ETHICS IN ARTIFICIAL INTELLIGENCE

Vaishali Deshwal
Assistant Professor, Department of CSE, Ajay Kumar Garg Engineering College, Ghaziabad, U.P, India
deshwalvaishali@akgec.ac.in

Abstract—**Within its larger context, the interaction between artificial intelligence and ethics in healthcare is an important and multifaceted subject. Making sure that medical artificial intelligence (AI) technologies are fair, safe, and respect patient privacy is an ethical task. This includes worries about the fairness of patient care, the accuracy of diagnoses made by artificial intelligence, and the security of private health information.**

***Keywords*— Ethics based Artificial Intelligence, Artificial Intelligence, Natural Language Processing, AI governance**

I. INTRODUCTION

AI is a fast-growing field of study that includes a range of technological abilities that offer new opportunities, resources, and uses fo3r society and businesses. AI has set off a disruptive force in the always changing healthcare industry, offering unmatched improvements in patient care. AI has improved patient data and diagnostics, clinical decision-making, predictive medicine, and clinical diagnosis. Large and frequently unstructured datasets, sophisticated sensors, natural language processing (NLP), machine learning (ML), and, more just now, large language models (LLMs) are some of the tools used to give computers and software reasoning and data processing powers comparable to those of human intelligence.

While artificial intelligence (AI) has several definitions, it usually refers to the expanding field of computing as well as an information system's capacity to understand, learn from, and adapt to accomplish goals. Governments, businesses, and public institutions are investigating AI to create new services and products and increase efficiency. But the quick spread of AI systems has also raised concerns about algorithmic opacity, unethical behavior, and unforeseen effects, like gender and racial biases. These issues are commonly known as the sadness of AI. The idea of reliable AI and AI governance has been supported by the auditing of AI systems [1]. Large and frequently unstructured datasets, sophisticated sensors, NLP, ML, and, more recently, LLMs are some of the tools used to give computers and software reasoning and data processing powers comparable to those of human intelligence. AI's ML field has made it possible for algorithms to find patterns in data without explicit programming.

A subset of ML known as "deep learning" makes use of artificial neural network algorithms to gain knowledge from data and give predictions or judgments. It is especially helpful for tasks like voice and picture recognition [2]. NLP is a field of computer science and AI that studies how natural language is used in human-computer interactions. Computer techniques and algorithms are used to comprehend, analyze and generate human language.

Medical fields including radiology, pathology, and oncology, where images are crucial, has seized the chance to incorporate AI into healthcare, and substantial research and development attempts have been assembled to transfer the promise of AI to therapeutic requests [3-5]. AI applications in mental health care could provide benefits such as better patient responsiveness, novel treatment options, and opportunities to engage with marginalized communities. In clinical and research contexts, NLP can be used in addition to symptom evaluation for schizophrenia. It makes use of data that is probably more associated with compromised brain functions, including compromised reward processing, connection, and information processing. On the account of ChatGPT's capabilities, which include simulated training, language translation and curriculum development, it is a valuable device for medical education. Furthermore, it can help with the retrieval of data for research and will enhance the accuracy of medical documentation in clinical settings[6-7].

Academics, institutions, and policymakers are realizing more and more how crucial it is to audit AI systems to make sure they are being used in a way that is both beneficial and socially acceptable. The development of AI could make addressing healthcare inequalities more difficult. It is important to comprehend and discuss the moral issues in depth because of the many advantages of artificial intelligence and the necessity to minimize any potential drawbacks. To lessen the ethical and legal challenges related to AI in healthcare, a multimodal approach comprising lawmakers, developers, medical experts, and patients is required [8].

a) Conceptual foundation: ethical values and AI audits

This research centres on the kind of AI auditing that seeks to verify that AI is morally sound when measured against accepted moral standards. The methodology has been referred to as "AI auditing based on ethics" and "ethical algorithm auditing". Assurances of legality, correctness, efficiency, or safety are a few examples of different approaches to AI auditing, in addition to ethics-based auditing [9]. Nonetheless, auditing based ethics is an important branch of AI auditing since conversations about AI ethics are becoming more widespread among practitioners and academics, which paves the way for regulation developments like the planned AI Act. Additionally, it has been determined that one of the key issues in guaranteeing responsible AI systems is the translation of ethical ideas into actions.

A consequentialist approach might be used to define ethics-based AI auditing. "evaluations of the algorithm's detrimental effects on stakeholders' rights and interests, along with a related identification of the circumstances and/or algorithmic elements that contribute to these effects" is how ethical algorithm audits are defined, emphasizing consequences.

From a deontological standpoint, it is described as "a methodical procedure wherein the behavior of an entity—whether past or present—is evaluated for conformity with pertinent standards or guidelines." is the definition of ethics-based auditing of automated decision-making systems.

We pay particular attention to ethical guidelines presented in the literature on AI ethics and AI audits in order to maintain the review's breadth reasonable and consistent with our research concerns. There is a wealth of literature on AI ethics principles, and several summaries of these concepts have been published. A reliable source of compiled expert information is the Ethical Guidelines for reliable AI, which was created by the EU-appointed High-Level Expert Group(HLEG) on AI. The European Commission established the autonomous EU AI HLEG to give recommendations for reliable AI.

b) AI governance:

The structures, regulations, and procedures that direct the use, creation, and utilization of AI technology are referred to as AI governance. Its goal is to guarantee that AI systems minimize possible risks while optimizing advantages by making them ethical, equitable, transparent, and accountable. Important elements of AI governance consist of:

- Ethical Guidelines: Defining values such as justice, nondiscrimination, and respect for human rights that will guide the ethical application of AI.
- Accountability and Transparency: Making sure AI systems have procedures in place to hold developers and users responsible for their actions, as well as transparency in how they operate and make choices.
- Security and privacy: safeguarding the confidentiality of data and making sure AI systems are safe from malevolent intrusions.
- Establishing industry best practices and standards for the creation and application of artificial intelligence.
- Research & Development: Fostering further investigation to identify and reduce the risks related to artificial intelligence (AI) while encouraging creativity.
- Regulatory Frameworks: drafting legislation and rules to handle the application and effects of AI while upholding human rights and current legal requirements.

II. FOUR CORE ETHICAL PRINCIPLE

Four fundamental ethical precepts [10] are outlined here: justness, explainability, prevention of damage, and respect for human autonomy.

a) Respect for human autonomy:

It is the idea that people must be capable to exercise their full autonomy and that AI systems should support them rather than control or subjugate the. Positive freedoms, such as Positive liberties and human wellbeing, or negative freedoms, such as freedom from coercion and monitoring, can be used to conceive autonomy [11]. Providing for a desired degree of human independence also involves purposefully finding a harmonize between the decision-making authority granted to artificial agents and that which is kept for humans.

b) Prevention of harm:

The concept of "prevention of harm" states that artificial intelligence (AI) systems shouldn't inflict, aggravate, or expose people to physical or mental pain. Safety, security, and the avoidance of predictable and inadvertent harm are all emphasized in the literature on AI ethics standards. Concerns about privacy and the possibility of an AI weapons race are frequently raised.

c) Fairness:

According to AI, fairness in AI systems is the ability to uphold procedural fairness, which is the right to challenge and seek redress, as well as substantive fairness, which is the liberty from bias and differentiation and uniform opportunity. In terms of results and procedures, as well as the potential for bias in AI systems, fairness and justice are related. Fairness also addresses concerns of equity, diversity, and inclusion that are relevant to the workforce and society at large.

d) Explicability:

Explicability is the idea that AI systems should be open and honest about their goals and capabilities, and that, to the greatest extent feasible, individuals who are impacted by their actions should be able to understand them Transparency demands address automated decision-making, data utilization, human-AI interaction [12], and data purpose. Furthermore, it has been suggested that explicability enhances the other principles since it is necessary to have a sufficient understanding of AI systems' behavior in order for them to respect autonomy, prevent harm, and be fair.

III. AI AND ETHICAL ISSUES

Important ethical dilemmas and associated subtopics that need to be taken into consideration have been discovered by studies on the ethical considerations associated with AI's role in healthcare. This review discusses these problems as well as the challenges presented by the latest creation of LLMs. Openness, Privacy, bias, Cybersecurity, data quality and responsibility are the main issues [13,14].

a) Openness:

Transparency has become a major ethical concern, particularly for complicated black box AI systems that perform extremely well but have opaque decision-making processes. Finding the right balance between explainability and accuracy is crucial, especially when making choices in situations with a lot of risk.

Explainability is the ability of an AI-driven system to communicate to a human why it made a specific prediction or decision. From a medical standpoint, it is imperative to differentiate between two levels of explainability in AI systems. In general, the first level focuses on comprehending how the system arrives at its conclusions, whereas the second level explains the training procedure that allows the system to take in information from examples and generate outputs.

Different types of explainable AI (XAI) are compiled and classified in a study centered on XAI requests in healthcare. The XAI types can be divided into two categories: There are two types of approaches: model-specific techniques are created specifically for a certain model type and cannot be applied to other models; and model-agnostic techniques are criteria-neutral and can be applied to a wide range of XAI models. Certain algorithms that improve medical imaging's interpretability. Post hoc interpretation strategies include algorithms like Class Activation Mapping, Local Interpretable Model-Agnostic Explanations (LIME), and Layer-wise Relevance Propagation, Backpropagation. Following analysis, these strategies offer comprehensible information about the model using outside sources[15].

b) Privacy:

With the use of DL and ML to make predictions based on user data, privacy becomes a crucial concern in the field of data-driven healthcare. Patients have faith that medical personnel would safeguard their personal information, especially delicate data like age, sex, and medical records. Private databases including medical histories and genetic sequences may make it more difficult to gather information and develop new diagnostic techniques. Data sharing is difficult because some companies may be used to defend hiding information under the guise of privacy protection [16].

Number of privacy and security flaws present in the context of the Internet of Medical Things (IoMT) and wearable technology in the healthcare industry, posing a risk to sensitive data. Attacks using denial-of-service and ransomware have the capacity to cause potentially fatal situations. Therefore, Users voice worries about the shortcomings of the technology and the usage of personal data [17].

c) Bias:

Biases found in healthcare data can affect AI algorithms. Biases may have an effect on AI system. Apart from commonly recognized research biases such as blinding and sampling, it is imperative to discern both explicit and implicit biases within the healthcare system. These biases may affect vast amount of data utilized in AI system training. Factors such as the eligibility restrictions for clinical trials and implicit biases in treatment decisions could impact clinical decision-making and hence the predictions provided by AI. AI has shown to be useful in identifying racial disparities in cancer results and inquire into the influence of financial standing and race on oncology-associated health results[18,19].

d) Cybersecurity:

Preventing unwanted entry, theft, destruction, or other detrimental attacks on networks, computer systems, and digital data is known as cybersecurity. AI systems' inability to be understood and explained can hide security problems. Using publicly accessible data from a variety of sources, open-source intelligence (OSINT) raises questions about political campaigns, national defense, the cyber sector, social challenges, criminal offense profiling, and cybercrimes.

The 3U's of cybersecurity: usage, user and usability serve as a basic foundation for comprehending the relationship among diverse cybersecurity parts from a comprehensive cybersecurity view[19].

Researchers are looking into using semi-supervised learning (SSL) with cybersecurity data repositories to create trustworthy models for computer security and cybersecurity systems. To create such models, SSL, a specific kind of ML, can use half labeled data or just a less amount of labels. To ensure the efficacy and applicability of these cybersecurity prototypes, it is imperative that the data utilized in their growth appropriately reflect real-world facts.

e) Data Quality:

Convolutional neural networks (CNNs) are becoming more and more popular for image-related tasks. With the help of contemporary machine learning techniques, its use has expanded to include non-imaging data. Because CNNs can convert non-imaging input into images, they can be employed for a wide range of jobs beyond traditional imaging tasks. This allows medical professionals to develop hybrid deep learning models using multiple data models that combine various patient data, including genetic, imaging, and clinical information. Language models have the potential to provide data or responses that are wholly made up or unsupported by real data. It is important to find and solve these delusional inclinations for language prototypes to be dependable and reliable, especially in the field of healthcare where accurate and factual information is crucial[21].

f) Responsibility:

There are important problems about whom or what should be accountable for the consequences of AI actions when it comes to AI responsibility attribution. Human responsibility in relation to AI systems emphasizes meaningful control as a result of diligence and warns against completely automated medical systems.

It might be difficult to assign responsibility precisely when there are various options and agents involved. This is known as responsibility dispersal. An AI-driven digital tumor board instance illustrates the idea by changing clinical decision-making and distributing accountability among multiple stakeholders[22,23].

Authorship is improper for AI tools because they cannot make decisions or contribute to research in the same manner that human writers can. However, proponents assert that AI tools may be very helpful in coming up with ideas and aiding in the writing process, therefore they should be given proper credit and acknowledgment for their contributions to paper authorship[24,25].

IV. AI ETHICS IN THE FUTURE

Future advances in AI may result in systems that exhibit enhanced autonomy or consciousness, which would have significant ethical and philosophical ramifications. Examining these possibilities requires thinking critically about basic issues such as the nature of identity, awareness, and ethics in regard to non-human animals.

V. ROBOT ETHICS, MEDICAL ETHICS AND ARTIFICIAL INTELLIGENCE

The conventional approach to addressing emerging technology is focusing the ethics of robotics and artificial intelligence on diverse types of “concerns.” Artificial intelligence and medical ethics: Although AI has a lot of potential in medicine, there are several obstacles to overcome. AI has a wide range of applications in medicine, necessitating the expertise of a diverse group of writers, editors, and reviewers. Commercial applications of AI experience necessitate disclosure regarding potential conflicts of interest.

Ethics of Machines: Machine ethics is the study of morality for machines, or “ethical machines,” as opposed to the usage of technologies as objects by humans. It’s often unclear whether this is purposeful to be a comprehensive or partial explanation of AI ethics [26].

Sometimes this has the consequence that we need to develop machine ethics if robots exhibit ethically meaningful behavior. Consequently, some define machine ethics more broadly, saying that it is ensuring that robot behavior toward humans and possibly other machines is morally acceptable.

VI. CONCLUSION

Analyzing the rapidly evolving topic of AI ethics closely reveals a number of difficult problems, including as privacy breaches, dishonest business practices, and the moral implications of machine behavior. As technology develops, worries about information ownership and the erosion of individual freedoms persist. An unstable environment is the result of both regulatory flaws and the relentless pursuit of profit by massive corporations.

Machine ethics, best exemplified by Isaac Asimov’s guidelines, casts doubt on AI’s ethical duty. Contact between humans and robots poses ethical dilemmas and emphasizes the need for stringent legislation. While traversing this challenging terrain, a careful balance must be struck to ensure that technology advances responsibly and safeguards fundamental human rights and ideal. To develop a comprehensive and well-rounded strategy, effective AI [27] governance necessitates the participation of multiple stakeholders, including governments, the commercial sector, academia, and civil society.

REFERENCES

[1] Secinaro, Silvana, Davide Calandra, Aurelio Secinaro, Vivek Muthurangu, and Paolo Biancone. “The role of artificial intelligence in healthcare: a structured literature review.” *BMC medical informatics and decision making* 21 (2021): 1-23.

[2] Janiesch, Christian, Patrick Zschech, and Kai Heinrich. “Machine learning and deep learning.” *Electronic Markets* 31, no. 3 (2021): 685-695.

[3] Barragán-Montero, Ana, Umair Javaid, Gilmer Valdés, Dan Nguyen, Paul Desbordes, Benoit Macq, Siri Willems et al. “Artificial intelligence and machine learning for medical imaging: A technology review.” *Physica Medica* 83 (2021): 242-256.

[4] Hunter, Benjamin, Sumeet Hindocha, and Richard W. Lee. “The role of artificial intelligence in early cancer diagnosis.” *Cancers* 14, no. 6 (2022): 1524.

[5] Yasaka, Koichiro, and Osamu Abe. “Deep learning and artificial intelligence in radiology: Current applications and future directions.” *PLoS medicine* 15, no. 11 (2018): e1002707.

[6] Sallam, Malik. “ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns.” In *Healthcare*, vol. 11, no. 6, p. 887. MDPI, 2023.

[7] Karabacak, Mert, and Konstantinos Margetis. “Embracing large language models for medical applications: opportunities and challenges.” *Cureus* 15, no. 5 (2023).

[8] Prakash, Sreenidhi, Jyotsna Needamangalam Balaji, Ashish Joshi, and Krishna Mohan Surapaneni. "Ethical Conundrums in the application of artificial intelligence (AI) in healthcare—a scoping review of reviews." *Journal of Personalized Medicine* 12, no. 11 (2022): 1914.

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

[10] Floridi, L. and Cowls, J., 2022. A unified framework of five principles for AI in society. *Machine learning and the city: Applications in architecture and urban design*, pp.535-545.

[11] Akmal, Muhammad, Binod Syangtan, and Amr Alchouemi. "Enhancing the security of data in cloud computing environments using Remote Data Auditing." In *2021 6th International Conference on Innovative Technology in Intelligent System and Industrial Applications (CITISIA)*, pp. 1-10. IEEE, 2021.

[12] Cabrera, Á.A., Druck, A.J., Hong, J.I. and Perer, A., 2021. Discovering and validating ai errors with crowdsourced failure reports. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), pp.1-22.

[13] Canales, Cecilia, Christine Lee, and Maxime Cannesson. "Science without conscience is but the ruin of the soul: the ethics of big data and artificial intelligence in perioperative medicine." *Anesthesia & Analgesia* 130, no. 5 (2020): 1234-1243.

[14] Jeyaraman, Madhan, Sangeetha Balaji, Naveen Jeyaraman, and Sankalp Yadav. "Unraveling the ethical enigma: artificial intelligence in healthcare." *Cureus* 15, no. 8 (2023).

[15] Chaddad, Ahmad, Jihao Peng, Jian Xu, and Ahmed Bouridane. "Survey of explainable AI techniques in healthcare." *Sensors* 23, no. 2 (2023): 634.

[16] Price, W. Nicholson, and I. Glenn Cohen. "Privacy in the age of medical big data." *Nature medicine* 25, no. 1 (2019): 37-43.

[17] Sabry, Farida, Tamer Eltaras, Wadha Labda, Khawla Alzoubi, and Qutaibah Malluhi. "Machine learning for healthcare wearable devices: the big picture." *Journal of Healthcare Engineering* 2022, no. 1 (2022): 4653923.

[18] Hashimoto, Daniel A., Elan Witkowski, Lei Gao, Ozanan Meireles, and Guy Rosman. "Artificial intelligence in anesthesiology: current techniques, clinical applications, and limitations." *Anesthesiology* 132, no. 2 (2020): 379-394.

[19] Istasy, Paul, Wen Shen Lee, Alla Iansavichene, Ross Upshur, Bishal Gyawali, Jacquelyn Burkell, Bekim Sadikovic, Alejandro Lazo-Langner, and Benjamin Chin-Yee. "The impact of artificial intelligence on health equity in oncology: scoping review." *Journal of medical Internet research* 24, no. 11 (2022): e39748.

[20] Yadav, Ashok, Atul Kumar, and Vrijendra Singh. "Open-source intelligence: a comprehensive review of the current state, applications and future perspectives in cyber security." *Artificial Intelligence Review* 56, no. 11 (2023): 12407-12438.

[21] Azamfirei, Razvan, Sapna R. Kudchadkar, and James Fackler. "Large language models and the perils of their hallucinations." *Critical Care* 27, no. 1 (2023): 120.

[22] Coeckelbergh, Mark. "Artificial intelligence, responsibility attribution, and a relational justification of explainability." *Science and engineering ethics* 26, no. 4 (2020): 2051-2068.

[23] Bleher, Hannah, and Matthias Braun. "Diffused responsibility: attributions of responsibility in the use of AI-driven clinical decision support systems." *AI and Ethics* 2, no. 4 (2022): 747-761.

[24] Verdicchio, Mario, and Andrea Perin. "When doctors and AI interact: on human responsibility for artificial risks." *Philosophy & Technology* 35, no. 1 (2022): 11.

[25] Stokel-Walker, Chris. "ChatGPT listed as author on research papers: many scientists disapprove." *Nature* 613, no. 7945 (2023): 620-621.

[26] Sachin Jain, "Deep Learning obstacles in madical image analysis: Boosting trust and explainbility". Available online: Jan-June2024_4.pdf (akgec.ac.in).

[27] Vikas,"Machine Learning methods in software engineering". Available online: Jan-June2024_10.pdf (akgec.ac.in).

ABOUT THE AUTHOR

Vaishali Deshwal received the M.Tech. degree in Computer Science and engineering from GCET, Greater Noida. Her research interests are Artificial Intelligence, Machine Learning and deep learning. She is currently working as Assistant Professor, AKGEC Ghaziabad.