Consumer
Technology
Association™

ANSI/CTA Standard

The Use of Artificial Intelligence in Health Care: Trustworthiness

Approved American National Standard
ANSI

ANSI/CTA-2090

February 2021

NOTICE

Consumer Technology Association (CTA)™ Standards, Bulletins and other technical publications are designed to serve the public interest through eliminating misunderstandings between manufacturers and purchasers, facilitating interchangeability and improvement of products, and assisting the purchaser in selecting and obtaining with minimum delay the proper product for his particular need. Existence of such Standards, Bulletins and other technical publications shall not in any respect preclude any member or nonmember of the Consumer Technology Association from manufacturing or selling products not conforming to such Standards, Bulletins or other technical publications, nor shall the existence of such Standards, Bulletins and other technical publications preclude their voluntary use by those other than Consumer Technology Association members, whether the standard is to be used either domestically or internationally.

Standards, Bulletins and other technical publications are adopted by the Consumer Technology Association in accordance with the American National Standards Institute (ANSI) patent policy. By such action, the Consumer Technology Association does not assume any liability to any patent owner, nor does it assume any obligation whatever to parties adopting the Standard, Bulletin or other technical publication.

This document does not purport to address all safety problems associated with its use or all applicable regulatory requirements. It is the responsibility of the user of this document to establish appropriate safety and health practices and to determine the applicability of regulatory limitations before its use.

(Formulated under the cognizance of the CTA **R13 Artificial Intelligence Committee**.)

Published by
©CONSUMER TECHNOLOGY ASSOCIATION 2019
Technology & Standards Department
www.cta.tech

FOREWORD

This standard was developed by the Consumer Technology Association's R13 Artificial Intelligence and R13 WG 1 Artificial Intelligence in Health Care.

(This page intentionally left blank.)

TABLE OF CONTENTS

(This page intentionally left blank.)

The Use of Artificial Intelligence in Health Care: Trustworthiness

1 SCOPE

Artificial Intelligence (AI) is quickly becoming a pervasive tool in the health care industry. This standard identifies the core requirements and baseline for AI solutions in health care to be deemed as trustworthy. Additionally, it explores the impact of the trustworthiness of AI in health care through the lens of the end user (e.g., physician, consumer, professional and family caregiver, public health, medical societies, and regulators) and will identify the unique challenges and opportunities for AI in the health care sector.

1.1 Background

The nature of AI can make people suspicious of product performance, specifically in health care applications. Many factors go into earning and sustaining trust in an AI health care product or application and these factors vary depending on the type of use and end user/stakeholder.

This standard identifies three major expressions of how trust is created and maintained: Human Trust, Technical Trust, and Regulatory Trust. Human Trust focuses on fostering humanistic factors that affect the creation and maintenance of trust between developer and user. In other words, a product that is difficult to use or to understand may affect the ability to create trust between individuals. Technical Trust focuses on the technical execution of the design and training of AI systems to deliver as expected. Finally, the principles of Human and Technical trust are embodied in law and regulations intended to prevent potential harm to the end user. The compliance and enforcement of the law and the regulations by institutions are required to foster Regulatory Trust. Systemic trust is a positive side effect of designing AI systems with an understanding of and attempt to balance the circular nature of these concepts.

It should be noted that the factors identified in this standard can affect all three categories. For example, building Regulatory Trust will be reliant on successful Technical Trust.

2 REFERENCES

2.1 Normative References

The following standards contain provisions that, through referenced in this text, constitute normative provisions of this standard. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this standard are encouraged to investigate the possibility of applying the most recent editions of the standards listed here.

2.2 Normative Reference List

1. ANSI/CTA-2089, *Definitions and Characteristics of Artificial Intelligence*, February 2020, https://cta.tech/standards
2. ANSI/CTA-2089.1, *Definitions and Characteristics of Artificial intelligence in Health Care*, February 2020, https://cta.tech/standards

2.3 Informative References

The following references contain provisions that, through referenced in this text, constitute informative provisions of this standard. At the time of publication, the edition indicated was valid. All standards are subject to revision, and parties to agreements based on this standard are encouraged to investigate the possibility of applying the most recent edition of the standard indicated below.

2.4 Informative Reference List

3. *Ethically Aligned Design, First Edition: A vision for Prioritizing Human Well-being with Autonomous and Intelligence Systems*, 2018, standards.ieee.org
4. Gilkson, Ella and Wooley, Anita. “Human trust in artificial intelligence: Review of empirical research.” *The Academy of Management Annals* (2020). https://journals.aom.org/doi/10.5465/annals.2018.0057.
5. J3016_201806, *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, June 2018, sae.org
6. *Health Insurance Portability and Accountability Act,* 1996, https://www.hhs.gov/hipaa/index.html
7. *Federal Trade Commission Act Section 5: Unfair or Deceptive Acts or Practices*, 2016, www.ftc.gov
8. *Guiding Principles for the Privacy of Personal Health and Wellness Information*, 2019, https://cta.tech
9. *User Experience Basics*, https://www.usability.gov/what-and-why/user-experience.html
10. Morville, Peter. “*User Experience Design”* (June 2004). https://semanticstudios.com/user_experience_design/
11. *Software as a Medical Device (SaMD): Clinical Evaluation*, September 2017, imdrf.org/documents
12. *AMA Privacy Principles,* May 2020, ama-assn.org
13. *Consumer Data Privacy Legislation,* 2019, https://www.ncsl.org/research/telecommunications-and-information-technology/consumer-data-privacy.aspx

14. *American Heart Association ECG Database*, https://www.ecri.org/american-heart-association-ecg-database-usb
15. *OpenfMRI public MRI image datasets*, https://openfmri.org/
16. *Software as a Medical Device: Possible Framework for Risk Categorization and Corresponding Considerations*, September 2014, www.imdrf.org.

3 HUMAN TRUST

This section addressing human trust looks specifically at topics related to the human interaction and perception of the AI solution. While these topics outlined below reside in the human trust section, it is important to note that they are not issues that impact just the human factors of AI trust, but can also be factors impacting technical and regulatory trust.

3.1 Human Interaction

Human interaction with AI, engendered by human trust, is integral to the very operation of AI and the interconnected role of security and privacy features, transparency, reliability and the immediacy of behaviors in developing that trust. One key aspect to consider is, the role of "anthropomorphism [regarding cognitive and] …..emotional trust"[Reference 3].

Trust in AI is important as trust has been shown to be a predictor of technology acceptance. "Review of the empirical research on cognitive trust in AI demonstrates that AI representation and the level of machine intelligence play an important role in the nature of the trust people develop." [Reference 3]. Due to the low visibility of embedded AI, [in particular], its impact on human emotions and emotion related trust is less clear, and as of this writing there is very little empirical research that addresses how tangibility, or immediacy impacts emotional trust in embedded AI." [Reference 3].

Therefore, these factors need to be considered and understood:

a. That the AI is appropriate for all intended users (includes disability testing (physical and cognitive) where relevant).
b. The extent to which humans verify the decisions made using AI is understood and communicated.
c. Safety mechanisms for AI features (including interface) are present
d. Additional instructions and information on what and how data are shared; readily accessible opt-in/out features are present.
e. The developers of an AI interface should take into account through measurement and testing the ability to produce an emotional response from users, either negative or positive. This should be considered in all phases of the AI product lifecycle.
f. Anthropometric features can influence trust both positively and negatively:

i. Cultural considerations.
ii. Manipulation for likeability.
iii. High expectations regarding product/technology introduction/ potential also for product/technology abandonment.
iv. Impact of self-created interface.
v. How the interface is framed/presented to the user (human likeness).
vi. How the AI interface/representation (robotic, virtual or embedded) such as physical robot, chatbot, or avatar impacts interaction with users.

3.1.1 Requirements

The model developer should:

- Make clear what the system can and cannot do. - Help the user understand what the AI system is capable of doing.
  - Example applications: [Activity Trackers, Product #1] "Displays all the metrics that it tracks and explains how. Metrics include movement metrics such as steps, distance traveled, length of time exercised, and all-day calorie burn, for a day. The application may not accurately measure the activity of individuals with mobility impairments."
- Make clear how well the system can do what it can do. Help the user understand how often the AI system may make mistakes.
- Time services based on context. Time when to act or interrupt based on the user's current task and environment.
  - Example applications & Example violations: [Activity Trackers, Product #2] "Context is very basic, it notifies me when I approach my goal; hit my goal; or exceed my goal. The timing of it is not clear, however. The timestamps are varied, too... It feels pretty arbitrary; my interpretation of the reasoning behind the notification can't be described by my activity or proximity to the goal."
- Show contextually relevant information. Display information relevant to the user's current task and environment.
  - Example applications: [Activity Tracker, Product #2] "The product knows I am about to wake up for the day and provides a morning reminder to take my medicine with breakfast."
- Match relevant social norms. Ensure the experience is delivered in a way that users would expect and accept, given their social and cultural context.
  - Example applications: [Activity Trackers, Product #1] "Provides a reminder to stand up without understanding my social context (e.g., in a meeting, having lunch with a friend). Does not consider the social context prior to sending notification for activity and does not use tone appropriately – just says "time to stand!" no matter what."

- Mitigate social biases: Ensure the AI system's language and behaviors do not reinforce undesirable and unfair stereotypes and biases.
  - [Activity tracker, Product #1] "Product assumes user is healthy and that he/she is able to take 10,000 steps per day."
- Support efficient invocation. Make it easy to invoke or request the AI system's services when needed.
- Support efficient dismissal. Make it easy to dismiss or ignore undesired AI system services.
- Scope services when in doubt. Engage in disambiguation or gracefully degrade the AI system's services when uncertain about a user's goals.
- Make clear why the system did what it did. Enable the user to access an explanation of why the AI system behaved as it did. If possible, link to authoritative sources or trustworthy guidelines used to inform the AI.
- Enable trained accessible and assistive technology support personnel to interact with device/software users who live with cognitive or physical disabilities; correct/update/reboot AI device/software remotely; provide additional guidance and options to device/software users for remote health and safety monitoring methods.

The model developer, shall:

- Support efficient correction. Make it easy to edit, refine, or recover when the AI system output is wrong. Consider if and how end users can remove or modify incorrect or inaccurate data.
  - Example applications: "If the product recognizes a senior that has fallen but the senior has not fallen. The health care provider can update the status and confirm the senior did not fall."
  - Example applications: "Use entered data based on the assumption weight being calculated in U.S. pounds but instead was being captured as kilograms."

3.2 Explainability

In addition to knowing what a product does and how well it does it, people are also interested in how something works. Certain AI technologies are straightforward to explain, such as decision trees that were created based on clinical practice guidelines. However, other technologies can be much more difficult to explain.

There are a variety of stakeholders, including regulators, health care professionals, public health agencies, and patients, among others. As a result, there may be a need for different strategies to fulfil the needs of this diverse group of stakeholders to understand how the technology works. What stakeholders care about may vary; some may require or expect more detailed explanation than others. Note that not all stakeholder types are defined and not every topic of interest is listed. These lists are intended to help the reader understand that there is a wide variety of stakeholders with different needs

Stakeholder categories and some examples of where AI could be helpful include:

- Regulators – including regulatory reviewers, auditors, etc.
- Health Care professionals – including physicians, nurses, etc.
- Health Care specialists – Health care professionals that have a deep knowledge of a specific field of medicine (e.g., radiologists, cardiologists).
- Health Care administrators managing supply needs [e.g., ventilators, dialysis machines, Extracorporeal membrane oxygenation (ECMOs)] all of which have been problematic to forecast during a pandemic or other public health emergencies.
- Lay users – including patients and caregivers confronting difficult personal health care decisions, etc.
- Public Health Agencies – AI can help predict population health issues and help guide vaccination campaigns.
- Insurance companies -- population health surveillance is an important topic and AI can help identify gaps in care.

3.2.1 Requirements

The model developer, shall:

- Provide a clear description of what is being predicted (intent) and what is the expected output of the model.
- To the extent known, provide an explanation of the key clinical parameters the model will look for in the data source.
- Explain any limitations of, the training data set especially those might restrict the applicability of the AI model to certain groups, regions, or time periods (see bias Section 4.1.1).
- Describe the processes, frequency, and controls to monitor performance of model after deployment including performance metrics used.
- Set parameters for product performance, including rates of failure and success and communicate what those parameters are to stakeholders.
- Set variation limits where if the product performance is outside those limits, appropriate stakeholders are notified.
    - For purposes of transparency, these variation limits should be communicated to the stakeholders.

This model developer, should:

- Provide a description of the AI model type (e.g., Neural Network, Decision Tree, Logistic Regression, Continuous Learning)
    - The model developer should discuss the relative advantages of the chosen method vs. alternatives (e.g. Recurrent neural net model

performance was demonstrated to be better than a Logistic Regression model.).

- Inform stakeholders, at key intervals, when the model has been updated and why.

3.3 User Experience

AI is defining a new paradigm for human and machine interactions and how we use interfaces is becoming increasingly important to conveying trust through a cohesive user experience. Historically, health information technology (health IT) has been developed to accommodate the requirements of various entities (e.g., regulators, payers, and insurers), sacrificing provider and patient needs as a result. Many in the medical community have experienced the use of health IT as a requirement rather than an opportunity. Implementation challenges along with the lack of incorporating user-centered design in application development have also negatively impacted clinical workflows. Hence, it is important to be aware of the history of health IT while developing new solutions. Developing AI for clinical decision support should start with the needs of the end user.

Communication from the AI system to the end user will need to flow as if two humans are exchanging relevant medical information through a conversation i.e., medical diagnoses, treatments, and check-ups. How medical Information that is exchanged between AI and patient will also need to be designed to “stick” to the patient in order to approximate the effective conveyance of that same information by a medical professional i.e., engaged listening, body language, and non-verbal signals.

A user may interact with an application in multiple ways. The application itself may run on a personal computer, a mobile device, a wearable device, and/or a speech unit, and the interface itself may be based around structured text (e.g., answer these 10 multiple choice questions) or it may be a free-form chatbot. This interaction may occur directly with software running on the medical device hardware (known as SiMD “Software in a Medical Device”) or it may be software running independently of any dedicated hardware (known as SaMD, “Software as a Medical Device”).

An interface that is difficult to use will frustrate the user. The quality and appropriateness of the interface will impact the user’s trust in the system.

- "User experience (UX) focuses on having a deep understanding of users, what they need, what they value, their abilities, and also their limitations."[Reference 10]
- In order for the information to have value, it must be: [Reference 10]
  - Useful: Your content should be original and fulfill a need
  - Usable: Site must be easy to use

- Desirable: Image, identity, brand, and other design elements are used to evoke emotion and appreciation
- Findable: Content needs to be navigable and locatable
- Accessible: Content needs to be accessible to people with disabilities
- Credible: Users must trust and believe what you tell them

UX defines the functionality of a product and UI relates to the design and end user interface with which the consumer or patient interacts. An application's "usability" must consider both UI design and the user's cognitive task support (e.g., workflow design, data visualization, and functionality).

The interface must be appropriate for its target audience. An application that interacts directly with a patient needs to consider their education level, health literacy, cognition (the ability to understand and remember what is being said), dexterity if manual entry is required to interact with the software, speaking ability if speech recognition is used for data entry, likelihood of misspelling data that they enter, etc.

The vocabulary of what is being said should also consider that English might not be the user's primary language – this can impact both the instructions given to the patient and regional accents in the patient's reply.

3.3.1 Requirements

The model developer and AI solution provider, shall:

- Design a user interface that is appropriate for the target audience. This includes consideration of the user's education level, cognitive abilities, language skills, mobility/dexterity in the use of the interface, visual impairment, etc.
- Design the interface to be fault tolerant. This can include:
  - Detection of user interface errors (e.g., spellcheck feature).
  - Functions that allow errors to be corrected.
  - Tolerance for extraneous signals (e.g., background noise in a speech-recognition interface).
- Make sure that a particular use case of AI is explained to ensure proper understanding of model and its strengths and weaknesses
  - State recommendations clearly to ensure proper understanding of the model and supporting evidence upon which they are based.
  - Specifically list the known limitations of data and the model as well as the scenarios where the model should or should not be used.
  - Explain the logic behind the model and recommendations.
  - Provide clear definitions (data dictionary) of features.
  - Explain model accuracy in clear and easy to understand language.
- Where possible, AI is integrated in existing workflows or is used to enhance existing workflows.

3.4 Levels of Autonomy

There is a wide range of applications where AI could help improve clinical care including providing information and clinical guidance for management of specific conditions, or even undertaking certain tasks previously performed exclusively by physicians such as interpretation of imaging studies. AI health care applications could also provide valuable patient decision support tools for patients and population health surveillance tools for public health agencies, payers, and regulatory agencies. However, physicians, patients, and other stakeholders must first trust an AI application to do its job, understand how to correctly use the application, and know its limitations. We will discuss in this section how the level of autonomy of an AI application is a key determinant of the level of trust that must be established before the application will achieve broad acceptance in health care.

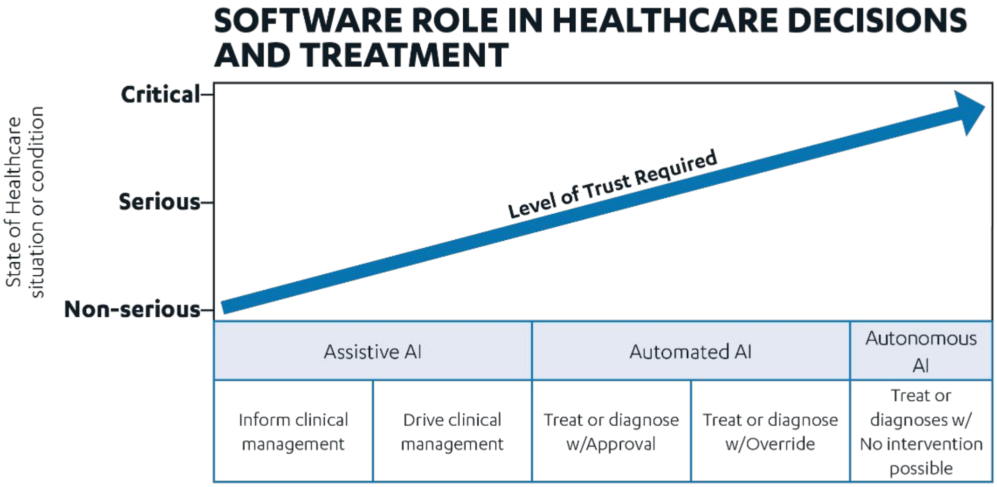
When considering the levels of AI autonomy in health care applications, it is worthwhile to review standards developed for other AI applications which provide a framework for classifying and understanding different levels of autonomy. For example, the Society of Automotive Engineers (SAE) J3016 *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles* has defined 6 levels of autonomy for driving automation.

In the SAE scheme, software provides a range of driving assistance – in Level 0, for example, features are limited to warnings and momentary assistance. This progresses from one level to the next, with software providing additional assistance (e.g., steering and/or brake support for the driver), reaching a level at which the vehicle can drive itself under limited conditions, and finally achieving a level at which the vehicle is fully autonomous in all situations.

It should be emphasized that AI systems will interact with humans even when their function is autonomous. An autonomous vehicle will still need to obtain input on where to go, and the various needs of the individual passenger. The interaction with the ML/AI entity is in fact a social interaction and AI developers need to address the behavioral influence of the interface design.

One difference between the SAE standard and AI standards in health care is that the SAE is a standard that applies to a very narrow set of use cases – driving a vehicle. While there are a variety of road conditions and traffic situations, the basic task of using a steering wheel, accelerator, and brakes to navigate is a much simpler use case than the totality of potential application of AI in health care. Apart from the many different intended uses, there are drastically different use environments where AI might be applied in health care (e.g., an AI application used in the ICU to improve care for critically ill patients versus an AI application that provides guidance to patients for disease prevention wherever the patient is located).

In 2014, the International Medical Device Regulators Forum (IMDRF) published a paper on a risk-based classification scheme for Software as a Medical Device applications. In considering how to adapt the IMDRF table from the paper to broaden its applicability to better address levels of AI autonomy in health care for applications used by physicians, we further break down the “Treat or diagnose” category in the IMDRF [Reference 16] table into three levels of autonomy and consider the levels of trust required for successful deployment by practicing physicians. As shown in the figure below, AI applications that provide information or recommendations as to treatment or diagnosis with the physician making the final decision are examples of Assistive AI. A higher level of autonomy is Automated AI where the application selects the treatment or diagnosis either subject to the physician’s approval or in a higher level of autonomy implements a treatment or establishes a diagnosis unless the physician overrides the application. Finally, completely autonomous AI applications function in disease treatment or diagnosis without any physician intervention required or possible.



SOFTWARE ROLE IN HEALTHCARE DECISIONS AND TREATMENT

Assistive AI		Automated AI		Autonomous AI
Inform clinical management	Drive clinical management	Treat or diagnose w/Approval	Treat or diagnose w/Override	Treat or diagnoses w/ No intervention possible

As suggested by this table, the level of trust that a health care provider must have in a health care AI application product before adopting its use is proportional to the level of application autonomy and to the severity of the clinical condition being treated. We consider that fully autonomous AI systems used in health care for treatment or diagnosis will require the high levels of trust regardless of the perceived severity of the clinical condition involved since interventions even for minor, non-critical situations can on occasion have life-threatening consequences over the longer term. As an example, instituting low dose aspirin as a preventive measure in an asymptomatic patient at risk of cardiovascular disease can on occasion result in life threatening gastrointestinal hemorrhage.

As the level of autonomy of health care AI solutions increases from assistive to fully autonomous systems, higher levels of trust will rely strongly on meeting higher expectations of reliability. This can be established through trials during validation of the ML/AI and monitoring the results during use once the product is released. The resulting reliability and associated risks can use statistical techniques to predict expected performance and risks of failure accurately. This statistical reliability can then be described in understandable ways to all people who interact with the ML/AI.

Humans are also highly influenced by appearance which can be very different from reliability and risk. This leads to one side effect of any technology, the risk of "overtrust," that is, individuals can get into the habit of approving whatever the software recommends without thinking critically about the situation. And it should not be assumed that health care providers are immune to this risk.

In view of this risk, it may be appropriate, in certain applications, to monitor health care providers for overtrust behaviors. If there is a high-risk application that has a variety of factors that affect performance, and if the provider always agrees with the recommendation of the application, a notification would be sent to the user reminding them of the limitations of the application or a requirement could be put in place to require re-training on the safe use of the application.

As the level of trust required increases in parallel with the level of risk associated with an AI application, so too does the level of "explainability" need to increase (see Section 3.2). This is especially true for AI applications that influence or even control health care decisions for individual patients where the health care provider continues to be responsible, and potentially liable, for adverse consequences. In some instances, where regulatory approval is required before deployment of a health care AI application, the fact that approval has been obtained may diminish, but not remove, the need for users of the technology to understand how it works.

3.4.1 Requirements

To support transparency the product documentation (e.g., operator's manual) shall be clear regarding the:

- Level of autonomy of the application. This includes a description of what functionality the application will automatically perform, what functionality is dependent on user-approval, and what functionality is provided for informational purposes only.
- Requirements for safe use of the application by the user regarding supervision of the technology, including specific safeguards to be employed when using fully autonomous AI technology or when using AI at any autonomy level when providing clinical care in clinical situations where the risk of harm is elevated.
- Controllability of AI functions. It is important that humans working with the AI function and who are responsible for results are able to interact with the ML/AI to get optimal results and reduce risks. This requires:
    - Explaining how to invoke or request the AI system's services when needed.
    - Explaining how to dismiss or ignore undesired AI system services.
    - Explaining how to edit, refine, or recover when the AI system is wrong.

4 TECHNICAL TRUST

This section addressing technical trust specifically considers topics related to data usage, including access and privacy. While the topics outlined below reside in the Technical Trust section, it is important to note that they do not just impact the technical application of AI trust; they can also impact Human and Regulatory Trust.

4.1 Data Quality & Integrity

The quality and integrity of data used to develop AI algorithms and train machine learning systems directly impacts the trustworthiness of the system when applied in practice. There is no shortage of data created by patients, providers, and the health care industry generally, but harnessing it is a complex endeavor for developers and end-users. Issues rooted in data provenance may replicate certain systemic biases or magnify errors which ultimately may result in patient safety and health equity issues. A focus on data quality, security, and access is essential to establishing and maintaining the trust necessary to further the growth and use of AI in health care.

4.1.1 Bias

For users to trust an AI solution, one essential pillar of a trusted AI solution is fairness. Fairness of an AI solution can typically be achieved with minimization of bias in the individual components and at every step of the process of building the AI solution. As data is the foundational building block of an AI solution, minimizing data bias is paramount for creating a trusted AI solution.

Data bias could be introduced into an AI system at least from the following processes:

- Collecting new data:
    - Without a good understanding of the intended use and use cases for an AI solution to be built, the data collected could have bias that will impact the model such as missing data for certain groups of the population. For instance, a Canadian company developed auditory tests for Alzheimer's disease based on a training data set containing speech samples collected from native English speakers only. As a result, the model interpreted pauses or different pronunciations as indicators of this disease. As a result, it falsely identified more non-native English speakers than native English speakers having early signs of this disease.
- Selecting data from existing data sets:
    - Selecting data to train a model without knowing the details/specifications of the existing data sets could result in the bias in the trained model. For example, data from a geriatric hospital might not be appropriate for use in a ML system for pediatric patients.
- Combining/joining data sets:

  - Combining/joining two or more data sets from different studies without understanding and mitigating the bias from each data set. For instances, there are multiple data sets of a specific object in the visual world. To form a large data set for this specific object, simply combining these data sets could introduce bias (e.g., if a data set-specific-model can perform well on the corresponding data set but not generalize well to other data sets).
- Using a data set improperly:
  - The way a data set is used to build a model could lead to bias. For example, the method of splitting data randomly into training and validation sets results in the data used for validation having the same bias as the data used for training.

4.1.1.1 Requirements

The model developer, owner of the AI solution, and domain expert shall

- Understand the motivation or intended use of the AI solution to be created.
- List the potential use cases for an AI solution to be built.
- Make the right decisions on dataset composition, such as, data instance types (e.g., text, images, and specific demographic group(s) or general population), number of each type, etc. so that the AI solution built based on the new data will generate fair results.
- Determine if the existing data set are "raw" data or pre-processed data.
  - For pre-processed data, find out what kind(s) of pre-processing has been performed so that the same preprocessing software/method can be applied to the input data during inference.
  - If there is need to capture additional new data, it is important to know how the existing data was collected (e.g., hardware/sensor, environment condition) so that the new data can be collected under similar conditions.
- When combining or joining multiple existing data sets:
  - Learn and/or model the bias for each data set.
  - Mitigate or undo the associated bias from each data set.
  - Find out the commonalities to all the data sets (e.g., through modeling) to achieve cross-dataset generalization.
- When splitting a collected data set into training, validation, and testing datasets, make sure each of them is randomly selected by applying certain techniques (e.g., data shuffling with a random number generator). By doing so, it can reduce the potential bias introduced in this process.

4.1.2 Data security:

For users to trust an AI solution, one essential pillar of a trusted AI solution is data security. Users want to understand that the data included in the AI solution is trusted and secure. Relevant considerations include:

- HIPAA details some requirements for the secure storage of data and notification in the event of a security breach, but many vendors may not be covered by HIPAA.
- Consumers may not understand what is covered by HIPAA and be unaware that the information they share is not protected by law. Developers should consider data security whenever information is collected and stored. Organizations should assess their risk of breach and take steps to minimize the risk of exposure.
- For vendors not covered by HIPAA, FTC and breach notification rules of some States apply.
- Data security involves safeguarding the training and validation datasets, model training process, and model deployment tools used by an organization. Furthermore, if the product is an adaptive system that continues to learn during operational use, it is important to have appropriate data provenance and systems monitoring in place. The lack of safeguards throughout the model training and deployment process model can lead to critical vulnerabilities, which if exploited, may decrease trust in software, the company implementing the software, and across all other uses within health care setting.

4.1.2.1 Requirements

The model developer shall comply with all relevant jurisdiction rules related to data security.

4.1.3 Privacy

Keeping personal health information private and protected is a core component of trust in the relationship between patients and clinicians. This makes privacy and data security critical for AI in the health care setting where personal health information is the raw material for most AI systems. Consumers expect that personal data will be protected and want an assurance that organizations will keep their information confidential.

Personal health data falls into two categories. The first category is data generated and collected by health plans, health providers, and clearinghouses, as well as each of these entities' business associates, which is directly protected by HIPAA. The second category is comprised of such health information which has been generated and collected by individuals, such as wellness data they have generated using fitness monitors and personal health applications. While this second category is not protected by HIPAA, protection is critical to trust in both cases [Reference 6].

By providing clear and specific information on how personal health information is used, shared, stored, and managed, organizations can promote the consumer trust necessary to encourage AI use.

- Non-HIPAA Protected Information:

  - Section 5 of the Federal Trade Commission (FTC) Act [Reference 7] prohibits misleading consumers through unfair or deceptive acts or practices. The FTC has determined that businesses without reasonable data practices or deceptive privacy statements may violate the FTC actions. This includes failing to disclose key information regarding an organizations' privacy policy or hiding or burying relevant information. The FTC has resources to help ensure protection of consumer privacy and security.
  - To promote privacy and security and engender patient trust, organizations gathering health data should limit the information they collect to what is necessary for development and use. If information is no longer needed for its originally stated purpose, it should be deleted.
  - The privacy landscape is changing rapidly, and organizations should consider privacy policies that grant consumers the ability to delete personal information held by the organization or opt-out of the use or sharing of their information.
    - Organizations such as the Consumer Technology Association (CTA) [Reference 8]and the American Medical Association [Reference 12] have developed voluntary privacy principles for health data not covered by HIPAA.
- HIPAA
  - Regulates the use and disclosure of protected health information (PHI) for the purpose of the administration and provision of clinical care.
    - Important to note that much of the consumer-generated health data (e.g., via consumer wearables) is not protected under HIPAA
  - Covered entities are permitted to share or sell health data, but protected health information must first be de-identified prior to sharing with a third party when data sharing is used for purposes other than providing clinical care.
    - HIPAA standardizes de-identification as the removal of a demographic information that could be used to identify an individual, such as names, telephone numbers, email address, and specific dates and geographic regions
  - Third parties may de-identify protected health information on behalf of a covered entity when acting as a business associate. Under this arrangement, business associates must meet the requirements of the HIPAA Privacy and Security Rule, which requires data protection policies and employee training.
- Additional Considerations:

  - Additionally, developers are encouraged to review and identify any relevant state laws that might impact privacy and data usage [Reference 13].
  - It should be noted that the General Data Protection Regulation (GDPR)[1] and those regulations should be considered for EU products.

4.1.3.1 Requirements

Organizations shall take steps to ensure that the personal health data used in development, testing, and commercial use of AI systems is kept private and secure. This includes:

- Be transparent about the personal health information they collect and why. The public disclosure of data sharing agreements is not required by law but provides additional transparency and can provide patients and the public with an understanding of how health data would be used.
- For HIPAA covered entities and their business associates, organizations are legally required to ensure the confidentiality of PHI they create, receive, maintain, or transmit. Organizations must take steps to prevent the unauthorized disclosure, alteration, or destruction of PHI. All organizations should take steps to ensure that personal health data is protected and stored in a secure manner.
- For data not covered by HIPAA, developing a privacy policy that is understandable to a user and includes the reason the data is being collected or used. These disclosures must be displayed clearly and conspicuously.
  - Specifically, individuals have the right to know whether their data will be used to develop and/or train machines or algorithms. The opportunity to participate in data collection for these purposes must be on an opt-in basis.
- The de-identification of data promotes patient privacy by removing demographic and other information that could tie data to an individual. Before using data in AI development, organizations should utilize a de-identification process (using techniques that are demonstrably robust, scalable, transparent, and provable) to minimize the risk of patient re-identification through personal health information and make such processes and techniques publicly available.

4.1.4 Source

Users want to be reassured that the source data that was used to train the system was good data and is relevant to their situation. There are many sources of health care data that can be used to train a system. Information could be sourced from Electronic Health Records, hospital billing data, published literature, patient registries, etc. There is considerable variation in health care and in the resulting health care data, and users will want to know that the impact of this variation is minimized for their application.

One way to earn users trust is for developers to only use data that is complete, correct, and relevant to the situation in which it is going to be used. For example, a system intended to diagnose breathing problems that was trained using an elderly population may not be appropriate for pediatric patients.

Factors that earn users trust include:

1. Is the data from a reliable source? Is information from the patient or is this from a health care professional?
2. How was the data collected? How large a sample was used? How many independent sources were used (e.g., was this from a single hospital or a network of hospitals?)
3. Is the training set representative of the target situation? Have the specific characteristics been confirmed, such as the population from which the training data was derived match the target or proposed use?
4. When was the data collected? Has clinical practice changed over time? Has the system been re-trained (or otherwise updated) since its initial training?

4.1.4.1 Requirements

Data sources used for development of AI shall be documented and described in sufficient scope and detail to facilitate the evaluation for trustworthiness. The following attributes of the data source need to be addressed:

- Data acquisition:
  - Define the technology and methodology for acquiring source data. For example, are the source data derived from a medical device or other instrumentation? Was it derived in accordance with approved device use instructions or were unique methods applied? Was the data acquisition conducted by qualified persons, lay persons or directly by the human subject?
- Data construct:
  - Describe the construct and attributes of the data set used for AI development. How does this dataset compare to the typical data presented to a health care user for the same medical purpose (e.g., diagnosis or treatment decisions)?

        - Is the data unstructured, but in natural language form that may be reused for other purposes?[2]
- Data merging:
    - If the dataset was derived from multiple original sources, describe the methods used to merge the data. If data source attributes differed, describe what was different and what techniques were used to homogenize the dataset? What risks associated with this approach were considered? How were risks mitigated and algorithm results validated?
        - Interoperability of data -- Describe the data in terms of its adherence to standards and ability to be used for broader purposes. For example, is the data derived from standardized physiological measurement instruments and in a standard form [Reference 14 and 15]?
- Post-acquisition processing:
    - Describe any methods used to further process the dataset prior to use in AI development. Is the dataset modified in any way from its original content or form? For example, does the dataset contain simulated records based on original data for the purpose of increasing the dataset for machine learning purposes? Describe any methods applied to combine or merge data from multiple sources or origins.
- Data quality and integrity:
    - Describe the methods used to verify and validate the accuracy and completeness of the dataset, as defined by the data acquisition and curation plan. Describe any exceptions taken.
- Training data vs. test data:
    - Describe the origins and relative segregation between these data sets, as it relates to the bullets above.

4.1.5 Access to Data

For data used for development of AI, describe the accessibility to the data in sufficient scope and detail to facilitate the evaluation for trustworthiness.

The model developer shall:

- Describe the data in terms of access to others for purposes of validation or development. Is the dataset publicly available? If so, describe the process for obtaining it.

[2] Lionbridge directory of general life sciences, healthcare and medical datasets (https://lionbridge.ai/datasets/18-free-life-sciences-medical-datasets-for-machine-learning/)

- Describe any independent third party that has assessed the data for integrity and trustworthiness (e.g., a test authority or regulatory agency). Is the data proprietary? Is the assessment report publicly available?

5 REGULATORY TRUST

This section addressing topics of interest to regulators. While these topics reside in the regulatory trust section, it is important to note that they do not just impact the regulatory AI trust. These factors can also impact human and technical trust.

5.1 Regulatory Impacts

AI innovation and integration may be challenging because of the highly regulated nature of the health care industry. Accordingly, the health care and life science industries should pay particular attention to governmental pronouncements on policy related to AI. The public expectation is that a regulatory system exists that ensures safe, effective, and reliable AI products in the clinical setting, as well as development and deployment that is explainable, protects privacy, and promotes responsible use of data. The institutions and agencies at the state and federal level with oversight authority over various aspects of health care cannot be treated as afterthoughts and the applicable rules, regulations, and guidance, should be considered at every stage of the machine learning pipeline to further public trust.

Additionally, the development and lifecycle management of AI-based medical device products should comply with the approaches described in internationally recognized standards such as IEC 62304 (software lifecycle processes), ISO 14971 (risk management), and IEC 62366 (usability). While some international health authorities require compliance to such standards in order to commercialize a medical device product, adherence also promotes technical, human, and regulatory trust.

For example, medical device regulatory frameworks take into account both premarket development and post market performance of medical device products. In the context of AI-based medical devices, change management is particularly important, especially for adaptive artificial intelligence and machine learning technologies. Regulators have and continue to evolve approaches for reviewing and clearing post-launch modifications to software as medical device products, placing an emphasis on the significance of the risk associated with the change.

Additional resources are available in Annex A.

ANNEX A: REGULATORY RESOURCES

(informative)

A.1 Data Privacy Laws

The development of AI requires voluminous amounts of health information to properly discern the language of medicine, and to find meaning and structure and meaning in electronic health record (EHR) and information provided directly by a patient. Compliance with federal and state privacy laws addressing health care information is a key factor in establishing and maintaining the systemic trust necessary for continued growth of artificial intelligence in health care. State laws could also be more restrictive than the federal laws and both sets of laws should be reviewed during the development and use of health care AI.

A.2 FDA

The FDA regulates drugs and medical devices through powers granted to it in the Food, Drug and Cosmetic Act. The focus of this regulation is the safety and efficacy of drugs and medical devices to promote safe delivery of medicine and to prevent use of harmful medical devices. The FDA monitors and regulates medical devices only in the design and development, and its enforcement authority does not extend to uses that would be considered the practice of medicine.

AI has been classified both as a device and as a product. The current FDA approach on regulation is focused on accuracy and relevancy of the data inputs and model outputs, the marketing of AI systems and the transparency of AI performance.

A.3 FTC

In efforts to prevent fraud, deception and unfair business practice, the Federal Trade Commission exercises enforcement, and administrative responsibilities under more than 70 laws. The FTC indicates that the application of AI in health care has the potential for unfair or discriminatory outcomes and the perpetuation of existing socioeconomic disparities. FTC may review the use of artificial intelligence on grounds of false or misleading health claims, representations of a software's performance, and claims affecting consumer privacy and data security. The FTC's Health Breach Notification Rule also requires businesses to provide notifications to their consumers after a breach of personal health record information

A.4 State Medical Boards

State medical boards exercise authority in furtherance of public protection by setting the professional standard of care and ethical standards of professional conduct within that state and ensuring compliance by disciplining physicians whose actions or manner of delivering care fall outside the scope of those standards. In general, state medical boards have discretion to act on the “practice of medicine”, general defined as “diagnose, cure, advise, or prescribe for any human disease, ailment, injury, infirmity, deformity, pain or other condition, physical or mental, real or imaginary, by any means or instrumentality.”

One initial challenge that state regulators will face the delineation between a clinical decision support tool and a tool which, under current state law definitions would be engaging in the practice of medicine. If an algorithm is found to be practicing medicine, and does so to the detriment of public safety, state regulators could exert oversight and act to enforce to ensure a proper standard of care for its use in a clinical setting.

Sources of general guidance on the current standards of care may be obtains from the Federation of State Medical Boards, which issues nonbinding policy statements for medical boards, or professional societies such as the American Medical Association and American Osteopathic Association, as well as specialty boards that issued guidance to practitioners on how to integrate emerging technologies into the delivery sector of health care. Alignment with the principles of this guidance helps maintain public trust in the development and use of AI in health care.

A.5 Tort Law

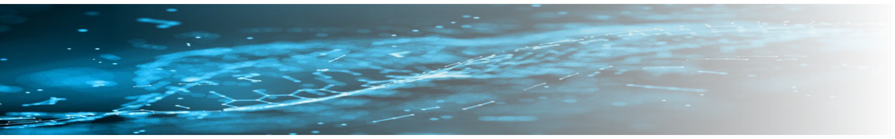
State tort laws may help develop regulations and practices in the scope of health care AI. A patient may seek damages from a provider when the behavior of any actors along the chain falls below the standard of care. Transparent efforts in the development of AI may create additional considerations in this area, as explainable algorithms will make assessing liability easier, but also create incentives for developers to avoid being transparent.

A.6 NIST

Although the National Institute of Standards and Technology does not regulate products, it does coordinate with various government agencies and supports the development of standards, including AI. NIST has held multiple workshops about AI, and is mentioned in this paper as a possible resource for the development of AI in health care. https://www.nist.gov/topics/artificial-intelligence

Table 1: Sources of Guidance and Standards

Area of Use	Examples	Data Privacy Laws	FTC	FDA	State Medical Boards	Tort Law
Assistive Intelligence in Clinical Setting/Clinical Decision Support System	AI used to assist physicians by giving treatment, diagnosis or screening device, but ultimately relies on physician interpretation to direct patient care		X	X	X	X
Autonomous AI	Autonomous AI that provides treatment, diagnoses, or screens without physician interpretation.		X	X	X	X
Health care Administration and Operational Efficiency	AI used to enhance clinical operations	X	X			
Patient Decision Support System	AI used directly by consumers to manage health conditions	X	X	X	X	
Research	Use of human subject data to develop, train, refine AI	X	X			



Consumer Technology Association Document Improvement Proposal

If in the review or use of this document a potential change is made evident for safety, health or technical reasons, please email your reason/rationale for the recommended change to standards@CTA.tech.

Consumer Technology Association
Technology & Standards Department
1919 S Eads Street, Arlington, VA 22202
FAX: (703) 907-7693 standards@CTA.tech

Consumer
Technology
Association™