ST. TERESA'S COLLEGE(AUTONOMOUS), ERNAKULAM

AFFILIATED TO MAHATMA GANDHI UNIVERSITY

ETHICAL CONCERNS OF USING AI IN HEALTHCARE AND THE CRIMINAL JUSTICE SYSTEM

PROJECT REPORT

In partial fulfilment of the requirements for the award of the degree of

BACHELOR OF SCIENCE IN COMPUTER

APPLICATIONS [TRIPLE MAIN]

Submitted By:

POOJA SANTHOSH

III B.Sc. Computer Applications [Triple Main]

Register No: SB22CA029

Under the guidance of

Ms. RAJI S PILLAI

Assistant Professor

Department of Computer Applications

2022 - 2025

ST. TERESA'S COLLEGE(AUTONOMOUS), ERNAKULAM

AFFILIATED TO MAHATMA GANDHI UNIVERSITY

CERTIFICATE

This is to certify that the project report entitled "Ethical Concerns of Using AI in Healthcare and the Criminal Justice System" is a bona fide record of the work done by POOJA SANTHOSH (SB22CA029) during the year 2022 – 2025 and submitted in partial fulfilment of the requirements for the degree of Bachelor of Science in Computer Applications (Triple Main) under Mahatma Gandhi University, Kottayam.

17/03/2025

Head of the Department

Internal Examiner

ST. TERESA'S COLLEGE DEPT. OF COMPUTERS ERNAKULAM

17/3/'25

External Examiner

Dr. Rebecca Mol. NA

Date: 14/03/25

DECLARATION AND ACKNOWLDEGEMENT

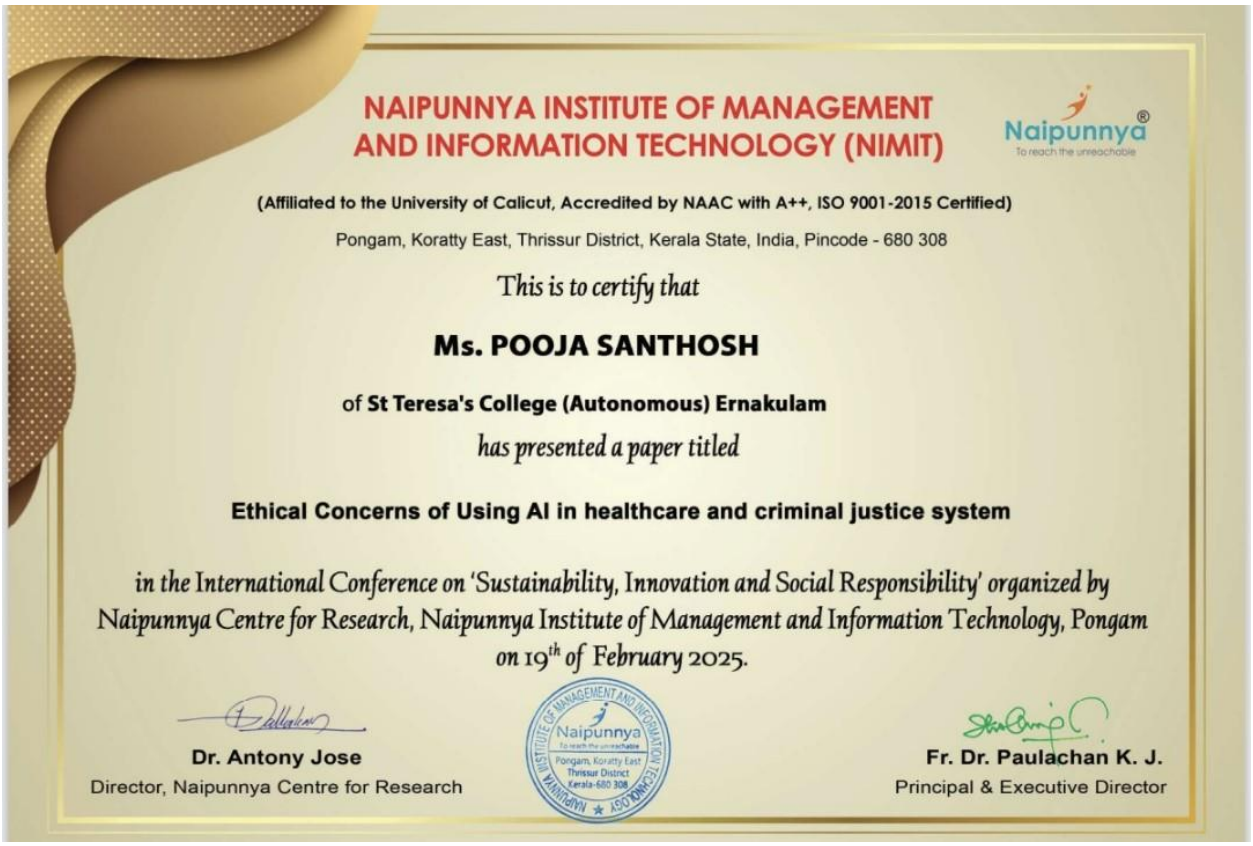
I, POOJA SANTHOSH (SB22CA029), BSc Computer Applications student of St. Teresa's College (Autonomous), Ernakulam, hereby declare that my work is original.

I would like to thank Rev. Dr. Sr. Nilima (CSST) Manager, Directors Rev. Sr. Tessa (CSST), Rev. Sr. Francis Ann (CSST), Principal Dr. Alphonsa Vijaya Joseph, Class Teacher and Assistant Professor Ms. Sheeba Emmanuel for the wonderful opportunity to work on my research project titled Ethical Concerns of Using AI in Healthcare and the Criminal Justice System. I am eternally grateful to Ms. Raji S Pillai for her tremendous insights and mentoring throughout the course of my research. My time working with her was a lesson in commitment. None of this would have been possible without the endless and enduring support of my close family and friends who through example taught me the importance of ethics. It is under their wings that my paper took shape. I dedicate my paper to them. Lastly, I express my sincere thanks to everyone who has directly or indirectly contributed to my paper.

Pooja Santhosh

CERTIFICATION

The paper was presented at the International Conference on 'Sustainability, Innovation and Social Responsibility' organised by Naipunnya Institute of Management and Information Technology, Pongam on 19th of February 2025. The certificate issued is attested.



NAIPUNNYA INSTITUTE OF MANAGEMENT AND INFORMATION TECHNOLOGY (NIMIT)

Naipunnya®
To reach the unreachable

(Affiliated to the University of Calicut, Accredited by NAAC with A++, ISO 9001-2015 Certified)

Pongam, Koratty East, Thrissur District, Kerala State, India, Pincode - 680 308

This is to certify that

Ms. POOJA SANTHOSH

of **St Teresa's College (Autonomous) Ernakulam**

has presented a paper titled

Ethical Concerns of Using AI in healthcare and criminal justice system

in the International Conference on 'Sustainability, Innovation and Social Responsibility' organized by Naipunnya Centre for Research, Naipunnya Institute of Management and Information Technology, Pongam on 19th of February 2025.

Dr. Antony Jose
Director, Naipunnya Centre for Research

Naipunnya
To reach the unreachable
Pongam, Koratty East
Thrissur District
Kerala-680 308

Fr. Dr. Paulachan K. J.
Principal & Executive Director

Ethical Concerns of Using AI in healthcare and criminal justice system

Pooja Santhosh, Student, Department of Computer Applications, St. Teresa's College (Autonomous)

Raji S Pillai, Guide, Department of Computer Applications, St. Teresa's College (Autonomous)

"The potential benefits of AI are enormous, but so too are the potential risks. We have to think very carefully about how we design these systems to ensure they align with human values." — Stuart Russell

Abstract

The use of artificial intelligence in areas like healthcare and criminal justice system where mistakes are highly paid for requires extensive debate and elaborate discussions. They contribute significantly by improving diagnoses and pacing verdicts. Biases in data, privacy concerns, opaque processes and zero accountability are unfortunate side effects. Vulnerable populations and high-stake decisions are a common characteristic of these critical sectors. Relying on AI for justice dehumanizes the criminal justice system reducing it into a puddle of algorithms. AI often becomes the black box where lack of transparency makes it hard to understand and appreciate their decisions. The need for fairness and equity demands that AI doesn't override human ethics. Another significant worry is whether everyone will have equal access to AI.

This study traverses these challenges by understanding the ethical concerns and addressing them in parallel with real world scenarios and observations. Our aim is to emphasize how AI must be used with care and caution in the process of decision making. We research ways by which AI can be put to use to make complex and complicated decisions without sacrificing ethics and human dignity. The following frameworks are construed in the duration of the study. It is important to acknowledge biases, inequities and absence of transparency. Responsible and ethical deployment of AI calls for strict regulations and training on data that reflects real life. There should be proper regulations, awareness and understanding around the use of AI. It is suggested that AI in welfare should be standardized to avoid conflicts. We propose that the deployment of AI should be a multi-disciplinary collaboration of science and ethics. Thus, fairness, accountability, and human rights will take the front seat. We end the paper on the note that the touch of human judgement is irreplaceable.

Keywords: Biases, ethics, fairness, decisions

Methodology

This study is qualitative research based on findings from published academic works and non-academic literature. The paper incorporates theoretical analyses and real-world scenarios. The concerns are described both industry specific and cross industry based. Some common

concerns are noticed and mitigations are discussed. The limitations may stem from biases in secondary sources.

Analysis

Ever since the debut of AI, efficiency has taken a leap for good. It was this argument that led to AI being introduced in critical and time sensitive sectors like healthcare and criminal justice system. However, AI falls short when matters of ethics are considered. To balance the pros and cons, it is necessary to address the shortcomings. In this section of the paper, we discuss how AI comes with its own set of biases and errors.

Healthcare

When machines are fed data with biases, decisions are also biased. Historically, the misrepresentation of minority groups in data has contributed to significant disparities in medical studies. Biases in AI can further accelerate this problem leading to gaps in diagnoses and treatment. To make matters worse, AI can miss diverse symptoms in particular ethnic groups. This can culminate in inaccurate recommendations. AI does not provide a substitute for human biases.

The molecule of AI is data. However, personal health information is a sensitive territory. AI needs tremendous access to sensitive data which poses a security threat. Data breaches are an eminent concern. Data scientists are arguing the possibility of anonymous data to be re-identified by AI. Whether personal data is worth risking for the promise of collective reward is unanswered.

AI is an enigma to mere mortals like us. The lack of transparency of decision making can corrode the process. When the diagnostician is unable to explain the diagnose, the black box is more a curse than a blessing. The patients can lose faith in the system or in the doctor altogether.

The autonomy of the patient and healthcare provider is persistently questioned in a system guided by AI. The care can feel algorithmic and unattached. There is a lack of social and emotional bonding. The doctors are forced to trust a machine over their instincts. It is an intimidating change.

Informed consent is an important variable in the equation. AI is changing every day. It is hard to consent to a system one doesn't fully understand. AI errors are an omnipresent danger. The ethics of it is paradoxical. Would you rather trust a machine to do your surgery over an experienced surgeon?

Law suits in medicine are hardly rare. The dimensions increase when AI is included. It is important for accountability in critical care. It is undecided who takes the blame for mistakes caused by AI. This can lead to irresponsible administration. Misdiagnoses, wrong prescriptions and patient deaths are all liable mistakes.

AI requires vast resources. If AI is not subsidized, wealth can make AI an unfair advantage. AI can form a two-tier health industry with urban areas having more access to it. This will wrongly impact rural areas with less access to electricity and infrastructure. It is important to ensure equity in AI.

Misuse and dual use of AI is an ethical red flag. Insurers can use AI to discriminate high risk patients. Marketing and law enforcement can also exploit this data for non-ethical purposes. Genetic data is another high value resource that can be exploited.

AI needs less administration staff and more tech savvy staff. This can cause job displacements affecting labor force. People can lose their jobs. More people will be forced to change their disciplines. This might have an adverse effect on elderly staff who have less experience with technology. Ageism is a valid ethical concern.

Over reliance on AI, drifts in algorithm, new biases, human oversight, dehumanization and lack of empathy in healthcare are some long-term ethical concerns.

Criminal Justice

Racial and demographic biases disproportionally affect black defendants while underestimating the risks for white defendants. This is reflected by COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) system. This simply is an example for systematic racism in historical data.

One of the applications of AI in criminal justice system is predicting recidivism. AI does this by analyzing historical data. When assessing the likelihood of an individual to reoffend, it is crucial to factor in human complexity, social, economic and political factors. Lack of this can lead to injustices. AI prediction fails to consider rehabilitation, education and family support. This can endanger Blackstone's principle.

Most often, the concerns are related to fairness and equity. Biases can affect marginalized communities amplifying the existing racial injustice. People of color are unfairly impacted. There is an erosion of civil liberties.

The algorithms of AI are obscure. The propriety nature of AI makes it difficult for people to access the underlying models. Defendants and legal representatives may struggle for recourse. This clearly is a violation of the rights of an individual. It is important for individuals to understand how decisions are made.

Predictive models used to determine bail and sentencing do not consider information about the life of the individual. The models use historical data instead of assessing character or life situations. The automated nature of AI compromises fair judgement.

The widespread surveillance of individuals raises privacy concerns. Face recognition systems used without consent pile onto these concerns. Marginalized communities experience more policing and discriminatory surveillance. It poses a classic example of over policing due to AI. Thus, AI can contribute to systematic oppression.

There are several AI models in use today. Each model uses its own set of algorithms. When AI is not standardized, decisions can vary across courts. It is imperative to standardize AI in criminal justice system to have a consistent judicial process.

AI in criminal justice system can be used for risk assessment, predictive policing and parole decisions. However, there is no meaningful human oversight. The justice system becomes an assortment of unclear algorithms. Individuals lose the right to contest life altering decisions. An automated AI system undermines the autonomy of people. Fairness and justice always require a human eye.

Shared Concerns

1. Bias and Discrimination: Non-inclusive biased data can reinforce demographic biases leading to unfair treatment of marginalized communities
2. Accountability and Transparency: Lack of transparency and accountability acts as an obstacle in understanding and contesting decisions made by AI
3. Privacy and Autonomy: Sensitive data triggers privacy concerns impacting individual autonomy
4. Fairness and Equity: The absence of human oversight endangers the due process of fairness

Discussion

Case Study 1- IBM Watson for Oncology

IBM Watson is one of the best examples of an AI healthcare model. The purpose of Watson was to analyze medical records and suggest customized treatments. The decision was to be taken based on available medical literature, recorded clinical trials, new medical studies and patient's health records. The accuracy of Watson was found to be unreliable. The evaluations at Memorial Sloan Kettering Cancer Center discovered that Watson was found to make unsafe recommendations about 30% of the time. Watson had prescribed a treatment plan that was not compatible with the patient's health condition. This is particularly worrying considering Watson was designed for customized treatment plans. Despite the fact that Watson was trained on an enormous volume of diverse data, it had failed to deliver accurate results. It is also worth noting that Watson's decisions were mostly based on incomplete data. It had not considered crucial patient information that typically would be prioritized by a doctor (Choi, 2017).

Case Study 2- The Algorithmic Bias in Health Risk Prediction

To understand the effect of algorithmic bias of AI, we can look at the following example. This instance sheds light on how AI amplifies systematic oppression. In 2019, a study was focused on analyzing the algorithm used by UnitedHealth to predict health risks. It was determined that the algorithm had a major racial bias. The algorithm tended to assign different health risk scores to Black and White people with the same health condition. Black people were assigned a lower health score than White people despite the same condition. This was due to the historical data biases. Since black people were historically denied healthcare, there were less records of them in the data. The 2019 study presents a good case of why AI data should be monitored.

Case Study 3- COMPAS Algorithm in Predictive Policing

The COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) is an algorithm that was designed to find the likelihood of reoffenders. The algorithm is used by several courts in the US. Regardless, many have accused COMPAS of exhibiting racial biases. ProPublica found the algorithm to be biased in 2016. The algorithm was unfair to Black people. It predicted that Black people had a higher risk of committing future crimes than White people. The Black defendants were labelled as future criminals twice as much as the White defendants (Vaccaro, 2019).

Case Study 4- Predictive Policing with Palantir

Predictive policing is one of the most significant applications of AI in criminal justice system. Palantir Technologies is a notable data analytics company that helped develop predictive policing tools for law enforcement. The tools are designed to analyze vast amounts of social media data and historical crime records to predict possible crimes and criminals. The tools have however come under scrutiny for over policing in certain regions and discriminating against minority groups. The Palantir predictive policing system deployed in Los Angeles led to more police patrolling of areas with Black and Latino neighborhoods. The Center for Policing Equity determined that the Palantir predictive policing system can amplify the racial biases already existing in policing. This can be attributed to the fact that the fodder data comes from previous policing activities that were racially targeted (Oatley, 2022).

Mitigation Strategies

Diverse and Representative Data: Considering the fact that one of the biggest ethical concerns of using AI in healthcare and criminal justice system is biases, it is important to monitor the data. The data should be made diverse accommodating all minorities of the society. This will ensure fairness and equity. Demographical and social biases can be controlled if data is properly represented.

Bias Audits: There must be regular audits for data biases. This serves as a method to avoid demographical biases. If audits bring forward disparities, the model can be corrected accordingly. This also allows the data sources to be handled with caution. The logistics of this operation can involve an independent third party.

Explainability and Transparency: The algorithms and working of AI must be designed in such a way that they are interpretable and explainable. This allows the people to trust and rely on AI.

Training and Awareness: The people must be made aware on how AI makes decisions. This will allow people to assess biases and make critical judgements. It also helps them contest AI decisions.

Fairness Constraints in Model Development: We can accommodate fairness variables to ensure due process. This means incorporating certain social and political factors into the AI modelling. This will ensure that the decisions aren't only based on data.

Use of Multiple Models: Sometimes, a single model can cause omissions. If we subscribe to multiple models, the output of the models can be compared to ensure fair outputs. The models can be fed on diverse data.

Transparency in Risk Assessment: Risk Assessment systems like COMPAS should be made transparent so people can understand how the scores are generated. This makes them able to challenge unfair accusations.

Impact Assessments: Objective assessments should be made by independent parties to understand the impacts AI can have on fairness and equity. Potential harms, mistakes and biases must be assessed prior to AI modelling.

Clear Documentation of AI Decisions: All the decisions made by AI should be clearly documented to enhance understanding. Step by step documentation also provides a path to ensuring accountability.

Human Systems: It is important to ensure than AI doesn't replace human judgement. All critical decisions must by overseen by a human counterpart.

Regulatory Oversight: We can build a regulatory framework to ensure that AI models are always updated and consistent. An independent third party can oversee the regulations in order to ensure quality AI standards.

Public Disclosure of AI Methods: The developers of AI systems should disclose the algorithms, methods and databases they use. This will allow for corrections and accountability.

Public Accountability: There must be laws that dictate that the AI systems should be disclosed to the public. Public accountability can prevent misuse.

Auditability: In addition to bias audits, the AI models should also be audited to ensure ethical deployment. They can also check for security threats.

Clear Responsibility for Decisions: It must be made clear who takes the accountability in case of a mistake. Regulations must be made between developers and users.

Data Encryption and Secure Storage: There must be efficient protective mechanisms like end-to-end encryption to protect sensitive data. This reduces the chance for data breaches and unauthorized access.

Anonymization and De-identification: The data that is used to train AI should be made anonymous so that outcomes are not traced to individuals. This also allows the data to be secure. Furthermore, there must be accommodations to prevent re-identifying data.

Strict Data Use Policies: Institutions must follow guidelines on how data is collected and used. The patients and defendants must be made aware of these guidelines. The guidelines should be updatable according to the changes in the AI models.

Data Minimization: Unnecessary data should be dismissed as this can lead to future misuse. Data minimization ensures that only necessary data is collected from the people.

Clear Data Usage and Retention Policies: In addition to collecting limited data, there must also be established time limit to how long the data will be stored by the AI systems.

Safety Against Surveillance: Surveillance activities like predictive policing and facial recognition systems must be limited to prevent misuse. They can be made to require warrants. There should also be an objective third party to oversee surveillance activities.

Transparency in Data Usage: People should have a right to access the data that is being used in decision making. This will allow people to understand the fairness behind the decision.

Clear Communication: Clear communication is required for informed consent. People must be made aware of the benefits and harms of using AI. They must fully understand the extend of AI.

Patient Empowerment: There must be right of choice. People should be able to choose whether to use AI. They should also be allowed to change their decisions anytime.

Human Oversight: It is important to have a human oversight on AI processes. Final decisions must be overseen by a human. People should also be able to request for human assistance throughout the AI process.

Right to Challenge AI Decisions: Individuals that are affected by AI decisions must have the right to challenge them. They should be allowed access to all the necessary information. Appealing AI generated decisions must be encouraged.

Continual Training: Workers must be provided continual training on AI. The limitations of AI should be properly studied to avoid over reliance.

Validation and Testing: AI systems must be tested aggressively before use. They must be constantly updated and tested to ensure optimum functionality. Errors must be studied thoroughly.

Reskilling Workers: To prevent job displacements, existing workers can be trained to use AI. This will ensure that people with traditional jobs do not lose out on opportunities.

Job Creation in AI-related Fields: In addition to filling existing jobs, AI can also create more jobs leading to a more flourishing economy.

Human-AI Collaboration: The best parts of AI can be combined with the best parts of humanity to form a superior system. Humans can use AI to make better decisions concerning health and justice instead of simply relying on AI for decisions.

Multi- disciplinary collaboration: Scientists, developers and ethicists must come together to form an AI model that ensure efficiency and fairness. AI data also should reflect social parameters and ethical concerns.

Conclusion

The use of AI in healthcare and criminal justice systems can be beneficial if the ethical concerns are addressed. The mitigation strategies provided in the study formulate a way to use AI without compromising human values. However, it is important to note that AI should never replace human decisions.

References

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. John Wiley & Sons.

Berk, R. A. (2021). Artificial intelligence, predictive policing, and risk assessment for law enforcement. *Annual Review of Criminology*, *4*(1), 209-237.

Casillas, J. (2024). Bias and discrimination in machine decision-making systems. *Ethics of Artificial Intelligence*, 13-38.

Choi, Y. S. (2017). Concepts, characteristics, and clinical validation of IBM Watson for oncology. *Hanyang Medical Reviews*, *37*(2), 49-60

Fazel, S., Burghart, M., Fanshawe, T., Gil, S. D., Monahan, J., & Yu, R. (2022). The predictive performance of criminal risk assessment tools used at sentencing: Systematic review of validation studies. *Journal of Criminal Justice*, *81*, 101902.

Hardyns, W., & Rummens, A. (2018). Predictive policing as a new tool for law enforcement? Recent developments and challenges. *European journal on criminal policy and research*, *24*, 201-218.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature machine intelligence*, *1*(9), 389-399.

Kassan, P. (2020). Ten Years Away... and Always Will Be: A review of Artificial Intelligence: A Guide for Thinking Humans by Melanie Mitchell. *Skeptic (Altadena, CA)*, *25*(1), 56-60.

Oatley, G. C. (2022). Themes in data mining, big data, and crime analytics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *12*(2), e1432.

O'neil, C. (2017). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Polevikov, S. (2023). Advancing AI in healthcare: a comprehensive review of best practices. *Clinica Chimica Acta*, 117519.

Vaccaro, M. A. (2019). *Algorithms in human decision-making: A case study with the COMPAS risk assessment software* (Doctoral dissertation).

Wang, F., & Preininger, A. (2019). AI in health: state of the art, challenges, and future directions. *Yearbook of medical informatics*, *28*(01), 016-026.

Wickramarathna, N. A., & Edirisuriya, E. A. T. A. (2021). Artificial Intelligence in the Criminal Justice System: A Literature Review and a Survey.

Whitcomb, C. G. (2020). Review of Shoshana Zuboff (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power: New York: Public Affairs. 704 pp. ISBN 9781781256848 (Hardcover). *Postdigital Science and Education*, *2*, 484-488.