Stanford MEDICINE

Stanford | Center for Health Education

EVALUATIONS OF AI APPLICATIONS IN HEALTHCARE STUDY GUIDE

MODULE 5: THE REGULATORY ENVIRONMENT FOR AI IN HEALTHCARE

LEARNING OBJECTIVES

- Understand why most AI solutions in healthcare have not received regulatory approval, to date
- Describe best practices for AI development, in particular good machine learning practices
- Recognize the risk framework used to classify AI solutions in healthcare that is used by the Food and Drug Administration (FDA)
- Know the 3 concepts of model properties that can be regulated
- Understand main differences between EU, China and US regulations on AI solutions

OVERVIEW

International Medical Device Regulators Forum (IMDRF): An international workgroup composed of AI regulators that come together to develop a path for standardized AI regulations.

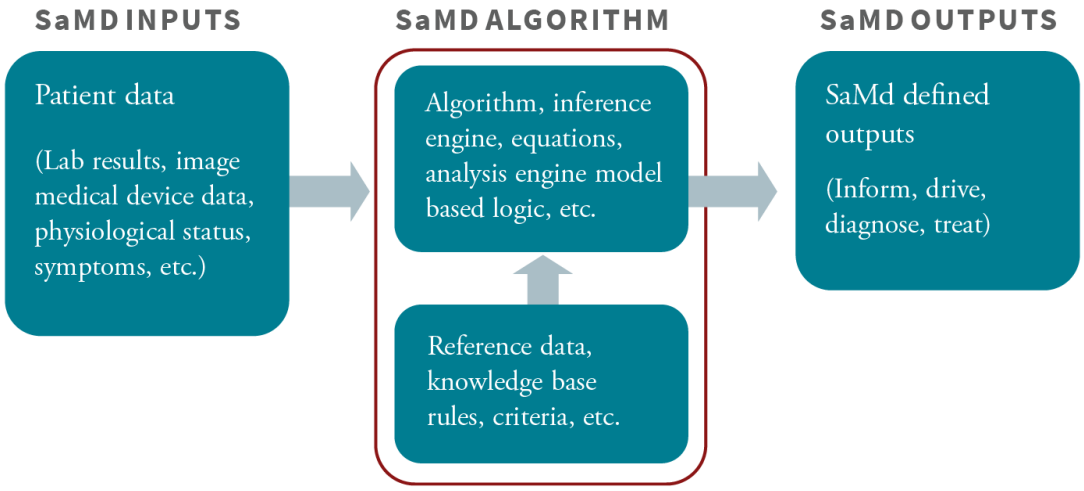
Software as a Medical Device (SaMD): Software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device.

- Medical purpose: Intended to treat, diagnose, cure, mitigate, or prevent disease or other conditions
- SaMDs are NOT part of hardware

COMPONENTS OF REGULATION

The SaMD can be described in 3 components: SaMD inputs, SaMD algorithms, and SaMD outputs.

Copyright 2020 © Stanford University



This is the framework used by the US Food and Drug Administration to regulate AI Solutions / SaMD.

The FDA's regulatory framework starts with a Market Application, which includes a definition statement and category (I, II, III, or IV). The category is based on the risk associated with the use of the proposed AI solution.

Depending on the category defined, data requirements necessary for regulation may include:

- A premarket notification (or 510(k) statement of equivalency)
- De Novo request (no similar product exists on the market for comparison)
- Premarket Application (PMA), which is reserved for high risk AI applications

Definition Statement is required for every SaMD application and is used to identify the submission category, which defines risk and subsequent data requirements.

The definition statement must:

- Clearly identify the intended medical purpose of the model (treat, diagnose, drive clinical management, inform clinical management)
- State the healthcare situation or condition that the AI model is intended for, which includes critical, serious, and non-serious conditions
- Include the intended population for the application
- Identify the intended users (or stakeholders) of the model

Using information from the definition statement, the SaMD Category is defined, which is based on a risk framework developed by the IMDRF.

Copyright 2020 © Stanford University

STATE OF HEALTHCARE SITUATION OR CONDITION	SIGNIFICANCE OF INFORMATION PROVIDED BY SaMD TO THE HEALTHCARE DECISION		
	Treat or Diagnose	Drive Clinical Management	Inform Clinical Management
Critical	IV	III	II
Serious	III	II	I
Non-Serious	II	I	I

SaMD N12[2] Framework

In the framework, risk is set by the intended use of the SaMD and the state of the health situation it targets.

- The columns in the grid represent the **Significance of information provided for a healthcare decision**. This is the **ACTION** in the **outcome-action pairing** framework used for AI evaluation.
- The rows represent the **State of healthcare situation or condition**, which identifies the state of the healthcare situation or condition as critical, serious, or non-serious

An example that demonstrates the definition statement and risk category:
A SaMD AI application that "**drives clinical management**" in a "**critical healthcare situation or condition**" (Category 3):

- In a hypothetical situation, an AI application is developed for ICU patients that receives electrocardiogram, blood pressure, and temperature signals from a patient monitoring system. The physiologic signals are processed and analyzed to detect patterns that occur at the onset of patient instability and deterioration, a threshold set by a utility analysis. When physiologic instability is detected, an alarm is generated to indicate that immediate clinical action is needed to prevent potential harm to the patient.

SaMD regulations place devices into four categories based on the risks associated with the use of the device. Category I being the lowest risk; Category IV being the highest risk.

- Category I devices require general controls
- Category II devices are medium to moderate risk and require the same general controls and some additional controls
- Category III and IV applications require "general controls" and pre-market approval which is the most stringent regulatory category. To date, there are very few Category IV SaMDs that

Copyright 2020 © Stanford University

are approved in the US. These are high risk, generally life-supporting, life-sustaining AI applications.

All applications must include general controls. General controls require that all AI solutions comply with three components:

1. Quality system regulations
2. Current good manufacturing practices
3. Properly labeled - the label of the device follows FDA guidelines and regulations

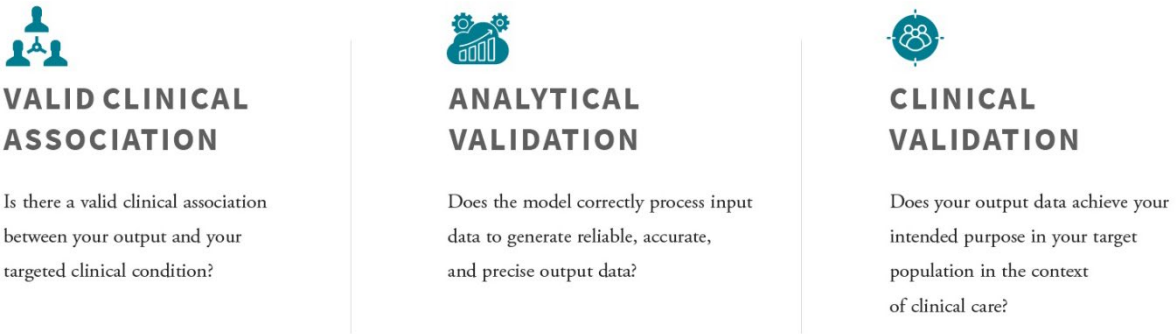
Any adverse event associated with the AI solution must be reported.

Quality System Regulations: Manufacturers are required to have processes in place for controlled bug resolution, incident reporting, standardized design processes and overall risk management.

CLINICAL EVALUATION PROCESS

All AI applications needing regulatory approval must include general controls. As part of the **general control**, **the Clinical Evaluation Process** is a framework used by regulators to understand quality system regulations. The IMDRF defines the clinical evaluation process as **ongoing activities** conducted for the assessment and analysis of a SaMD's clinical safety, effectiveness and performance.

The Clinical Evaluation Process includes three components.



Valid Clinical Association (Category I): Is there a valid clinical association between your SaMD output and your SaMD's targeted clinical condition? Scientific validity of the AI solution or the extent to which the SaMD's output is clinically accepted or well-founded (based on an established scientific evidence), and accurately corresponds to the healthcare situation and condition identified

Copyright 2020 © Stanford University

in a real-world setting. An indicator of the level of clinical acceptance and confidence one can have of the SaMD's output

Minimum evidence to support the clinical association could include:

- Literature searches
- Original clinical research
- Professional society guidelines
- Examples of how your model can generate new evidence
- Secondary data analysis
- The performance of a clinical trial based on your AI solution
- Required for AI regulation and ensures the clinical acceptance and uptake of the AI solution in the healthcare setting

Novel associations - associations that are newly discovered by your AI:

- As literature and existing randomized clinical trials do not exist for this association, there are other solutions to regulate this software, which may include performing a clinical trial or secondary data analyses

Analytical Validation (Category II): Evaluates whether your AI solution correctly processes input data to generate accurate, reliable, and precise output data. Part of the verification and validation phase that should be performed by the AI manufacturer. Provides objective evidence that the AI solution was correctly constructed and the data processing is reliable.

May come as part of your good software engineering practices or from generating new evidence through use of curated databases or previously collected patient data

Clinical Validation (Category III): Does the use of your AI's output data achieve your intended purpose in your target population in the context of clinical care? Related to the positive impact of an AI Solution on the health of an individual or population.

Clinical Validation should be:

- Measurable, patient-relevant clinical outcome(s)
- Including outcome(s) related to the function of the model

Copyright 2020 © Stanford University

- A positive impact on an individual or public health

Prior to product launch of the AI product (pre-market), evidence must exist on the following:

- AI accuracy
- Specificity
- Sensitivity
- Reliability
- Usability
- Limitations
- Scope of use in the intended use environment with the intended user

After launching the product (post-market) the product must:

- Continue to collect real world performance data to
- Further understand the healthcare needs to ensure the AI solution is meeting those needs
- Monitor the product's continued safety, effectiveness and performance in real-world use

The IMDRF identifies that clinical validation is a necessary component of regulation and that it can be demonstrated through several paths, including:

- Referencing existing data from studies conducted for the same intended use
- Referencing existing data from studies conducted for a different intended use, where extrapolation of such data can be justified
- Generating new clinical data for a specific intended use

The Clinical Evaluation Process is an important component of the regulatory environment.

FDA APPLICATION

In addition to the **general control process**, there are other regulatory control requirements (data requirements) that accompany an SaMD application that depend on the application's **category of risk (1 - 4)**, which can include one of these regulatory components:

- Premarket Notification 510(k)
- De Novo Notification
- Premarket Approval (PMA)

Copyright 2020 © Stanford University

Premarket Notification 510(k) is the simplest data requirement, if the SaMD is similar to a product already on the market. The intent is to inform the regulatory agencies that the AI solution is safe and effective, which is determined by demonstrating the AI solution is equivalent to a legally marketed device, often known as a "predicate".

To determine equivalence, the AI solution must have:

- The same intended use as the predicate AND have the same technological characteristics, OR
- The same intended use as the predicate and a different technology that will not raise safety or efficacy questions AND the AI solution is at least as safe and effective as the predicate

As more and more applications become approved, pre-market notifications will become an easier and efficient pathway towards regulation.

De Novo Notifications: Submitted when there is no "predicate". Limited to Category I and Category II SaMDs and are a risk-based classification process.

The de Novo notification should include:

- Clinical data (if applicable) that are relevant to support the assurance of the safety and effectiveness of the AI solution
- Non-clinical data including bench performance testing
- Description of the probable benefits of the AI solution when compared to the probable or anticipated risks when it is used as intended

Premarket Approval (PMA): Required for high risk SaMDs (Category III and IV). The most stringently regulated application required by the FDA. Includes rigorous technical studies, non-clinical laboratory studies, laboratory studies, and clinical investigations.

Before PMA approval, the applicant must provide valid scientific evidence demonstrating reasonable assurances of safety and effectiveness for the device's intended use.

Copyright 2020 © Stanford University

After an AI solution receives regulatory approval, a modification may be required in certain circumstances. A modification request must be submitted if there is new risk or a change of an existing risk.

A modification may be required if there is:

- A change to risk controls to prevent significant harm
- A change that significantly affects clinical functionality or performance. A change in clinical functionality could include
    - New indications for use
    - New clinical effects
    - Significant technology modifications that affect performance characteristics

SaMD modifications generally fall into three categories:

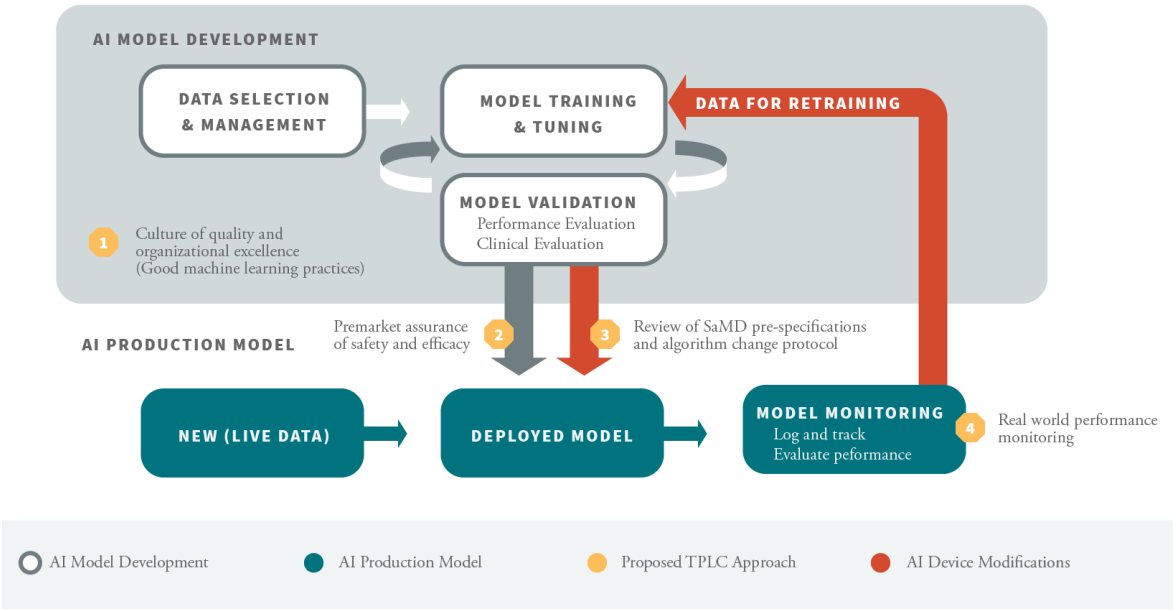
1. Change in performance
    - Example: The incorporation of new training data or a change in AI architecture which could alter performance
2. Change to the model Input
    - Example: The incorporation of different sources of the same input or adding new inputs that were not previously considered
3. Change of the intended use of the output
    - Example: Change in disease (apply to new condition)

Software modifications are common and essential in the total life cycle of the AI Solution.

PRODUCT APPROVAL

In line with the framework proposed by IMDRF, the FDA has developed the following diagram related to the **total product life-cycle (TPLC)** for an AI workflow.

Copyright 2020 © Stanford University



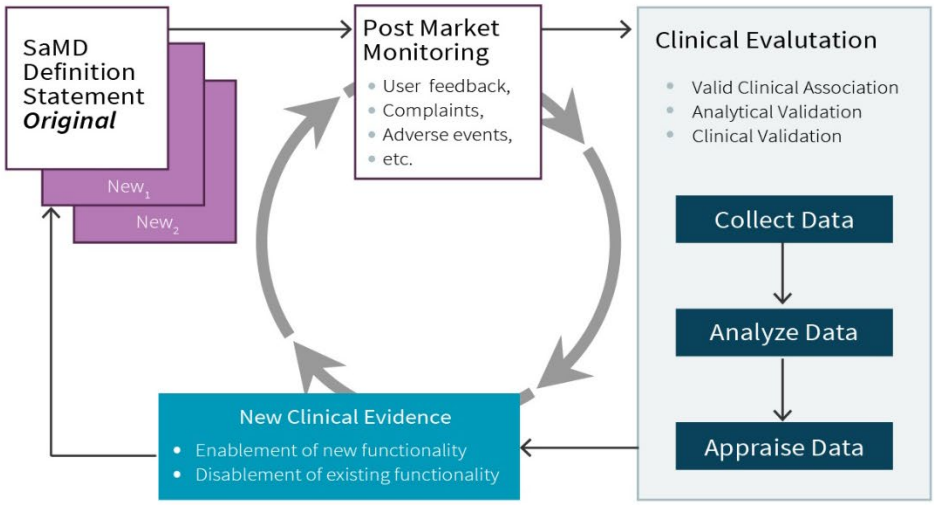
There are 4 distinct components in the total product life-cycle:

1. The culture of quality and organizational excellence, which is also referred to as Good Machine Learning Practices. This includes all components that must be considered when developing an AI solution, such as
    - Data selection and management
    - Model training and tuning
    - Model validation
2. Premarket assurance of safety and efficacy of the AI solution. It is expected that safety and effectiveness are continually monitored throughout the life cycle, including patient risks and patient safety. Regulators expect a manufacturer to perform a risk assessment and evaluate that the risks are reasonably mitigated throughout the TPLC.
3. Regular monitoring of safety and intended use, which is used to identify when a software modification is required. The regular monitoring of the deployed model is necessary and should include the ability to log and track model performance.
4. The Continuous Learning expected from an SaDM that Leverages Real World Data. “Continuous learning” is not “machine learning.” It refers to collecting post-market information to update and evaluate your existing AI solution.

Copyright 2020 © Stanford University

The TPLC diagram demonstrates how regulators are thinking about using real world data for continuous learning, the basis of the learning healthcare system.

Locked algorithms are AI solutions that provide the same results each time the same input is provided. Generally fixed functions to a given set of inputs, and may use a manual process for updates and validation.



Pathway for Continuous Learning – Use of real World SaMD Performance Data in Ongoing SaMD Clinical Evaluation

How to regulate "adaptive" AI solutions or machine learning applications?

- Following deployment, these types of adaptive models may provide a different output in comparison to the output initially approved for a given set of inputs. These automated changes use a well-defined process, which aim to improve outcomes based on new data or additional data that is taken as an input. There are two stages to an adaptive algorithm:
    - Learning stage: The algorithm "learns" how to change its behavior, based on the addition of new input types or new cases to an already existing training set
    - Update stage: The algorithm will update when the new version of the algorithm is deployed

This is a paradigm shift in the regulatory process and requires a new total product lifecycle that allows these devices to continually improve while providing effective safeguards.

Considerations for the regulatory environment for an AI model:

Copyright 2020 © Stanford University

1. Developers should be aware of SaMD Risk Classifications, total product life-cycle (TPLC) and Good Machine Learning Practices (GMLP)
2. The population, performance and intended use are aspects of the model that can be regulated and any change to these characteristics after approval would require a notice of modification
3. The intended use and state of the healthcare situation of the AI model drives the level of regulation and risk category
4. Safety and effectiveness must be continuously monitored post-market using real-world data to ensure a learning health system

Example - Arterys

- Indications for Use: Arterys is a software that uses cardiovascular images acquired from magnetic resonance (MR) scanners. It analyzes blood flow from the heart using the MR images. The output is intended to be used to support cardiologists, radiologists, and other healthcare professionals for clinical decision making.
- Risk Classification: Class II
    - Significance of information is to "inform clinical management"
    - State of healthcare situation or condition is "critical"

Example - IDx-DR

- Indications for Use: IDx-DR is a retinal diagnostic software device that incorporates an adaptive algorithm to evaluate ophthalmic images for diagnostic screening to identify retinal diseases or conditions
- Risk Classification: Class II
    - Significance of information is to "drive clinical management"
    - State of healthcare situation or condition is "serious"

Example - Guardian Connect (Medtronic)

- Indications for Use: The Guardian Connect system is indicated for continuous or periodic monitoring of glucose levels in the fluid under the skin, in patients (14 to 75 years of age) diagnosed with diabetes.
- Risk Classification: Class II
    - Significance of information is to "drive clinical management
    - State of healthcare situation or condition is "serious"

Copyright 2020 © Stanford University

The FDA does not regulate certain types of clinical decision support (CDS) tools under 21st century cures act.

Three criteria determine whether CDS are regulated as an SaMD:

1. The software cannot receive, analyze or otherwise process a medical image or signal from an in vitro diagnostic device or from any other signal acquisition system
2. A healthcare professional must be able to understand the basis of its recommendations
3. The software cannot be intended as the sole source of recommendations regarding treatment, diagnosis or prevention of a disease

The FDA doesn't regulate AI solutions that are "laboratory-developed tests" designed, developed and deployed within a single health care setting.

FDA's Digital Health Software Precertification Program (Pre-Cert):

- Streamlines regulation of AI solutions
- Organizations may become "approved" which would allow them to bring products to market without a premarket review, provided the product is "lower risk"
- Once certified, organizations can make certain minor changes to its AI products without having to submit a modification request

An organization must demonstrate the FDA's five quality and organizational excellence principles in order to be considered for the Pre-Cert program:

1. Product quality
2. Patient safety
3. Clinical responsibility
4. Cybersecurity responsibility
5. Proactive culture

GLOBAL ENVIRONMENT

It is important to note that there are some differences across the globe regarding regulatory guidelines for AI in healthcare.

EU's General Data Protection and Regulation (GDPR)

- Outlines a comprehensive set of regulations for the collection, storage, and use of personal information which may be used in AI solutions

Copyright 2020 © Stanford University

- Describes the right of citizens to receive an explanation for algorithmic decisions. The implications would exclude the use of many types of algorithms used today in advertising and social networks - and eventually healthcare.
- Provides protection of data from EU citizens but given the global Al momentum, the laws may impact companies from the US and worldwide
- A critical component of this regulation is Article 22: "Automated individual decision making, including profiling." Article 22 requires explicit and informed consent before any collection of personal data.
- The "right to explanation"
    - Requires that meaningful information about the AI logic as well as the potential significance and consequences of the data-driven system are provided upon request
    - Refers to the use of the black box algorithms
    - Could potentially limit the types of models that manufacturers are able to use in health-related applications
    - Holds the manufacturers of AI-based technologies more accountable

While there are some differences between the US and EU regulations, a common theme from both entities is the protection of the individual, their data, and their right to information.

China leads in the number of AI patents as a result of this favorable environment. AI governance in China is aimed at the development of "responsible AI" and focused on the societal beneficiary rather than the individual beneficiary. China has also put a focus on regulating education so that the nation produces more STEM workers.

China requires businesses and private citizens to share their data with the government – almost the opposite of US and EU regulations. The incentives for data sharing and elimination of data silos could catapult China in clinically meaningful AI technologies.

Principles for AI governance released by China's Ministry of Science and Technology (MOST) include:

1. Harmony and friendliness
2. Fairness and justice
3. Inclusivity and sharing
4. Respect privacy
5. Secure/safe and controllable
6. Shared responsibility

Copyright 2020 © Stanford University

7. Open collaboration
8. Agile governance

China is rapidly developing important AI solutions in healthcare and these governing policies are essential to balance both the innovation of technology as well as the safety of healthcare delivery.

The White House Office of Management and Budget (OMB) document provides Guidance for Regulation of AI Applications. The regulations are not specific to healthcare, but they provide the umbrella of regulations applied to AI solutions.

The ten guiding principles are aligned with the FDA's regulatory processes:

1. There must be public trust in AI
    - Public trust and validation is critical to the adoption and acceptance of these technologies in the healthcare sector
    - Privacy and other risks must be addressed with appropriate mitigation strategies that are well documented and transparently reported
2. There must be public participation in AI development
    - There should be ample opportunities for the public to provide information and participate at all stages possible of the "rule-making" process
3. AI solutions must be based on scientific integrity - clinical validation in the clinical evaluation process proposed by the IMDRF
    - AI in healthcare should be based on scientific and technical information and processes that are likely to have a clear and substantial influence on public policy or private sector decisions - and these standards should be held to the highest level of quality, transparency, and compliance
    - Best practices include clearly stating the strengths, weaknesses, intended optimizations or outcomes, bias mitigation, and appropriate uses of the AI application's results
    - Data used to train the AI system must be of sufficient quality for the intended use
4. Every AI solution must have a Risk Assessment and Management component
    - A risk-based approach should be used to determine which risks are acceptable and which risks present the possibility of unacceptable harm, or harm that has expected costs greater than expected benefits
    - If an AI tool fails, the magnitude and nature of the consequences will inform the level and type of regulatory effort that is appropriate to identify and mitigate risks
5. Benefits and Cost

Copyright 2020 © Stanford University

- For an AI algorithm to be deployed, it must offer significant potential benefit
- Before implementing an AI solution, agencies must consider the full societal costs, benefits, and effects related to the development and deployment of these applications

6. AI solutions must be flexible
    - Performance based and flexible approaches should be easily adapted and updated - this would include the continuous monitoring of real world evidence to improve upon the AI solution
7. AI solutions must be fair and non-discriminatory
    - Transparency regarding potential biases and discriminatory aspects of the algorithm is becoming more and more important as we see potential harm due to biases promoted or exacerbated through AI
8. AI solutions that provide Disclosure and Transparency will promote public trust
    - Healthcare systems should disclose when AI solutions are in use and how these applications can impact patients and decisions
9. All AI solutions must address Safety and Security issues
    - Safety and security should be considered throughout the design, development, deployment, and operation process
    - There should be controls in place to ensure confidentiality, integrity, and availability of the information processed, stored, and transmitted by AI systems
10. AI solutions must include multidisciplinary stakeholder involvement or Interagency Coordination
    - All sectors affected by the AI solution should coordinate and share experiences and challenges of AI solutions

Safety, transparency and multidisciplinary aspects are key ingredients to a successful and well-thought out AI solution.

CITATIONS AND ADDITIONAL READINGS

FDA, U. 2019. “Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD).” FDA.

US Department of Health and Human Services. Software as a medical device (SAMD): Clinical evaluation. Guidance for industry and Food and Drug Administration Staff. 2017.

Copyright 2020 © Stanford University