Formalizing Ethical Design in Prostate Cancer Image Analysis: A Preliminary Case Study

Sadie Lee
Cognitive Systems and Data Science
University of British Columbia
Vancouver, Canada
slee103@student.ubc.ca

Adam Resnick
College of Medicine and Science
Mayo Clinic
Rochester, United States
Resnick.Adam@mayo.edu

Nasibeh Zanjirani-Farahani
College of Medicine and Science
Mayo Clinic
Rochester, United States
ZanjiraniFarahani.Nasibeh@mayo.edu

***Abstract*—Although artificial intelligence (AI) has shown potential to revolutionize healthcare, adoption in clinical practice has been limited due to concerns regarding trustworthiness and patient outcomes. Several ethical guidelines have been issued in recent years, however, implementation is often challenging. This study utilizes a preliminary case study to implement the Coalition for Health AI (CHAI) ethical guidelines in the development of a prostate abnormality detection system, using formalized specifications to address the challenge of translating abstract principles into design requirements. Initial evaluation indicates that such formalism is effective for this translation, and that implementation of ethical guidelines is necessary for trustworthy AI systems in healthcare.**

***Index Terms*—Formal specification, ethical design, medical image analysis, artificial intelligence, translation to practice**

I. Introduction

With the transition of the healthcare industry to a digital domain, artificial intelligence (AI) has shown immense potential in the transformation of healthcare by enabling the analysis of large volumes of medical data. However, adoption in clinical practice has been limited due to concerns regarding the trustworthiness of AI systems. Issues with data privacy, bias, lack of transparency, safety, and reliability have been of concern [1]. In a medical context, there is a prominent ethical imperative for AI systems given the high-risk nature [2], and thus the implementation of ethical guidelines is necessary.

Several ethical AI guidelines and recommendations have been published worldwide in recent years [3]. While consisting of similar themes, there is little collective consensus for which should be implemented in a given setting, especially in high-risk settings such as healthcare, and thus many have not been implemented in practice. Different stakeholders may also apply different ethical frameworks that do not contain formal definitions or specifications, resulting in a lack of cohesiveness with vague responsibilities of each stakeholder [4].

The Coalition for Health AI (CHAI) has collaboratively worked to develop practical quality standards for the ethical development and deployment of AI solutions with the aim of improving trustworthiness and increasing adoption of these solutions in clinical practice [4]. However, challenges to the implementation of such ethical guidelines have led to limited uptake [5].

As such, the contributions of this study are the implementation of the CHAI ethical guidelines for medical AI development with a formalization of specifications to improve translation to design requirements. Resulting from the differences that are expected for various AI systems in healthcare, we focus on the case of medical imaging, specifically prostate abnormality detection from magnetic resonance (MR) images. Moreover, given the preliminary nature of the case study, the scope of this paper focuses on the design and preparation stages of the AI lifecycle.

II. Related Work

A. Coalition for Health AI Guidelines

The goal of the CHAI quality standards is to increase the reliability, safety, and trustworthiness of AI systems in healthcare throughout the end-to-end lifecycle [4]. The six-stage lifecycle for AI systems in healthcare developed by CHAI encompasses processes from initial problem identification and solution planning to the evaluation and deployment of an AI solution, and how it is integrated in the clinical workflow. This lifecycle is built on a set of core principles that are relevant to each stage and are, 1) usefulness, usability, and efficacy; 2) fairness and equity; 3) safety and reliability; 4) transparency, accountability, intelligibility; and 5) security and privacy [4].

B. Challenges to Implementation

Despite the ubiquity of ethical AI guidelines, challenges to their implementation into an AI system have been categorized by [5] into five levels: ethical principles, design, technology, organizational, and regulatory. We focus on the first two challenges, ethical principles and design, given the scope of the paper.

The challenge of ethical principles is the relationships between them: conflicts and contradictions may occur. For example, if an AI system is trained on retrospective data, historical biases are embedded such that certain groups are naturally favored. Other groups then need to be protected to ensure fairness in the dataset and thus in the decisions made by an AI system. However, fairness is often achieved by an optimal balance of impact, performance, and resources which inherently requires tradeoffs. Optimal performance, for example, may lead to unequal outcomes for a group with

certain protected characteristics due to lack of testing in the population [5].

At the design level, the translation of abstract ethical principles into tangible design requirements, features, and functionalities of an AI system in healthcare can be challenging to implement [5]. The abstract ethical principles such as fairness and transparency are open to interpretation, and formal definitions are uncommon in a healthcare context. Conflicts may also occur at the design level, whereby ensuring full transparency, for example, could compromise patient privacy.

III. Case Study – Prostate MR Image Analysis

A. Project Background

While designing and implementing processes for AI development, we were tasked with exploring ethical practices and how these could be addressed specifically for medical image analysis at a large healthcare organization in the Midwest. We viewed this as an opportunity to apply the CHAI ethical guidelines and develop formal specifications to assist in translating these guidelines into design requirements. The use case focuses on an abnormality detection system for prostate MR images. Stages of development with the principles implemented are outlined in Figure 1.

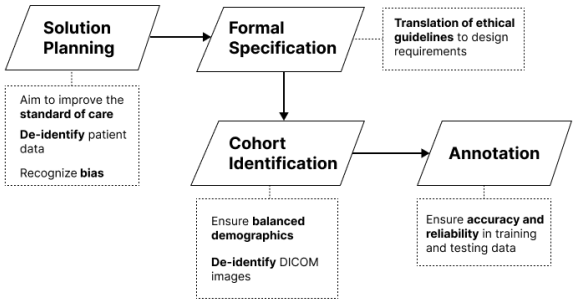

Fig. 1. Stages of development integrating the CHAI ethical guidelines in the context of the prostate abnormality detection use case.

B. Prostate Image Analysis

Prostate cancer (PCa) is one of the most commonly diagnosed malignant carcinomas, and second leading cause of mortality from cancer in men worldwide [6]. Diagnostic accuracy is thus significant to the prognosis of patients with PCa. While systematic tissue biopsies have previously remained the standard of care for diagnosis, advances in MRI have enabled non-invasive, lower-risk procedures to be performed in addition or replacement to systematic biopsies [6].

Algorithms have been developed to assist diagnostic radiologists who interpret MR images looking for prostate abnormalities. These algorithms can perform image analysis at various stages and have the potential to improve the PCa screening workflow where the AI system has the potential to indicate abnormalities that may not have been identified by a radiologist, enabling patients to benefit from quicker, more precise results [7].

C. Solution Planning

The solution planning stage intended to address the CHAI principles of usefulness and efficacy by considering the necessity of the AI system in relation to current standards of care and potential integration into the clinical workflow. It was assessed that the prostate abnormality detection system aimed to improve the standard of care by enabling radiologists as the intended end users to identify abnormalities that may not have been previously seen. Considering potential sources of bias, as per CHAI guidelines in this stage, we identified that our dataset was primarily comprised of patients from one region (Midwest) and the majority identified as white, likely due to the demographics of patients who have received care from the healthcare organization. This has the potential to reduce overall generalizability, and thus ensuring fairness was necessary in the cohort identification stage.

Furthermore, the use case was assessed to require clinical unstructured data (MR images and radiology reports), as well as structured data (e.g. patient demographics and diagnoses) to identify the relevant images. Given the technical need of structured patient data from the electronic health record (EHR), albeit retrospectively, the CHAI core principle of privacy was identified as necessary to ensure that the patient data was protected. This was done through a comprehensive data de-identification process to ensure personal health information (PHI) was not revealed. PHI includes but is not limited to name, medical record number, date of birth, age, ethnicity, date of visits, and locations where services were received according to the Health Insurance Portability and Accountability Act (HIPAA) [4]. Patients from the healthcare organization gave consent to de-identified data collection.

As the use case aimed to detect prostate abnormalities, risks to patients were recognized, particularly regarding forms of cognitive bias such as automation bias [8], which could occur due to over-reliance on indications made by the AI system and less attention to images not flagged by the system.

Assessing integration into the clinical workflow, we determined that the detection system should assist radiologists in interpreting images from multi-parametric MRI (mpMRI) procedures alongside AI output in the form of image annotation where the AI system annotates images from the mpMRI exam to indicate potential abnormalities.

D. Formal Specifications

With the lack of formalisms for ethical guidelines leading to vaguely defined responsibilities among stakeholders, we focus on applying formal specifications to the design and preparation stages of the AI lifecycle. Formal specifications use mathematical notation to unambiguously define the characteristics an information system must have [10]. Z notation was chosen for its schema calculus, which allows for greater modularity [10]. The aim was to develop formal specifications for the CHAI ethical principles in the context of the prostate abnormality detection system.

1) Usefulness, usability, efficacy: Usefulness, usability, and efficacy are specified where the detection system must improve the standard of care (i.e. *ImprovedCare* must be true), the set of annotated abnormalities cannot be empty, and the set of users who are assisted (e.g. radiologists) must be the same set for end users, meaning that all users receiving assistance are the end users.

UsefulnessUsabilityEfficacy

$[ABNORMALITY, USER]$
$ImprovedCare : BOOL$
$AnnotatedAbnormalities : \mathbb{P}\, ABNORMALITY$
$AssistedUsers : \mathbb{P}\, USER$
$EndUsers : \mathbb{P}\, USER$

$ImprovedCare = TRUE$
$AnnotatedAbnormalities \neq \emptyset$
$AssistedUsers = EndUsers$

2) Fairness and equity: Fairness and equity are specified as treating all users without bias based on protected attributes such as race, gender, or their equivalents in the identification of a cohort; i.e. for every user x and y, and for every pair of protected attributes a and b associated with x and y, if x and y have different protected attributes a and b, then x and y must receive the same treatment.

FairnessEquity

$[USER, ATTRIBUTE]$
$ProtectedAttribute : USER \nrightarrow ATTRIBUTE$
$Treatment : USER \rightarrow TREATMENT$

$\forall x, y : USER;\ a, b : ATTRIBUTE \bullet$
$(x \in \mathrm{dom}\, ProtectedAttribute \land y \in \mathrm{dom}$
$ProtectedAttribute \land$
$ProtectedAttribute(x) = a \land ProtectedAttribute(y)$
$= b \land a \neq b) \Rightarrow (Treatment(x) = Treatment(y))$

3) Safety and reliability: Safety and reliability are specified such that for every dataset d, if d is part of the data used for training and validation, then d must be diverse and representative.

SafetyReliability

$[DATASET]$
$TrainedOn : \mathbb{P}\, DATASET$
$ValidatedOn : \mathbb{P}\, DATASET$
$Diverse : DATASET \rightarrow BOOL$
$Representative : DATASET \rightarrow BOOL$
$AccuracyEnsured : BOOL$
$BiasMinimized : BOOL$

$\forall d : DATASET \bullet (d \in TrainedOn \lor$
$d \in ValidatedOn) \Rightarrow$
$(Diverse(d) \land Representative(d))$
$AccuracyEnsured = TRUE$
$BiasMinimized = TRUE$

4) Transparency, intelligibility, and accountability: Transparency, accountability, and intelligibility are specified where for every user u who is an intended user of the system, user oversight must be allowed (i.e. *Oversight* must be true), and the system must implement mechanisms to mitigate risks to patients (i.e. *RiskMitigation* must be true).

TransparencyAccountabilityIntelligibility

$[USER]$
$IntendedUsers : \mathbb{P}\, USER$
$Oversight : BOOL$
$RiskMitigated : BOOL$

$\forall u : USER \bullet (u \in IntendedUsers) \Rightarrow$
$(Oversight = TRUE)$
$RiskMitigated = TRUE$

5) Privacy and security: Privacy and security are specified such that patient data must be kept confidential, cannot be shared without consent, and must be de-identified in both images and structured data; i.e. for every patient p, and for every piece of structured data d or image i associated with p, if d or i is to be shared (i.e. included in $Consent(p)$), then d and i must be de-identified.

PrivacySecurity

$[PATIENT, DATA, IMAGE]$
$PatientData : PATIENT \leftrightarrow DATA$
$PatientImages : PATIENT \leftrightarrow IMAGE$
$Consent : PATIENT \rightarrow \mathbb{P}(DATA \cup IMAGE)$
$DeIdentified : (DATA \cup IMAGE) \rightarrow BOOL$

$\forall p : PATIENT;\ d : DATA;\ i : IMAGE \mid$
$(d \in PatientData[p] \lor i \in PatientImages[p]) \bullet$
$(d \in Consent(p) \lor i \in Consent(p)) \Rightarrow$
$(DeIdentified(d) = TRUE \land$
$DeIdentified(i) = TRUE)$

E. Cohort Identification

As identified in the first stage, mitigation of bias was necessary to ensure fairness based on the CHAI core principles. Data bias was minimized when identifying the patient cohort to be used as the dataset by ensuring balanced demographics with the available structured patient data. Patients were identified with Structured Query Language (SQL) from a harmonized de-identified database containing retrospective structured and unstructured clinical data. With the onset of prostate cancer typically occurring in men ages forty and older [7], criteria for inclusion were patients who identified as male, were forty and older, and had a prostate MRI examination at the healthcare organization. Potential outliers were identified who had prostate MRI exams but did not meet these criteria, although given the preliminary nature of the case study, only a small subset of patients was able to be used as part of the dataset.

In addition to metadata in Digital Imaging and Communications in Medicine (DICOM) images, PHI may also be

embedded in images themselves and should be removed to ensure patient privacy [8]. Once images were identified as relevant based on the inclusion criteria, images were de-identified.

F. Annotation

With the de-identified prostate MR images, abnormalities were annotated by radiologists to ensure the CHAI principles of accuracy and reliability. Abnormalities were defined as markedly hypointense or a potential area of extraprostatic extension (EPE). For imaging data, the region of interest (ROI) is labeled either by 1) marking the approximate centroid of the target; 2) drawing a bounding box around the target; or 3) drawing the contour of the target (pixel-based annotation) depending on the modeling task. Annotation with the prostate MR images consisted of manually drawing binary mask contours (Method 3) to overlay the specific location of the abnormality.

IV. Results

A. Cohort Identification

The final dataset for the preliminary case study based on associated structured patient data was comprised of 5 patients, 7 studies where studies are all the images acquired in a given imaging protocol, 47 T2-weighted imaging series where imaging series are the specific type of data captured by an imaging modality (in this case, MR) given a pre-determined set of acquisition parameters, and 1181 images where an image is defined as one slice in an imaging series.

TABLE I
Cohort Identification

Patients	Studies	Series	Images
5	7	47	1181

V. Discussion

The formal specifications with Z notation assisted in understanding and implementing ethical guidelines as design requirements of the prostate abnormality detection system. With the design and preparation scope of the case study, the CHAI principles primarily focused on were usefulness, usability, and efficacy in solution planning, fairness and equity in cohort identification, privacy and security in de-identification of data, and safety and reliability in annotation.

Furthermore, the study highlighted the need to address potential sources of bias throughout the AI lifecycle, including data bias and automation bias. Particularly in the high-risk setting of diagnostic radiology, it is crucial to mitigate these risks. Identifying bias in the solution planning stage and the utilization of structured patient data to identify MR images relevant to the detection task enabled increased transparency for performance in varying populations.

Given the preliminary nature of the case study, the primary limitations were scoping and time constraints–the CHAI ethical guidelines were unable to be implemented into the modeling and deployment stages of the AI lifecycle. Furthermore, a major challenge in evaluating the implementation of the ethical guidelines in the case study was the absence of clearly defined measures for successful implementation, as also noted by [5]. Future work is needed to establish clear measures of success and concretely evaluate whether the implementation of the ethical guidelines was effective.

A. Conclusion

The CHAI ethical guidelines were implemented in a preliminary case study for a prostate abnormality detection system which is beneficial to better understand sources of potential bias and ensure ethical development. We addressed the challenges to implementation of ethical guidelines by providing formal specifications for the CHAI principles as design requirements. Using these specifications to guide the design and preparation processes beginning with evaluating requirements and identifying a cohort of patients with structured data may thus assist in ensuring the ethical use of AI systems in medical imaging workflows, although further work is needed to evaluate its impact in clinical practice.

References

[1] J. R. Geis et al., "Ethics of Artificial Intelligence in Radiology: Summary of the Joint European and North American Multisociety Statement," Radiology, vol. 293, no. 2, pp. 436–440, Nov. 2019, doi: 10.1148/radiol.2019191586.

[2] D. Peters, K. Vold, D. Robinson, and R. A. Calvo, "Responsible AI—Two Frameworks for Ethical Design Practice," IEEE Transactions on Technology and Society, vol. 1, no. 1, pp. 34–47, Mar. 2020, doi: 10.1109/TTS.2020.2974991.

[3] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines," Nat Mach Intell, vol. 1, no. 9, pp. 389–399, Sep. 2019, doi: 10.1038/s42256-019-0088-2.

[4] N. Economou, M. Elmore, A. Callahan, J. McCall, and R. Baig, "Assurance Standards Guide Coalition for Health AI (CHAI)," CHAI, https://chai.org/assurance-standards-guide/ (accessed Aug. 2, 2024).

[5] M. Goirand, E. Austin, and R. Clay-Williams, "Implementing Ethics in Healthcare AI-Based Applications: A Scoping Review," Sci Eng Ethics, vol. 27, no. 5, p. 61, Sep. 2021, doi: 10.1007/s11948-021-00336-3.

[6] F.-J. H. Drost et al., "Prostate MRI, with or without MRI-targeted biopsy, and systematic biopsy for detecting prostate cancer," Cochrane Database of Systematic Reviews, vol. 2019, no. 4, Apr. 2019, doi: 10.1002/14651858.CD012663.pub2.

[7] S. Wang and R. M. Summers, "Machine Learning and Radiology," Med Image Anal, vol. 16, no. 5, pp. 933–951, Jul. 2012, doi: 10.1016/j.media.2012.02.005.

[8] P. H. Gann, "Risk Factors for Prostate Cancer," Rev Urol, vol. 4, no. Suppl 5, pp. S3–S10, 2002.

[9] B. Kocak et al., "Bias in artificial intelligence for medical imaging: fundamentals, detection, avoidance, mitigation, challenges, ethics, and prospects," Diagnostic and interventional radiology (Ankara, Turkey), Jul. 2024, doi: 10.4274/dir.2024.242854.

[10] J. M. Spivey, "An introduction to Z and formal specifications," Software Engineering Journal, vol. 4, no. 1, p. 40, 1989. doi:10.1049/sej.1989.0006