



September 25, 2024

Brenton Hill, JD MHA
Head of Operations and General Counsel
Coalition for Health AI

Re: Coalition for Health AI (CHAI) Assurance Standards Guide and Assurance Reporting Checklists

Dear Mr. Hill,

Thank you for the opportunity to comment on CHAI's Assurance Standards Guide and Reporting Checklists. We look forward to supporting CHAI products to be the best they can be for consumer use. AMIA applauds the vast effort CHAI undertook to create this comprehensive document and the associated Checklist.

The American Medical Informatics Association (AMIA) is the professional home for more than 5,500 informatics professionals, representing frontline clinicians, researchers, public health experts, and educators who bring meaning to data, manage information, and generate new knowledge across the research and healthcare enterprise. As the voice of the nation's biomedical and health informatics professionals, AMIA plays a leading role in advancing health and wellness by moving basic research findings from bench to bedside, and evaluating interventions, innovations and public policy across care settings and patient populations.

First, AMIA requests clarity as to the CHAI's intention for the Guide. Is the Guide exclusively intended to inform best practices, or is it also intended to inform regulatory efforts?¹ If it is the latter, AMIA recommends including clear intersections with existing regulations (FDA, ISO, etc.).

Second, and most saliently, while the comprehensiveness of the Guide and its promotion of assurance standards for AI is a positive, AMIA urges CHAI to amend the Guide to be much more accessible, redundant language removed, and connections between various areas clarified prior to being finalized. The majority of AMIA's feedback stems from the concern that the Guide is inaccessible as it is too lengthy to be consulted as a reference guide. In addition to its length, the guide's redundant

¹ See Gottlieb: [Congress Must Update FDA Regulations for Medical AI | JAMA Forum | JAMA Health Forum | JAMA Network](#) and Blumenthal: [The Regulation of Clinical Artificial Intelligence | NEJM AI](#)

language in sections makes it hard to evaluate and connect relevant pieces. AMIA also urges expanding and updating the healthcare examples contained in the appendices.

It is appropriate for themes to come up in multiple Stages, but the way it is currently written makes it burdensome and confusing to those reading the entire guideline. AMIA suggests providing clarity as to the recommended use of the Guide. Is it intended to be consumed as whole, as a primer to the Checklist, or as a reference guide that can be thumbed through to find answers to specific questions? For example, if the redundancy is purposeful to allow readers who will only be responsible for a specific Stage to have detailed guidance without having to read the detail for every Stage, then it may be helpful to clarify that in the executive summary and introduction. At the same time, individuals responsible for a specific concept rather than a Stage will likely find the table useful to focus on associated or connected tasks to be considered or completed.

Importantly, the healthcare examples are limited and require updating in the appendices. More are needed to ensure this is helpful to all interest holders who will be implementing AI in their provision of care. AMIA suggests linking the principles with the use cases so that the examples are clearer. AMIA has examples available from our annual [AI Showcase](#). **We suggest discussing the most current examples with CHAI following AMIA's upcoming AI Showcase this November.** More immediately, we offer several examples from a quick PubMed search published in 2023-2024:

- [artificial intelligence in clinical care jama - Search Results - PubMed \(nih.gov\)](#)
- [artificial intelligence in clinical care jamia - Search Results - PubMed \(nih.gov\)](#)
- [artificial intelligence new england journal of medicine - Search Results - PubMed \(nih.gov\)](#)
- [artificial intelligence lancet - Search Results - PubMed \(nih.gov\)](#)

Third, AMIA suggests CHAI address decommissioning and archiving of historic AI model deployments for historical knowledge.

Fourth, and final, more specific comments to the Guide are listed below.

Section 4

- 1) The AI Lifecycle section ends at “Deploy and Monitor.” While the authors do discuss the need to monitor over time to detect drift or other issues, there does not seem to be anything explicit about the re-assessment and re-validation that would need to be done should the solution be ported to a new context of use, new population, new geographic area, or in the event of any other major system change that could have

an impact on the characteristics of the population to which the AI would be applied - e.g. a hospital consolidation or merger.

- 2) More generally, while there is some discussion of data reliability and assessment, it is unclear that there is sufficient practical guidance to the end-user to understand how to assess this in a context-specific application, or how to mitigate potential gaps, particularly with respect to equity considerations.

Section 5

- 1) The AI Lifecycle on page 14 is informative and a solid foundation for organizing use of innovative AI systems in healthcare. However, the value and fundamental element to the success of AI tools in healthcare is the healthcare data. It is suggested that the proposed AI Lifecycle move Stage 3 as Stage 1, or at least emphasize the need to concurrently determine data requirements as AI tools are being developed. This bottom-up approach requires that AI developers (at the onset of AI model development) identify health care data types and sources that are uniquely defined by the complex patient and population characteristics that unfold in the healthcare ecosystem in various settings. See:
<https://academic.oup.com/jamia/article/30/7/1323/7163188?searchresult=1>.
- 2) Regarding the first Core Principle for Trustworthy Health AI on page 17, a key element of governance is not only highlighting risk of AI systems, but also quantifying risk. A suggested consideration for Core Principles of Trustworthy Health AI should be organizing resources and experts within the AI community to define and provide tools to better assess and quantify risk. CHAI can lead these efforts to better clarify within the AI community what levels of risk will undergo regulation and evaluation for AI tools used in healthcare settings. See:
<https://arxiv.org/pdf/2209.06317>.
- 3) Additional comments on page 17, under “Core Principles for Trustworthy Health AI”:
 - a. Item one is titled “usefulness, usability, and efficacy”, which appears redundant in its use of usefulness and usability. The section attempts to define “useful” but uses “usable” in that definition, which makes it ever harder to distinguish between the two. We encourage CHAI to use standard definitions for these terms. “Usefulness” may be more clearly defined by the value a user finds in an AI product, whereas “usability” may be defined by how easy it is for a user to complete a task using the AI product.
 - b. Alternatively, “usefulness” may be replaced with “utility”
 - i. In the same section, the Guide states “an AI solution must provide a specific benefit to patients and/or healthcare delivery”. Healthcare is the provision of health to patients; it is not necessary to distinguish “benefit to patients” and “healthcare delivery”. It may be acceptable, however, to state a specific benefit toward reducing the cognitive or workflow burden on healthcare providers”.

- c. At the end of the first paragraph under this section, it is worth noting that the increased efficiency in workflow will also result in user satisfaction, in addition to reduced costs.
- 4) “5.1 Core Principles in Context” beginning on page 20 discusses the necessity of understanding AI tools in clinical settings. We suggest review of “Recommendations for the safe, effective use of adaptive CDS in the US healthcare system: an AMIA position paper” to inform this section.²
- 5) At the bottom of Page 20, NIST is referenced, but not cited.

Section 7

AMIA has several questions to point out here from the perspective of healthcare organizations (HCOs) who may try to implement the guidelines.

- 1) What portion of HCOs have the expertise and experience to implement the AI best practices? If they do not all have this ability, are there concerns that the implementation may be biased to only the best-funded HCOs, further leaving behind smaller, community, or rural HCOs? It may be important to call out the need to “level the playing field” when it comes to the availability of resources required to implement AI driven tools in a healthcare setting to support equitable healthcare delivery.
- 2) What is the financial, human resource, and burden cost to implement all the recommendations in this guideline? Is it realistic to expect all of these will be followed? If not, what are the most essential aspects that must be followed to achieve assurance? AMIA recommends an estimate of risk, burden, cost, and feasibility to understand the final recommendations are realistic.
- 3) AMIA supports the helpful approach stated in the introduction:
Elaborating on considerations from each core principle described above, it discusses how to implement those principles in the form of assessments at each stage of the AI Lifecycle, while also demonstrating use case-dependent variations at the end of each stage. Considerations may be repeated with nuances at different stages or under different principles.

Fairness and Equity: Stage 1

- 4) Under “Define and apply *fairness and equity* in the context of the problem and its AI solution,” on page 24, does CHAI have a recommendation as to who defines the problem? Is this left vague on purpose?

² Petersen C, Smith J, Freimuth RR, Goodman KW, Jackson GP, Kannry J, Liu H, Madhavan S, Sittig DF, Wright A. Recommendations for the safe, effective use of adaptive CDS in the US healthcare system: an AMIA position paper. J Am Med Inform Assoc. 2021 Mar 18;28(4):677-684. doi: 10.1093/jamia/ocaa319. PMID: 33447854; PMCID: PMC7973467.

- 5) AMIA suggests comparing the risk of deploying with the risk of *not* deploying the AI solution, such as continuing with the status quo or an alternative approach to the problem.
- 6) AMIA recommends an approach for what happens when computation methods cannot fully avoid bias. What is an acceptable amount of bias? How is such bias mitigated? How are biases made transparent to the users?
- 7) Under “Determine whether externally acquired AI solutions comply with privacy and data security policies,” on page 24, consider whether impermissible information blocking may impair the function of the AI solution and whether a permissible exception is helpful or harmful to fairness and equity.

Safety and Reliability: Stage 1

- 8) Under “Ensure that the developer and the implementer organizations are responsible for the safety, effectiveness, and performance of the AI solution throughout its lifecycle,” on page 26, we encourage CHAI to include the impact of data drift.
- 9) Under “Consider ethical and legal challenges and how they will be handled,” on page 26, AMIA encourages CHAI to consider the further question of whether patients will be made aware *how* AI will be used, not just that AI is being used. For example, whether AI is being used autonomously versus as a virtual assistant in clinical decision-making. This must be presented in terms that are clear and understandable to patients and all end users.

Transparency, Intelligibility, and Accountability: Stage 1

- 10) Under “Determine what types of information should be documented,” on page 27, CHAI should encourage the documentation to include data sources, data quality, and known biases in the training data in the details regarding data.
- 11) Under “Consider how to communicate potential risks of an AI solution to end users and/or patients,” on page 27, it is not only important to determine how risks will be evaluated, but also how they will be identified, the likelihood of the risk’s occurrence, the risk’s impact, and possible risk mitigation tactics. It may be necessary at times to differentiate between the chance of failure (risk) vs. the consequences of that risk (level of outcome).
- 12) Under “Assess factors pertaining to impact on patients,” on page 28, cost should be listed with “risks and benefits” as well as part of determining whether a patient can opt out of having the AI solution. AMIA cautions the Guide from assuming it may be a simple “yes” or “no” for all AI uses. However, it may be burdensome or even prohibitive to discuss every component of care that employs AI to accomplish care-related information management tasks in a clear way to all various end users. In addition to prioritizing transparency, which is imperative, AI implementation must not add additional excess burden to healthcare providers, or the overall system. AI has the potential to ameliorate the glaring issue of documentation burden that is currently

devastating our healthcare workforce and patient access to quality care, while aligning with quality workflows. AI use must enhance accuracy of data capture, improve quality of patient information collected, streamline documentation needs for reimbursement purposes and eliminate requirements for duplicate data entries. Focusing on eliminating documentation redundancies while maintaining patient data accuracy is direly needed to avoid preventable errors, address liability concerns, and positively impact patient care.

- 13) Under “Establish specific goals, standards, terms, and conditions,” on page 28, AMIA encourages that there should be a determination of AI solution’s deployment *and outcomes* goals. Indeed, the outcome goals may be a more important measure of success.
- 14) On page 28, AMIA suggests CHAI change “Consider stakeholder engagement” to “Ensure stakeholder engagement in the final consideration under the Transparency, Intelligibility, and Accountability section.

Security and Privacy: Stage 1

- 15) The consideration “Define the proposed use of AI systems in relation to specific mission/business objectives,” on page 29, can be deleted as it is redundant with content elsewhere in the section.

Fairness and Equity: Stage 2

- 16) Under “Consider appropriate and effective channels for end user feedback related to bias and fairness,” on page 31, we encourage CHAI to add that such feedback will be “replied to” in a timely manner and that users are not discouraged from providing feedback due to a lack of responsiveness to their concerns.

Safety and Reliability: Stage 2

- 17) On page 32, the final sentence under “Ensure that a process is in place to manage ethical and legal challenges” is redundant with the following section, “Plan risk assessment methods from conception through to deployment of the health AI solution.”
- 18) On page 33, AMIA suggests adding a category of “errors without harm” under “Plan a monitoring process for adverse events (AEs) and serious adverse events (SAEs)”.
- 19) On page 33, AMIA strongly supports the explanation to the end user under “Ensure that AI models are labeled with transparent information about their development and limitations.” The AI solution should be able to respond to a user request to “show your work” (and “show your training data demographics”) underlying any particular decision. AMIA believes that the “Enable clinical intervention and override by ensuring that the AI system is intelligible to end users” is redundant.

Transparency, Intelligibility, and Accountability: Stage 2

- 20) In consideration of the decision thresholds, AMIA suggests (1) a threshold to test and (2) a threshold to treat. These are based on pre-test and post-test probabilities for specific patients on populations as well as the nature of the test and desired outcomes. These decisions require the input of those skilled in this sort of determination and AMIA is happy to partner with CHAI to develop further recommendations in this area.

Stage 5: Pilot (starting on page 51)

- 21) It would be helpful to see a discussion about how to select users and settings for the pilot. User engagement and readiness are critical to pilot effectiveness, regardless of whether the pilot is a success. This Stage deserves a discussion about how pilot users and settings are selected, what expectations of the users are, and how the implementation team will support them in their trailblazing use of the new or modified AI solution.
- 22) AMIA also suggests a discussion on criteria for determining whether a pilot was successful and ready to be deployed more broadly as in Stage 6; how this would be communicated, including lessons learned and changes made as a result of the pilot; and go/no-go factors regarding proceeding to Stage 6 deployment.
- 23) AMIA suggests CHAI leverage their interdisciplinary stakeholder membership to create a staged deployment model based on risk that could be emulated by regulators.

Stage 6: Deploy & Monitor

- 24) Under “Consider how the anticipated benefits, risks, and costs of the AI solution compare with the actual benefits, risks, and costs when used in the deployment environment,” on page 58, we suggest having a process for determining whether a variation depending on the setting in which the AI solution is deployed is due to differences in people, processes, or technology.

Thank you for your consideration of these comments and suggestions. If you would like to discuss further, please contact Reva Singh, AMIA’s Vice President of Public Policy, at rsingh@amia.org.

Sincerely,



Genevieve Melton-Meaux, MD, PhD, FACMI
AMIA Chair/President