EUROPEAN ACADEMIC RESEARCH
Vol. XI, Issue 2/ May 2023

Impact Factor: 3.4546 (UIF)
DRJI Value: 5.9 (B+)

ISSN 2286-4822
www.euacademic.org

The Challenges on Implementing Artificial Intelligence in the International Criminal Justice System

Dr. IVAS KONINI
Lecturer, Department of Criminal Law, Faculty of Law, University of Tirana, Albania
E-mail: ivas.konini@fdut.edu.al
Dr. IV. ROKAJ
Lecturer, Department of Criminal Law, Faculty of Law, University of Tirana, Albania
E-mail: iv.rokaj@fdut.edu.al

Abstract:

Artificial Intelligence (AI) has the potential to revolutionize the legal system by improving efficiency, reducing costs, and enhancing decision-making processes. However, implementing AI technology in the legal system presents several challenges that stem from the complex nature of the legal system and the ethical considerations that must be taken into account.

The legal system is complex, with intricate regulations and diverse legal principles, posing a significant challenge to the use of AI. Additionally, AI must possess a deep understanding of legal doctrines and analytical reasoning to interpret and apply legal principles accurately.

Ethical considerations must be addressed to ensure that AI technology does not result in biased or unfair outcomes. AI algorithms rely on data, and if the data is biased, the algorithm will be biased, leading to discriminatory outcomes against certain groups of people. Therefore, mechanisms must be implemented to promote fairness and accountability.

There is also a risk of dehumanization and reduced human judgment when implementing AI in the legal system. This may lead to a reduction in the role of human judgment and decision-making in legal proceedings, which may be unfavorable for the overall functioning of the legal system.

Successfully integrating AI in the legal system requires a significant effort to overcome the various challenges posed by the system's complexity and ethical considerations involved. The integration of AI technology in the penal legal system should aim to enhance human judgment, reduce bureaucratic delays, and promote fairness in legal proceedings.

Keywords: Artificial Intelligence, penal law, legal system, ethics, transparency.

INTRODUCTION

Artificial intelligence[1], is a field of computer science focused on the development of algorithms and software that can perform tasks that typically require human intelligence, such as problem-solving, decision-making, and language understanding.

[1] From now on, AI.

Artificial intelligence has been recognized as a highly effective resource for multiple sectors. The ability of AI technology to analyze large volumes of data and identify patterns that are difficult to detect by humans has made it one of the most sought-after technologies in recent years. AI technology has the potential to revolutionize many sectors, including the criminal justice system. The implementation of AI in the penal rights system promises to improve decision-making, speed up processes, and enhance overall efficiency. Like any novel technology, the adoption of AI also comes with its own set of risks and difficulties.

The use of AI in the criminal justice system is largely centered around the analysis of data to make informed decisions. Machine learning, natural language processing, and computer vision represent some of the essential technologies utilized in this field of study. These tools allow for the analysis of vast amounts of data, including criminal records, incidents of recidivism, and past patterns of criminal behavior. By automating these analyses, AI can help identify potential risks and opportunities for early intervention, essentially supporting crime prevention measures[2].

One of the significant advantages of AI in the penal rights system is its ability to remove human biases from decisions. Human decision-making is often influenced by personal biases, which can lead to wrongful judgments[3]. The implementation of AI technology can help eliminate these biases and promote objectivity in decision-making. It can also help standardize processes and ensure that decisions are made based on consistent and objective data.

However, the implementation of AI technology in the criminal justice system is not without challenges. One notable issue is the potential for algorithmic bias. When biases are present in the data used for AI system training, the algorithm will replicate these biases in its decision-making process[4]. This can lead to unintended consequences, such as the perpetuation of racial and gender inequalities. There have been concerns that an algorithm may unfairly label individuals as high-risk solely based on their demographics, leading to disproportionate sentencing and further marginalization of already vulnerable communities[5].

It is also essential to consider the ethical implications of using AI in the penal rights system. The use of AI technology in the criminal justice system raises questions around transparency, accountability, and privacy. The use of AI systems can undermine the ability of individuals to challenge decisions made about them. The lack of transparency in how these systems operate and make decisions can lead to a lack of trust in the criminal justice system[6].

Furthermore, there is a risk that the use of AI may lead to a shift towards risk management over rehabilitation. Risk algorithms may focus on identifying individuals who pose a risk rather than understanding the root causes of criminal behavior or providing rehabilitation solutions. If the primary focus is on risk management, there is a risk that the criminal justice system will fail to address the underlying issues that contribute to criminal behavior.

[2] Megan Stevenson, "Machine Learning and Criminal Justice," Annual Review of Criminology, vol. 3, no. 1 (2020).
[3] Andrew Ferguson, "The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement," New York University Press (2017).
[4] Skeem, J. L., Monahan, J., & Lowenkamp, C. T. (2015). Risk, Race, and Recidivism: Predictive Bias and Disparate Impact. Criminology, 53(4), 680-712.
[5] Virginia Eubanks, "Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor," St. Martin's Press (2018).
[6] Kate Crawford and Ryan Calo, "There is a blind spot in AI research," Nature, vol. 538, no. 7625 (2016), https://doi.org/10.1038/538311a.

While the implementation of AI technology in the penal rights system has the potential to revolutionize the criminal justice system, there are also significant challenges that must be considered. The risk of algorithmic bias, ethical concerns around transparency and accountability, and the potential for a focus on risk management over rehabilitation, are only a few of these challenges. Therefore, it is crucial to address these issues proactively to ensure the ethical and effective use of AI technology in the criminal justice system.

1. THE ETHICAL IMPLICATIONS OF AI IN THE PENAL LEGAL SYSTEM

In recent times, the effect of AI on the criminal justice system has become a subject of growing attention and anxiety. AI has the potential to improve the efficiency and fairness of the justice system, but it also raises ethical and legal issues that need to be addressed[7]. Some of the ways in which AI is being used in the penal justice system include:

a. *Risk assessment*: AI algorithms can be used to analyze large amounts of data to predict the likelihood of a defendant reoffending or failing to appear in court[8]. These risk assessments can inform decisions about bail, sentencing, and parole.
b. AI has the potential to analyze crime data and forecast crime hotspots through predictive policing, enabling law enforcement agencies to optimize resource allocations[9].
c. *Sentencing and parole decisions*: AI can be used to analyze factors such as a defendant's criminal history, socioeconomic status, and education level to inform sentencing and parole decisions[10].

However, there are concerns that AI may perpetuate and amplify biases that already exist in the justice system, such as racial or socioeconomic biases[11]. There is also a lack of transparency around how these algorithms are developed and how they make decisions, which can make it difficult to ensure that they are fair and unbiased[12]. As a result, it is important to carefully consider the ethical implications of using AI in the penal justice system and to ensure that adequate safeguards are in place to address potential biases and protect the rights of defendants.

It is understandable that the use of artificial intelligence (AI) in the penal justice system raises significant ethical implications. Some of the most notable ethical implications include:

I. *Bias:* The extent of AI algorithm bias is directly related to the quality and fairness of the data sets used for their training, and there is a risk that these algorithms may perpetuate and amplify existing biases in the justice system such as racial, gender, or socioeconomic biases.

[7] Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2). Retrieved from https://doi.org/10.1177/2053951716679679 .
[8] Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. Science Advances, 4(1). Retrieved from https://doi.org/10.1126/sciadv.aao5580 .
[9] Lum, K., & Isaac, W. (2016). To predict and serve?. Significance, 13(5), 14-19.
[10] Veron, R. (2019). AI in Criminal Justice: Five Risks to Address. IEEE Technology and Society Magazine, 38(2), 56-65.
[11] Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks: ProPublica. Accessed 01.05.2023 from: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
[12] Kleinberg, J., Mullainathan, S., & Raghavan, M. (2017). Inherent trade-offs in the fair determination of risk scores. Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT*), 5-7.

In the context of the penal justice system, AI algorithms are only as unbiased as the data they are trained on[13], and there is a risk that these algorithms may perpetuate and amplify existing biases in the justice system such as racial[14], gender[15], or socioeconomic biases. This is because AI systems rely on historical data to identify patterns and establish correlations, and if historical data is biased, then the AI system is likely to reproduce these biases in its decisions[16]. For example, if an AI system is trained on the historical sentencing data of a judicial system which has a record of being biased against people of color, then the AI system is likely to recommend harsher sentences for people of color. It is important to address this issue of bias in AI systems in order to ensure that they do not perpetuate existing injustices in the criminal justice system. One way this can be addressed is through the development of AI systems that are explicitly designed to be fair and unbiased[17]. However, this requires a thorough understanding of the sources of bias, and the development of methods to identify and mitigate these biases as they arise.

II. *Lack of transparency*: The development process and decision-making mechanism behind AI algorithms are often not transparent or adequately disclosed. This lack of transparency raises concerns about due process, fairness, and accountability.

In the context of the penal justice system, it is important that AI systems are transparent in their decision-making processes[18]. When an AI system makes a decision, it can be difficult to understand how it came to that decision, which can create issues of accountability and due process. With traditional human decision-making processes, it is possible to question and scrutinize the decision-making process, including the reasons behind the decision. However, AI systems can be opaque and difficult to comprehend, which can lead to questions about the legitimacy of the decision. One possible solution to this issue is to develop AI systems that can explain their decision-making processes in a way that is understandable by humans[19]. This can also improve trust in AI systems and increase their adoption in the justice system. Transparency is essential to ensuring that AI systems are accountable and consistent with ethical principles. Much work needs to be done to develop methods to evaluate and measure the transparency of AI systems. It is essential that we continue to work towards creating transparent AI systems to ensure that they are trustworthy and reliable tools to support important decisions in the criminal justice system.

[13] Hao, K. (2019). Why AI is a threat to democracy—and what we can do to stop it. MIT Technology Review .
[14] Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. Big Data, 5(2), 153-163
[15] Crawford, K., & Calo, R. (2016). There is a blind spot in AI research. Nature, 538(7625), 311-313. Retrieved from https://doi.org/10.1038/538311a .
[16] Barocas, S., Hardt, M., & Narayanan, A. (2013). Fairness and machine learning. In V. Mayer-Schönberger & K. Cukier (Eds.), Big Data: A Revolution That Will Transform How We Live, Work, and Think (pp. 175-192). Boston: Houghton Mifflin Harcourt.
[17] Kearns, M., with Roth, A. (2020). The Ethical Algorithm: The Science of Socially Aware Algorithm Design. New York: Oxford University Press.
[18] Lichtenberg, J. (2018). Algorithmic decision-making and the control problem. Ethics and Information Technology, 20(1), 5-14.
[19] Datta, A., Sen, S., & Zick, Y. (2016). Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 598-610.

III. *Privacy:* AI systems may rely on extensive data collection and analysis, raising ethical concerns about the privacy and confidentiality of defendants and others involved in the criminal justice system[20].

There is a concern that sensitive information may be collected by these systems and used in ways that could harm individuals' reputations or livelihoods[21]. Furthermore, the collection and analysis of data may be subject to bias, as algorithms are only as unbiased as the data they are fed[22]. As such, it is crucial that the privacy and confidentiality of individuals involved in the criminal justice system are respected when adopting AI technologies.

IV. *Accuracy:* AI algorithms may not be as accurate as advertised, which could lead to unfair decision-making or wrongful convictions.

AI algorithms are trained using data and patterns, but the accuracy of the algorithms may not always be as reliable as advertised. For instance, algorithmic bias can result in incorrect predictions and decisions, leading to potential wrongful convictions or release of dangerous criminals[23]. A significant concern is the accuracy of risk assessment tools which are used to determine the likelihood of a defendant committing a future crime or violating their parole. These tools rely on complex algorithms that take into account a variety of factors, such as the defendant's past criminal record and social background. However, inaccuracies in these algorithms can result in injustices such as overestimating the likelihood of recidivism for some defendants, leading to harsher sentences or parole restrictions[24]. Moreover, the accuracy of AI systems can be hampered by incomplete or biased data. Data can be incomplete if certain groups, such as minorities or low-income individuals, are underrepresented in the training data. In addition, if AI algorithms perpetuate existing biases in the criminal justice system, the decisions of the algorithm may be no more accurate than the existing system[25]

V. *Autonomy:* There is a concern that the increasing reliance on AI in the criminal justice system could undermine the autonomy of judges, lawyers, and defendants, potentially reducing the ability of human beings to exercise discretion and make nuanced judgments.

There is a concern that the increasing reliance on AI in the criminal justice system could undermine the autonomy of judges, lawyers, and defendants, potentially reducing the ability of human beings to exercise discretion and make nuanced judgments[26]. As AI systems become more advanced, there is a risk that they may be given undue weight and influence in decision-making, ultimately reducing the role and agency of human actors in the criminal justice system[27]. It is therefore crucial that the

[20] Richardson, R., Schultz, J., Crawford, K., & Calo, R. (2019). Predictive policing and reasonable suspicion. Geo. LJ, 107, 123-152.
[21] Kroll, J. A., Huey, J., Barocas, S., Felten, E., Reidenberg, J. R., Robinson, D. G., ... & Schultz, J. (2017). Accountable algorithms. University of Pennsylvania Law Review, 165, 633-705.
[22] O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Books.
[23] Berk, R., & Bleich, J. (2013). Statistical procedures for forecasting criminal behavior. Criminal Justice and Behavior, 40(12), 1320-1337.
[24] Alexander, M. (2018). Algorithms, Criminal Sentencing, and the Wider Implications of an Error Rate. South Carolina Law Review, 69, 717-742.
[25] Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumeé III, H., & Crawford, K. (2018). Datasheets for datasets. arXiv preprint arXiv:1803.09010. Retrieved from https://arxiv.org/abs/1803.09010 .
[26] Garfinkel, S. L. (2018). Law and artificial intelligence. Annual Review of Law and Social Science, 14, 385-404.
[27] Calo, R. (2017). AI and the death of Decisional Autonomy. In Robot Law (pp. 261-282). Edward Elgar Publishing.

use of AI in the criminal justice system is balanced with the ability of human actors to exercise discretion and make independent judgments.

VI. *Stigmatization:* The use of AI in the criminal justice system could lead to the stigmatization of certain groups or individuals, particularly those who have been classified as high-risk or dangerous. This could occur if the AI system incorrectly identifies certain individuals as high-risk or dangerous, leading to them being treated unfairly or subjected to unnecessary scrutiny. One example of this is the COMPAS[28] (Correctional Offender Management Profiling for Alternative Sanctions) system used in some US jurisdictions, which has been criticized for disproportionately labeling Black defendants as high-risk.

According to a report by the AI Now Institute, the use of predictive risk assessment tools in the criminal justice system may reinforce and exacerbate biases by relying on historical data that reflects and perpetuates discriminatory practices. The report also notes that labeling individuals as "high-risk" can have negative consequences such as limiting their access to employment and housing[29].

It is important for developers and users of AI in the criminal justice system to be aware of these potential stigmatization effects, and to prioritize fairness and non-discrimination in the development and implementation of these systems. This may include ongoing monitoring of the impact of AI tools and implementing safeguards to prevent their misuse.

VII. *Human rights violations*: There is a risk that the use of AI in the criminal justice system could violate fundamental human rights such as the right to a fair trial, the right to privacy, and the right to non-discrimination.

There is a concern that the use of AI in decision-making processes such as pretrial risk assessment or sentencing recommendations may lead to arbitrary or discriminatory outcomes, undermining the principles of due process and equal protection under the law.

As per the special rapporteur on extreme poverty and human rights of the United Nations, the use of predictive algorithms in the criminal justice system may violate human rights by denying individuals their right to individualized determinations and denying them the opportunity to present mitigating factors. The Special Rapporteur also notes that AI systems may exacerbate existing inequalities by relying on biased data and reinforcing societal prejudices[30]. To prevent these potential human rights violations, it is crucial to ensure that AI algorithms in the criminal justice system are designed and implemented in a manner that is consistent with international human rights standards. This may require increased transparency, accountability, and oversight in the development and use of AI tools, as well as regular assessments of their impact on human rights.

These and other ethical implications of AI in the criminal justice system need to be carefully considered as AI continues to play an increasingly prominent role in the administration of justice.

[28] Retrieved from Ethical Conversation Intelligence - Managing Bias | Symbl.ai
[29] AI Now Institute, "AI Now Report 2019", accessed online https://indiaai.gov.in/research-reports/ai-now-2019-report
[30] UN General Assembly, "Report of the Special Rapporteur on extreme poverty and human rights", accessed June 7, 2021, https://undocs.org/en/A/74/480.

Needless to say, there are established ethical guidelines that all AI developers should rely upon. Below are listed only a few of these ethical principles, such as:

- The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems offers guidance for the ethical development and deployment of AI and autonomous systems through a framework[31].
- The Asilomar AI Principles[32] outline a set of principles for AI to be aligned with human values and ethical considerations.
- UNESCO has also published a set of guidelines for ethical AI[33].

Moreover, there are some legal frameworks and regulations that are relevant to AI in the penal rights system.

a. General Data Protection Regulation (GDPR)[34] - Introduced by the European Union to regulate the collection, processing, and storage of personal data.
b. Convention on Cybercrime[35] - Developed by the Council of Europe to prevent and combat cybercrime, including computer-related fraud, forgery, and hacking.
c. UN Guiding Principles on Business and Human Rights[36] - Outlines the responsibility of companies to respect human rights, including the potential negative impacts of AI on human rights.
d. Universal Declaration of Human Rights[37] - Guarantees certain fundamental human rights, including the right to privacy, which should be considered in the development and use of AI applications.
e. Rome Statute of the International Criminal Court[38] - Outlines the jurisdiction and powers of the international criminal court, which could potentially prosecute individuals or companies responsible for AI-related crimes.
f. Convention on the Rights of Persons with Disabilities (CRPD)[39] - Guarantees equal treatment and non-discrimination for persons with disabilities, including with regards to access to technology and digital services.
g. United Nations Convention against Transnational Organized Crime[40] - Outlines measures to prevent and combat organized crime, which could be related to the development and use of AI for illegal activities.
h. United Nations Guiding Principles for Artificial Intelligence[41] - Developed to assist with ethical considerations in the development and use of AI, including accountability, transparency, and human rights.

[31] https://ethicsinaction.ieee.org/
[32] https://futureoflife.org/ai-principles/
[33] https://unesdoc.unesco.org/ark:/48223/pf0000377897
[34] "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)".
https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN
[35] https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/185
[36] https://www.ohchr.org/sites/default/files/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf
[37] https://www.un.org/en/universal-declaration-human-rights/
[38] https://www.icc-cpi.int/sites/default/files/RS-Eng.pdf
[39]https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities/convention-on-the-rights-of-persons-with-disabilities-2.html
[40] https://www.unodc.org/documents/treaties/UNTOC/Publications/TOC%20Convention/TOCebook-e.pdf
[41]https://unsceb.org/sites/default/files/2022-09/Principles%20for%20the%20Ethical%20Use%20of%20AI%20in%20the%20UN%20System_1.pdf

i. The European Convention on Human Rights (ECHR)[42] - Guarantees certain fundamental rights, including the right to a fair trial, which could be impacted by the use of AI in judicial decision-making.
j. Convention on the Prohibition of Military or any other Hostile Use of Environmental Modification Techniques (ENMOD)[43] - Prohibits the military or hostile use of environmental modification techniques, which could potentially include AI-assisted climate modification.

2. THE IMPACT OF AI IN THE PENAL LEGAL SYSTEM

There are various forms of AI that are proved useful in the penal system to predict an offender's likelihood of reoffending. These AI systems rely on a legal framework that helps judges make informed decisions about the offender's risk level and to determine their suitability for parole or early release.

Here are some examples of different types of AI used in the penal system:

i. *Recidivism prediction models*: Machine learning algorithms that analyze an offender's criminal history, demographic characteristics, and other factors to estimate the likelihood of them committing another crime.

The use of recidivism prediction models is a growing trend in the criminal justice system, particularly in the United States. These models use sophisticated machine learning algorithms to analyze a range of data points - such as an offender's age, gender, prior offenses, history of substance abuse, and employment status - to determine the likelihood of that person committing another crime in the future[44]. This approach is often used to inform decisions around pretrial detention, plea bargaining, and post-release planning, with the goal of reducing the overall rate of recidivism[45]. However, there are concerns about the fairness and accuracy of these algorithms, particularly with respect to potential biases in the data used to train them and issues with transparency and accountability[46].

For example, a study by the non-profit organization Upturn found that some commonly used risk assessment tools had significantly higher error rates when assessing risk for Black defendants, potentially leading to unjustified detention and harsher sentences[47]. As such, the use of AI in the judicial system is often subject to debate and scrutiny, with questions around the potential benefits and drawbacks of relying on automated decision-making tools in such a high-stakes context.

One example of an app for recidivism prediction models AI is COMPAS (Correctional Offender Management Profiling for Alternative Sanctions)[48].

[42] https://www.echr.coe.int/Documents/Convention_ENG.pdf
[43] https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XXVI-1&chapter=26&clang=_en
[44] Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks: ProPublica. Accessed 01.05.2023 from: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
[45] Pew Charitable Trusts. (2011). The State of Recidivism: The Revolving Door of America's Prisons. Retrieved from https://www.pewtrusts.org/en/research-and-analysis/reports/2011/04/12/state-of-recidivism-the-revolving-door-of-americas-prisons
[46] Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks: ProPublica. Accessed 01.05.2023 from: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
[47] Rudin, C., & Carlson, D. (2019). The Disparate Impact of Risk Assessments. University of Chicago Law Review, 86, 1-40.
[48] COMPAS is a software tool that assesses the risk of recidivism for individuals who have been arrested or convicted of a crime. The system uses an algorithm to analyze a variety of factors, including criminal history, age, gender, education level, employment status, and substance abuse history, and produces a score that predicts the likelihood of

ii. Sentencing recommendation systems: AI tools that assist judges in determining the appropriate punishment for a given offense based on legal guidelines, previous case decisions, and other data.

Sentencing recommendation systems are another example of AI being used in the penal system. These systems use algorithms to analyze vast amounts of legal data, including previous cases and sentencing guidelines, to help judges determine appropriate sentences for offenders[49]. By providing more consistent and unbiased sentencing decisions, these tools could help reduce disparities in sentencing outcomes between different judges, defendants, and groups[50]. There are also potential drawbacks to relying solely on AI in determining sentences. Critics of these systems argue that they may not take into account the unique circumstances of each individual case, or that they could be subject to bias issues similar to those found in recidivism prediction models[51]. Additionally, some argue that the use of AI tools in the justice system could reinforce existing structural inequalities, by using data that already reflects systemic biases and perpetuating these biases in decision-making. As such, the use of AI in sentencing decisions is often debated among legal experts and policymakers, with questions around how to balance the benefits of increased consistency and efficiency against concerns about fairness, transparency, and due process.

One example of an app for sentencing recommendation systems AI is the "Risk Management System" (RMS) by Corrections Victoria in Australia[52].

iii. Biometric identification and monitoring technologies: Facial recognition, voice analysis, and other tools that can identify individuals and track their movements within correctional facilities or in the community.

Biometric identification and monitoring technologies are becoming increasingly prevalent in the penal system. Some examples include:

- *Facial recognition*: This technology is used to identify individuals in correctional facilities or in public areas, such as near parole offices or probation centers. It is also used to monitor the movements of visitors, staff, and inmates in and out of these facilities. However, critics have raised concerns about the accuracy of facial recognition technology, as it can disproportionately misidentify people of color and women[53].
- *Voice analysis*: Voice recognition software can be used to verify the identity of inmates making phone calls or participating in virtual visits[54]. It can also be used to detect changes in an offender's vocal patterns that may indicate changes in their mental state.

reoffending within two years. COMPAS has been used by a number of U.S. states and other jurisdictions to help inform decisions on pretrial release, sentencing, and parole.

[49] Goh, D. (2021). Inside AI's impact on the justice system. World Economic Forum.

[50] Stolbach, B., Gunderson, J., & Roberts, J. V. (2020). Artificial Intelligence-Based Sentencing Recommendations: An Overview of the Values and Ethical Issues. The American Journal of Bioethics, 20(10), 7-16.

[51] Vazquez, J. (2019). Artificial Intelligence and Sentencing: A Primer. American Bar Association.

[52] This system uses AI to analyze data and provides recommendations to the courts on the level of risk a person may pose to the community and what support is required to reduce that risk. The RMS analyzes numerous factors such as past criminal history, the seriousness of the offense, age, and gender, among others, to create a profile of the offender. From there, the system generates a risk score and presents options for sentencing and treatment based on the individual's risk level.

[53] Buolamwini, J. and Gebru, T. (2018) 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification', Proceedings of the 1st Conference on Fairness, Accountability and Transparency, pp. 77-91.

[54] Bradford, B., Schuilenburg, M., and Van der Leun, J. (2014) 'The Emergence of Smart Surveillance Technologies in the Fight Against Crime: A European Perspective', Policing: A Journal of Policy and Practice, 8(2), pp. 175-187.

- *Biometric tracking devices*: GPS-enabled ankle monitors and other wearable technologies are increasingly used to monitor low-risk offenders in the community, rather than keeping them in jail or prison. Some devices can track the wearer's location in real-time, while others measure their physiological responses, such as heart rate and skin conductivity[55]. However, critics argue that these devices can be intrusive and stigmatizing for those wearing them, and that they can sometimes be inaccurate or malfunction[56].

Below are listed some applications that use biometric identification and monitoring technologies AI in the penal rights system:

- The "Gang Intelligence Application"[57] ;
- the "Biometric Identity Management System"[58];
- the "Guardian"[59] app.

iv. *Natural language processing and sentiment analysis*: AI techniques that analyze written or spoken communications from offenders, either to assess their mental state or to flag potential threats or violations.

Natural language processing and sentiment analysis are becoming increasingly important tools in the penal system. Some examples include:

- Sentiment analysis of inmate communications: Some correctional facilities are using sentiment analysis to monitor written or spoken communications from inmates, looking for indicators of mental health issues, threats, or other potential problems[60]. This can allow staff to intervene before a situation escalates.
- Language analysis in writing samples: Artificial intelligence can analyze the language patterns and word choice in written correspondence from inmates, which may provide insights into their thought processes and potential behavioral issues[61].
- Translation services: AI-powered translation tools can be used to help non-English speakers communicate with prison staff or navigate legal documents, ensuring that they have access to the same resources as other inmates[62].

[55] Simonite, T. (2019) 'Electronic Ankle Monitors Were Already Invading Our Privacy. They're About to Get Worse', Wired Magazine.
[56] La Vigne, N. et al. (2018) 'Electronic Monitoring in the Criminal Justice System: The Promise and the Perils', RAND Corporation.
[57] This app uses facial recognition to identify gang members and associates who may be entering the facility to visit inmates.
[58] This app is used by the Federal Bureau of Prisons, which includes both facial recognition and fingerprint scanning technology to identify inmates and staff members. The system is used for a range of purposes, from managing inmate movements and reports to tracking employee attendance and managing access to restricted areas.
[59] This app, designed by a company called Vigilant Solutions, provides a comprehensive overview of all inmate movement, scheduling, and activity within the correctional facility. The Guardian app uses data from a variety of sources, including video cameras, RFID sensor systems, and other tracking technologies, to compile a real-time picture of inmate behavior and facility operations. This data is then analyzed by AI algorithms to identify potential security risks or other issues that require attention. The app can be accessed by correctional officers, administrators, and other authorized personnel on their mobile devices, allowing them to monitor the facility and respond quickly to any emerging situations. The Guardian app also includes features such as automated alerts, customizable reporting, and analytics tools, making it a powerful tool for managing complex correctional environments.
[60] Crouch, D. J., & Supangan, R. (2019). Predictive analytics in correctional facilities: An overview of the landscape. The Prison Journal, 99(4), 451-474.
[61] Marder, J. (2018) 'AI Is Changing the Game for Writing Analysis and Prediction', Forbes.
[62] Chaudhary, N. and Chaudhary, V. (2018) 'Machine Translation for Legal Domain: A Comparative Study', Proceedings of the 3rd International Conference on Intelligent Computing and Control Systems, pp. 440-445.

Natural language processing and sentiment analysis can help correctional staff quickly identify potential issues and intervene before they become serious problems. However, there are also concerns about privacy and the potential for biases in AI algorithms used in this context.

One example of an app for natural language processing and sentiment analysis AI used in penal rights systems is "Prison Grievances"[63].

v. *Prison management systems*: Automated platforms that handle tasks like inmate classification, scheduling, and resource allocation to improve the efficiency and safety of correctional institutions.

Prison management systems are a type of AI used to streamline operations within correctional institutions. These platforms automate tasks like inmate classification, scheduling, and resource allocation, which can reduce the workload of prison staff and help prevent errors or security breaches. Some specific examples of prison management systems include:

- Offender management systems: These platforms track information about inmates, including their offense history, medical records, and work assignments. This information can be used to determine housing assignments and to identify inmates who may require specialized care or monitoring.
- Electronic monitoring systems: These platforms use GPS monitoring technology to track the movements of offenders who are released into the community on parole or probation. This allows correctional staff to monitor compliance with the conditions of release and to respond quickly if an offender violates those conditions.
- Automated scheduling systems: These platforms assign work and other activities to inmates based on their skill level, availability, and other factors. By automating this process, prison staff can ensure that each inmate is assigned to the most appropriate activities and that sufficient staffing is available to provide security and support.

While prison management systems can improve the efficiency of correctional institutions, there are also concerns about the potential for these systems to perpetuate biases against certain groups of inmates or to violate inmates' privacy rights[64].

One example of an app for prison management systems AI used in penal rights systems is "SmartPrison"[65].

[63] This app, uses AI to analyze and classify grievances filed by inmates within correctional facilities.The app collects data from grievance forms and other sources, then applies natural language processing algorithms to extract key information such as the subject of the grievance, the severity of the issue, and the sentiment of the inmate's complaint. The app also uses sentiment analysis to categorize grievances as positive, negative, or neutral, providing correctional staff with an overview of inmate attitudes and opinions. Correctional staff can access the app on their mobile devices to review, track, and respond to grievances filed by inmates. The app includes features such as customizable reporting, automated notifications, and a centralized database of grievances, making it easier for staff to manage and address issues within the facility.

[64] Mackey, T. K. (2019) 'Big Data and Corrections: Assessing the Emerging Role of Technology in Penal Reform', Big Data & Society, 6(1).

[65] This app uses AI-powered technologies to improve the efficiency and security of prisons. It provides a range of features for correctional staff, including facial recognition, biometric identification for access control, and video surveillance. It also uses predictive analytics to detect potential security threats and monitor inmate behavior in real-time. One of the key benefits of the SmartPrison app is that it can help to reduce incidents of violence and other security breaches in correctional facilities. By providing staff with accurate and timely information about potential risks, it allows them to take proactive measures to prevent incidents from occurring. The app also includes features for managing inmate health and wellbeing, such as monitoring medication schedules and tracking mental health

3. WHAT ARE THE POTENTIAL IMPACTS OF AI VULNERABILITIES ON THE PENAL LEGAL SYSTEM?

There are several AI vulnerabilities that can impact the penal rights system.

Firstly, biased algorithms can lead to discrimination against certain groups, such as minorities and low-income individuals, in decision-making processes such as risk assessment and sentencing. This can result in unfair treatment and violation of their fundamental rights[66].

Biased algorithms can have significant impacts on the penal rights system, potentially leading to significant unfairness and violating fundamental rights. One key issue is that many AI algorithms used in these systems are often trained on biased datasets, which can lead to discriminatory results. This is often due to historical biases and discrimination in the criminal justice system itself, which can then be reinforced and amplified by AI systems. For example, risk assessment tools may be biased against certain groups due to factors such as unequal access to legal representation or over-policing in certain communities[67].

Another significant concern is the lack of transparency and accountability in many AI systems used in the penal system. This can make it difficult to detect and correct biases, as well as to hold accountable those responsible for creating and implementing these systems. As a result, even unintentional biases may go unchecked and contribute to significant injustice[68].

More broadly, there is a risk that AI systems could lead to the mass surveillance of individuals, which could also violate fundamental rights. For example, facial recognition technology has often been criticized for its potential to infringe on privacy and civil liberties, particularly if used without appropriate safeguards or oversight. There is also the risk that use of these systems could be extended beyond their intended purpose, leading to overreliance on AI systems and potentially allowing them to replace human decision-making entirely. This could undermine important legal principles such as due process and the right to a fair trial[69].

Secondly, lack of transparency and accountability in AI systems can make it difficult to understand and challenge decisions made by these systems. This can also result in violation of due process rights[70].

One key issue is that these systems often rely on complex algorithms that can be difficult to understand or interpret, making it challenging for defendants or lawyers to challenge any decisions made by the system. This is particularly problematic when it comes to risk assessment tools or pretrial detention decisions, which can have significant impacts on a defendant's rights and freedoms[71]. Moreover, these systems are often developed without sufficient input from affected communities or stakeholders,

indicators. This can help to improve the overall quality of life for inmates, while also reducing the workload for correctional staff.

[66] Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks: ProPublica. Accessed 01.05.2023 from: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[67] Tong, Y., & Ribeiro, S. (2021). A systematic review of bias in algorithmic risk assessment tools in the criminal justice system. ACM Computing Surveys (CSUR), 54(1), 1-39.

[68] Selbst, A. D., Boyd, D., Friedler, S. A., Kinney, B., & West, S. M. (2019). Fairness and Abstraction in Sociotechnical Systems. Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19.

[69] Custers, B., Calders, T., & Schermer, B. (2020). The limits of dataprotection law in regulating profiling in the gig economy. Computer Law & Security Review, 39, 105394.

[70] Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389-399. Accessed 01.05.2023 from: https://www.nature.com/articles/s42256-019-0088-2 .

[71] Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. ProPublica.

which can lead to significant distrust and lack of legitimacy in the criminal justice system. This can be exacerbated by the fact that many AI systems used in the penal system are proprietary and owned by private companies, making it difficult for outside experts or civil society groups to evaluate their efficacy or fairness[72]. In addition to these challenges, there is also the risk that AI systems could reinforce or exacerbate existing biases and discrimination in the criminal justice system. For example, risk assessment tools may be trained on historical data that reflects biases or inequities in the criminal justice system, leading to unfair and discriminatory outcomes for certain groups. Without appropriate oversight and accountability mechanisms, these types of biases may go unchecked and contribute to further inequality and injustice[73].

Thirdly, cybersecurity threats can compromise the integrity and confidentiality of sensitive information, such as criminal records and personal information of defendants and victims. This can lead to identity theft and fraud, as well as wrongful conviction and sentencing[74].

Cybersecurity threats pose a significant risk to the integrity and confidentiality of sensitive information in the criminal justice system. One key concern is the potential for data breaches or hacking attacks that could compromise the security of criminal records and personal information of defendants and victims. This could include sensitive data such as fingerprints, DNA samples, and other biometric information that is used to identify suspects and support criminal investigations. If this data falls into the wrong hands, it could be used for identity theft, fraud, or other malicious purposes that can lead to financial harm and reputational damage[75]. Moreover, data breaches or unauthorized access to criminal justice data can also undermine the credibility and fairness of the legal system itself. If sensitive information is leaked or misused, it could result in wrongful conviction and sentencing, as well as other miscarriages of justice. This is particularly concerning given the increasing reliance on digital tools and AI systems in the criminal justice process, which could amplify the impact of any security breaches or data leaks[76]. One issue worth considering is the risk of insider threats and misuse of criminal justice data by individuals within the system. Such misuse may involve illicit access, unauthorized disclosure of sensitive information, or even corruption by law enforcement officials and others with authority. If there are no suitable safeguards or monitoring mechanisms in place, it can be challenging to identify and prevent such misconduct, which can further erode public trust in the legal system.

Fourthly, the use of AI in policing activities, such as predictive policing, can lead to unethical profiling and surveillance of individuals, which may also violate their right to privacy[77].

---

[72] Boyd, D., Lassiter, B., & Bagus, A. (2019). Promoting responsible AI in the criminal legal system through interdisciplinary research, public engagement, and advocacy. Philosophy & Technology, 33(3), 409-423.

[73] Raji, I. D., & Buolamwini, J. (2019). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. Conference on Fairness, Accountability and Transparency. Accessed 01.05.2023 from https://www.media.mit.edu/publications/actionable-auditing-investigating-the-impact-of-publicly-naming-biased-performance-results-of-commercial-ai-products/ .

[74] Edwards, L. J., Holt, T. J., & Turner-McGrievy, G. M. (2020). Cybersecurity and criminal justice: International perspectives. Routledge.

[75] Mark, G. (2018). Cybersecurity and data protection in law enforcement: A review of current practices. Computer Law & Security Review, 34(5), 997-1008.

[76] Collier, R., & Bond, J. (2018). Cybersecurity and the criminal justice system: Understanding the risks, threats, and implications for the public and private sector. Security Journal, 31(4), 1036-1055.

[77] Selinger, E., & Hartzog, W. (2018). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. Harvard Kennedy School. Accessed 01.05.2023 from https://www.turing.ac.uk/sites/default/files/2019-08/understanding_artificial_intelligence_ethics_and_safety.pdf

The use of AI in policing activities, such as predictive policing, has raised concerns about potential ethical issues and violations of individuals' rights to privacy[78]. Predictive policing involves using AI algorithms to analyze data and predict where crime is likely to occur in the future. The data used may include information on past criminal activity, demographics, and other socio-economic factors[79]. However, this approach can lead to unethical profiling and surveillance of individuals based on age, race, gender, or other characteristics, leading to discriminatory practices within law enforcement. Predictive policing also raises concerns about the accuracy of the AI algorithms used, as they may be based on biased datasets and can perpetuate historical discrimination and inequalities[80]. The use of AI in policing activities can also lead to violations of individuals' right to privacy. The data analyzed by AI systems may include sensitive personal information, such as biometric data or online activity, leading to intrusive surveillance and potential abuse by law enforcement. This can create a chilling effect on freedom of speech and expression, as individuals may feel hesitant to express their opinions or engage in certain activities for fear of being monitored. In conclusion, the use of AI in policing raises important ethical and privacy concerns that must be carefully considered and addressed. Law enforcement agencies must ensure that the use of AI is based on unbiased and accurate data, and that individuals' rights to privacy and non-discrimination are protected[81].

Lastly, AI systems can be hacked just like any other software system. Hackers can manipulate the input or output of an AI system to affect its behavior, which can lead to biased or incorrect predictions or decisions.

AI systems used in criminal justice rely on large amounts of data and proprietary algorithms, which if hacked, could be manipulated or corrupted, potentially leading to inaccurate or biased results and unfair sentencing outcomes. Moreover, AI-based attacks could be used by hackers to target vulnerable points in the criminal justice system, including the identification of security protocol weaknesses and the execution of sophisticated phishing attacks on specific individuals or departments. In addition to these technical risks, the use of AI in the penal rights system also raises important ethical and legal concerns. For example, some experts have raised concerns about the lack of transparency and accountability in many AI systems used in criminal justice, as well as the potential for these systems to exacerbate existing biases and inequalities. If hackers gain access to these systems, they could amplify these risks, further eroding trust in the legal system[82]. Some of the most relevant crimes a person can commit while hacking AI include:

- Unauthorized access: Gaining access to an AI system without proper authorization is considered a crime. This includes exploiting vulnerabilities to circumvent security measures and accessing data or functionality that the hacker is not authorized to use[83].

[78] Electronic Frontier Foundation. (2019). Artificial Intelligence and Law Enforcement: Opportunities and Risks. Retrieved from https://www.eff.org/wp/artificial-intelligence-and-law-enforcement-opportunities-and-risks
[79] Sengupta, S. (2019). The Problem With Predictive Policing. The New York Times. Retrieved from https://www.nytimes.com/2019/02/08/opinion/predictive-policing-black-people.html
[80] Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Conference on Fairness, Accountability and Transparency, 2018. Retrieved from https://doi.org/10.1145/3270101.3270103
[81] Electronic Frontier Foundation. (2021). Police Use of Artificial Intelligence: 2021 in Review. Accessed on 01.05.2023 from https://www.eff.org/deeplinks/2021/12/police-use-artificial-intelligence-2021-review
[82] Amoore, L., de Goede, M., Piotukh, V., & Bachmann, L. M. (2020). (Re)configuring the spaces of algorithmic governance: From security to public reason. Public Administration, 98(3), 524-539. Published by Routledge.
[83] U.S. Department of Justice. (2020). Computer Crime and Intellectual Property Section (CCIPS). Retrieved from https://www.justice.gov/criminal-ccips/computer-crime-and-intellectual-property-section

- Theft of IP: Theft of intellectual property (IP) is also a common crime committed during AI hacking. This includes stealing algorithms, models, or datasets that the hacker can use to train their own AI systems or sell them on the black market[84].
- Cyberstalking: Cyberstalking involves targeting an individual or group and using AI systems to monitor and track their activities, such as online behavior, location, and personal information. This can lead to harassment, threats, or other forms of cybercrime[85].
- Financial fraud: AI hacking can also involve financial fraud, such as using AI to generate fake data or transactions, manipulate stock prices or other financial markets, or conduct phishing attacks[86].

These crimes are serious offenses that can result in severe consequences, including fines, imprisonment, and other legal penalties. It is essential to report any suspected AI hacking incidents to the proper authorities and take steps to prevent further attacks.

Overall, these AI vulnerabilities can negatively impact the fairness and effectiveness of the criminal justice system, as well as the protection of individual rights and freedoms.

CONSIDERATIONS AND RECOMMENDATIONS

It is immensely important to make sure that AI decisions are transparent, explainable, and subject to review[87].

This is particularly relevant when it comes to the use of AI in making decisions that have a significant impact on individual rights and freedoms, such as sentencing, bail decisions, and parole decisions[88]. In these cases, the lack of transparency in the decision-making process can lead to a loss of trust in the legal system and an increased risk of discrimination and bias[89].

To address these concerns, many legal systems are implementing requirements for transparency and explainability in AI-based decision-making[90]. For example, the General Data Protection Regulation (GDPR) of the EU includes a "right to explanation" provision, which requires organizations using automated decision-making systems to provide individuals with a clear explanation of how decisions are made[91].

Moreover, some jurisdictions have implemented guidelines stating that AI systems must be open to audit and review by human experts to ensure that their

[84] Techopedia. (2019). Artificial Intelligence (AI) Crime. Retrieved from https://www.techopedia.com/definition/30229/artificial-intelligence-ai-crime

[85] U.S. Department of Justice. (2020). Computer Crime and Intellectual Property Section (CCIPS). Accessed on 01.05.2023 from https://www.justice.gov/criminal-ccips

[86] Dakalbab, F., Abu Talib, M., Abu Waraga, O., Bou Nassif, A., Abbas, S., & Nasir, Q. (2022), Artificial intelligence & crime prediction: A systematic literature review Social Sciences & Humanities Open, Volume 6, Issue 1, 2022, 100342.
Accessed on 01.05.2023 from https://www.sciencedirect.com/science/article/pii/S2590291122000961

[87] Makridakis, Spyros, and Ilias G. Tatsiopoulos. "The ethical challenges of AI in criminal justice." Big Data & Society 5.2 (2018)

[88]Spohn, Cassia, and Natasha Madon. "Probation and parole: The evolution of supervision." The Handbook of Community Corrections (2013): 55-75.

[89] Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks: ProPublica. Accessed 01.05.2023 from: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[90] Sengupta, Anindya, et al. "Towards accountable AI: Hybrid human-machine predictions for the legal system." arXiv preprint arXiv:1807.05556 (2018).

[91] Article 22(3) GDPR - Automated individual decision-making, including profiling. https://gdpr-info.eu/art-22-gdpr/

decisions are fair and unbiased[92]. For instance, the UK government has produced a set of guidelines on the use of AI in the criminal justice system, which states that any AI system used must be "transparent, understandable and subject to appropriate oversight and accountability"[93].

Given the significant impact that AI can have on individual rights and freedoms in penal legal systems, it is critical that these systems are developed and implemented in a way that ensures transparency and accountability[94].

By following these recommendations, law enforcement agencies can help to ensure that the use of AI in penal rights systems is conducted in a manner that is transparent, fair, and respectful of individuals' rights.

- Use unbiased and representative data: Law enforcement agencies should prioritize the use of unbiased and diverse data in AI systems used in penal contexts. They must ensure that the data sets used are representative of the population and do not perpetuate past discrimination and inequalities.
- Conduct regular audits and evaluations: Regular audits and evaluations of AI systems can help to identify and address any potential biases or discriminatory practices. These audits should be conducted by external, independent experts with appropriate technical expertise and subject-matter knowledge.
- Develop clear guidelines and standards: Law enforcement agencies must develop clear guidelines and standards for the use of AI in penal contexts. These guidelines should outline what data is used, how it is collected, and how it is analyzed, as well as specifying appropriate protocols for transparency and accountability.
- Ensure accountability and transparency: Law enforcement agencies must ensure that the use of AI in penal contexts is transparent and accountable. This includes providing clear and easily understandable explanations of how AI is being used, as well as ensuring that decisions made by AI systems can be explained and challenged where necessary.
- Foster public engagement and dialogue: Finally, it is essential that the use of AI in penal contexts is subject to open and informed public engagement and dialogue. This involves engaging with affected communities, civil society, and other stakeholders to ensure that there is ongoing dialogue and feedback on the use of AI in penal rights systems.

The increasing use of AI in the penal legal system raises important ethical and legal questions. Here are three key recommendations on the impact of AI in the penal legal system:

- Ensuring transparency and explainability in AI systems used in the penal legal system is a critical issue that requires a range of measures. Here are some ways in which this can be achieved:
    - *Standardization*: Develop standardized practices and benchmarks that establish clear criteria for AI-based

[92] Office of the Privacy Commissioner of Canada. "Guidelines for the use of artificial intelligence in human resources management." Government of Canada,
[93] . UK Ministry of Justice. "Algorithmic decision-making and the use of artificial intelligence in the criminal justice system: Interim guidance." Government of the United Kingdom,
[94] Green, Ben, and Brent Mittelstadt. "Artificial intelligence and accountability in the legal industry." Artificial Intelligence and Law 24.3 (2016): 285-305.

decisions in the penal legal system. By standardizing these practices, it will be easier to ensure consistency and transparency in how AI systems operate.

- *Auditing:* Implementing auditing processes that can identify and analyze any irregularities in the AI systems used in the penal legal system. Regular auditing can ensure that the AI systems' decisions are fair, unbiased, and transparent.
- *Clear documentation*: Ensure that AI systems used in the penal legal system include comprehensive documentation that clearly outlines how the software makes its decisions. This will enable legal professionals to track and analyze the decision-making process.
- *Independent reviews*: Conduct independent third-party reviews of AI systems used in the penal legal system to assess their transparency, bias, and accuracy. Independent reviews can provide valuable insights into the performance of these systems and help to identify areas that need improvement.
- *Training and education*: Train and educate legal professionals, including judges, prosecutors, and defense attorneys, about how AI systems work and their decision-making criteria. This will help legal professionals understand the limitations and benefits of AI systems and make informed decisions about their use in the legal system. Create appropriate oversight and regulation: To protect the rights of individuals, AI systems used in the penal legal system must be appropriately overseen and regulated to ensure that they do not discriminate or violate human dignity. This includes developing specialized standards, standardized testing, and independent reviews of such systems.

- Guard against discrimination and inequality: There is a risk that AI systems will perpetuate existing discrimination and inequality. A robust and effective antidiscrimination policy must be established to ensure that the use of AI in the penal legal system does not result in discrimination against historically marginalized groups. In addition, strategies need to be developed proactively to diversity recruitment and address AI system limitations in including underrepresented and disadvantaged groups.
- Create appropriate oversight and regulation: To protect the rights of individuals, AI systems used in the penal legal system must be appropriately overseen and regulated to ensure that they do not discriminate or violate human dignity. This includes developing specialized standards, standardized testing, and independent reviews of such systems. The following are some ways we recommend to guard against discrimination and inequality in the use of AI in the penal legal system:
    - *Collect and analyze data*: Collecting data on a wide range of factors, including race, gender, age, and socio-economic status, can help identify any biases in AI systems. This data can then be analyzed to detect and correct any patterns of discrimination that may emerge.

- *Test for fairness and bias*: AI systems can be tested for fairness and bias through various methods such as adversarial testing, sensitivity analysis and human testing. These tests can help identify and address any existing biases or unfairness in AI-based decision-making.
- *Establish explicit rules and guidelines*: Establishing clear and explicit guidelines and rules for the use of AI in the justice system can help ensure that decisions made by AI systems are free from discrimination and bias. This can include developing restrictions on the use of certain types of data, such as race or ethnicity, that may be unfairly used in determining outcomes.
- *Foster diversity*: Diverse teams involved in the development of AI systems can help identify different biases and perspectives, which can ultimately help ensure that AI systems are more comprehensive and unbiased. Strategies for diversity recruitment and inclusion may include partnerships with diverse organizations, recruiting from underrepresented groups, and supporting community outreach programs.

Lastly, here are some recommendations on how to prevent AI vulnerabilities affecting the penal legal system:

- *Regular testing and auditing of AI systems*: AI systems used in the penal legal system should be regularly tested and audited to identify any vulnerabilities. This can help prevent unintended consequences or errors caused by machine learning algorithms. Regular testing can also ensure that the system is operating as intended, and adjustments can be made to prevent errors or bias from affecting outcomes.
- *Ethical and legal framework for AI use*: The use of AI in the penal legal system should be governed by an ethical and legal framework that ensures that the technology is used responsibly and transparently. This framework can include protocols for collecting and processing data, guidelines for decision-making algorithms, and methods for ensuring accountability and oversight.
- *Collaboration between experts and stakeholders*: Collaboration between experts in AI and stakeholders in the penal legal system, including lawyers, judges, and policymakers, can help ensure that the technology is implemented in a way that meets the needs of the system and addresses potential biases or vulnerabilities. Open communication and transparency can also help stakeholders better understand the benefits and limitations of AI in the penal legal system, ultimately improving outcomes for all parties involved.