Trends and Convergence (including Standards) – Lessons Learned from other Fields

Pat Baird, Sr. Regulatory Specialist

Pat.Baird@Philips.com

February 27, 2024

I want to thank Patty Krantz-Zuppan (Meditronic) for her slides regarding IEC TC62 & SNAIG

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

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.

THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS

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

Levels of Autonomy (LOA)

How much freedom do we give the software? How much oversight does it need?

There are various levels of autonomy, and the level of autonomy drives risk assessments, trustworthiness levels, concerns over liability, etc.

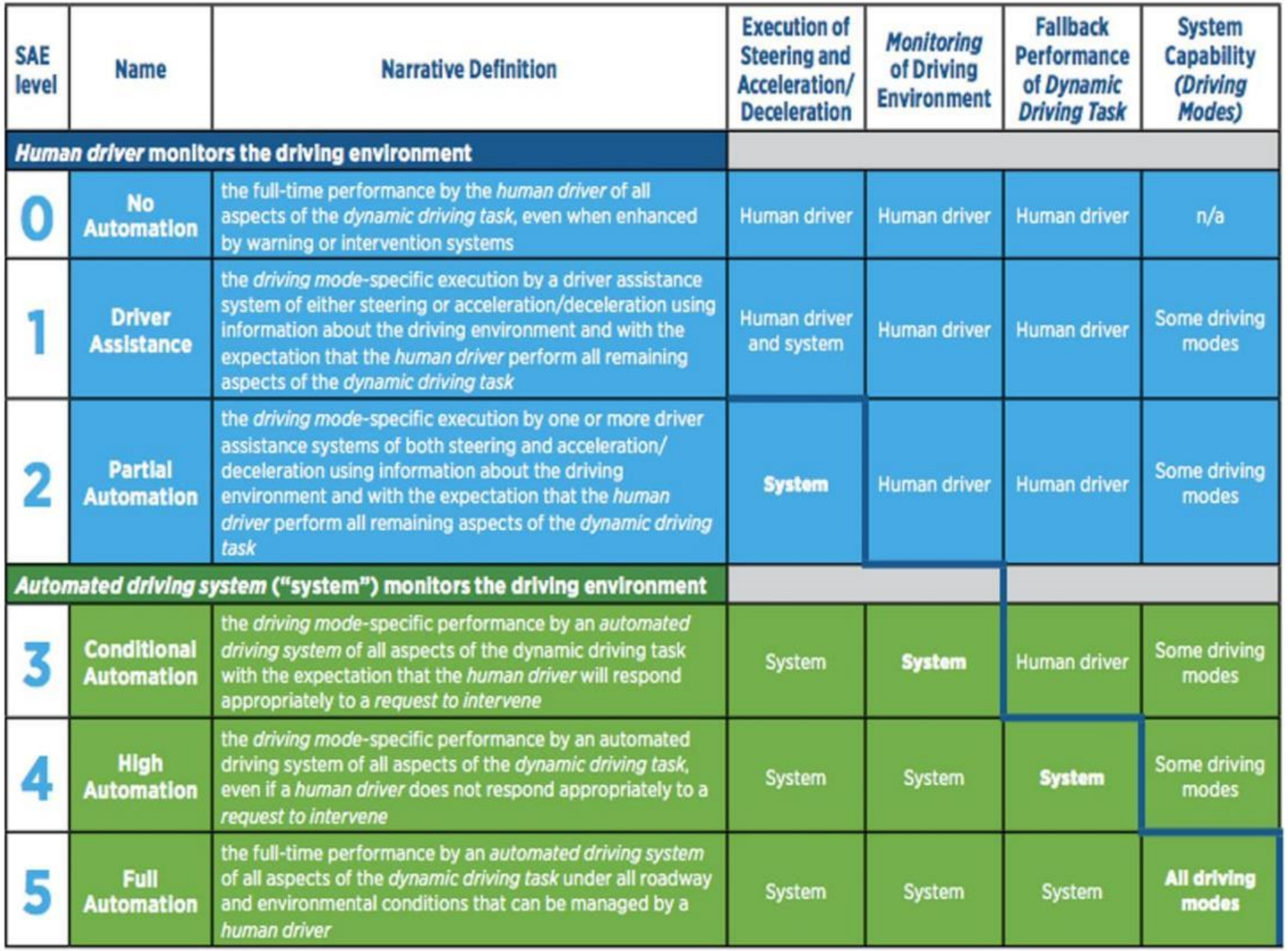
In **1978** a paper was published regarding automation and teleoperation, and it outlined 10 levels of automation:

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

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

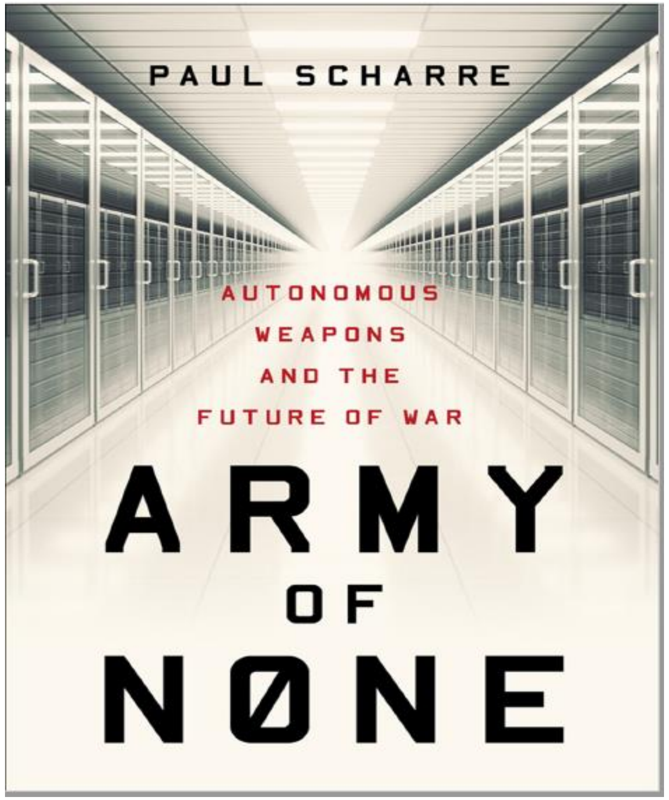
Levels of Autonomy

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

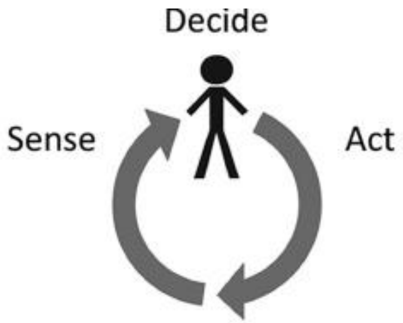


SAE level	Name	Narrative Definition	Execution of Steering and Acceleration/ Deceleration	*Monitoring* of Driving Environment	Fallback Performance of *Dynamic Driving Task*	System Capability *(Driving Modes)*
***Human driver* monitors the driving environment**						
0	No Automation	the full-time performance 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**

To what degree are “humans in the loop” ??

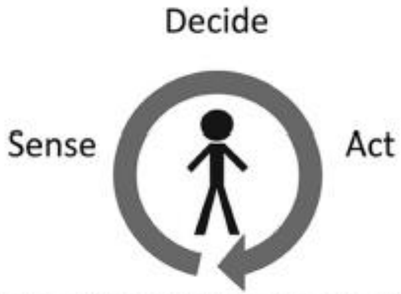


First level needs a human to complete the task



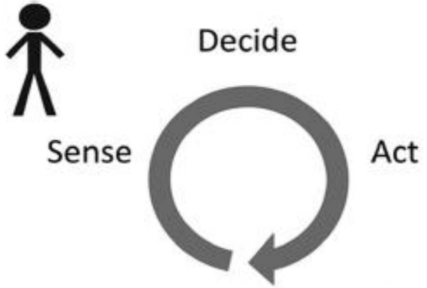
The machine performs a task and then waits for the human user to take an action before continuing.

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.

We could consider different levels of risk based on the level of autonomy – systems that are fully automated might be higher risk..

Standards Overview

- 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.**
- IEEE also developing a number of AI standards.
- AAMI & BSI have collaborated to develop healthcare AI standards.

My point is that there are many organizations looking at this technology and are committed to defining and sharing good practices.

Topics discussed in AI Standards

There will be a giant family of standards for ML systems, including:

- Definitions
- Governance
- Risk management
- Trustworthiness
- Security
- Managing Bias
- Verification & Validation
- Data Management
- Postmarket considerations

And others!

One of the major success factors will be keeping this at a manageable level.

ISO/IEC JTC1 SC42 has a lot of projects…

4213 - Information technology — Artificial Intelligence — Assessment of machine learning classification performance
4213 - 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
5471 - Artificial intelligence — Quality evaluation guidelines for AI systems
6254 - Objectives and methods for explainability of ML models and AI systems
8200 - Controllability of automated artificial intelligence systems
12791 - Treatment of unwanted bias in classification
12792 - Transparency taxonomy of AI systems
20546 - Big Data - Overview and Vocabulary
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
20547.4 - Information technology — Big data reference architecture — Part 4: Security and privacy
20547.5 - Big Data reference architecture - Part 5: Standards roadmap
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
24029.1 - Assessment of the robustness of neural networks - Part 1 Overview
24029.2 - Formal methods methodology
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
22100-5 - Safety of machinery — Relationship with ISO 12100 — Part 5: Implications of artificial intelligence machine learning
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

Some of these topics are not what you think -- note that "risk management" is enterprise-risk, not safety-risk. If you want safety, look at 5469

Some (relatively) new projects include "Oversight"; another is exploring the positive use-cases for AI applications.

Also be aware that some law makers assume that horizontal standards can apply to all industry – after all, they are horizontal!

Pre-Standards: AFDO/RAPS (was Xavier) AI Working Groups

- Started in late August 2017 at the Xavier University AI Summit.
- Composed of a group of experts from medical device and pharmacology industries, academia, government
- Purpose: Maximize the advantages of artificial intelligence in advancing patient health by identifying how to provide a reasonable level of confidence in the performance of continuously learning systems in a way that minimizes risks to product quality and patient safety
- Three sub-committees: 1) GMLP, 2) AI Operations, 3) AI at the Point of Care
- GMLP has published papers on Good Machine Learning Practices (a supplement to 62304), trustworthiness, data quality, bias management, and is currently working on postmarket.

Consumer Technology Association (CTA)

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

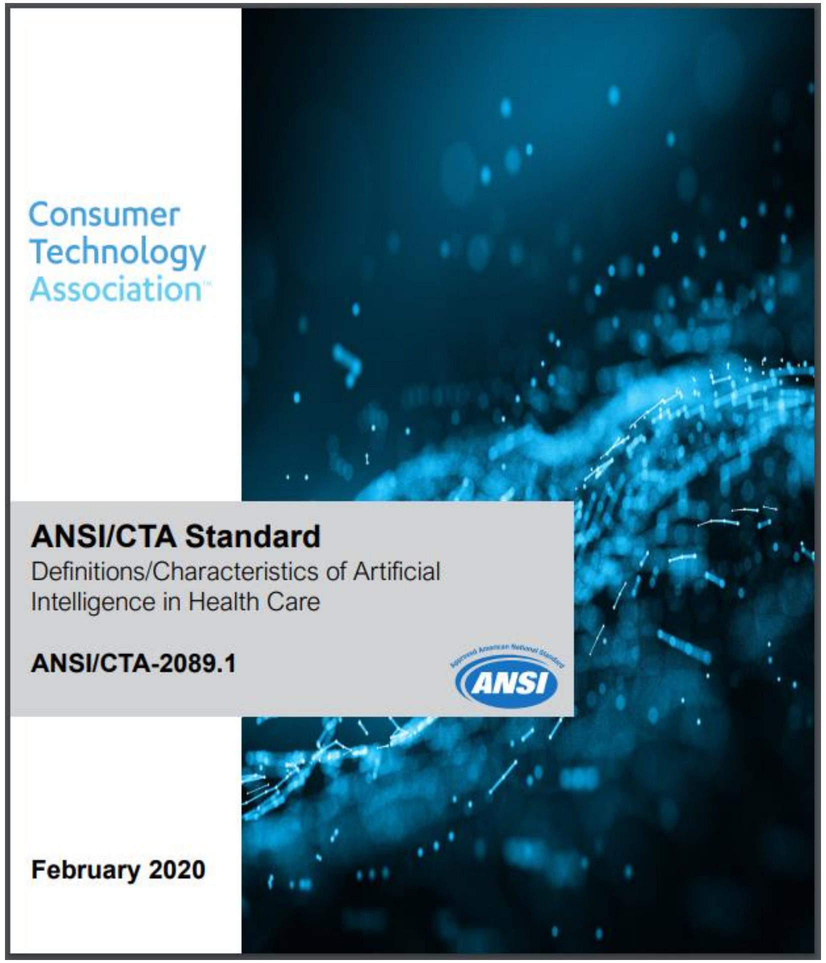
AI standards committee (R13) & Health Care working group (R13 WG1) have published:

- “Definitions / Characteristics of AI in Health Care (ANSI/CTA-2089.1)”
- “The Use of AI in Health Care: Trustworthiness (ANSI/CTA-2090)”
- “The Use of Artificial Intelligence in Health Care: Managing, Characterizing and Safeguarding Data (ANSI/CTA-2107)”
- “Artificial Intelligence in Health Care: Practices for Identifying and Managing Bias (ANSI/CTA-2116)

The current project is a guide on a user-facing “Nutrition Label”

(What I like about CTA is that ***<u>CTA Get’s Things Done!</u>***)

(And their standards are available for free)



Consumer Technology Association

ANSI/CTA Standard
Definitions/Characteristics of Artificial Intelligence in Health Care

ANSI/CTA-2089.1

ANSI

February 2020

https://shop.cta.tech/collections/standards/artificial-intelligence

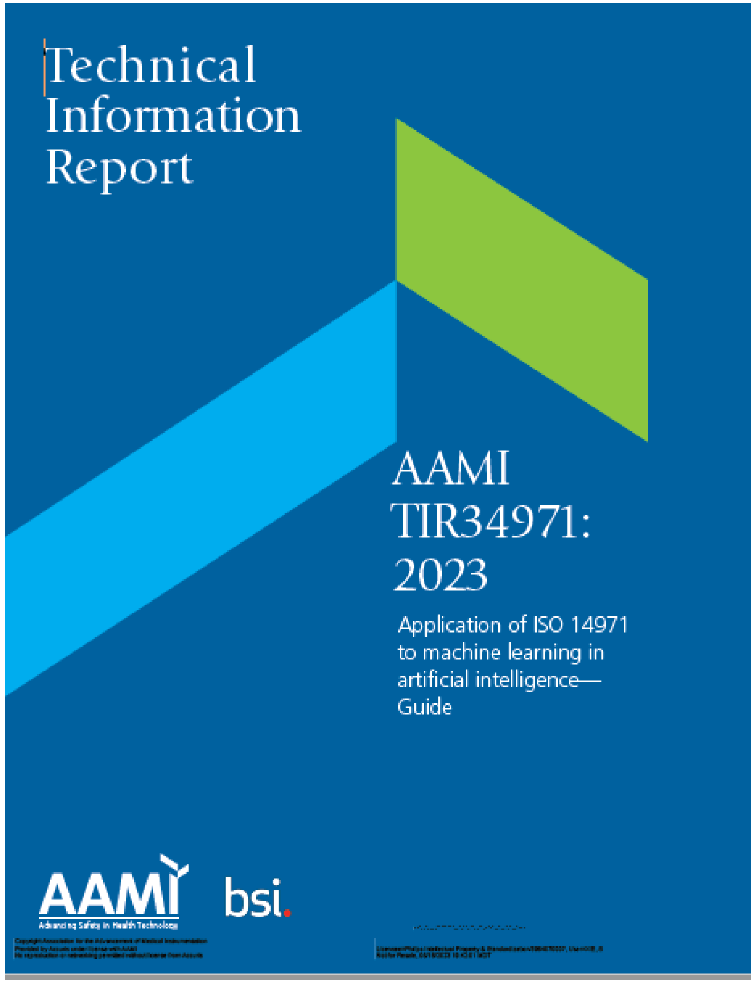
AAMI Artificial Intelligence

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 would be more useful…”

ISO 14971 is a commonly used risk management standard for medical devices. We used that to create a list of new ML-related hazards and possible controls. This was published in 2023 as “TIR 34971”.

We are expanding this to ISO/IEC; we had our first meeting in December (and our next meeting will be next week!) ISO/IEC JWG 1 will be naming this “24971-2.” We will be clarifying some of the language from 34971, and we will be adding in Large Language Model (LLM) related risks. LLMs were not a significant topic when we were developing 34971, but they have become very popular in the last year.



Technical Information Report

AAMI TIR34971: 2023

Application of ISO 14971 to machine learning in artificial intelligence—Guide

AAMI Advancing Safety in Health Technology

bsi.

Some ML-related Hazards from 34971

- Incorrect data
- Incomplete data (eg empty fields in a database)
- Subjective data (eg patient reporting pain scales)
- Inconsistent data
- Atypical data – the quality of the data during development might not represent the quality of the data in actual use (e.g. high resolution vs low resolution mammograms.)

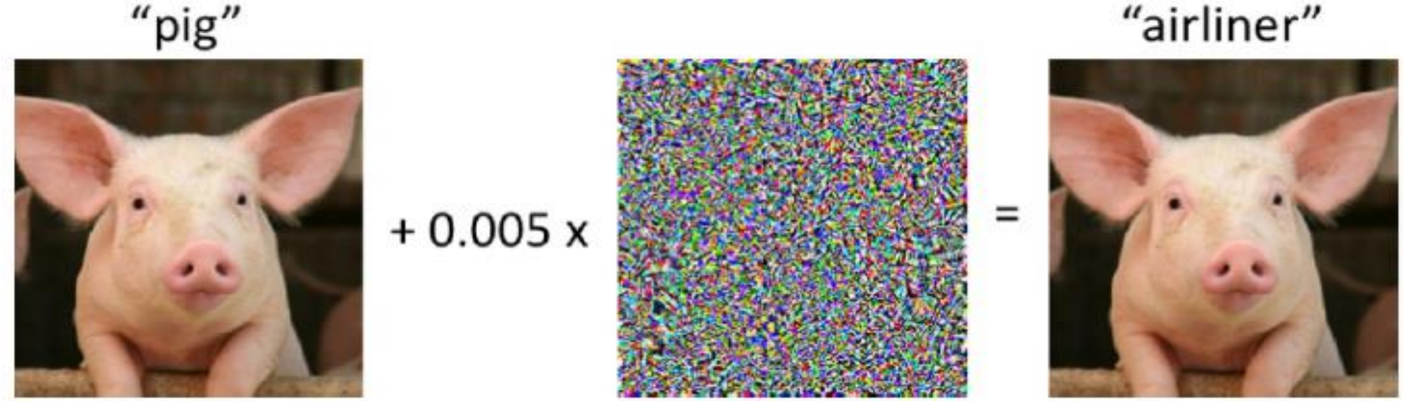
- Over-confidence – the user trusts the system too much and believes it will work in all situations.
- Perceived risk – user might perceive the risks to be lower than they really are and are more likely to trust or delegate to ML
- User workload – people are busy and don't have time to stop and think about the application; a busy user is more likely to trust the software.
- Self-confidence – the user could defer to a product's "superior judgement"
- Variation in social trust – different user populations (including different professions, different cultures) have varying levels of trust and the developers might not be aware of these differences.
- User policies: company policies may put their trust in the ML software, forcing users to agree with the ML application.

- Privacy failures – information might be disclosed to unauthorized persons. Although the data can be anonymized, the anonymization process can fail. Additionally, personally identifiable information might contain critical information for the algorithm and anonymizing the data may destroy this critical information.
- Bias due to privacy – not all patients are willing to share their data and this can vary by patient demographics. For example, older patients might be reluctant to share their data, resulting in a bias towards younger populations.

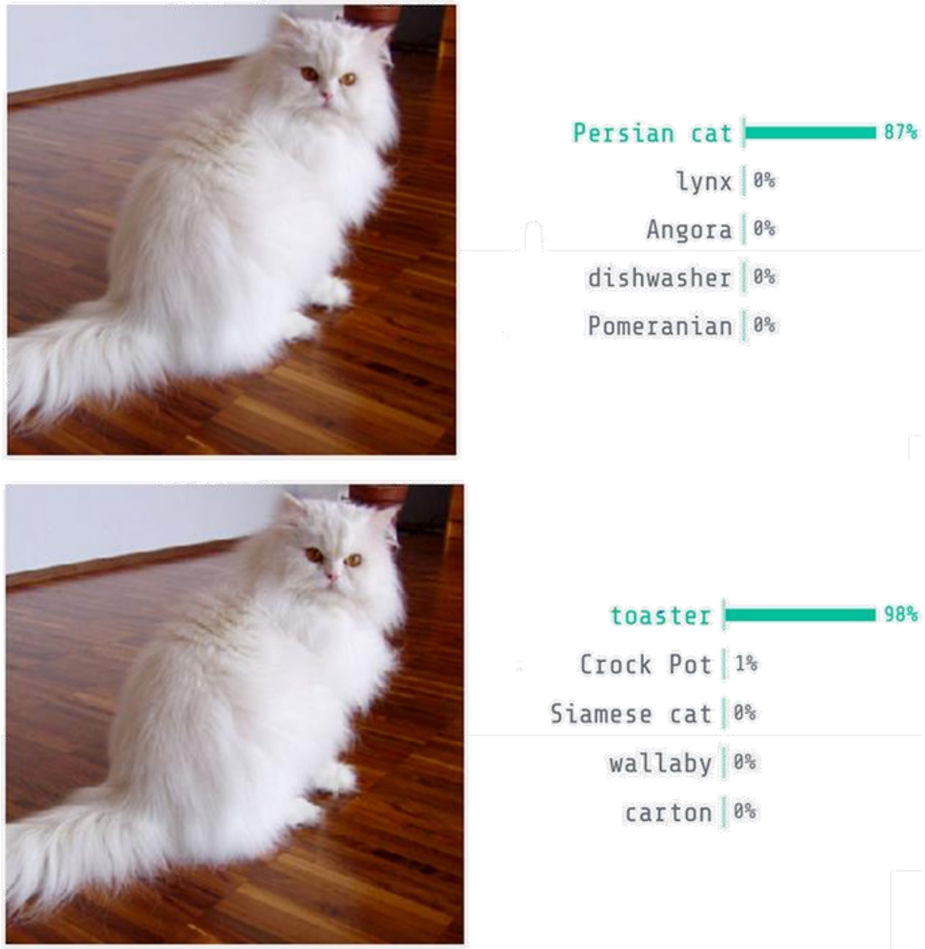
Cybersecurity Considerations..

ML systems have a lot of data. Potentially very attractive data. This data is often handled by multiple stakeholders as it is passed from one system to another.

Due to the nature of the data & how ML systems work, it might not be obvious that there has been a security issue…



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



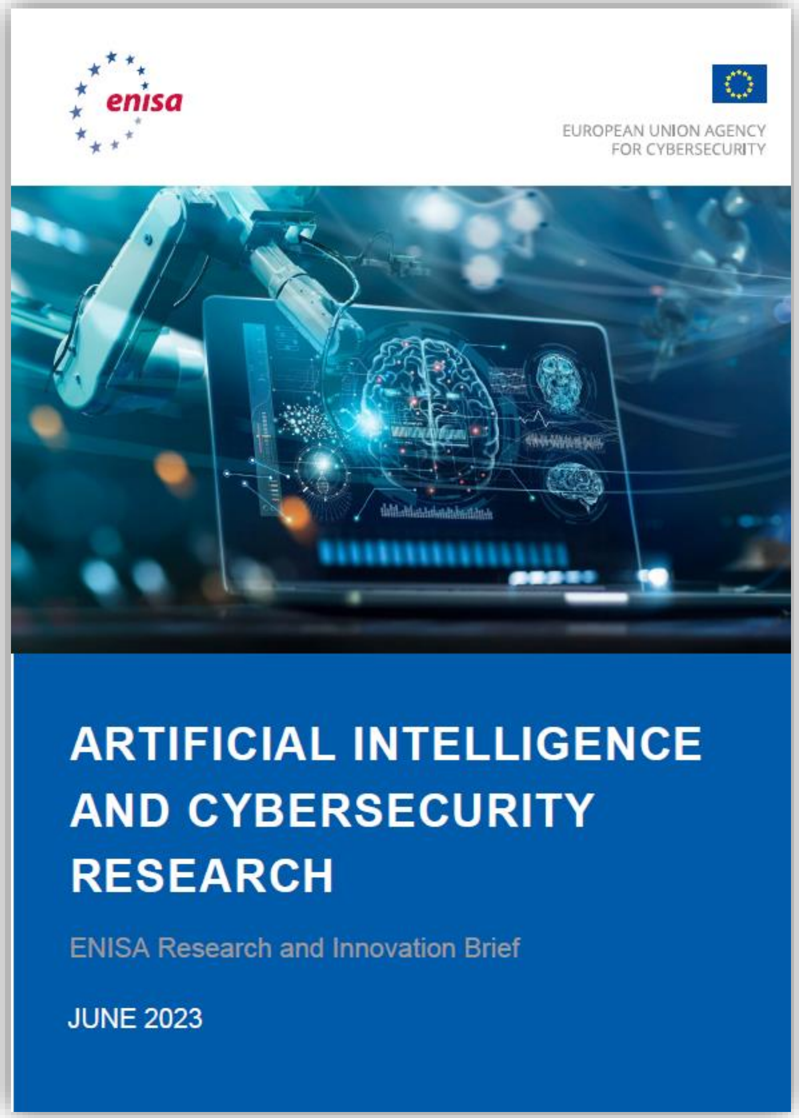
Source: "Artificial Intelligence and Medical Algorithms"
Berkman Sahiner, FDA, International

ENISA Reports…

ENISA has published several reports in the past several years about AI & cybersecurity:

- AI CYBERSECURITY CHALLENGES – published in 2020, it identifies a series of threats for AI systems
- SECURING MACHINE LEARNING ALGORITHMS – summarizes a literature search of 228 publications! Includes threats, vulnerabilities, and controls
- ARTIFICIAL INTELLIGENCE AND CYBERSECURITY RESEARCH looks at using AI as a tool for managing cybersecurity activities
- CYBERSECURITY AND PRIVACY IN AI – MEDICAL IMAGING DIAGNOSIS – really good list of threats, vulnerabilities, and controls
- MULTILAYER FRAMEWORK FOR GOOD CYBERSECURITY PRACTICES FOR AI – (based on ISO/IEC standards, IEEE, etc.)

Three of those reports came out in June!



enisa

EUROPEAN UNION AGENCY
FOR CYBERSECURITY

ARTIFICIAL INTELLIGENCE AND CYBERSECURITY RESEARCH

ENISA Research and Innovation Brief

JUNE 2023

Cybersecurity & Standards

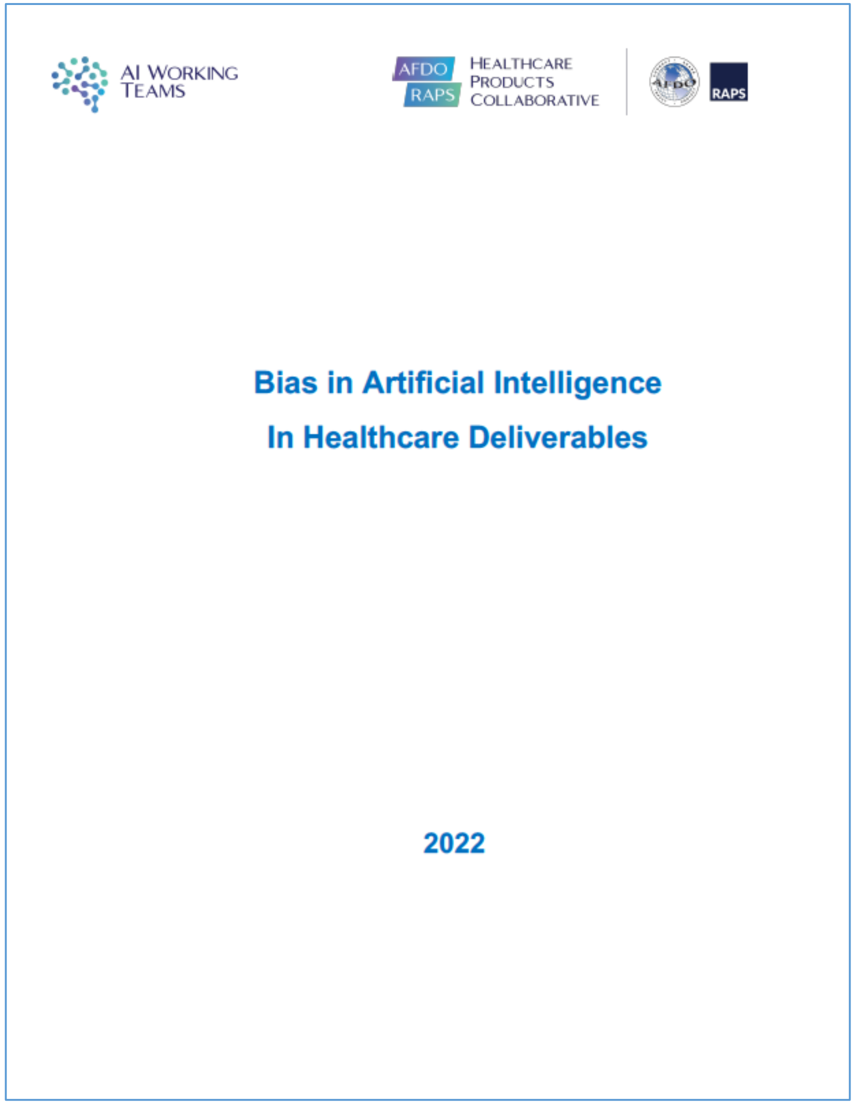
- ISO/IEC SC27 & SC42 have joined forces to work on an international standard for security of ML systems; however, this is a horizontal standard across all sectors
- AAMI has started a new Consensus Report to catalog some of the unique threats and vulnerabilities of ML systems in healthcare. Building on existing processes in TIR57 and TIR97, this is intended to raise awareness about potential problem areas.
- We are examining all of the phases of the ML lifecycle (e.g. data acquisition, cleaning, annotation, etc. all the way through to decommissioning) and are looking at what might be trouble in each of those phases, as well as potential risk controls.
- We hope to have it published by the end of September of this year.

Upcoming Project: Bias Management

One new AAMI project is to create a bias management standard, using the AFDO/RAPS whitepaper as a starting point.

Again, rather than invent something new, the AFDO/RAPS team uses the risk management process from 14971 to manage bias – even though not all types of bias result in harm:

1. Identify potential bias
2. Assess impact
3. Put controls in place where needed
4. Ensure controls are effective
5. Continue to monitor post-market



AI Working Teams

AFDO RAPS Healthcare Products Collaborative

AFDO RAPS

Bias in Artificial Intelligence

In Healthcare Deliverables

2022

IEC TC 62 Advisory Group SNAIG

TC62 formed a group to make recommendations about what standards might need to be updated or created. The group came up with a process for identifying and prioritizing future projects – this was not simply a collection of people's opinions.

Software Network and Artificial Intelligence Advisory Group's charter:

Monitor and analyze available information from outside sources and advise IEC TC 62

- Monitor other SDO's work programs (CEN/CENELEC, IEEE EMBS, SC42, etc.)
- Monitor worldwide regulatory requirements
- Provide actionable recommendations to the officers for new work and collaboration partners and liaisons

Assist TC 62 in realizing its vision for Emerging Technologies

(based on the TC 62 Business Plan)

Overview

Existing standards sufficient; some additions for the application to AI-MD software different partners

(all sorts of MD) QMS, RMS, PMS, Security, Privacy	Clarification where additional factors are to be considered - basic standards are basically sufficient		Impact of patient as operator in the private space and cloud are hardly considered at present
Traditional ME (Hardware + non AI Software)	covered		
SAMD	Extension/addition to 62304 for ML	clinical validation for SAMD	

New standards are necessary; IEC TC62 is in the lead or works in a JWG with partners

(ML – MD, maybe AI - MD in general) Logic component	Quality metrics for external data component	clinical validation (data aspects, statistical effectiveness)	Techno-Vigilance necessary
	Quality criteria and docu req. for „logic“ component		Clinical utility (effective benefit at the site)
	Methods for building test sets		Transparency in Usage and operation
(AI – MD) Bias (Ethics?)	Overarching process standard and consideration in all columns for symbolic AI + ML		

New standards are necessary; ISO TC215 is in the lead or works in JWGs; IEC TC 62 contributes

(ML – MD) Data component	Data Lifecycle process standard (incl. Selection, collection, vetting, documentation etc.)	
	Quality standards in design and development	Including changes based on PMS data or continuous learning
	Quality assurance methods and quality metrics	
	Methods for validation	

Other IEC TC62 partnership work

- Cooperation with ISO/IEC JTC1/SC42/JWG3 together with ISO TC215 Current project: ISO/IEC AWI TR18988 “Artificial intelligence - Application of AI technologies in health informatics”
- Investigating a double logo publication with the liaised partner IEEE EMB-SC regarding “P3191 Recommended Practice for Performance Monitoring of Machine Learning-enabled Medical Device in Clinical Use”

IEC 63450 Testing of AI/ML-enabled Medical Devices

SCOPE

(per approved new work item proposal 62/411/NP):

Establish methods for medical device manufacturers to verify and validate artificial intelligence/machine learning-enabled medical devices (AI/ML-MD), i.e. medical devices that use artificial intelligence, in part or in whole, to achieve their intended medical purpose.

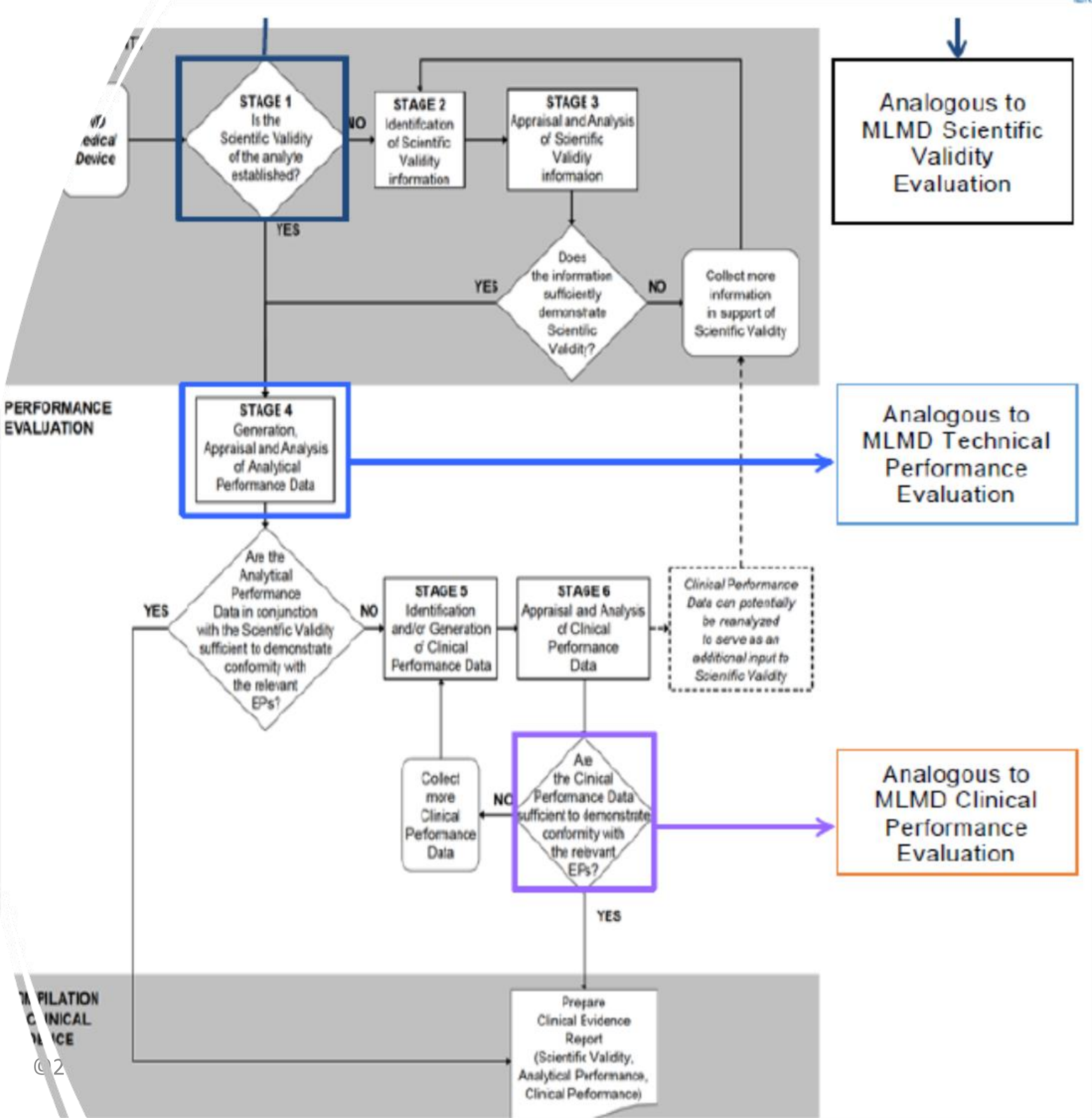
Includes verification and validation activities for the model of the artificial intelligence as well as selection, metrological characterization and management of the data sets.

Work is based on the analysis of JTC1/SC7's ISO/TR 29119-11, Software and systems engineering - Software testing - Part 11: Guidelines on the testing of AI-based systems

Expect a CD in 2024

IEC 63521 – MLMD – Performance evaluation process

- This project, unlike the ISO/IEC JTC1 SC 42 and CEN/CENELLEC JTC 21 standards, is based on the medical device sector foundational standards.
- It is based on the concept of valid clinical association, technical and clinical validation as contributing aspects of performance evaluation. Concepts already known from IMDRF SaMD N41 SaMD Clinical Evaluation or MDCG 2020-1 Guidance on Clinical Evaluation / Performance Evaluation of Medical Device Software.



edical
Device
STAGE 1
Is the Scientific Validity of the analyte established?
NO
STAGE 2
Identification of Scientific Validity information
STAGE 3
Appraisal and Analysis of Scientific Validity information
YES
Does the information sufficiently demonstrate Scientific Validity?
YES
NO
Collect more information in support of Scientific Validity
Analogous to MLMD Scientific Validity Evaluation
PERFORMANCE EVALUATION
STAGE 4
Generation, Appraisal and Analysis of Analytical Performance Data
Analogous to MLMD Technical Performance Evaluation
Are the Analytical Performance Data in conjunction with the Scientific Validity sufficient to demonstrate conformity with the relevant EPs?
YES
NO
STAGE 5
Identification and/or Generation of Clinical Performance Data
STAGE 6
Appraisal and Analysis of Clinical Performance Data
Clinical Performance Data can potentially be reanalyzed to serve as an additional input to Scientific Validity
Collect more Clinical Peformance Data
NO
Are the Clinical Performance Data sufficient to demonstrate conformity with the relevant EPs?
YES
Analogous to MLMD Clinical Performance Evaluation
FILATION
CLINICAL
CE
Prepare Clinical Evidence Report (Scientific Validity, Analytical Performance, Clinical Peformance)
©2

Previously published reference ☺

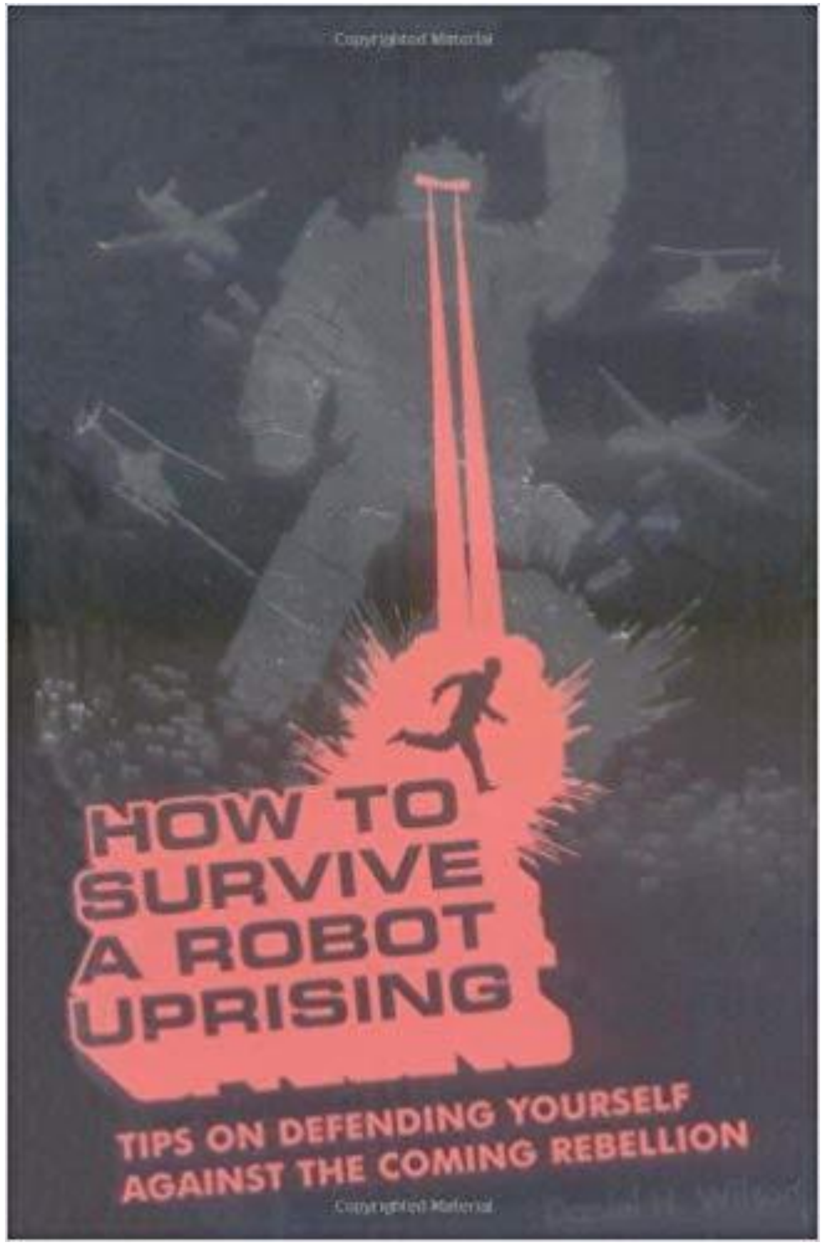
Although published in 2005, the advice is still applicable today

- Organized as a series of situations & advice on how to survive
- Humanoid robots
- Self-driving ground vehicles
- Module robots
- Smart Houses
- Thermal Imagers

It also provides great general advice

- How to treat a laser wound
- Using cybernetic implants
- Establishing a human base of operations in robot territory

Originally I bought this book as a joke (and I mention it in the presentation as a joke) – but it actually has good ideas. For example, when a swarm of robot insects is attacking you, use a fire extinguisher or a can of spray paint to disable the robots.


Copyrighted Material
HOW TO SURVIVE A ROBOT UPRISING
TIPS ON DEFENDING YOURSELF AGAINST THE COMING REBELLION
Copyrighted Material