CHAI Assurance Standards Guide

AI THAT SERVES ALL OF US.

CHAI

CHAI ASSURANCE STANDARDS GUIDE

The CHAI Assurance Standards Guide provides comprehensive guidance on quality and ethics for AI in healthcare. This guide was created by patient advocates, technology developers, clinicians, and data scientists, who worked together to create a common framework for AI standards in healthcare, based on real-world practices. This guide is designed for a wide audience, including stakeholders who are involved in the AI design, development, deployment, and usage processes.

The purpose of this guide is to ensure that AI technologies used in healthcare are reliable, safe, and effective. It combines existing standards into a coherent framework, providing practical considerations for applying these standards in day-to-day operations. By covering key principles such as usefulness, fairness, safety, transparency, and security, the guide aims to support the ethical development and implementation of AI solutions in healthcare.

CHAI

INTRODUCTION

AI has the potential to revolutionize healthcare by enabling sophisticated analysis of vast datasets, transforming patient care and administrative processes. However, AI also carries risks, including bias and the potential to perpetuate social inequities. Despite numerous guidelines, actionable standards are needed to ensure that AI in healthcare is safe, effective, and equitable.

Healthcare has long used data-driven algorithms to assist clinical decision-making and enhance administrative processes. With recent progress in AI, a new range of opportunities has arisen, allowing more advanced analysis of large datasets. AI in healthcare can be described as the use of algorithmic systems for a variety of tasks, such as decision support, diagnosis, treatment planning, medical imaging analysis, patient monitoring, clinical notetaking, precision medicine, and various administrative processes.

However, the rapid advancement of AI technology has also brought new challenges. Without shared standards of practice, there is a risk that AI could exacerbate existing disparities in healthcare, deepening the divide in accessibility and outcomes. The lack of standardization also poses challenges for technology developers and health systems, who must navigate a fragmented landscape of guidelines and regulations.

The CHAI Assurance Standards Guide is a significant step toward greater collaboration and alignment across the landscape of health AI. By translating core principles into practical considerations and anchoring those considerations to real-world use cases, the guide provides a concrete approach to bridging the gap between standards and practice.

THE AI LIFECYCLE

The AI lifecycle is central to understanding and implementing CHAI's Assurance Standards in healthcare. The six-step lifecycle outlines the essential stages and processes involved in developing, deploying, and maintaining AI systems.

By systematically addressing each phase of the lifecycle, the framework ensures that AI systems adhere to the highest standards of safety, efficacy, fairness, transparency, and security. This structured approach supports risk mitigation, managing biases, and promotes accountability and trustworthiness in AI applications.

1

- Engage stakeholders to define the problem & perform root-cause analysis
- Identify solution & plan future state
- Gather business requirements
- Assess feasibility, potential for impact, & prioritization
- Make procure/build/partner decision

2

- Select/understand model task & architecture
- Capture design & technical requirements or determine best solution to meet business requirements
- Design solution application & system workflow according to human-centered design principles
- Design deployment strategy with end users
- Design risk management, monitoring & reporting plan

3

- Access data
- Prepare data
- Develop data management plan
- Train & tune model

4

- Conduct installation qualification (when applicable)
- Validate local system performance (when applicable)
- Execute prospective, silent evaluation
- Establish risk management plan
- Train end users
- Test usefulness
- Ensure compliance with applicable healthcare regulations & standards

5

- Implement small-scale pilot to assess real-world impact
- Execute and update risk management plan
- Educate & train users on AI application reporting
- Assess usefulness and adoption

6

- Deploy at a larger scale on a general population
- Audit AI system to inform whether to maintain, refine or sunset
- Conduct ongoing risk management

CHAI

BRIEF DESCRIPTION OF THE AI LIFECYCLE

Define Problem and Plan

Identify the problem, understand stakeholder needs, evaluate feasibility, and decide whether to build, buy, or partner.

In this stage, the goal is to understand the specific problem an AI system is addressing. This involves conducting surveys, interviews, and research to find root causes. Teams will then decide whether to build a solution in-house, buy it, or partner with another organization.

Design the AI System

Capture technical requirements, design system workflow, and plan deployment strategy.

During design, the focus is on specifying what a system needs to do and how it will fit into a healthcare workflow. This involves defining requirements, designing the system, and planning for deployment & monitoring to ensure it meets the needs of providers and users.

Engineer the AI Solution

Develop and validate the AI model, prepare data, and plan for operational deployment.

This stage involves building an AI solution. The team will collect and prepare data, train AI models, and develop an interface for users. The goal is to create a functional AI system that can be tested and evaluated for accuracy and effectiveness.

Assess

Conduct local validation, establish a risk management plan, train end users, and ensure compliance.

The assessment stage tests AI systems to decide if they're ready for a pilot launch. This includes validating the system, training users, and ensuring it meets healthcare standards and regulations. The aim is to confirm that the system works correctly and is safe to use.

Pilot

Implement a small-scale pilot, monitor real-world impact, and update risk management.

In this stage, the AI systems are tested in real-world settings at a small scale. The goal is to evaluate its performance, user acceptance, and overall impact. Based on the results, the team will decide whether to proceed with a larger-scale deployment.

Deploy and Monitor

Deploy the AI solution at scale, conduct ongoing monitoring, and maintain quality assurance.

The final stage involves deploying AI systems at a larger scale and monitoring their performance. This ensures systems stay effective and can be adjusted as needed – maintaining high quality and reliability in healthcare.

CHAI

THE CORE PRINCIPLES FOR TRUSTWORTHY HEALTH AI

CHAI

Rationale

Before we describe the core principles of trustworthy AI, it's important to discuss why they are needed in the first place. In healthcare, ethics and quality are translated into practice through standards that guide risk-benefit assessments. These standards help ensure that each AI solution is thoroughly evaluated. By following these guidelines, organizations can perform comprehensive testing and local validation to assess the usefulness of AI solutions.

The importance of weighing risks and benefits can vary depending on the AI solution's context and purpose. For example, an AI tool for diagnosing life-threatening conditions like breast cancer requires more rigorous safety checks compared to one designed for health administrative tasks. Factors such as patient population, the severity of health issues, and the importance of the healthcare decision play a role in this assessment. Additionally, if users lack necessary knowledge to understand its output, this presents a risk that must be carefully managed.

Organizations will weigh these risks and benefits differently based on their risk tolerance, which is the level of risk they are willing to accept to achieve a desired outcome. For instance, higher risks might be acceptable in cancer treatment trials if the potential benefits are significant. Risk tolerance varies based on an organization's environment, culture, and values.

THE CORE PRINCIPLES FOR TRUSTWORTHY HEALTH AI

CHAI

Governance

AI governance within organizations ensures that standards based on quality and ethics are implemented effectively. Beyond local oversight, governance structures ensure independent reviews of AI solutions, meeting ethical standards and achieving high-quality outcomes. These structures drive accountability through standard operating procedures (SOPs) and clear role definitions.

Safety and security oversight should be placed high enough within the organization to enable prompt decision-making regarding resources, risk mitigation, incident response, and potential rollback of AI systems. Organizational stakeholders should establish clear governance policies to manage risks and changes, considering the organization's mission, risk tolerance, and legal obligations. Additionally, organizations need a bias management framework to evaluate AI fairness and equity throughout its lifecycle. This involves clear role definitions and accountability, including independent audits.

Proper training for those involved in AI selection, development, and deployment is crucial, focusing on ethics, quality, intended use, risks, and limitations. Risks should be framed by organizational standards for adverse events, allowing for appropriate risk assessment. Finally, organizations should ensure transparency and AI intelligibility to the public to enhance trust and awareness.

THE CORE PRINCIPLES FOR TRUSTWORTHY HEALTH AI

Usefulness, Usability & Efficacy

AI solutions should be beneficial, reliable, and improve user experience. They must solve specific problems and show clear benefits for patients and healthcare providers, such as better clinical outcomes and patient satisfaction. Usability means the AI should be easy to use and fit well into existing workflows. Efficacy ensures the AI achieves its goals and continues to perform well through ongoing testing and monitoring.

Fairness, Equity & Bias Management

AI solutions must be fair and work equally well for all demographic groups. Fairness means the AI's performance should be consistent across different groups, and outcomes should not depend on protected attributes like race or gender. Equity involves ensuring that AI solutions help reduce health disparities. Bias management includes regularly checking and correcting any biases in the data or AI system to promote fairness and equity.

Safety & Reliability

AI solutions should not harm patients or healthcare providers. This involves thorough testing and risk assessments before implementation, and continuous monitoring to detect and address any safety issues. Clear accountability and governance structures must be in place to ensure the AI system remains safe and reliable throughout its use.

Transparency, Intelligibility & Accountability

Stakeholders need clear and understandable information about AI systems and their outputs. Transparency involves sharing how the AI system works and its limitations. Intelligibility ensures stakeholders can understand the AI's decision-making processes. Accountability means being responsible for minimizing harm and addressing any negative impacts of the AI system.

Security & Privacy

AI systems must protect data confidentiality and integrity with strong security measures. This includes preventing unauthorized access and data breaches, and ensuring personal data is handled in compliance with privacy regulations. Organizations should have protocols for monitoring security and privacy, and for addressing any incidents, to keep data safe and maintain trust.

CHAI

IMPORTANCE OF INDEPENDENT REVIEW

Independent quality assurance is critical for ensuring the safety, efficacy, and trustworthiness of AI systems in healthcare. This process uncovers technical flaws, biases, and unintended behaviors that developers may overlook. By adhering to clear standards and benchmarks, AI solution developers can ensure transparency and accountability, ultimately building public trust.

Healthcare has a strong history of quality assurance and independent review standards for medications and devices, governed by the FDA in the US. This rigorous process of external review protects public health and safety by setting quality standards, ensuring safety and efficacy, and providing independent oversight.

To achieve the goals of independent quality assurance, AI solution developers must adopt clear standards and benchmarks for evaluating AI solutions on measures of safety, reliability, bias, fairness, and efficacy. Transparency regarding data used to develop models, AI methods, and validation processes are essential for accountability and public trust. Additionally, iterative review and data sharing support continuous learning and evaluation, promoting the long-term safety and efficacy of AI systems.

Independent review involves engaging third-party organizations to assess the performance and safety of AI systems. These organizations can provide an objective perspective, identifying potential issues that developers may have missed. Regular independent reviews can help ensure that AI systems continue to meet quality standards and operate safely and effectively.

Organizations should establish processes for conducting independent reviews at various stages of the AI lifecycle. This includes pre-implementation reviews to assess the design and development of AI systems, as well as post-implementation reviews to monitor ongoing performance and address any issues that arise. Independent reviews should be conducted by experts with the necessary technical and domain knowledge to evaluate the AI system thoroughly.

Transparency in the review process is essential for building trust in AI systems. Organizations should provide clear documentation of the review process, including the criteria used for evaluation, the findings of the review, and the actions taken to address any issues identified. This documentation should be accessible to all stakeholders, including patients, healthcare providers, and regulators.

IMPLEMENTING BEST PRACTICES FOR TRUSTWORTHY HEALTH AI

Stage 1 - Define Problem & Plan

Usefulness, Usability & Efficacy	Fairness, Equity & Bias Management	Safety	Transparency, Intelligibility & Accountability	Security & Privacy
Clearly explain the problem and why the AI solution is necessary	Ensure the AI solution does not disadvantage any groups	Identify potential harms and risks	Ensure there is a clear reason for using AI over non-AI solutions	Maintain complete documentation of AI systems and data
Assess how the AI will fit into existing workflows	Establish how fairness will be evaluated	Establish clear criteria for the patient population	Document the intended use and users of the AI solution	Establish and maintain policies to manage AI privacy and security risks
Evaluate the benefits, risks, and costs of the AI solution	Develop a strategy to monitor and mitigate biases	Ensure both developer and implementer are responsible for safety	Make project and model information accessible to all stakeholders	Clearly define the AI's purpose and ensure it aligns with organizational goals
Determine if end users will trust the AI solution	Identify socio-demographic groups at risk of bias	Ensure compliance with federal and local regulations	Keep thorough documentation of the AI solution	Conduct initial privacy and security risk assessments
Involve clinical experts in the AI's development and validation	Identify potential types and sources of bias	Address ethical and legal challenges	Clearly communicate potential risks to end users and patients	Regularly update risk as

CHAI

Stage 2: Design the AI System

Usefulness, Usability & Efficacy	Fairness, Equity & Bias Management	Safety	Transparency, Intelligibility & Accountability	Security & Privacy
Ensure usability is considered and documented	Ensure real-world outcomes are fair across all groups	Ensure users can control and override AI recommendations	Compare the AI system to the benchmarks and document predictors and validation methods	Trace AI system requirements to privacy and security risks
Document robustness testing and trust-building measures	Identify and document limitations and risks	Establish processes for error disclosure and legal considerations	Define clear and understandable decision thresholds	Implement user access control policies
Assess differences between development and implementation environments	Create easy and effective feedback mechanisms	Plan risk management from conception to deployment	Ensure documentation considers end user knowledge	Use privacy-enhancing technologies (PETs) to mitigate privacy and cybersecurity risks
	Ensure all stakeholders review and approve implementation processes	Determine if deployment constitutes HSR and meet IRB requirements	Assess performance across demographic groups and ensure explainability	Consider privacy preferences and contextual factors in design
		Establish a monitoring process for AEs and SAEs		
		Label AI models with development and limitation information		
		Define procedures for reporting flaws and safety concerns		
		Ensure the AI system allows for human oversight and intervention		
		Include end users in the design process		

Stage 3 - Engineer the AI Solution

Usefulness, Usability, & Efficacy	Fairness, Equity, & Bias Management	Safety	Transparency, Intelligibility, & Accountability	Security & Privacy
Assess data quality and integrity	Justify use of protected attributes	Ensure training data represents the deployment population	Plan data security and scalability	Implement controls for privacy and security requirements
Consider bias and fairness during feature extraction	Address disparities between training data and target population	Monitor data quality and dataset drifts	Ensure transparency in data monitoring	Ensure data management policies address privacy and cybersecurity risks
Ensure data availability for model training matches deployment	Define and assess socio-demographic subgroups	Trace complaints, ethical concerns, and safety risks	Include socio-demographic information and diversity details	Protect against unauthorized access and data leaks
	Assess data quality by socio-demographic factors	Apply clear inclusion/exclusion criteria	Document data provenance and limitations	Ensure data inputs and provenance support accuracy and manage bias
	Evaluate proxies and composite scores for bias	Implement proper access controls and audit trails	Document data lineage	Protect development and production environments with secure user access
	Examine robustness of data representation	Ensure stakeholders understand roles in data quality	Implement version control for datasets	
	Ensure local data is representative for model tuning	Establish safety monitoring for adverse events	Assess patient impact and need for consent	
	Document training and test data for fairness and bias	Label AI models with development information	Ensure transparency in data manipulation rationale	

Stage 4: Assess

Usefulness, Usability, & Efficacy	Fairness, Equity, & Bias Management	Safety	Transparency, Intelligibility, & Accountability	Security & Privacy
Ensure AI integrates into workflows	Evaluate fairness and bias across subgroups	Evaluate local performance and safety	Report AI effectiveness to users and stakeholders	Train workforce on cybersecurity and privacy roles
Reassess if the AI addresses the problem	Ensure training and test datasets are independent	Implement risk management and assessment methods	Establish goals, standards, terms, and conditions	Assess performance of implemented controls
Reevaluate AI usability	Assess model performance and parity across subgroups	Triage and report risks to the implementer and developer	Define roles to foster trust and transparency	Identify third-party providers and ensure their compliance
Facilitate trust through risk-benefit assessment	Consider broader measures of performance and impact	Conduct verification and validation activities	Plan for data security and scalability	Perform risk assessment on third-party providers
Tailor AI to specific work contexts		Ensure transparency of validation methods and results	Ensure accessibility, equity, and explainability	Maintain third-party audit records
			Consider downstream impacts of AI	Ensure documentation of third-party systems
			Incorporate user feedback and documentation	Implement processes for third parties to report vulnerabilities
			Report on performance metrics and fairness audits	
			Test data and generalization contingencies	

Stage 5: Pilot

Usefulness, Usability, & Efficacy	Fairness, Equity, & Bias Management	Safety	Transparency, Intelligibility, & Accountability	Security & Privacy
Communicate AI capabilities to end users	Assess real-world outcomes for bias	Implement risk management and mitigation methods	Evaluate system's capacity to handle errors and data volume	Include stakeholder privacy preferences in algorithm design
Compare anticipated and actual benefits, risks, and costs	Ensure representativeness of pilot site and approach	Maintain monitoring for adverse events and serious adverse events	Provide education/training for end users	Implement and review audit log records
Re-assess usability in clinical environment	Evaluate human interaction and workflow impact	Implement a structured, transparent decision-making process	Identify ongoing audit monitoring methods	Establish configuration change control processes
Manage clinician disagreements with AI output		Mitigate automation bias	Assess end user experience	Ensure an incident response plan is in place
Assess user actions after AI interaction		Establish robust reporting and recall procedures	Consider continuous reporting methods	Establish delivery and resilience requirements for critical AI services
		Continue human factors evaluation	Communicate model limitations to users and patients	Examine and document privacy and cybersecurity risks
		Regularly review AI solution's relevance and obsolescence	Ensure transparency in clinical trials	Incorporate contextual factors into AI design

Stage 6: Deploy and Monitor

Usefulness, Usability, & Efficacy	Fairness, Equity, & Bias Management	Safety	Transparency, Intelligibility, & Accountability	Security & Privacy
Re-assess usability in the clinical environment	Monitor data drift impacts on bias	Implement risk management and assessment methods	Report effectiveness to end users and stakeholders	Support incident response plans with impact assessments
Evaluate AI integration in workflow	Identify responsible parties for monitoring bias	Maintain monitoring for adverse events and serious adverse events	Ensure patients are aware of AI use	Notify stakeholders about cybersecurity incidents or privacy events
Monitor AI solution performance over time	Mitigate model drift impacts on fairness	Regularly review AI relevance and obsolescence	Maintain access to project-related and model information	Continuously evaluate privacy risk
Manage clinician disagreements with AI output	Monitor system bias impacts effectively	Establish robust reporting and recall procedures		Assess and communicate compliance with legal requirements
Solicit and use end user feedback	Facilitate feedback from impacted populations	Implement proper access controls and audit trails		
Compare anticipated and actual benefits, risks, and costs	Assess risks of performance drift from pilot to full deployment	Report unintended uses of AI solution		
Assess user actions after AI interaction	Monitor AI system performance and parity	Conduct impact analysis on safety and benefit measures		
Support user trust building	Clarify accountability for data/model breaches	Mitigate automation bias		
Define inclusion/exclusion criteria for patients	Evaluate transition impacts from pilot to full deployment	Ensure AI solutions are labeled with development information		
	Inform affected groups about AI role	Ensure end-of-life (EOL) processes are clear		
	Provide end-user feedback loops	Use assurance techniques for supply chain risk management		
		Ensure updates maintain safety and effectiveness		

USE CASE PROFILES

Predictive EHR Risk
Pediatric Asthma Exacerbation

A. Privacy and security measures are crucial as the AI application integrates with patient data from the EHR.

B. The solution predicts asthma exacerbations in pediatric patients using EHR data.

C. Additional or specific privacy and security measures, such as requiring authentication for access, may be needed.

Imaging Diagnostic
Mammography

A. AI solutions in medical imaging, such as mammography, require FDA clearance.

B. Organizations must ensure compliance with federal regulations throughout the development and deployment process.

C. Clear documentation of compliance and adherence to regulations is essential.

Generative AI
EHR Query and Extraction

A. This use case addresses the challenge of navigating vast, unstructured EHR data.

B. Compliance with local privacy laws and international standards, such as GDPR and HIPAA, is essential.

C. Establishing specific goals and outcome measures for the AI tool, and handling adverse events or system failures, is crucial.

Claims-Based Outpatient
Care Management

A. End users, including physicians and nurses, must be able to control and override AI recommendations.

B. Monitoring model drift and ensuring instructions are tailored to different user types are critical considerations.

C. Assigning responsibility for monitoring to the appropriate individuals is essential for effective implementation.

Clinical Ops & Administration
Prior Authorization with Medical Coding

A. This use case involves automating prior authorization processes using AI.

B. Ensuring transparency and understanding of model outputs among end users is vital.

C. Providing clear guidelines for using AI outputs and integrating human oversight is necessary.

Genomics
Precision Oncology with Genomic Markers

A. Integrating clinical data, genomic insights, and clinical trial findings to identify the best treatment for patients.

B. Transparency about biases in datasets and clear documentation are crucial.

C. Addressing underrepresentation of certain demographics in clinical trials is essential for fairness and accuracy.