AI Ethical Framework: Final Report

By Athika Fatima - 101502209

Introduction

Artificial Intelligence is changing industries, but it is developing with unnerving speed and unleashing ethical topics. An AI ethical framework would offer guidance for ethical development and deployment while minimizing the attendant risks and maximizing social value. This paper presents essential elements of the AI ethical framework used for AI in healthcare, demonstrating their applicability in medical decision-making and patient care.

1. Transparency

Transparency is one of the essential pillars of the AI ethics framework. It points to the necessity of straightforward and open communication regarding the operation of AI systems, including decisions taken therein. Transparency in AI-powered diagnostics means providing understandable explanations for how the AI model works, what data it has been trained on, and the reasoning behind its conclusions. Consequently, when carrying out medical image analysis, the AI needs to explain exactly which features in the image led to that particular diagnostic result to afford clinicians the confidence to trust that diagnostic output when it is making an informed decision based on any AI recommendation.

Lack of transparency can engender suspicion against the reliability and fairness of the system. Concealing the logic behind the decision taken by the system may prevent the extensive uptake of AI tools in healthcare, and may, on the flip side, raise doubts in the clinician's mind about the validity of AI recommendations, thus subsequently wreaking havoc in patient diagnosis. Transparency in diagnostic AI allows health professionals to have insight into the reasons behind a specific diagnosis conducted by the AI. For instance, in the case where the system flags a possible tumor on the chest X-ray of the patient, the system must indicate which features had got to that conclusion.

2. Accountability

Accountability means clearly defining the person who can be held responsible for the actions and results of the AI systems. Whereas the meaning of accountability varies across sectors, the significance of accountability in the clinical setting is magnified, as every medical decision affects the lives and wellness of patients. While employing AI in treatment decisions, issues are raised with respect to determining who is responsible in so far as the AI has relied on a decision which is wrong, and by doing so caused great damage and even death to a patient. Thus, in

health care, accountability must necessarily be extended out to the AI system developers, the health care providers utilizing it, and the patients exposed to their decisions. Developers ensure that the AI systems learn from diverse, representative datasets free from biases that might interfere with diagnosis accuracy, while healthcare providers interpret the AI suggestions and decide on treatment or diagnosis.

In AI diagnostics, it allows the tracing of responsibility, thus ensuring that mistakes made or harms inflicted through AI are trackable back to a responsible party. If a condition is misidentified by an AI tool and results in patient harm, healthcare providers are accountable for how they chose to use the tool.

3. Fairness

Fairness is the process of avoiding discrimination among individuals of any social groups in AI systems based on characteristics such as race, sex, or socio-economic status. Honesty becomes all of the more important in health since diagnosis decisions can have such an immense impact on a patient's condition. Biases in the training data sets may lead AI systems to arriving at wrong decisions. Because of this, marginalized groups would disproportionately bear the brunt of such decisions. For example, suppose there is an AI diagnosis system, such as skin cancers developed from images of lighter-skinned patients, that wrongfully diagnoses skin cancers in darker-skinned patients. Such bias would breed insecurity, misdiagnoses, and possibly poorer health outcomes for underrepresented patients on the part of doctors.

Fairness in AI diagnostics calls for the use of diverse and inclusive datasets representing the patient population across multiple demographics. There will also be a need for constant monitoring and auditing of the AI system to spot any emerging biases that might affect the resting performance capacity. It is necessary to train the system on many medical images demonstrating all possible skin tones to detect conditions in all patients correctly.

4. Privacy

Privacy is one of the foremost ethical concerns in AI, suffering the brunt of violations within the healthcare sector that apply AI systems to deal with sensitive patient data. AI support for diagnostics history usually involves extensive clinical data processed over personal health items, such as medical history, genetic information, and diagnostic images. The confidentiality of patient information must be respected in using AI methods dictated by data protection laws, such as the Health Insurance Portability and Accountability Act in the US and the General Data Protection Regulation of the European Union. This practically means, firstly, that all personal health data used by AI systems must be anonymized, secondly, that such data be encrypted during storage, and thirdly that patients have control over how their data will be used by the AI system. If an AI diagnostic tool is incorporated within a clinic or hospital, it would avert unauthorized transients of sensitive data. In the practice of medicine, patients should be informed

about the way they're being diagnosed with AI and, therefore, be given options to opt out of any data sharing with the AI systems should that sensation invoke the slightest discomfort in them. To provide an example, if an AI tool is to act on medical images, the images must be anonymized before they are processed in a manner that protects identity.

5. Robustness and Safety

Robustness and safety indicate the trustworthy and secure operation of AI systems in particularly demanding or odd conditions. In this respect, the AI-supported diagnostic systems in healthcare, for example, must be reliable, safe, and accurate across a vast array of medical contexts. An AI-based pneumonia diagnostic tool would have to be robust enough to take account of variation in X-ray quality, patient demographics, and medical conditions.

AI systems ought to be safe for clinical use. They must undergo thorough testing for reliability, accuracy, and safety while deploying them to avoid possible harm. For example, an AI system that misdiagnoses a terminal Disease could delay treatment and result in tremendous injury to the patient. Additionally, in AI-enabled diagnostics, robustness and safety ensure the system's ability to deliver proper, satisfactory, and dependable outputs across a wide range of medical care contexts and patient demographics. Should the performance of an AI system deteriorate or it is prone to false negatives or false positives, patient care and safety may be at risk with grave consequences occurring.

6. Human Oversight

Human oversight serves as a guarantee that AI systems can be employed with responsibility with regard to human values and goals. In healthcare, human oversight is paramount as a safety net so that the AI tools supplement, not replace, human experts. AI may assist in diagnostics, but ultimately the healthcare provider should have the final say in diagnosis and treatment decisions. It also involves the competent training of healthcare personnel to utilize AI tools, knowing the limits of AI decisions, and when to disregard an AI recommendation. It is imperative that healthcare professionals apply their academic and clinical judgment in actually using AI in clinical decisions. Regarding AI diagnostics, one of the functions of human oversight is to help doctors and medical staff intervene when AI tools send back results that are questionable. For example, if the tool suggests a diagnosis at odds with the doctor's experience, the physician would be able to ignore the recommendation or suggest further testing.

Bibliography

- Reid Blackman (2020). *A Practical Guide to Building Ethical AI*. Harvard Business Review.
- Anna Jobin, Marcello Ienca & Effy Vayena (2019). *The Global Landscape of AI Ethics Guidelines*. Nature Machine Intelligence.
- Luciano Floridi and Josh Cowls (2019). *A Unified Framework of Five Principles for AI in Society*. Harvard Data Science Review.
- Sandra Wachter, Brent Mittelstadt & Chris Russell (2017). *Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR*. Harvard Journal of Law & Technology.
- Paula Boddington (2017). *Towards a Code of Ethics for Artificial Intelligence*. Springer.
- Brent Mittelstadt, Patrick Allo, Mariarosaria Taddeo & Sandra Wachter (2016). *The Ethics of Algorithms: Mapping the Debate*. Big Data & Society(Research.