[18]

Artificial intelligence and the prism of ethics

Chelsea Barabas's influential critique of algorithmic ethics in criminal justice explores how mainstream efforts to create "fair" AI—whether through statistical criteria or institutional protocols—often miss the deeper power dynamics embedded in data-driven punishment. Her work dismantles the idea that mathematical adjustments alone can produce equitable outcomes in a legal system already shaped by historic surveillance and racialized control.[19]

Challenging statistical remedies

Barabas begins by dissecting popular fairness formulas—like demographic parity or equalized odds—that attempt to make algorithms more neutral. Through analysis of widely debated tools such as Correlational Offender Management Profiling for Alternative Sanctions (COMPAS), which forecasts repeat offenses, she illustrates how these numerical standards can not compensate for structurally biased data. It is said that such frameworks are limited to surface-level recalibration, overlooking how historical injustices are baked into the datasets themselves. Metrics may shift outcomes marginally, but they don't interrogate the roots of over-policing or criminalization that created the disparities.[20]

Unpacking the compliance toolkit

Barabas then applies the lens of scrutiny to the common institutional responses—transparency reports, performance benchmarks, external audits—which are widely seen as ethical safeguards. Yet she brings out a fine revelation that these practices often function as technocratic practices, giving an illusion of control while impeding the imponderable question: should these technologies be in existence at all? Instead of critically evaluating the purpose and impact of predictive surveillance, these protocols polish its legitimacy and embed it deeper within bureaucratic routines.[21]

---

17 *Supra* note 11.

18 *Ibid.*

19 Chelsea Barabas, "Beyond Bias: Re-Imagining the Terms of 'Ethical AI' in Criminal Law," *SSRN Electronic Journal* (2019), *available at:* https://papers.ssrn.com/abstract=3392824 (last visited on June 24, 2025).

20 *Ibid.*

21 *Ibid.*

Power, prediction, and punishment

The current discourse in algorithmic sentencing demonstrates that both the mathematical and managerial approaches interfere with the individual considerations and reinforce the idea of preemptive control in the criminal justice system, reducing the sentencing analysis to mere calculations and hypothetical predictions relating to the future criminal proclivity of the offender. Unfortunately, the algorithmic sentencing systems fail to treat the offender as a complex human being in his 'specific social context' and ultimately reduce the offender to the status of 'risk profiles' to be managed. Algorithm-based sentencing approach, therefore, neglects the social context judgment. This risk-oriented logic, unfortunately, expands carceral reach under the guise of neutrality. True ethical inquiry must go beyond reforming tools and ask whose interests these systems serve, and at what societal cost.[22]

Instead of tuning bias metrics or enhancing documentation, Barabas espouses for an abolitionist re-imagining of AI's role in criminal law: Reframe outcome measures to center community-defined indicators of safety and well-being rather than recidivism rates. Decouple data regimes from punitive logics by refusing to treat predictive analytics as neutral "tools" and instead interrogating what social controls they reinforce. There is a need to develop participatory methodologies that surface gaps in carceral data (for instance, qualitative harm narratives) and allocate resources to non-carceral interventions first. This abolitionist stance shifts the critical question from "How do we make AI less biased?" to "What practices and power structures are we entrenching by embedding predictive models in law enforcement and sentencing?"[23]

Nascent attempts to "de-bias" large language systems often obscure the fact that many inequities are encoded long before any data scientist writes a line of code. Ruha Benjamin deliberates that what she believes the "New Jim Code" is, in fact, a technical rebranding of historic patterns of racial subordination. Instead of mere fairness audits, she exhorts us to adopt what she calls reparative frameworks—approaches entrenched in collective stewardship of technology and a consignment to dissembling punitive architectures that mirror logic rather than simply tweaking parameters.[24] Cathy O'Neil's analysis complements this by exposing how large-scale risk scores and predictive models become self-reinforcing "black boxes." She shows how these tools, once deployed, tend to amplify mistakes and entrench feedback loops that disproportionately punish vulnerable groups. Her conclusion is uncompromising: incremental adjustments to scoring thresholds will never suffice; problematic systems must be removed, not finetuned.[25]

---

22 *Ibid.*

23 *Ibid.*

24 Ruha Benjamin, Race after Technology: Abolitionist Tools for the New Jim Code (Polity, 2019).

25 Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown, 2016).

These commentators produce evidence in the context of the United States to indicate that the historical data-based discrimination against disadvantaged sections is demonstrated to the front lines of social service delivery, welfare measures, housing, and justice settings, systematically surveil and penalize the poor. It is further demonstrated that without deep structural change, such as empowering communities to govern their own data and offering non-punitive alternatives, no ethics handbook can prevent algorithmic programs from perpetuating cycles of poverty and exclusion.[26]

Building on these insights, Rashida Richardson, Jason Schultz, and Kate Crawford dissect the "dirty data" feeding predictive policing platforms. They reveal how datasets steeped in discriminatory policing practices infect every stage of algorithmic design. Their prescription is a rigorous datajustice paradigm: cleanse records at the source, establish enforceable oversight rights, and halt deployments until foundational harms are addressed.[27] Finally, Kristian Lum and William Isaac turn a critical eye to the mathematical metrics frequently proposed as fairness solutions—demographic parity, equalized odds, and their kin. They demonstrate that these metrics are often mutually exclusive and ignore the broader social contexts in which algorithms operate. Their call is for a socio-technical approach that couples formal evaluations with policy reforms and participatory design methods, moving beyond the narrow confines of statistical definitions.[28]

IV Role of AI in sentence determination: Scenario in the United States

In many countries, algorithmic techniques are presently applied at all stages of the criminal justice system, from the inception of criminal prosecution to the final stage of decisions in court and even at the appellate stage. In fact, the risk assessment and the future re-offending algorithms, such as the (COMPAS)[29] have been used in the US for a long time to assist the courts about the probable chances of an offender

26 Virginia Eubanks, Automating Inequality: *How High-Tech Tools Profile, Police, and Punish the Poor* (St. Martin's Press, 2018).

27 Rashida Richardson, Jason M. Schultz, and Kate Crawford, "Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice," 94(1) *New York University Law Review* 192–233 (2019).

28 Kristian Lum and William Isaac, "To Predict and Serve?" 13(5) *Significance* 14–19 (Oct 2016).

29 Northpointe, Inc., COMPAS: A Fourth-Generation Risk-Need Assessment System [COMPAS is a fourth-generation (4G) risk-need assessment system that incorporates a range of theoretically relevant criminalistics factors and key factors emerging from meta-analytic studies of recidivism. It is a risk-need assessment tool designed by Northpointe, Inc., to provide decisional support for "the Department of Corrections" when Executing placement decisions, overseeing offenders, and strategising treatment. The COMPAS risk assessment is rooted from information gathered from the defendant's criminal record and an interview with the defendant. A COMPAS report consists of a risk assessment designed to predict recidivism and a separate needs assessment for identifying program needs in areas such as employment, housing and substance abuse. Risk assessment component of COMPAS produces risk ratings illustrated as a bar chart, featuring 3 bars that denote pre-trial, general, as well as violent recidivism risk. Each bar represents defendant's risk level on a scale of 1-10] (2009).

falling back into crime after release from prison. If a greater role is accorded to AI in sentence determination, a more radical instance would be the use of algorithms designed to provide sentence recommendations in individual criminal cases.[30] It is noteworthy that systems that determine sentences in cases of serious crimes such as rape and drug possession have already been put into practice in some of the Asian countries as well.[31]

It seems that AI is everywhere. Computer-run algorithms have permeated the sentencing process in various ways across many jurisdictions, and these instruments are likely to play an even more influential role soon.[32]Some jurisdictions are using AI to assist judges in making decisions about sentencing. These algorithms analyze data on similar cases and provide recommendations to judges based on factors such as the severity of the crime, the defendant's criminal history, and mitigating circumstances. AI algorithms are being used to analyze data and predict the likelihood of a defendant committing future crimes. These risk assessment tools use various factors such as criminal history, demographics, and socio-economic indicators to produce a score that can help judges make decisions about sentencing. It is also used to develop sentencing guidelines that consider a wider range of factors and circumstances than traditional guidelines. These guidelines can help judges make more informed and consistent sentencing decisions.

The first concrete and relatively detailed proposal for computer-assisted sentencing was made in 1971 by John Hogarth who envisioned an algorithm based on an accumulating database for the determination of sentence and subsequent outcomes in terms of re-offending.[33]He also predicted that the computer would then provide a court with a range of information to guide sentencing decisions, including sentences imposed in similar cases by previous courts, reconviction statistics for specific sentences, a risk score for the offender currently being sentenced, and a sentence recommendation that matched offender profiles with specific dispositions. The database would grow over time and would incorporate "new methods of correctional treatment."[34]

It has been noteworthy of mentioning Jasper Ryberg and Julian Roberts have undertaken comprehensive research on AI utilization for the sentencing process, and they say that AI has become a million-dollar industry. Various techniques that could

30 *Ibid.*

31 Khazanah Research Institute, "This is the case in Malaysia" (2021).

32 Ryberg, J., "Sentencing and Algorithmic Transparency", in *Sentencing and Artificial Intelligence* (Eds by Ryberg, J. and Roberts, J.V., Oxford University Press, 2022)13.

33 Hogarth, J. (1971), *Sentencing as a Human Process* (University of Toronto Press, Toronto).

34 Ryberg, J. and Roberts, J.V. (Eds.), "Sentencing and Artificial Intelligence" Oxford University Press(2022).

be relevant if AI were to become more prominent in sentencing are already in use. The most obvious example is risk assessment algorithms, which are currently employed in many US States.[35] However, various uses of AI raise a fundamental question about whether it will promote or undermine values such as fairness, transparency, mercy, and integrity in sentencing.

Many countries around the world are regularly applying AI in their administrative and judicial work but the Indian judicial system is yet to make use of AI tools inthe judicial and administrative setup on the lines in which it is employed in the United States and other countries. Probably, on an experimental basis, AI algorithms are presently used in India only for the purpose of translation of judgments and disposal of traffic challans. Observing how different nations are using AI algorithms in sentencing procedure, one can find that it is used in sentencing decisions through the method of predictive analytics. The experience in these countries tells us that algorithms analyze huge amounts of data, including demographic information, criminal history of the convict, and other factors, to predict the future chances of a convict re-offending or the severity of their crime. It is a profound claim of mathematicians and computer scientists that AI algorithms are technically equipped to identify patterns and trends in crime and re-offending that humans may not identify during sentence determination providing additional information to judges that they may use in their decisions.

Examinining the AI's functioning in judicial decision-making particularly in quantum calculation of sentences. An important consequence of introducing computational methods to judicial decision-making is that judges will find themselves having to acknowledge the decisions and predictions made by their AI-powered aids. As technology advances and begins to incorporate not only legal input but also social science data as well, judges may find themselves having to explain disagreements between their own judicial opinions and the predictions made by legal prediction tools. Judges cannot now afford to simply ignore or dismiss the determinations and predictions made by these computational tools without suffering the potential loss of perceived legitimacy of their decisions.[36]

However, the bigger question is the quality and accuracy of computational approaches, as a lot of apprehensions about the fairness of data are expressed in some States in the US where racially colored predictions as to re-offending have been made by algorithms used in the sentencing process.Utah employed RANAT (Risk and Needs Assessment Tool) to aid judges in arriving at sentencing decisions.AI system utilized in quantum computation of sentence analyses variables including substance abuse,

35 *Ibid.*

36 Abdi, S. and Alarie, B. *Legal Singularity: How Artificial Intelligence can make Law Radically Better,* University of Toronto Press (2023).

criminal history, along with employment status to forecast future re-offending probability. Judges utilize these risk evaluations to customize punishment and rehabilitation strategies for individuals, with the objective of diminishing recidivism rates and facilitating seamless reintegration into society.

In Pennsylvania, the COMPAS technique is employed to evaluate risk levels of individuals pending trial/punishment. COMPAS develops risk scores by analysing many data points, comprising criminal history along with social circumstances, thereby assisting judges in making intelligent decisions over probation, bail, along with other sentence alternatives. In California, the PSA(Public Safety Assessment) program utilize AI to assess variables including previous arrests, criminal history alongside age, to forecast future criminal behavior probability. Judges utilise these risk evaluations to establish suitable conditions for pre-trial release and punishment, with the objective of enhancing fairness and efficiency within the justice system. These instances illustrate the growing integration of AI technologies into sentencing judgements, offering judges data-driven insights to enhance criminal justice outcome efficiency.

When we look at some particular cases in the US linked with AI software utilization in the sentencing process, 1st reference might be executed to *State of Wisconsin* v. Eric L. *Loomis, 2015*[37] where COMPAS utilization, an AI software, impacted sentencing decision. A large reoffending threat was predicted by the software for a defendant, factors judges evaluated in deciding punishment duration. This case ignited controversy as opponents expressed apprehensions regarding equality along with precision of employing AI algorithms in sentencing determinations. The discussion emphasised the ethical ramifications of depending on AI inside the criminal justice system.

Brief facts of the case

The State of Wisconsin contended that Loomis was legally responsible for being the driver in a drive-by shooting case. He was charged with five counts for committing the said crime and was also a repeat offender:

(i) Recklessly endangering the safety of first-degree;

(ii) Attempting to escape from the lawful command of a traffic officer;

(iii) Using a vehicle in the absence of the consent of the owner;

(iv) Possession of a firearm ;

(v) Possession of a short-barreled gun.[38]

Loomis refuted his participation in the drive-by shooting. He relinquished his "right to a trial" moreover entered 1 guilty plea to 2 lesser charges: attempting to evade 1

37 *State of Wisconsin* v. *Loomis*, 2016 WI 68, 2015 AP 157-CR (2016).

38 PATC refers to party to a crime.

traffic officer while employing a motor vehicle with no owner authorisation. The plea agreement indicated that additional charges would be dismissed but acknowledged in:

The other counts will be dismissed and read in for sentencing, although the defendant denied he had any role in the shooting, and only drove the car after the shooting occurred. The State believes that he was the driver of the car when the shooting happened. The State will leave any appropriate sentence to the court's discretion, but will argue aggravating and mitigating factors.

After accepting the plea of Loomis, the circuit court mandated a Pre-Sentence Investigation (PSI). The PSI comprises the attached COMPAS report.[39] As the PSI explains, Risk scores aim to forecast the probability that people having similar criminal histories are either less or more inclined to reoffend post-release from incarceration. Nonetheless, the COMPAS risk assessment does not ascertain the precise probability of an individual offender re-offending. Rather, it offers a forecast by contrasting the individual's information with a comparable data set.

The COMPAS risk scores for Loomis suggested a high recidivism risk across all three bar charts. His PSI contained an explanation of the appropriate application of the COMPAS risk assessment and warned against its improper use, emphasising that it should be utilised to identify offenders who might profit from interventions and to address pertinent risk factors during supervision.

During the sentencing phase, the State contended that the circuit court need to utilise the COMPAS report in establishing a suitable punishment. Eventually, the circuit court cited the COMPAS risk score alongside additional sentence considerations in ruling probation out moreover the court held that "the convict is identified, through the COMPAS assessment, as an individual who is at high risk to the community. In terms of weighing various factors, the court ruled out probation because of the seriousness of the crime and because of the criminal history, history of supervision, and the risk assessment tools that have been utilized, suggesting that the convict is extremely high risk to re-offend."

---

39 Northpointe, Inc., COMPAS: A Fourth-Generation Risk-Need Assessment System [COMPAS is a fourth-generation (4G) risk-need assessment system that incorporates a range of theoretically relevant criminalistics factors and key factors emerging from meta-analytic studies of recidivism. It is a risk-need assessment tool designed by Northpointe, Inc., to generate decisional support for the Department of Corrections during formulation of placement decisions, supervising offenders, alongside strategising treatment. COMPAS risk assessment relies on data obtained from the defendant's criminal record and an interview with the defendant. A COMPAS report consists of a risk assessment designed to predict recidivism and a separate needs assessment for identifying program needs in areas such as employment, housing and substance abuse. The risk assessment component of COMPAS produces risk ratings illustrated as a bar chart, featuring three bars that denote pre-trial, general, and violent recidivism risk. Each bar represents a defendant's risk level on a scale of 1 to 10] (2009).

As per the plea questionnaire/waiver of rights form, Loomis might be imprisoned for up to 17 years and six months for both crimes. The court imposed the maximum punishment for the two crimes to which he pleaded guilty. Loomis submitted a motion for post-conviction relief seeking a new sentencing hearing. He contended that the circuit court's evaluation of the COMPAS risk assessment during sentencing infringed upon his due process rights. He contended that the circuit court misapplied its discretion by incorrectly presuming the veracity of the factual foundations for the read-in charges.

The circuit court conducted two hearings in the post-conviction period. At the first hearing, the court addressed Loomis's claim that it had erroneously exercised its discretion in how it considered the read-in charges. After considering the relevant case law and legal standards, the court concluded that it had applied the proper standard and denied Loomis's motion on that issue.

Eric Loomis appealed the circuit court's denial of his post-conviction motion requesting a resentencing hearing. The Court of Appeal certified the specific questions of whether the utilisation of a COMPAS risk assessment during sentencing infringes upon the defendant's right to due process, either due to COMPAS's proprietary nature hindering defendants from contesting the assessment's scientific validity, or because gender factors are considered in the COMPAS evaluations.

Grounds of appeal

Firstly appellant Loomis asserted that the "COMPAS risk assessment" at sentencing by the Circuit Court violates "the defendant's right to due process" for the following reasons;

(i) It violates the "defendant's right to be sentenced" rooted upon precise information, partly due to the proprietary nature of COMPAS that inhibits evaluating its accuracy,

(ii) It violates "defendant's right to an individualized sentence", and

(iii) It improperlyemploys gendered evaluations in sentencing.

Secondly The "Circuit Court"improperly exploited its discretion by assuming that the factual bases for the read-in charges were true.

The Wisconsin Supreme Court observed that "the defendant is not challenging COMPAS risk assessment utilization for decisions other than sentencing, and he is not challenging utilization of COMPAS report portion needs at sentencing. Instead, Loomis challenges only the risk assessment portion utilization of COMPAS report at sentencing. The SC affirmed the order of circuit court on the following grounds;

i. A "Circuit Court" considering "COMPAS risk assessment" at sentencing does not violate a "defendant's right to due process" if it is used properly, observing the limitations and cautions set forth.

ii. The circuit court explained that it did not place sole reliance on the sentencing algorithm tool called COMPAS risk scores, but the same was bolstered by further independent variables. Moreover, its utilization hasn't been determinative in deciding the issue if Loomis might be monitored securely as well as efficiently inside community. Hence, it had been noted that the circuit court didn't exercise its discretion erroneously.

iii. The circuit court's consideration of the read-in charges was not an erroneous exercise of discretion because it employed recognized legal standards.

Although the "SC of Wisconsin" in the "Loomis case" (2016) stated risk assessment of COMPAS algorithm made in the sentencing process by "the Circuit Court" does not violate "the right of defendant" to due process, lot of discussion on this case concerning fairness, accountability moreover transparency issues in AI application has been there. It seems evident to dwell deeply into howthe COMPAS algorithm operates to offer an assessment of the reoffending risk of the convict in a particular case.Risk-need assessment transformation from first generation to fourth generation is discussed below.

V Origin and changing scenario in penitentiary appraisal

It has been remarkable to acknowledge that in history of AI application to sentencing process, the last three decades in penitentiary practice have seen an advancement from the initial mechanism of first-generation (1G) to currently working 4G appraisal methods. These improvements occurred as different generations of appraisal and categorization methods addressed the more obvious anomalies of previous phases.

The first-generation method placed reliance on clinical and professional judgment in the absence of any explicit or objective scoring rules. It dominated penitentiary appraisal analysis for many decades and still remains anoption on priority for many experts in correctional decision-making.[40] The problem with this method is its excessive subjectivity, inconsistency issues, bias, and probable stereotyping, legal weakness, and lower predictive legal value than structured objective methods.[41]

The greatest strength of second-generation (2G) appraisal is its empirical basis that primarilyrelieson simple additive point scales with only a few standardized factors. These methods mainly focused on risk prediction, brevity, and efficiency. The main criticisms of this method are a lack of theoretical background, restricted scope of risk and need factors, exclusion of dynamic risk factors, absence of treatment implications, vulnerable explanatory outcome, and questionable value in the case of female offenders. However, these sequence models are often surprisingly effective

40 Brennan, T., Dieterich, W., and Ehret, B. "Evaluating the Predictive Validity of the COMPAS Risk and Needs Assessment System", 36:1 *Criminal Justice and Behaviour* (2009).

41 *Ibid.*

in their predictive veracity and generally outperform professional judgment or the opinions of trained experts.[42]

The third-generation (3G) appraisal came in the late 1970s and 1980s, which introduced a clearer, based on empirical method, and theory-driven perspective, and a broader alongside more efficient selection of pervasive factors. In addition, somefactors were designed to be dynamically sensitive to change. Interestingly, the Level of Services Inventory- Revised (LSI-R) exemplified these trends and perhaps has become the most preferred risk and penitentiary appraisal mechanism. However, these methods, including the LSI-R, were criticized due to their narrow focus on theoretical premises (primarily social learning theory), missing gender sensitivity, an excessive focus on risk evaluation, and a near complete failure to assess and include the strengths of offenderor protective factors as considered in the "good lives" model.[43]

The previous experiences of trial and error, offered criticisms, further experiments, and attempts to bring accuracy, fourth-generation (4G) assessments came into being. It is noteworthy that many general factors characterize 4G methods, which include the following;

i. a broad-based selection of explanatory mechanisms,

ii. inclusion of a comprehensive range of risk and need parametersthat emphasize 'veracity and authenticity of the content',

iii. consideration of the strengths perspective,

iv. shift to more effective statistical modeling,

v. consistent integration of necessity or risk realms with the agency management information system, and

vi. increasing reliance on criminal justice database and web-based implementation of appraisal technology.

It is noteworthy that such integration gives an opportunity to users to keep track of the convicts from the beginning to end of the case to ensure sequential case management monitoring, feedback on information, and judicial decision-making. It is interesting to mention that COMPAS has incorporated all these features and the reliability of reoffending scales as well as predictive veracity for both male and female convicts is said to have been enhanced.

VI Automated decision-making, human rights concerns, and the European Union Artificial Intelligence Act

The EU has become the world's first organization to enact and adopt separate and exclusive legislation, creating a comprehensive legal framework for regulating artificial

42 *Ibid.*

43 *Ibid.*

intelligence across the entire European region. The legislation is aimed at ensuring that AI systems not only operate in a safe and trustworthy manner but also respect fundamental rights and European Union values. The primary goal of this important legislation is to build and promote "trustworthy AI" within the region to encourage innovation while ensuring safety and protection of fundamental rights within the EU.[44] It is also expected that the legislation will strengthen investment in AI, helping businesses and public services. The legislation operates at the touchstone of a risk-based approach and categorizes AI systems into unacceptable, high, limited, and minimal risk, with different rules for each of these categories.

The risk-based approach emphasizes the potential risk posed by AI systems to society, human rights, and safety. It works on the principle- the higher the risk, the stricter the rules. The AI Act of the EU sets out the requirements for AI service providers and algorithm developers, including strict obligations for high-risk systems, and prohibits certain AI practices deemed to pose an unacceptable risk.[45]

It is significant that before adopting and applying AI systems in Indian legal and judicial systems, there is an urgent need to evolve a comprehensive legal framework on somewhat the same lines as the EU AI Act to ensure safety, transparency, and accountability in our institutions and to protect the society and human rights.

VII Suggestions and way forward

While AI is having potential of improving the sentencing process by providing more accurate and consistent decision-making, there are concerns about biases in the algorithms and necessity of transparency along with oversight in their use. Judges can take several steps for assuring artificial intelligence tools applied in sentencing decisions are fair. One approach is to conduct regular audits and reviews of the algorithms to assess their accuracy and fairness. Judges can also seek transparency from the developers of the AI systems, requesting detailed explanations of how the algorithms work and the data they use.There is, indeed, critical necessity ofdeveloping a system of regulatory framework for regular audits and reviews of the algorithms used in administration of criminal justice. To address the impending threats of potential biases in AI algorithms utilized to sentence, the judges might utilize the following strategies:

Firstly, possibility that AI algorithms can inadvertently perpetuate bias and discrimination present has been there in historical data which has been pertinent to see that AI system doesn't unfairly penalize individuals over race, gender, socio-economic status, or any other protected characteristic. Judges may undertake continuously bias testing on the AI algorithms to identify any discrimination in

44 EU AI Act, *available at:* https://artificialintelligenceact.eu

45 *Ibid.*

outcomes based on factors such as caste, religion, gender, or socio-economic status *etc*. Judges can always use protective methods to tackle biasby analyzing the impact of the applied variables in sentencing process.

Secondly, the mechanism of decision-making with AI algorithm utilization should be transparent, and individuals needs to have "the right to understand"how their sentence was determined. There must be a well-placed regulatory framework to hold artificial intelligence technique accountable for their outcomes to ensure transparency. Judges can ensure non-discriminatory sentencing process by actively addressing biases in the application of AI algorithms.

Thirdly, AI algorithms require access to sensitive personal data to make sentencing decisions. It has been vital for assuring no unauthorized data access is made so and that privacy rights are protected.While AI systems could assist judges in making more informed decisions, there must constantly be one human overseeing procedure and having the final say in sentencing which has been crucial moreover necessary to avoid complete automation of the sentencing mechanism to safeguard against potential errors or ethical concerns.

Fourthly, AI intervention in sentence determination must be guided by ethical principles and respect for human rights. It is important to consider the broader societal implications of introducing AI into "the justice system" and ensure that it protects justice and fairness.It has been vital that policy-makers should engage a varied set of stakeholders, comprising legal specialists, ethicists, alongside community leaders, in deliberations over AI implementation in sentencing prior to its adoption. Then the judges should be adequately trained and fully equipped to identify and address potential biases in the algorithms while dispensing justice. A code of ethical principles in application of AI tools should be developed and handed over to the trial judges across the country before the permission to apply AI tools is formally granted. Then the trial judges should adhere to the documented ethical protocols for AI application in sentencing choices. By rigorously following ethical rules and standards of fairness, openness, alongside accountability, trial judges could guarantee that AI algorithms conform to legal and ethical principles within the administration of criminal justice.

Fifthly, AI algorithms can only be unbiased to the extent as the data they rely on, is unbiased. The biggest limitation of any AI algorithm is that it works on the already available data, and if that data contains bias or discrimination, AI algorithm is bound to carry the same in its results. To prevent biases and discrimination in sentencing, it is critical to assure training data utilized for the algorithms is diverse and representative of all demographics.This has been vital for regularly auditing alongside monitoring AI algorithms used in sentencing procedure to identify and address any biases or discriminatory patterns that may arise which could be accomplished by utilizing bias detection tools, conducting regular reviews, and involving diverse teams in the

development and implementation of algorithms.The policy-makers may request developers to ensure transparency concerning the data sources and criteria employed in the development of AI systems. Comprehending the data inputs and procedures enables both policymakers and judges for examining potential for algorithm-bias and make informed choices regarding their application from time to time.

Sixthly, AI algorithms utilized in quantum determination should be transparent and explainable so that their decision-making procedures could be understood by various stakeholders and can be scrutinized by the trial judges. This step to achieve transparency can help prevent discriminatory outcomes and increase accountability in the system. Continuous monitoring of AI algorithms in sentencing judgements to determine their effects on fairness and equity is a crucial necessity. By monitoring outcomes and assessing the efficacy of bias mitigation strategies, trial judges can preemptively resolve any emerging difficulties and maintain the integrity of the justice delivery system.

Finally, training judges, lawyers and other stakeholders involved in the process about the potential biases and limitations of AI tools and techniques can help prevent discrimination and can make sentence determination process fair and just.

VIII Conclusion

Undoubtedly, considering AI tools in legal research is very helpful to judges in judicial decision-making,as it leads not only to quick disposal but alsoassists in producing qualitative judgments as observed in the US, Malaysia, and elsewhere. AI algorithms may play a very positive role in deciding the length and severity of sentence in criminal cases. However, machine-based learning and analysis are dependent on the available data in which conscious or unconscious discriminatory practices may already be embedded, as a result, the socially unconscious but intelligent machine which functions on available data, is bound to give discriminatory, racial, unequal and unfair analytical-computational outputs.

Moreover, judges should be very attentive towards all possible biases in the already available data, which might unconsciously carry the unfair element in an AI-based sentencing algorithm. It is pertinent that judges have to carefully examine the sources and data quality to identify and address any intrinsic biases that might impact the algorithm's response. It is essential for judges to check the ethical implications of using AI in sentence determination and to prioritize authenticity and transparency in the use of technology. Furthermore, judges must use their experience and judicial acumen most effectively while exercising discretion in the decision-making process to complement the machine-based information provided by AI algorithms. Judges may guarantee that sentence decisions are informed, equitable, and compliant with legal criteria by integrating AI technology insights with their experience and judgment. Judges are essential in supervising the application of AI in sentencing, promoting

transparency, accountability, and equity to maintain the foundations of justice within the criminal judicial system.

The unreasonable prediction regarding re-offending by the convict in Loomis's case is just the tip of the iceberg, and many more such unreasonable examples of computational analysis might come in the future. Hence, it would be safe to use these algorithms, carefully verifying the re-offending chances indicated by the algorithm used in sentence determination. There is an urgent need to develop a comprehensive regulatory framework in India for the makers of such sentence determination algorithms before embarking on a journey to use the AI techniques in sentence determination in criminal trials failing which the AI algorithms will cause more harm to fairness and transparency than assisting the sentencing courts in dispensation of justice. It is a bounden duty of trial courts that justice should never become a casualty in the name of quality and quickness. A careful approach of a sentencing judge and an authenticated strategy in using AI tools can successfully overcome the potential threats of machine biases in AI algorithms used for sentence determination and may very well promote neutrality, accuracy, and transparency in the criminal justice system.

*Humayun Rasheed Khan**
*Masood Ahmad***

* Professor, National Judicial Academy India, Bhopal.

** Assistant Professor (Law) AMU Centre Murshidabad, West Bengal, India.