Navigating the Landscape of Global AI Regulations and Standards

Xavier AI Summit
November 12, 2021

Pat Baird, Philips pat.baird@philips.com

Inspiring collaboration. Leading innovation. Making a difference.

www.xavierhealth.org

XAVIER HEALTH

AI – Current State

XAVIER HEALTH

"Artificial intelligence constitutes a host of computational methods that produce systems that perform tasks normally requiring human intelligence. These computational methods include, but are not limited to, machine image recognition, natural language processing, and machine learning. However, in health care a more appropriate term is "augmented intelligence" (AI), reflecting the enhanced capabilities of human clinical decision making when coupled with these computational methods and systems." Augmented intelligence in health care (AMA Board of Trustees, 2018)

Many ways in which AI can help healthcare

XAVIER HEALTH

- "Machines will not replace physicians but physicians using AI will soon replace those not using it."
  AI-augmented multidisciplinary teams: hype or hope? (Di Ieva, 2019) The Lancet
- "Interest in artificial intelligence in healthcare soared in 2019 with investors pouring $4 billion into the sector across 367 deals… That's up from nearly $2.7 billion invested in healthcare AI in 2018 across 264 deals." "Healthcare led AI investment, topping the $2.2 billion raised by financial and insurance AI."
  Fierce Healthcare coverage of CB Insights Report
- AI will be critical in meeting the care needs of a growing, aging population facing projected physician shortages. However, concerted effort is needed to assure this tech advances the quintuple aim.
  National Academy of Medicine Report on AI (Matheny et al., 2019)

FIGURE S-1 | Advancing to the Quintuple Aim
SOURCE: Developed by publication editors

Example of FDA Cleared AI: IDx-DR

Screening retinal images to detect retinopathy

XAVIER HEALTH

- Diabetic retinopathy is when high blood sugar levels damage blood vessels in retina
- Retinal images are uploaded to a cloud server where IDx-DR software analyzes and tells doctor “more than mild diabetic retinopathy detected; refer to an eye care professional” or “negative for more than mild diabetic retinopathy; rescreen in 12 months”
- Note that this product provides a screening decision without having the clinical also interpret the results and is therefore usable by healthcare providers not normally involved in eye care.

Example of FDA Cleared AI: IDx-DR …continued…

XAVIER HEALTH

- Clinical study of 900 images from 10 primary care sites
- Product correctly identified “more than mild” 87.4% of the time and “not more than mild” 89.5% of the time.
- Demonstrated to be robust for the range of study participant characteristics
- Clear disclaimers on who NOT to use this tool – including pregnant patients where diabetic retinopathy progresses rapidly and the tool is not intended to evaluate rapidly progressive situations.

Source: https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm604357.htm

Example of FDA Cleared AI: IDx-DR …continued…

XAVIER HEALTH

Benefit	Risk
IDx-DR offers the important benefits of potential increased access to diabetic retinopathy screening for people with diabetes in a primary care setting. Earlier detection of Diabetic Retinopathy can help to enable the timely delivery of potentially sight saving interventions.	No risk of direct harm to the patient Risk of false negative mitigated by the slow progression of disease and the recommendation to go for annual eye exams. A false positive would lead to referral of the patient to a specialist for further examination.

Benefits outweigh device risks

Source: https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm604357.htm

Many of the Key Success Factors are things we already know – e.g. Supplier Quality

We traditionally think of supplier quality as only applying to raw materials, sub-assemblies, etc.

For Machine Learning, the training data is the “raw material” – bad raw material results in poor quality finished product.

Success Factor: Good Data Handling Practices

XAVIER HEALTH

One challenge is that AI seems mysterious and magical, and people think we need a whole new way of thinking about it.

I propose that we handle data according to these rules:

- Keep records / retain information on the origin of the sample
- Sourcing, processing, preservation, testing and handling should be done in a safe manner
- Protect against contamination, viruses

Note: these concepts are already captured in IMDRF GRRP WGN47 FINAL:2018 document – when talking about tissue samples !!

My point is that we already know many good practices that simply need to be adapted for AI. We don't need to re-invent the wheel..

Some of the projects & standards in later slides are about data quality…

Image source: https://xkcd.com/1838/

Terminology challenges

XAVIER HEALTH

Agreeing to definitions often takes time, even when people are from the same industry.

- “Narrow AI” vs “General AI” – does the software know a lot about a little, or a lot about a lot?
- “Locked”, “Continuous Learning / Adaptive Systems” – does the system continue to learn over time? If so, does it learn after every single use or in a batched-mode?

Artificial Intelligence practitioners have their own set of terminology that sometimes conflicts with what we think of in medical devices.

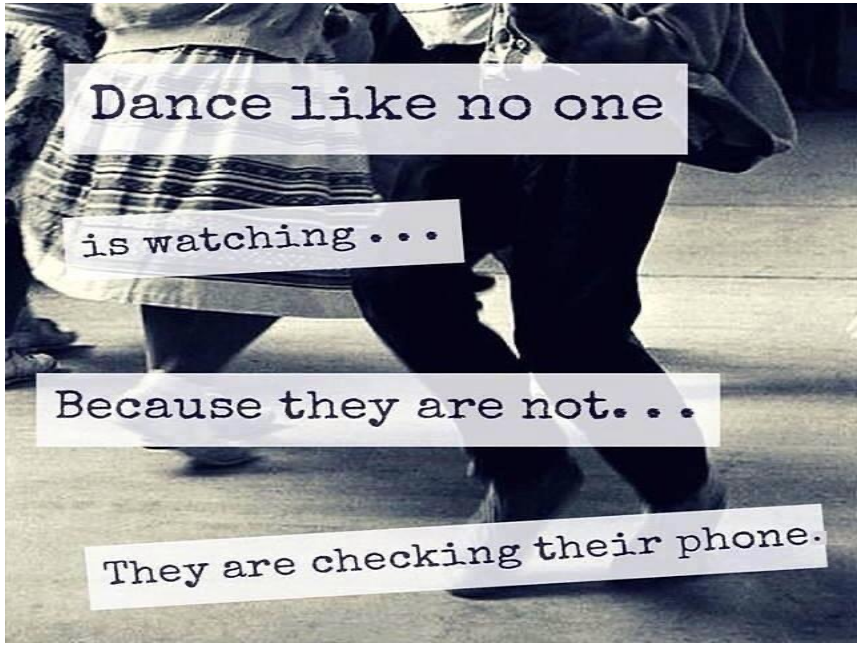
- “Validation” for medical devices often refers to meeting user needs; but “validation” in data science is making sure the data is valid (e.g. a negative heart rate is probably not a valid piece of data)
- “Bias” is something that data scientists try to eliminate, but I’ve talked to many caregivers that want algorithms to be biased towards their particular patient demographics.
- “Supervised” vs “Unsupervised” have very different meanings that you’d expect..

New Technology = New Risks

Pedestrian fatalities rose 11% in 2016, ‘distraction’ as a contributing factor.

As we gain new skills, what do we give up?

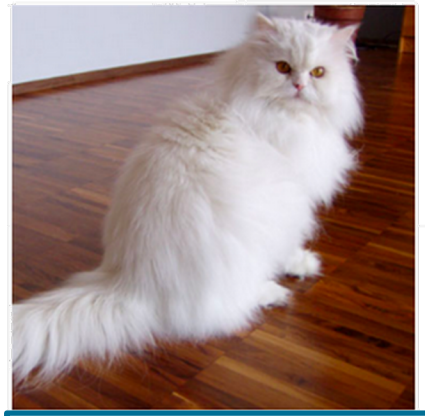
Need for standards to address unique failure modes for ML



What are the limitations ? What can go wrong?

- Will people blindly follow what the system says, even when they are given a choice?
  - Technology doesn't know everything - consider the 2017 California fires, the LA Times Reported "The Los Angeles Police Department asked drivers to avoid navigation apps, which are steering users onto more open routes — in this case, streets in the neighborhoods that are on fire. " http://www.latimes.com/local/california/la-me-southern-california-wildfires-live-firefighters-attempt-to-contain-bel-air-1512605377-htmlstory.html

AI Can Fail in Unexpected Ways



https://medium.com/@ageitgey

Source: "Artificial Intelligence and Medical Algorithms" Berkman Sahiner, FDA, International Conference on Medical Device Standards and Regulations, March 23, 2018

There Are New Hazards That We Never Had To Deal With Before…

Challenge – we forget items that are second nature to us

XAVIER HEALTH

- Draft UL standard on fully-autonomous vehicles: what does the vehicle do when approaching a stoplight that just turned yellow?
- Defeating facial recognition software
- Defeating speed limit sign

Takeaway: we are in an age of "Narrow AI"

https://www.businessinsider.com/hackers-trick-tesla-accelerating-85mph-using-tape-2020-2

Real World Example – Afib Prediction

XAVIER HEALTH

- Hospital developed AI software to identify patients at-risk for Afib.
- Software identified patients and reduced Afib – great!
- But software missed some patients as well – hospital looked into it
- Talked to floor nurse – MIL visited the afternoon of the event
- MIL are apparently a new hazard that have not previously been identified.
- Takeaway – EHR systems don't capture everything, there will always be extra information that affects the patient but software doesn't know. Data doesn't replace knowledge..

IMDRF

XAVIER HEALTH

- In 2020, the IMDRF formed an AI working group, and the first project it to develop a glossary of AI/ML-related terms.
- One very important discussion about scope – should the IMDRF efforts be looking at systems that use ML *anywhere* in the product, or only places where the ML impacts risk/benefit? E.g. ML is used to screen for breast cancer vs. ML is used to optimize workflow ?
- A draft version of the glossary is out for public comment; comments due in late November. http://www.imdrf.org/consultations/cons-aimd-mlmd-ktd.asp

!! Note this is just a glossary, it doesn't provide any guidance about ML product development

Levels of Autonomy (LOA)

XAVIER HEALTH

- How much freedom do we give the software? How much oversight does it need?
- Let's leverage what other industries are doing – for different levels of autonomy, what has already been published??
- There are various levels of autonomy, and the level of autonomy drives risk assessments, trustworthiness levels, concerns over liability, etc.

Sheridan, T. B., & Verplank, W. L. 1978. Human and computer control of undersea teleoperators. Cambridge, Mass: Massachusetts Institute of Technology, Man-Machine Systems Laboratory.

In 1978 a paper was published regarding automation and teleoperation, and it outlined 10 levels of automation:

1. Computer offers no assistance; human does it all
2. Computer offers a complete set of action alternatives
3. Computer narrows the selection down to a few choices
4. Computer suggests a single action
5. Computer executes that action if human approves
6. Computer allows the human limited time to veto before automatic execution
7. Computer executes automatically then necessarily informs the human
8. Computer informs human after automatic execution only if human asks
9. Computer informs human after automatic execution only if it decides to
10. Computer decides everything and acts autonomously, ignoring the human

Levels of Autonomy

XAVIER HEALTH

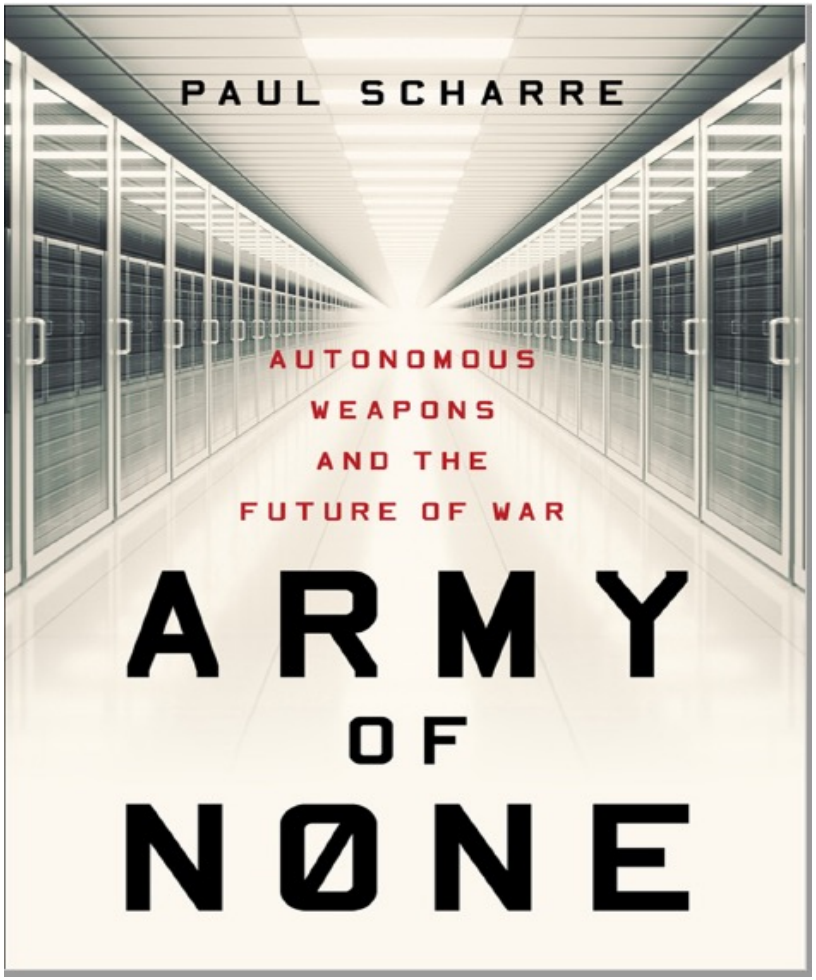
The automotive industry is also looking at autonomy -- this table is from an automotive standard, SAE J3016

SAE level	Name	Narrative Definition	Execution of Steering and Acceleration/ Deceleration	*Monitoring* of Driving Environment	Fallback Performance of *Dynamic Driving Task*	System Capability *(Driving Modes)*
***Human driver* monitors the driving environment**						
0	**No Automation**	the full-time performance by the *human driver* of all aspects of the *dynamic driving task*, even when enhanced by warning or intervention systems	Human driver	Human driver	Human driver	n/a
1	**Driver Assistance**	the *driving mode*-specific execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*	Human driver and system	Human driver	Human driver	Some driving modes
2	**Partial Automation**	the *driving mode*-specific execution by one or more driver assistance systems of both steering and acceleration/ deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*	**System**	Human driver	Human driver	Some driving modes
***Automated driving system* ("system") monitors the driving environment**						
3	**Conditional Automation**	the *driving mode*-specific performance by an *automated driving system* of all aspects of the dynamic driving task with the expectation that the *human driver* will respond appropriately to a *request to intervene*	System	**System**	Human driver	Some driving modes
4	**High Automation**	the *driving mode*-specific performance by an automated driving system of all aspects of the *dynamic driving task*, even if a *human driver* does not respond appropriately to a *request to intervene*	System	System	**System**	Some driving modes
5	**Full Automation**	the full-time performance by an *automated driving system* of all aspects of the *dynamic driving task* under all roadway and environmental conditions that can be managed by a *human driver*	System	System	System	**All driving modes**

Military applications & LOA

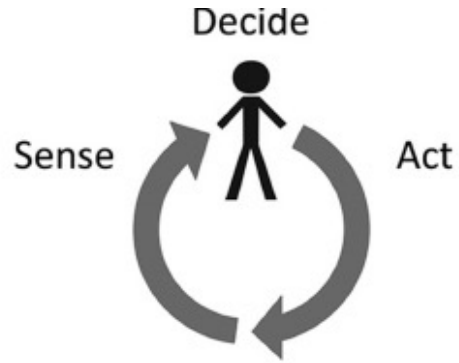
XAVIER HEALTH

The auto industry isn't the only one looking at autonomy…



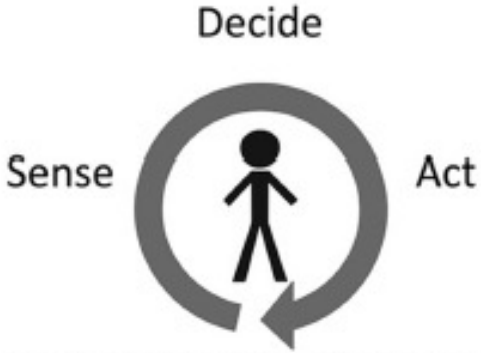
Loopy Humans - to what degree are "humans in the loop" ??

First level needs a human to complete the task



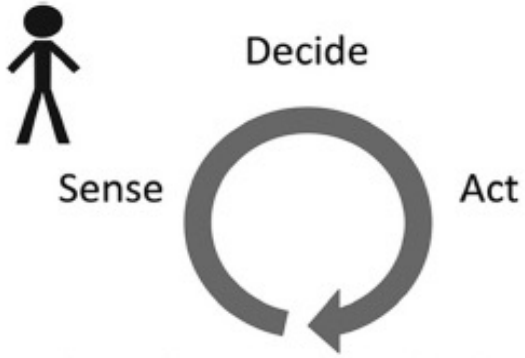
The machine performs a task and then waits for the human user to take an action before continuing.

Second level allows for human over-ride



The machine can sense, decide, and act on its own. The human user supervises its operation and can intervene, if desired.

Third level does not allow for intervention



The machine can sense, decide, and act on its own. The human cannot intervene in a timely fashion.

IMDRF Software Risk Framework – an idea I have…

XAVIER HEALTH

The International Medical Device Regulators Forum published a framework for assessing the risk of software, according to the significance of the software when making decisions, and the criticality of the patient state

State of Healthcare situation or condition	Significance of information provided by SaMD to healthcare decision		
	Treat or diagnose	Drive clinical management	Inform clinical management
Critical	IV	III	II
Serious	III	II	I
Non-serious	II	I	I

For AI, I think that "Treat or diagnose" has several levels, depending on how much freedom the AI is given – does the human have to "approve" before action is taken or does human have the ability to "override".

State of Healthcare situation or condition	Significance of information provided by SaMD to healthcare decision				
	Treat or diagnose w/ no intervention possible	*Treat or diagnose w/Override*	Treat or diagnose *w/Approval*	Drive clinical management	Inform clinical Management
Critical	???	??	IV	III	II
Serious	??	IV?	III	II	I
Non-serious	IV?	III	II	I	I
	Automated/Autonomous Intelligence			Assistive Intelligence	

This has been included in CTA standards and being considered by WHO.

IMDRF "Software as a Medical Device: Possible Framework for Risk Categorization and Corresponding Considerations"

Standards Overview

XAVIER HEALTH

- ISO/IEC JTC1, SC42, developing horizontal standards for all industries. Many simultaneous projects and even more are being created. Not likely that these horizontal standards would be required for medical devices, but they may contain ideas that we like and would carry to healthcare.
- ISO/IEC TC215 (health software standards like 62304, 82304, 80001-x series, + 200 more standards) has a report with some recommendations, and has proposed a Task Force to maintain a current landscape, collect use cases, etc.
- IEEE also developing a number of AI standards, but only a few are specific to healthcare.
- CTA is developing general AI standards as well as healthcare-specific AI standards.
- AAMI & BSI have also started an AI standards committee.

Click to edit Master title style

XAVIER HEALTH

ITU/WHO Focus Group AI for Health

- Artificial Intelligence for Health (A4IH) offers substantial improvements for public and clinical health, e.g. early detection, diagnosis and risk identification, treatment decision support, self-management, improved outcomes, …
- For world-wide adoption, need evaluation standards on effective AI for Health
- Focus Group AI for Health (FG-AI4H) created July 2018; open platform
- FG-AI4H goals: standardized framework for benchmarking and evaluation of AI solutions

World Health Organization & ITU-T

XAVIER HEALTH

Structure:

- Expert panel (clinical)
- Establishing benchmarking platform
- Data & solution quality assessment committee
- Data handling committee
- Regulatory committee

Example projects:

- Outbreak detection
- Fall predictions
- Dermatology
- Neuro cognitive disease
- Snake bite….

DAISAM Work Group: Data and AI Solution Assessment Methods

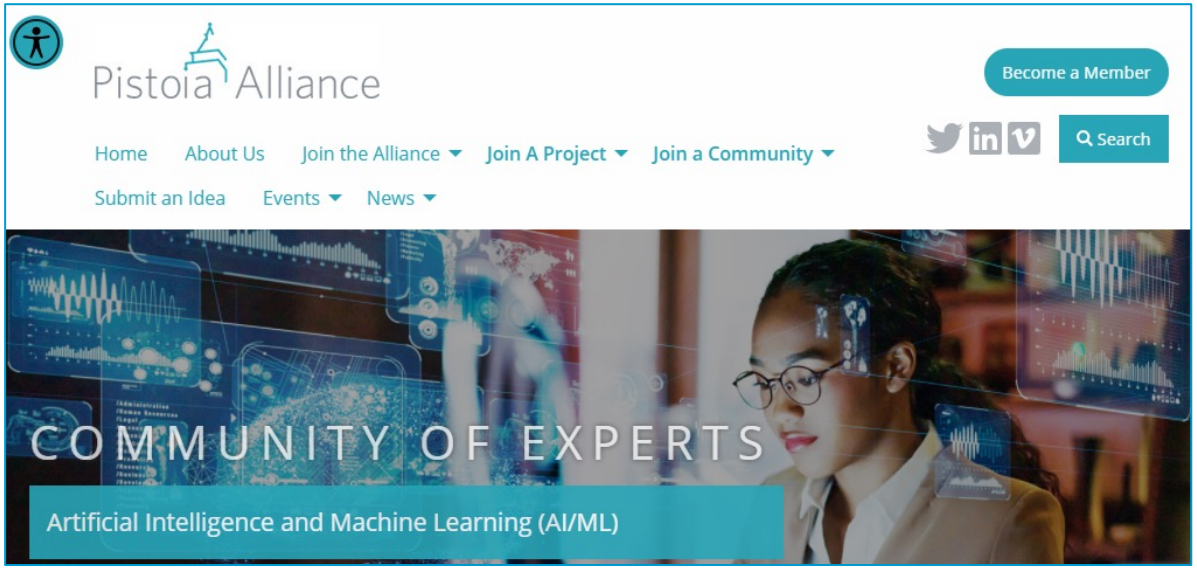
XAVIER HEALTH

- Charter is to develop a quality assessment approach for candidate AI applications. Includes both data quality and solution(algorithm) quality
- Reviewed 50+ existing standards and guidance documents and created a library of 200+ questions. This was the basis to develop a draft assessment.
- This is an iterative process – we have an initial set of assessment questions, topic teams will have topic-specific thoughts, individual projects will think of their own criteria, etc (e.g. identify risks that we didn't think of..)
- This is a collective learning environment & we are capturing this information for future use. Note that we are trying to continuously learn about Continuous Learning!

Pistoia Alliance

XAVIER HEALTH

- Over 140 member organizations in the Pistoia Alliance
- All 10 largest pharmaceutical companies (by $ volume) are members
- 8 currently active and independently funded projects and 5 completed projects



CURRENT AREAS OF FOCUS INCLUDE:

- Developing a Best Practices Toolkit for Machine Learning Ops in life sciences (2021-2022)
- Educating members of the community about use cases and applications through our webinar series and conference presentations
- Defining specific pre-competitive project ideas and spinning out new projects or communities of interest.

Pre-Standards: BSI-AAMI Joint Initiative

XAVIER HEALTH

- UK-U.S. collaboration, with support from MHRA and FDA
- Research, surveys, in-depth discussions
- Stakeholder workshops (autumn 2018):
  - challenges, alignment to regulatory and standards landscape, information gaps, proposed solutions
  - terminology & categorization, alignment to IMDRF principles, next steps and priorities
- Position Paper published February 2019

New Committee: AAMI Artificial Intelligence

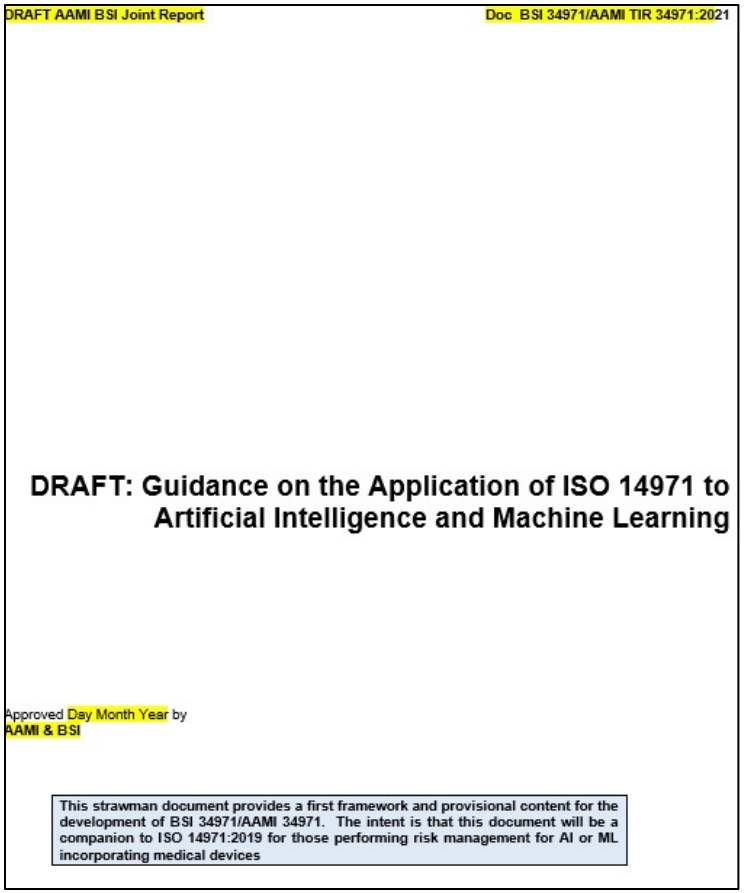
XAVIER HEALTH

After publishing a few whitepapers with BSI, AAMI/BSI started working on another whitepaper regarding AI risk management. Feedback we received on the whitepaper was "why are you doing another whitepaper? A standard or a TIR would be more useful…"

An official AI committee was formed, and the first project is TIR 34971 – Guidance on the Application of ISO14971 to Artificial Intelligence and Machine Learning.

You might remember that the ISO/IEC 24971 provides guidance on how to implement 14971, and includes an annex of special considerations for IVDs. The risk management process is the same, but there are some things you might not have considered when it comes to IVD risk management.

34971 is following that same pattern – the process is the same, but there are new ways to fail. There are different things to look for (e.g. bias), there are different risk controls to consider.



Consumer Technology Association (CTA)

CTA is the trade association for the consumer technology industry (all consumer industries – not just healthcare)

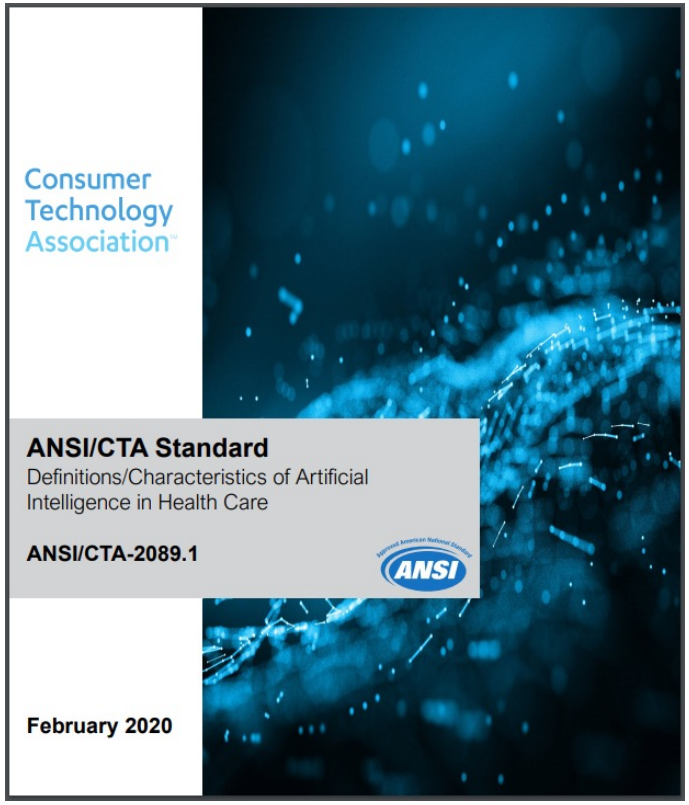
Established AI working group, published two whitepapers in 2018 – general introduction & use cases.

AI standards committee (R13) & Health Care working group (R13 WG1). Publications include:

- "Definitions / Characteristics of AI in Health Care (ANSI/CTA-2089.1)"
- "The Use of AI in Health Care: Trustworthiness (ANSI/CTA-2090)"

Current project is about Data Quality

Note: Scope includes consumer health, fitness, and wellness technology.



TC215 Ad Hoc Group on Application of AI to Health Informatics

XAVIER HEALTH

In 2019, ISO/TC 215 created an Ad Hoc Group (AHG 2) on Application of AI Technologies in Health Informatics, the charter was to:

1 - Conduct a landscape survey of the topic with respect to health informatics
2 - Identify key considerations for TC 215
3 - Provide a set of recommendations for future work

Resulted in ~ 30 recommendations, including the development of standards / TR for:

- Definitions
- Software lifecycle for AI solutions
- Evaluating performance and validity of AI health applications
- Explainability disclosures
- Security & privacy considerations
- Development of checklist of things to consider when developing or updating standards
- Consider educational projects
- Collaborate w/other organizations

Since many recommendations are ongoing (not one-time), a Task Force (TF5) was created to provide ongoing assistance (e.g. liaisons to other groups, keeping track of other related standards projects, etc.)

ISO/IEC JTC1 SC42 – AI Standards for all Industries

XAVIER HEALTH

Structure:
WG1 – Foundational standards (terminology, framework)
WG2 – Big Data (vocabulary, reference architecture)
WG3 – Trustworthiness (incl. risk, robustness, bias)
WG4 – Use cases and applications
WG5 – Computational approaches & characteristics of AI

JWG1 (SC40)– Governance implications of AI
AG1 – Management Systems Standard
AG2 – AI Systems Engineering

4213 - Information technology — Artificial Intelligence — Assessment of machine learning classification performance
5338 - AI system life cycle processes
5339 - Guidelines for AI applications
5392 - Information technology — Artificial intelligence — Reference architecture of knowledge engineering
5469 - Functional Safety **(this one is about safety..)**
5471 - Artificial intelligence — Quality evaluation guidelines for AI systems
6254 - Objectives and methods for explainability of ML models and AI systems
8183 - Data life cycle framework
20546 - Big Data - Overview and Vocabulary
22989 - AI Concepts and Terminology
23053 - Framework for AI using ML
23894 - Risk Management **(ISO 31000, not 14971)**
24027 - Bias in AI systems and AI aided decision making
24028 - Overview of Trustworthiness in AI
24030 - Use cases and application
24368 - Overview of ethical and societal concerns
24372 - Overview of computations approaches for AI systems
24668 - Process management framework for Big data analytics
25059 - Systems and software Quality Requirements and Evaluation (SQuaRE)
38507 - Goveranance implications of the use of AI by organizations.
42001 - Management system
82000 - Controllability of automated artificial intelligence systems
20547-1 - Big Data reference architecture - Part 1: Framework and application process
20547-2 - Big Data reference architecture - Part 2: Use cases and derived requirements
20547-3 - Big Data reference architecture - Part 3: Reference architecture
20547-5 - Big Data reference architecture - Part 5: Standards roadmap
24029-1 - Assessment of the robustness of neural networks - Part 1 Overview
24029-2 - Formal methods methodology
5259-1 - Data quality for analytics and ML — Part 1: Overview, terminology, and examples
5259-2 - Data quality for analytics and ML — Part 2: Data quality measures
5259-3 - Data quality for analytics and ML — Part 3: Data Quality Management Requirements and Guidelines
5259-4 - Data quality for analytics and ML — Part 4: Data quality process framework

ISO/IEC JTC1 SC42
has a lot of projects..

IEEE also has multiple (horizontal) AI standards, as well as a few specific to healthcare:

P2801 Recommended Practice for the Quality Management of Datasets for Medical Artificial Intelligence Recommendation (in progress)

P2802 Standard for the Performance and Safety Evaluation of Artificial Intelligence Based Medical Device: Terminology (published)

And although it's not healthcare-specific…

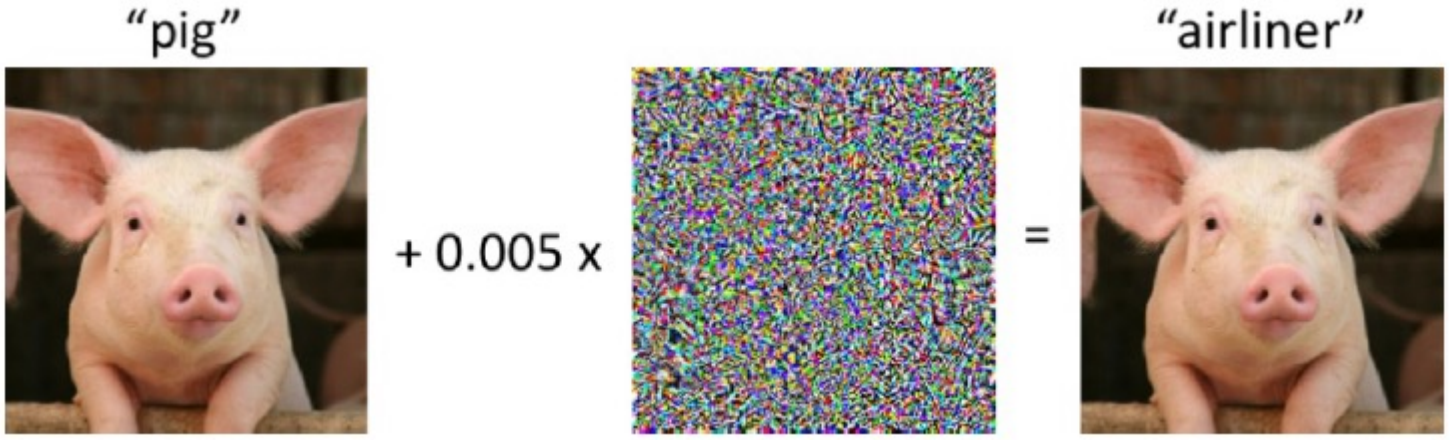
P7003 Algorithmic Bias Considerations (in progress)

New Topic! ML & Cybersecurity!

XAVIER HEALTH

Multiple organizations suddenly became interested in the unique challenges for cybersecurity for ML systems. What are the unique threats to ML, and do existing cybersercurity practices mitigate those threats? Projects include:

- Collaborative effort between SC27 and SC42
- Discussion in HSCC
- CTA exploring development of whitepaper
- WHO / FGAI4H



“Example of adversarial perturbation used to evade classifiers”;
Draft NISTIR 8269 A Taxonomy and Terminology of Adversarial Machine Learning

Questions?