Conducting an Effective IRB Review of Artificial Intelligence Human Subjects Research (AI HSR)

Presenter: Tamiko Eto, MS CIP

Tel: 415-635-7719

Email: Tamiko.Eto@kp.org

May 4, 2022

Conflicts of Interest Disclosures & Disclaimer

- I have no relevant personal /professional / financial relationship(s) with respect to this educational activity.

- **Disclaimer:** The information presented does not reflect my affiliation with my employer and is limited to my own opinions and personal interpretations of the regulations. This does not, in any way, serve to be a legal resource or alternative to IRB or Ethics Committee oversight. This presentation is for educational purposes only. You should consult with your organization's attorneys if you have questions or concerns about the relevant laws and regulations discussed.

Purpose

1. How AI helps and harms
2. Making AI HSR determinations
3. Using an AI HSR Reviewer Checklist
4. AI HSR involving FDA-regulated product.

What Can AI Do For Us?

Predictions

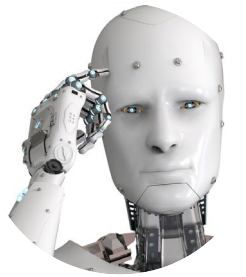
- Make risk or diagnostic predictions
- Provide insights
- Identify needs and solutions faster

Streamlining

- Drug development and target identification
- Elucidation of Drug Mechanisms of Action (MOA)
- Classify clinical documents

Automate Tasks

- Face recognition
- Product recommendations
- Virtual customer service providers
- Scheduling

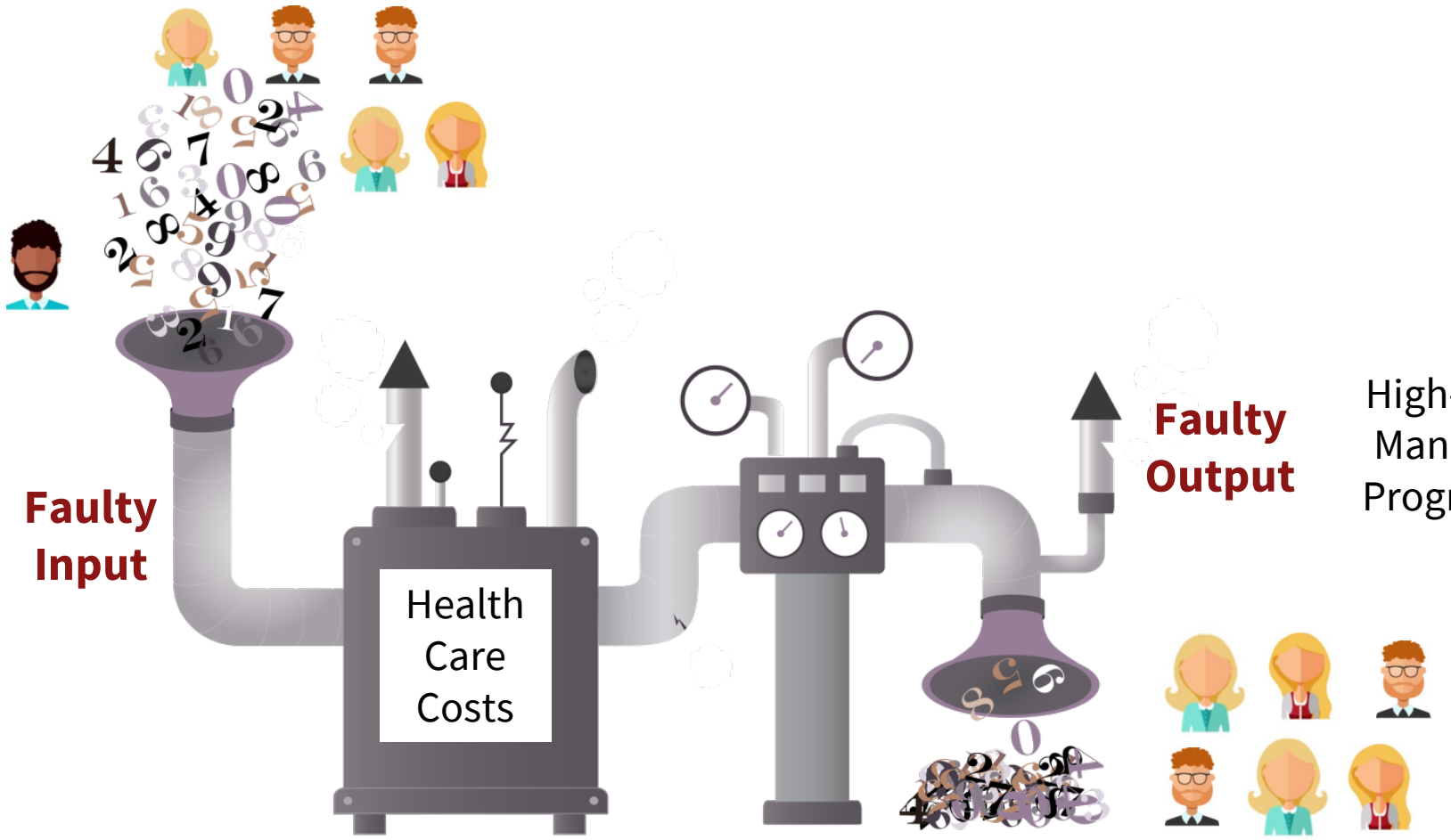


“Present-day AI is still not truly intelligent… ***because it is designed to solve the problems chosen by humans”***

-Daeyeol Lee, Birth of Intelligence: From RNA to Artificial Intelligence

G.I.G.O.
Garbage In. Garbage Out.

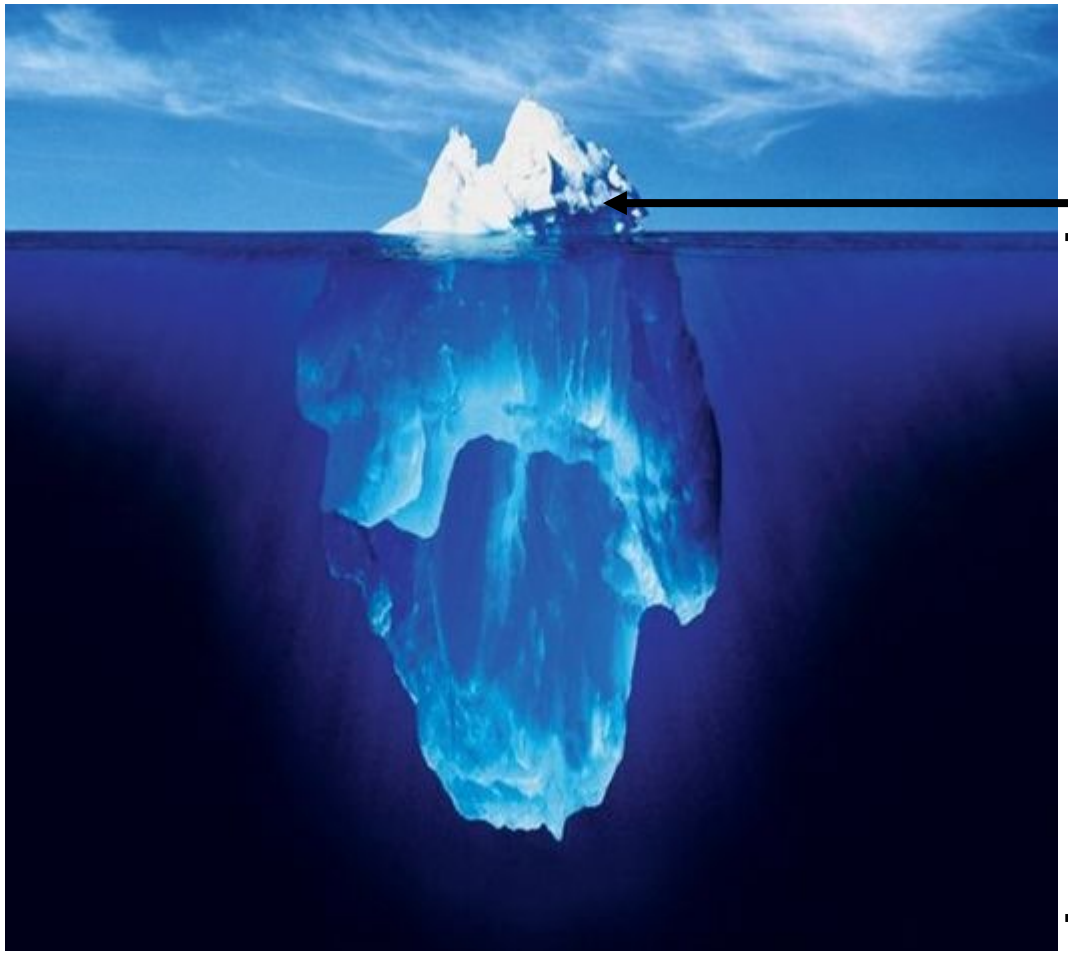
Patients in high-risk care management programs



High-risk Care Management Program Need

Examples of Promising AI

What Could Go Wrong?



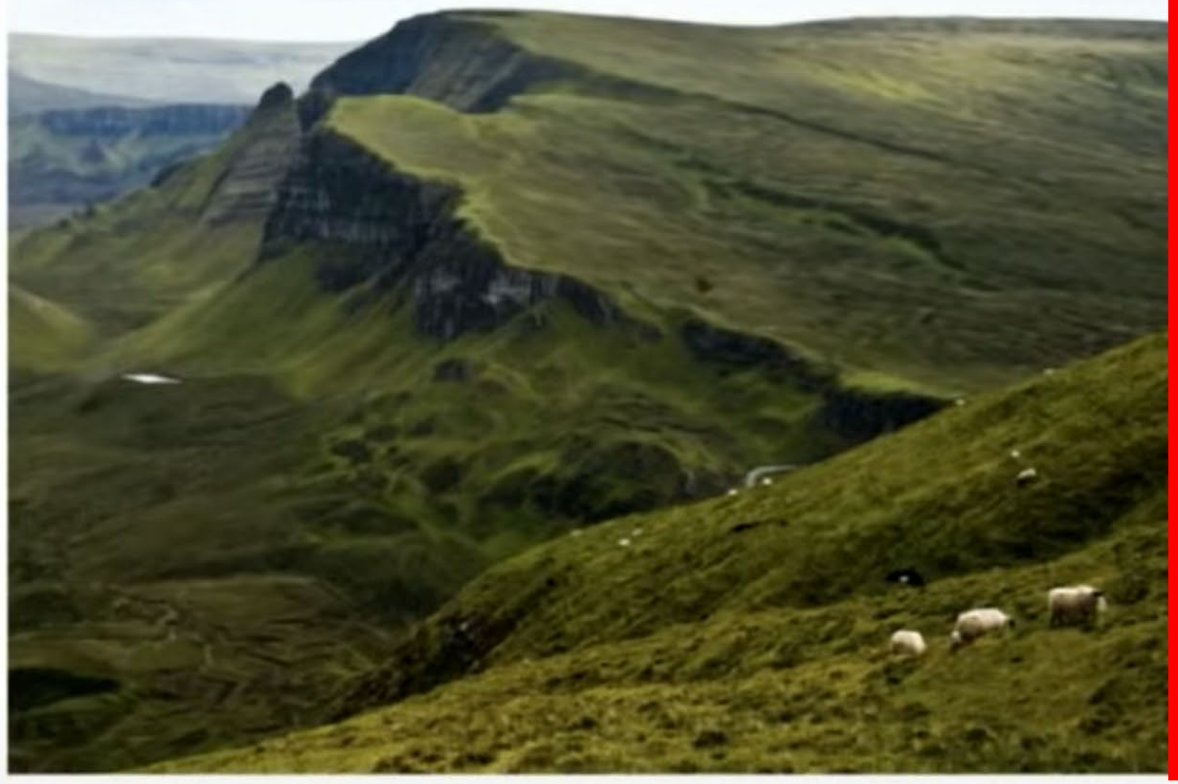
Product: 10% is what people see.

Product: 90% depends on numerous factors.

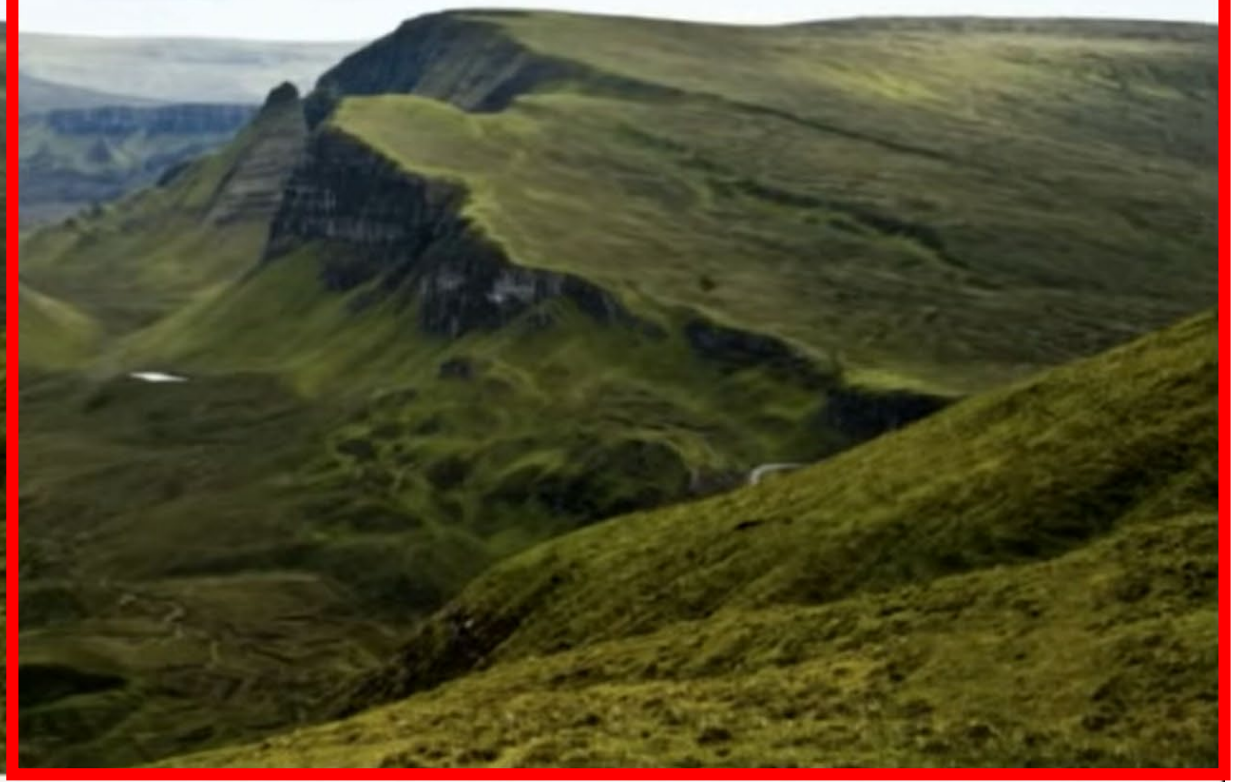
- **85%** of AI projects fail
- **87%** never make it into production
- **70%** of companies report minimal or no impact from AI (Rayome, 2019; Dilmegani, 2022)

Failed AI: Neural Net *Training Data*

A herd of sheep grazing on a lush green hillside with Quiraing in the background
Tags: mountain, grass, grazing, herd, sheep



A herd of sheep grazing on a lush green hillside with Quiraing in the background
Tags: mountain, grass, grazing, herd, sheep



Photos: Janelle Shane

azure.microsoft.com/en-us/services/cognitive-services/computer-vision/

https://tinyurl.com/sheepthread

Failed AI: *Training Data*

1 13 334

Janelle Shane @JanelleCShane · Mar 1, 2018
"A white cat is sitting on a windowsill"

1 4 59

Janelle Shane @JanelleCShane · Mar 1, 2018
"a dog and a horse are in a field"

https://www.aiweirdness.com/do-neural-nets-dream-of-electric-18-03-02/

https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/#overview

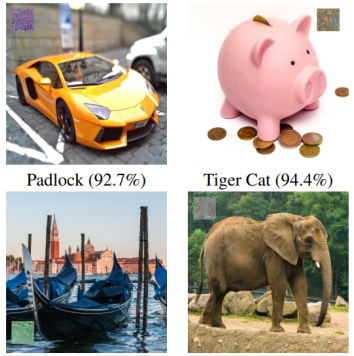
Failed AI: *"Adversarial Noise"*

LaVAN: Localized and Visible Adversarial Noise

Danny Karmon [1] Daniel Zoran [2] Yoav Goldberg [1]

Abstract

Most works on adversarial examples for deep-learning based image classifiers use noise that, while small, covers the entire image. We explore the case where the noise is allowed to be visible but confined to a small, localized patch of the image, without covering any of the main object(s) in the image. We show that it is possible to generate localized adversarial noises that cover only 2% of the pixels in the image, none of them over the main object, and that are transferable across images and locations, and successfully fool a state-of-the-art Inception v3 model with very high success rates.



Padlock (92.7%) Tiger Cat (94.4%)

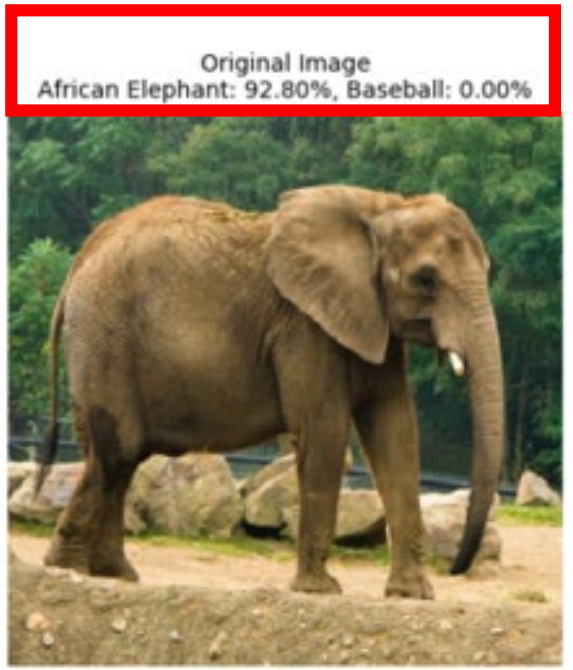
Car Mirror (94.5%) Stingray (90.5%)

1. Adversarial Noise

Deep neural-network architectures achieve remarkable results on image classification tasks. However, they are also susceptible to being fooled by adversarial examples: input instances which were modified in a particular way, and as a result, are misclassified by the network. Of course, for the adversarial example to be interesting, the change should be such that it does not confuse a human looking at the picture. Beyond the clear security implications, adversarial examples are also interesting as they may provide insights into the strengths, weaknesses, and blind-spots of these ubiquitous state-of-the-art classification models.

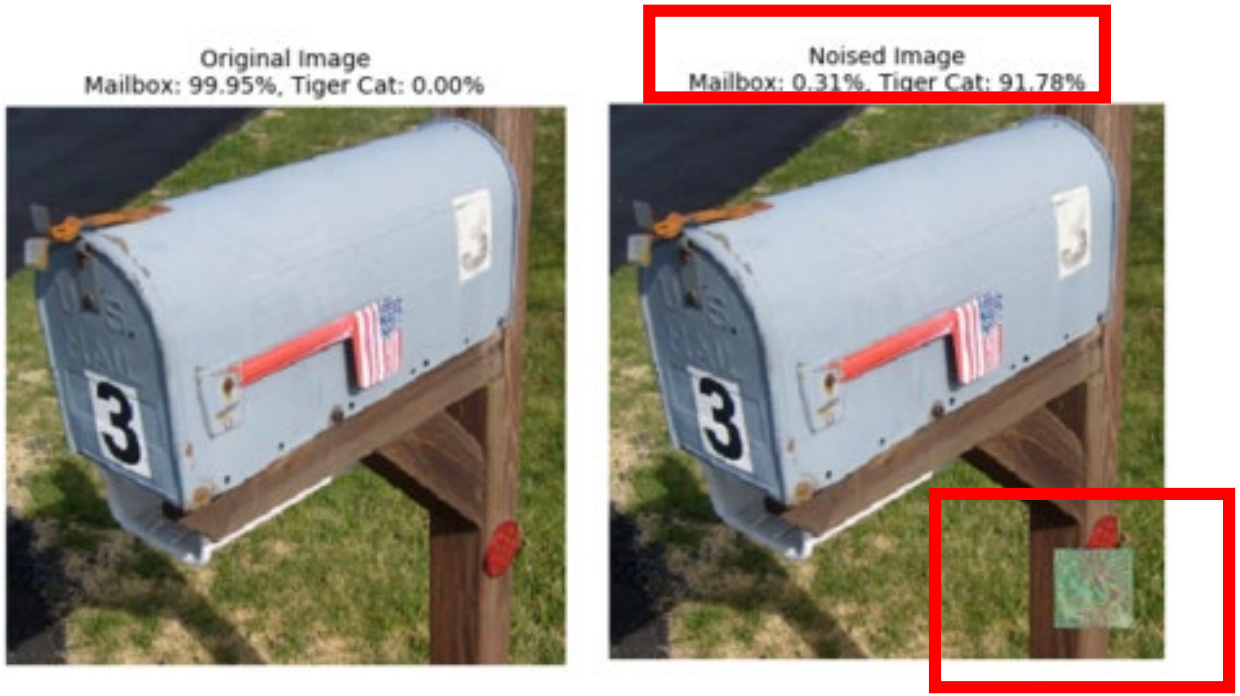
Most work on generating adversarial examples (we provide a more detailed review in section 5) focus either on noise which—while being imperceptible to humans—covers the entire image (Goodfellow et al., 2015; Szegedy et al., 2014), or on visible noise that covers prominent features of the main object in the image in a "natural" way (i.e., glasses with a specific pattern around a person's eyes in a face identification task (Sharif et al., 2016)). In contrast, we look at **visible noise** that is **localized** to a small area of the image (a bounded box with up to 2% of the pixels), and which **does not cover** the main object in the image. Figure 1 shows examples of such noised images that are misclassified by a state-of-the-art Inception V3 network with very high confidence.

A recent work by Brown et al (Brown et al., 2017) introduces a visible noise similar to ours. The works are complementary to a large extent. Their work focuses on the security implications and attempts to generate universal noise "patches" that can be physically printed and put on any image, in either a black-box (when the attacked network is unknown) or white-box (when the attacked network is known) setup. As a consequence, the resulting adversarial patches in (Brown et al., 2017) are relatively large (in a white-box setup, the generated noise has to cover about 10% of the image to be effective in about 90% of the tested conditions, and a disguised patch has to cover about 35% of the image for a similar result) and also visually resemble the target class to some extent. We do not attempt to produce a physical attack and are more interested in investigating the blind-spots of state-of-the-art image classifiers, and the kinds of noise that can cause them to misclassify.

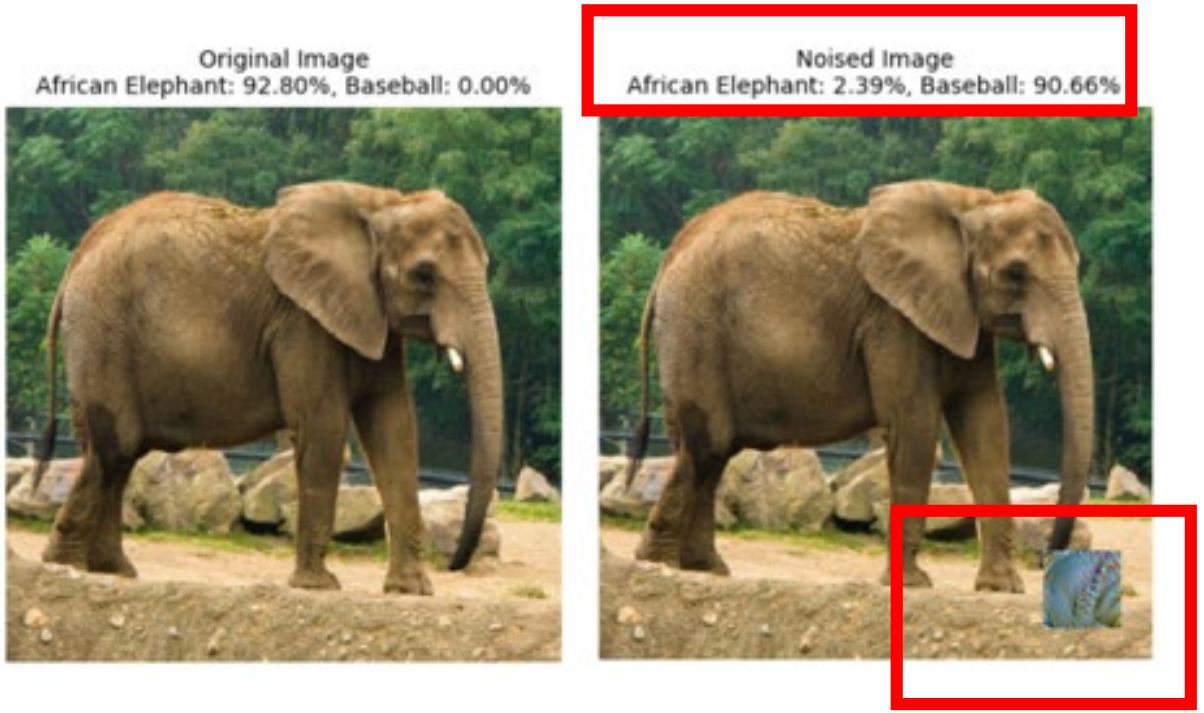


https://www.catalyzex.com/author/Daniel%20Zoran

Failed AI: *“Adversarial Noise”*



Mailbox (99.9%) → Tiger Cat (91.8%)



African-Elephant (92.8%) → Baseball (90.7%)

https://www.catalyzex.com/author/Daniel%20Zoran

Failed AI: Image Recognition "*Noisy Data*"

COMMENTARY

See related letters on pgs 2275 and 2277

Automated Classification of Skin Lesions: From Pixels to Practice

Akhila Narla[1], Brett Kuprel[2], Kavita Sarin[3], Roberto Novoa[4,5] and Justin Ko[3,5]

The letters "Interpretation of the Outputs of Deep Learning Model trained with Skin Cancer Dataset" and "Automated Dermatological Diagnosis: Hype or Reality?" highlight the opportunities, hurdles, and possible pitfalls with the development of tools that allow for automated skin lesion classification. The potential clinical impact of these advances relies on their scalability, accuracy, and generalizability across a range of diagnostic scenarios.

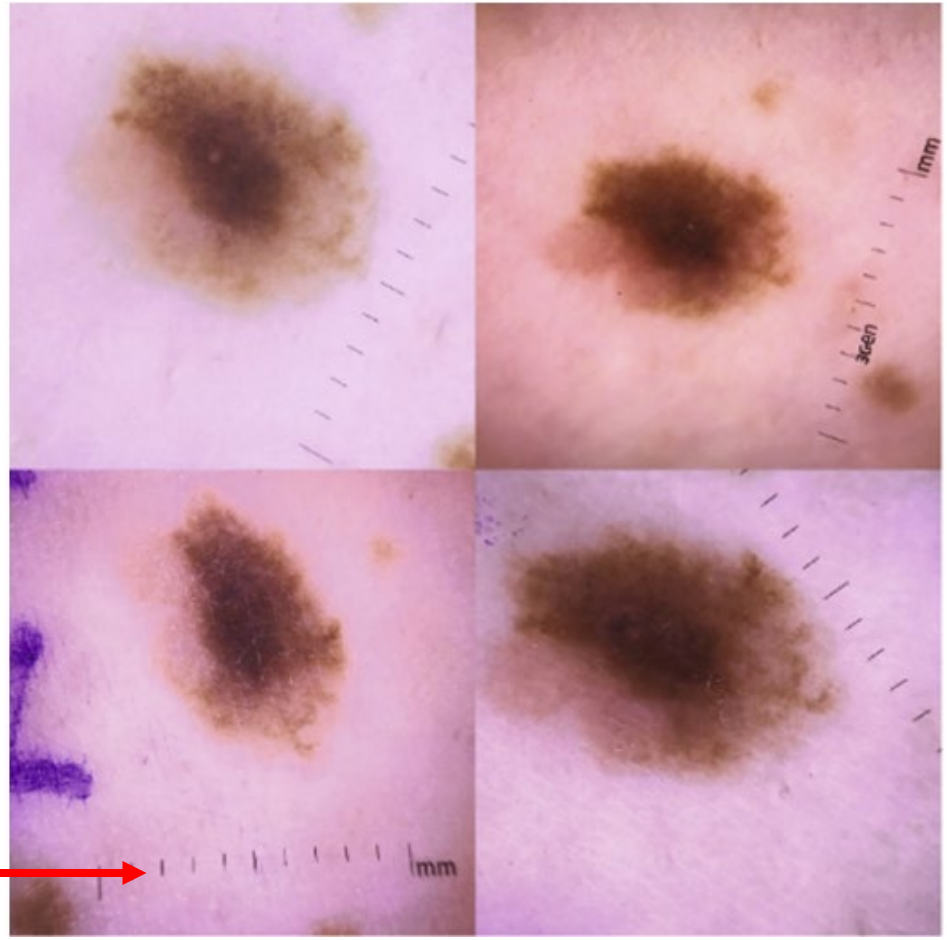
Journal of Investigative Dermatology (2018) **138,** 2108–2110. doi:10.1016/j.jid.2018.06.175

As researchers and clinicians delve into the medical applications of artificial intelligence (AI) and develop deep learning-based tools, dermatology's visually oriented tasks stand out as ripe for innovation. Both providers and patients have ready access to the tissue of interest, and with their smartphones, they possess the imaging devices needed to collect data at scale. We have seen a number of recent advances, including the work of Han et al. (2018), on the automated skin lesion classification tool, "ModelDerm." The dermatological applications of AI hold both opportunities and pitfalls as we cross from "pixels to practice," deploying these tools across diverse patient populations.

Contextual learning in lesion classification

A robust AI system of automated solitary lesion classification may be feasible for clinical integration and can augment clinical practice. However, the greatest utility would come from a one-system the body. Without multilesion change detection and classification capability, consumer-facing technology runs the risk of reassuring a hypothetical patient about the lentigo on her arm, while missing the melanoma on her leg. Lesion classification can also benefit from multimodal inputs such as age, gender, race, location on the body, or examples of other lesions on the body.

A one-system model may be capable of answering a number of clinical questions across a breadth of dermatological diseases, beyond the binary classification of benign versus malignant (Esteva et al., 2017), whereas from a logistical and usability perspective, it may be suboptimal to have a different model for each skin type or clinical classification task. Multiple models may worsen the performance of the algorithm on "edge" cases, such as patients with intermediate skin types or background skin disease (i.e., a patient with extensive psoriasis and squamous cell



Narla, A., et al. (2018). Automated Classification of Skin Lesions: From Pixels to Practice. Journal of Investigative Dermatology. Vol. 138. 10. 2108-2110

Making AI HSR Determinations

Is it AI Human Subjects Research or just a tool for research?

ROLE: AI as a Tool (Not Human-Centric)

- Form of data management
- Mining text records
- Record abstraction

Development ≠ Validation

ROLE: AI HSR (Human-Centric)

- Aim of the study is dependent upon the AI
- Testing efficacy of the AI
- Testing safety of the AI
- Testing feasibility of the AI

Is it Research?

Systematic Investigation?

1) *Having a hypothesis, or answering research question,*

   OR

2) *Research development, Testing, or Evaluation?*

Generalizable Knowledge?

- *Are you trying to contribute to the field?*
- *"research conducted with the intention of drawing conclusions that have some general applicability, or uses a commonly accepted scientific method"*
- *"when the technology developed can be applied to situations and populations beyond the current project."*

Develop a Product?

Is the study for the development of a product (even if the product will not be marketed)?

Is it "*Human Subjects*"?

Human-Centric / Human-Focused Datasets:
Datasets used or created to model human behavior, or understand humans or human conditions

NOT Human-Centric / Human-Focused Datasets:
Datasets that ***may involve human data*** but are NOT used to model human behavior, or understand humans or human conditions

Is it Human Subjects?

About whom?

Does the technology require collecting or using data (or specimens) ***from or about*** *"living" individuals*

Is it Human-Focused or Not Human-Focused?

Is the data identifiable?

Identifiable information
"…identity of subject…may be identified ***(or generated)*** *by the investigator (or a third-party).*

Is the data private?

Information about individuals' behavior occurring in a context with a ***reasonable expectation of privacy***

(e.g., medical records, school grades, personal posts or messages on social media or any other website where membership or special passwords/access privileges are required).

Level of Review

Does the study involve interventions?

Technology is used… to collect data in order to manipulate, manage, or influence a person, their environment or condition… including advising on a course of action as a result of the AI output

Direct or Indirect Interactions?

Example: *communication-only; change in treatment using predictions; or invasive procedure?*

Indirect: *person's data is used by the model ONLY.*

Direct: *person engages with AI model;*

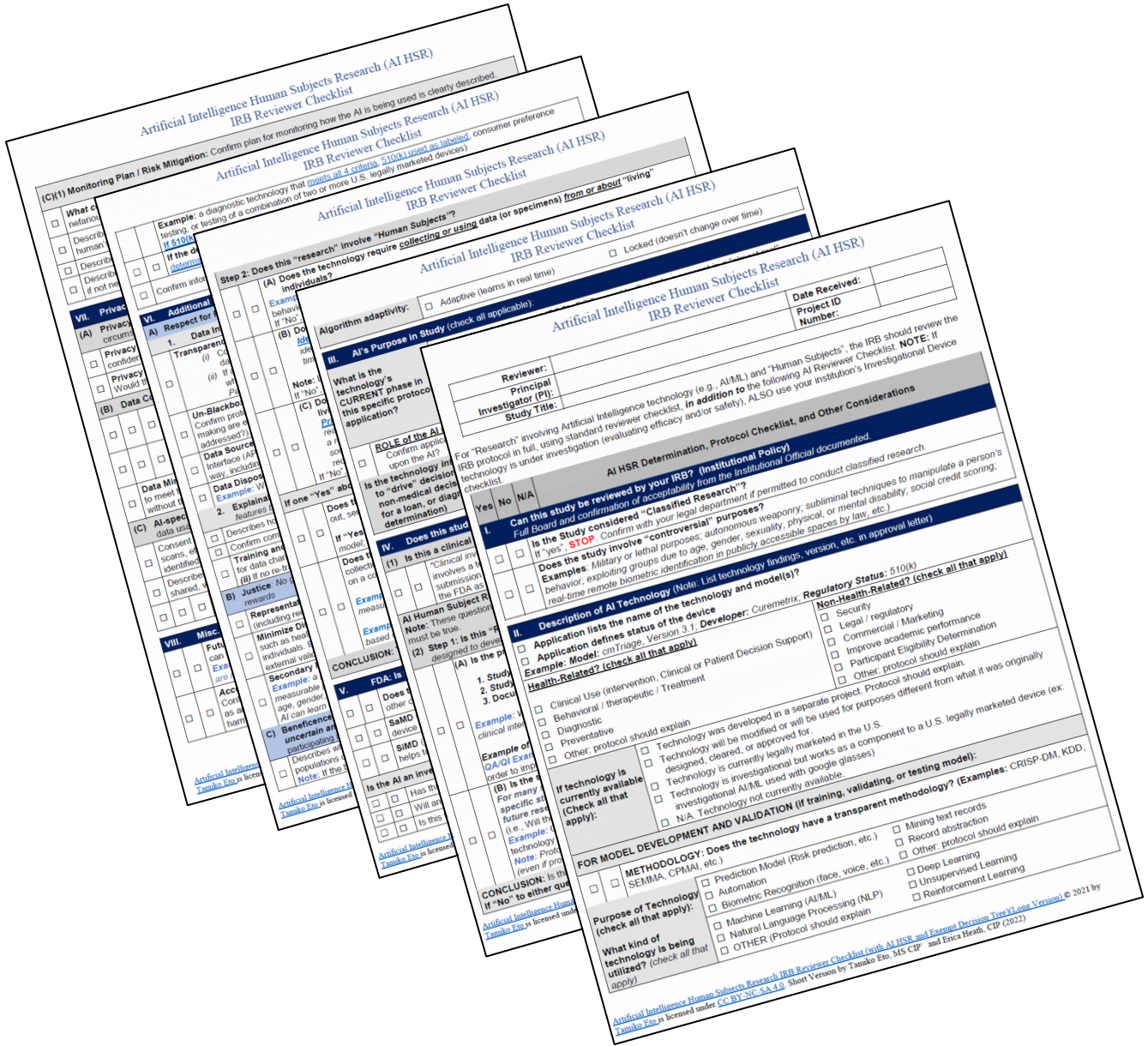
Intervention Examples

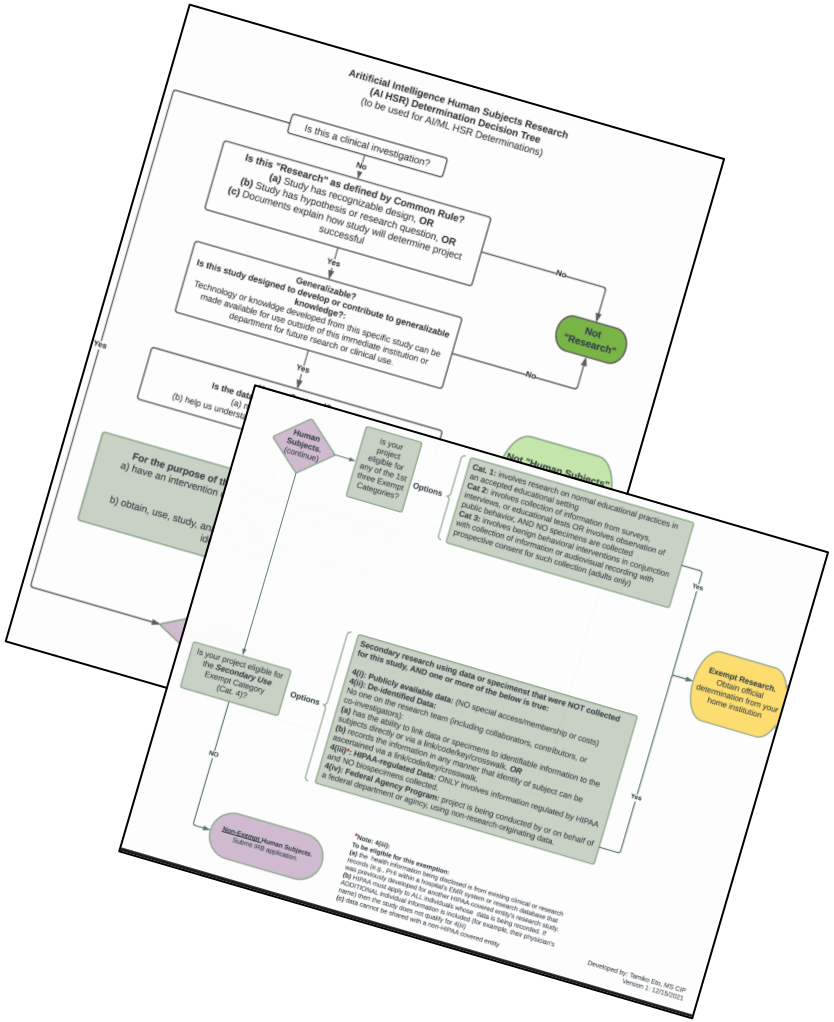
Example 1: *Prediction Model identifies someone who is at risk, informs physician, who would alter treatment based on output.*

Example 2: *Model inside a wearable device assess a person's condition and sends recommendations or alerts (action must be taken based on output of the model).*

Example 3: *Combination Product: AI-enhanced sensors on participants to obtain physiological measurements or biometric identifier.*

AI HSR Checklists & Decision Trees





PRIM&R This Month

Upcoming Events & Deadlines

Mark your calendar!

May 26

Upcoming Community Conversation: Artificial Intelligence Human Subjects Research

As artificial intelligence research continues to evolve, those working in human subjects protections are faced with new and unique challenges working under the current regulatory framework. Tamiko Eto, CIP, MS, has created an Artificial Intelligence Human Subjects Research (AI HSR) IRB Reviewer Checklist and Exempt Determinations Decision Tree that can help guide IRBs in reviewing AI research in both medical and non-medical scenarios. As part of Member Appreciation Month, Tamiko will host a Community Conversation for PRIM&R members on May 26, 3:00-4:00 PM ET where she'll walk through the checklist and facilitate discussion. Learn more

This event is open to PRIM&R members only.

"Any AI-driven software or CDSS that aims to have an impact on clinical decision making and is used as such in an existing clinical workflow fulfils the definition of software as a medical device (SaMD)"[1][2]

[1] *Becker K, Lipprandt M, Röhrig R, Neumuth T. (2019) Digital health: Software as a medical device in focus of the medical device regulation (MDR). Digit Health. 2019; 61(5–6): 211–218.*

[2] *Yaeger KA, Martini M, Yaniv G, Oermann EK, Costa AB. (2019). United States regulatory approval of medical devices and software applications enhanced by artificial intelligence. Health Policy Technology. Jun; 8(2):192-7.*

When is a project FDA-regulated?

Note: *Both DHHS (45 CFR 46) and FDA regulations (21 CFR 50, 56, 812) would be applied in IRB review*[1]

1 For drug or medical[2] devices needing a U.S. marketing permit.

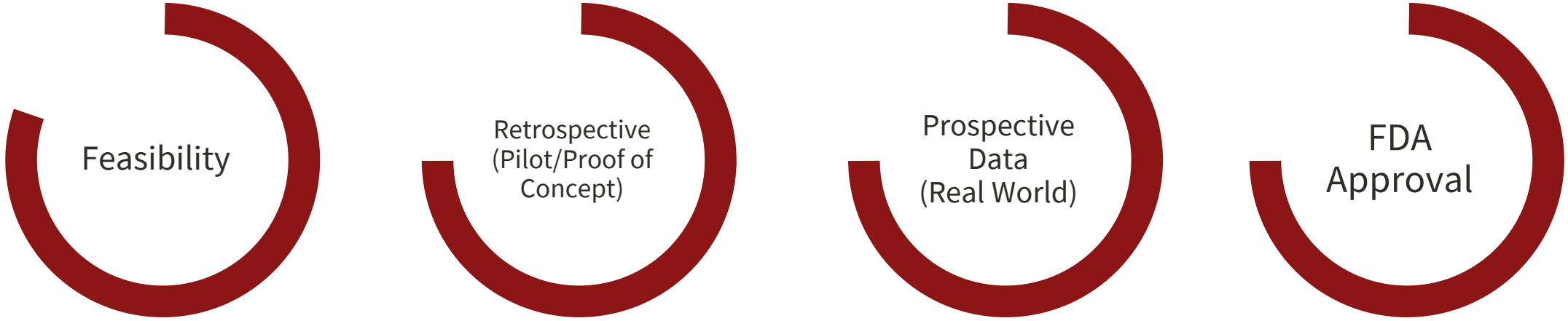
2 Drug or medical[2] device *clinical Investigations*[3]

(with or without need to market)

[1] *FDA defines "human subject" and "Research" differently from DHHS*
[2] *When software/technology is intended to cure, treat, mitigate, diagnose, or prevent a disease or other condition.*
[3] *FDA considers "research" synonymous to "clinical investigation"*

SaMD Implementation Into Clinical Workflow

Processes and Regulatory Approvals



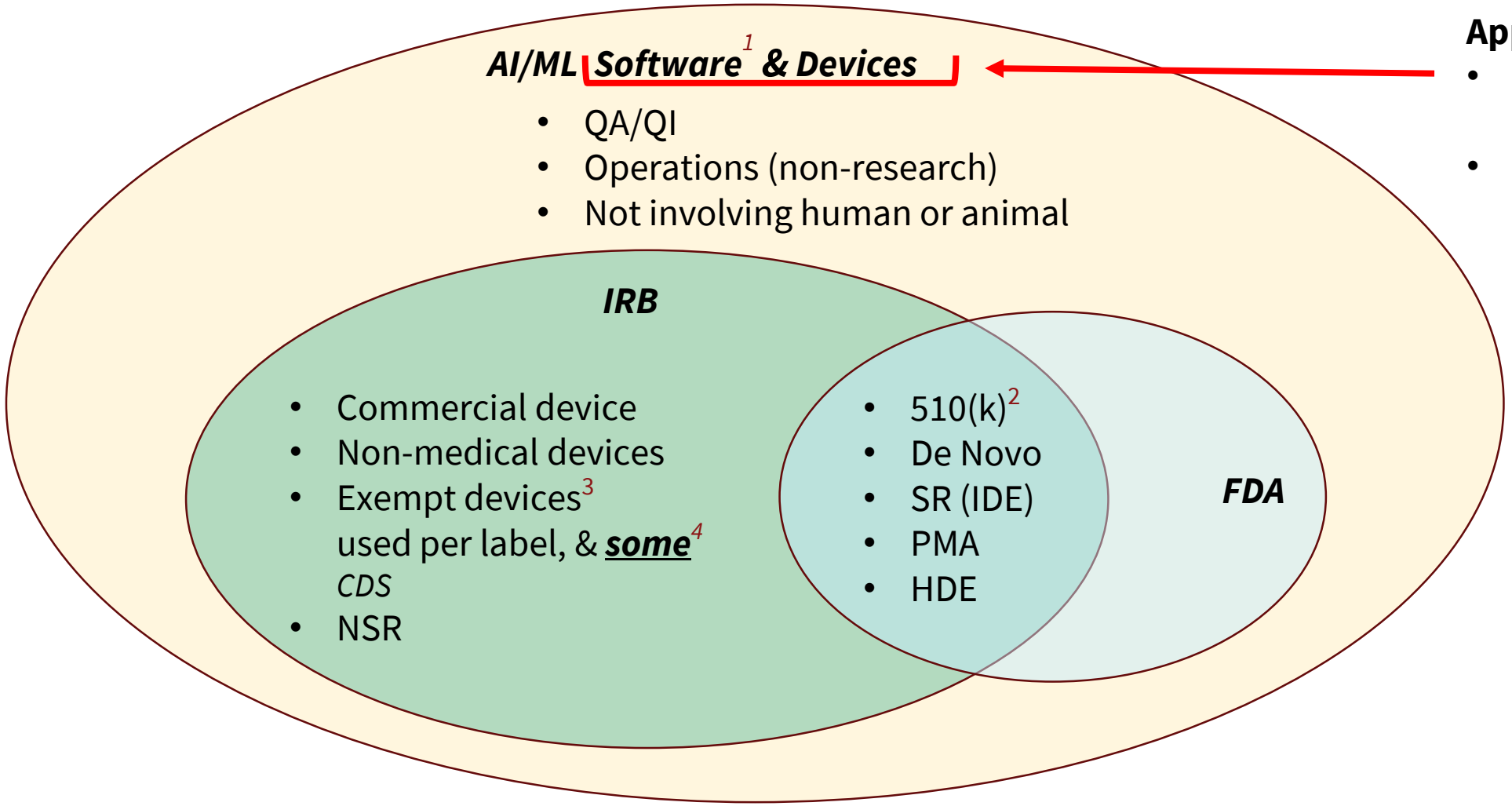
Evidence-based on Similar Systems & Literature

Training and Validation with Synthetic or Retrospective Data

Prospective Clinical Trial (RCT, multi-site)

Wide Clinical Adoption

AI-related Software & Devices and Corresponding Oversight



[1] A device software function may control a hardware device or be part of a hardware device (SaMD, or SiMD)
[2] Requires Premarket Notification
[3] IRB should review label to compare intended use in the protocol with the approved indications. Device approval indications can be found on FDA Device Approvals and Clearances databases.
[4] Low risk General Wellness (WG) products. Image processing not allowed (See Cures Act)

Non-IRB-Related (or slightly related) Considerations

Contractual Considerations

- Model ownership
- Data ownership
- Disposition of Training Data
- Comingling of Training Data
- Liability

Legal Considerations

- Avoid headlines!
- FTC actions
- Personal Injury Litigation

Bare Minimum Requirements

- Multi-site
- Prospective
- Post-market oversight
- Independent Validation
- Good Machine Learning Practices (GMLP) & Training and Education
- Clear labeling
- Non-FDA regulated AI/ML

Next Steps…

- Resources and References on following slides
- Please see Greg Manship for supplementary documents:
  - AI HSR IRB Reviewer Checklist
  - AI HSR Exempt Determination Decision Tree
  - AI HSR Human Subjects Research Decision Tree

Contact: Tamiko Eto, MS CIP

Tel: 415-635-7719 (KP Work Cell)
510-891-3529 (KP Office Phone)
Email: Tamiko.Eto@kp.org

Making AI HSR Determinations

Definitions

- **Research**
  - A systematic investigation including research development, testing, and evaluation designed to develop to contribute to generalizable knowledge
    - ***Generalizable knowledge*** is: information where the intended use of the research findings can be applied to situations and populations beyond the current project
- **Human Subjects**
  - A living individual ***about whom*** an investigator:
    - ***Obtains information*** or biospecimens *through intervention or interaction* with the individual, ***and*** uses, studies, or analyzes the information or biospecimens; ***OR*** **obtains, uses, studies, analyzes, or generates** identifiable private information or biospecimens

- **Artificial Intelligence (definition subject to change as AI evolves)**
  - An activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment (Nilsson 2010).
- **Artificial Intelligence System**
  - Software developed with one or more techniques that can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with (AI Act).

Artificial Intelligence Human Subjects Research (AI HSR)
IRB Reviewer Checklist

Reviewer:		**Date Received:**	
Principal Investigator (PI):		**Project ID Number:**	
Study Title:			

For "Research" involving Artificial Intelligence technology (e.g., AI/ML) and "Human Subjects", the IRB should review the IRB protocol in full, using standard reviewer checklist, ***in addition to*** the following AI Reviewer Checklist. **NOTE:** If technology is under investigation (evaluating efficacy and/or safety), ALSO use your institution's Investigational Device checklist.

Yes	No	N/A	AI HSR Determination, Protocol Checklist, and Other Considerations
I.			**Can this study be reviewed by your IRB? (Institutional Policy)** *Full Board and confirmation of acceptability from the Institutional Official documented.*
☐	☐		**Is the Study considered "Classified Research"?** If "yes", **STOP**. *Confirm with your legal department if permitted to conduct classified research.*
☐	☐		**Does the study involve "controversial" purposes?** **Examples**: *Military or lethal purposes; autonomous weaponry; subliminal techniques to manipulate a person's behavior; exploiting groups due to age, gender, sexuality, physical, or mental disability; social credit scoring; real-time remote biometric identification in publicly accessible spaces by law, etc.)*

II. Description of AI Technology (Note: List technology findings, version, etc. in approval letter)

☐ **Application lists the name of the technology and model(s)?**
☐ **Application defines status of the device**
Example: Model: *cmTriage, Version 3.1;* ***Developer:*** *Curemetrix;* ***Regulatory Status:*** *510(k)*

Health-Related? (check all that apply)	**Non-Health-Related? (check all that apply)**
☐ Clinical Use (intervention, Clinical or Patient Decision Support)	☐ Security
☐ Behavioral / therapeutic / Treatment	☐ Legal / regulatory
☐ Diagnostic	☐ Commercial / Marketing
☐ Preventative	☐ Improve academic performance
☐ Other: protocol should explain	☐ Participant Eligibility Determination
	☐ Other: protocol should explain

If technology is currently available (Check all that apply):	☐ Technology was developed in a separate project. Protocol should explain. ☐ Technology will be modified or will be used for purposes different from what it was originally designed, cleared, or approved for. ☐ Technology is currently legally marketed in the U.S. ☐ Technology is investigational but works as a component to a U.S. legally marketed device (ex: investigational AI/ML used with google glasses) ☐ N/A. Technology not currently available.

FOR MODEL DEVELOPMENT AND VALIDATION (if training, validating, or testing model):

☐	☐	**METHODOLOGY: Does the technology have a transparent methodology? (Examples:** CRISP-DM, KDD, SEMMA, CPMAI, etc.)

Purpose of Technology (check all that apply):	☐ Prediction Model (Risk prediction, etc.) ☐ Automation ☐ Biometric Recognition (face, voice, etc.)	☐ Mining text records ☐ Record abstraction ☐ Other: protocol should explain
What kind of technology is being utilized? *(check all that apply)*	☐ Machine Learning (AI/ML) ☐ Natural Language Processing (NLP) ☐ OTHER (Protocol should explain	☐ Deep Learning ☐ Unsupervised Learning ☐ Reinforcement Learning

Artificial Intelligence Human Subjects Research (AI HSR)
IRB Reviewer Checklist

Algorithm adaptivity:	☐ Adaptive (learns in real time) ☐ Locked (doesn't change over time)

III. AI's Purpose in Study (check all applicable):	
What is the technology's CURRENT phase in this specific protocol application?	☐ **ONLY Proof of Concept (POC):** POC meant to illustrate a concept in a "almost real" environment but does not get deployed into real-world (includes training, validation, and testing) ☐ **Pilot:** Real-world project uses technology in protected environment but NOT for use in real-world production. ☐ **Real-world Pilot:** Interventions/treatment may run in parallel with the training and re-training of model.
☐	**ROLE of the AI (in meeting the aims of the study):** Confirm application describes the portion of the project that requires AI. Is the aim of the study entirely dependent upon the AI?
Is the technology *intended to* inform or to "drive" decisions? (medical or non-medical decisions. Ex: eligibility for a loan, or diagnosis or treatment determination)	☐ **"Inform":** decision made (and confirmed) without the technology, but the technology can support the decision. ☐ **"Drive"** intended for use as an autonomous diagnostic system. May alert physician or patient of identified risk.

IV. Does this study require IRB review?		
(1) Is this a clinical investigation, as defined by FDA? If "Yes", SKIP to Section V.		
☐	☐	*"Clinical investigation"* is synonymous with *"research". "Clinical investigation"* means any experiment that involves a test article and one or more human subjects, and that either must meet the requirements for prior submission to the FDA...or the results of which are intended to be later submitted to, or held for inspection by, the FDA as part of an application for a research or marketing permit.
AI Human Subject Research (AI HSR) Determinations (2 steps) **Note:** These questions should be in your IRB checklist. They may be described differently for AI. At least one (A or B) must be true.		
(2) Step 1: Is this "Research", as defined by the Common Rule? *Research is defined as a systematic investigation designed to develop or contribute to generalizable knowledge.*		
☐	☐	**(A) Is the project a "systematic investigation"?** ***1. Study has a recognizable study design*** *(randomized, grounded theory, product validation, etc.),* ***OR*** ***2. Study has a hypothesis, or research question****?* ***OR*** ***3.* Documents explain how study will determine project was successful.** ***Example:*** *We hypothesize our model has strong diagnostic accuracy and can identify cancers missed in clinical interpretation.* ***Example of study design that may NOT constitute a systematic investigation:*** ***QA/QI Example:*** Hospitals use AI to identify hospital admission rates and wait times in an emergency room, in order to improve overall performance and/or services.
☐	☐	**(B) Is the study designed to develop or contribute to generalizable knowledge?** ***For many sites, this question may be reframed as "technology or knowledge developed from this specific study can be made available for use outside of this immediate institution or department for future research or clinical use (even if provided at no cost)".*** (i.e., Will the results be generalizable to any situation beyond the situation being studied?) ***Example:*** Obtaining new understanding about humans (to model human behavior) or developing a new technology that can be used broadly to learn more about, model, or predict human behavior. ***Note:*** *Protocol should explain if the study is intended (wholly or partially) for the development of a product (even if provided will not be marketed).*
CONCLUSION: Is this project research? (2 "yes" responses needed) **If "No" to either question, STOP. NOT "research". If "Yes" to both questions, Continue.**		

Artificial Intelligence Human Subjects Research (AI HSR)
IRB Reviewer Checklist

Step 2: Does this "research" involve *"Human Subjects"*?		
☐	☐	**(A) Does the technology require *collecting or using* data (or specimens) *from or about* "living" individuals?** **Example:** Data is "human focused" and used to either model human behavior, OR help us understand human behavior or human health conditions. If "No", STOP. Not "human subjects"; If "Yes", continue.
☐	☐	**(B) Does the study involve obtaining identifiable information *about or from* individuals?** ***Identifiable information*** *includes information about living individuals where the identity of the subject is identified or may be identified* ***(or generated)*** *by the investigator or a third-party in a reasonable amount of time through reasonable efforts.* **Note:** Limited Datasets containing health information are considered PHI and identifiable. If "No", STOP. Not "human subjects"; If "Yes", Continue.
☐	☐	**(C) Does the study involve obtaining PRIVATE information or Protected Health Information (PHI) about living individuals?** ***Private information*** *includes information about living individuals' behavior,* ***occurring in*** *a context with a reasonable expectation of privacy (e.g., activities in one's home or classroom), or information provided with a reasonable expectation of privacy (e.g., medical records, school grades, personal posts or messages on social media or any other website where membership or special passwords/access privileges are required).* If "No", STOP. Not "human subjects"; If "Yes", Continue.
If one "Yes" above, are there interactions or interventions?		
☐	☐	**Does the study involve any *interactions*** (communication, virtual, directly or indirectly; Ex: email, opt-in/opt-out, sending flyers, and/or via robots)? **If "Yes", protocol should describe AI's role in the interaction.** ***Example:*** ***Direct:*** *person engages with AI model;* ***Indirect:*** *person's data is used by the model ONLY.*
☐	☐	**Does the study involve any *interventions*?** (Includes procedures by which technology is used as a means of collecting data to manipulate, manage, or influence a person, their environment or condition, including advising on a course of action as a result of the AI output? ***Example 1:*** *Participants wear sensor, scanned by device, or perform tasks to obtain physiological measurements, or biometric identifiers.* ***Example 2:*** *Prediction Model identifies someone at risk; informs physician who would then alter treatment based on output/recommendations.*
CONCLUSION: The project is "Research" that involves "Human Subjects". Continue.		

V.		**FDA: Is the technology *possibly* regulated by FDA?** If No, **SKIP to Section VI.**
☐	☐	**Does this device meet the** definition of Medical Device**?** *"…intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals…"*
☐	☐	**SaMD** (Software *as* a Medical Device)? The software/AI/ML may be used in a medical device, but the medical device does not rely on the software to function.
☐	☐	**SiMD** (Software *in* a Medical Device)? Hardware/machine/device depends on AI to function (for example, AI helps to run a medical device; or AI is the primary way to view output)
Is the AI an investigational device? (Note: these are still subject to 21 CFR 50 & 56)		
☐	☐	Has this AI been cleared or approved by FDA for the same purpose as in this study?
☐	☐	Will any data need to be held for inspection by the FDA either now or later?
☐	☐	Is this technology exempt from the IDE requirements? (21 CFR 812.2(c))

		Example: a diagnostic technology that meets all 4 criteria, 510(k) used as labeled, consumer preference testing, or testing of a combination of two or more U.S. legally marketed devices) **If 510(k),** provide #: **Example: K123456**
☐	☐	**If the device study is NOT exempt from IDE?** If yes, technology requires the IRB to make an SR/NSR determination. Refer to your institution's SR/NSR SoP.
☐	Confirm information is included about the risks of the device as used in this study.	

VI. Additional Ethical Considerations	
A) Respect for Persons:	
	1. Data Integrity:
☐	**Transparency:** *(i)* Confirm the source and characteristics of data used to train the model are clearly explained (e.g., What datasets are going to be utilized? Will datasets be combined and why?) *(ii)* If applicable, confirm application and Informed Consent Form describe how participants will be notified when an AI product is part of their care or wellbeing, and what data that was trained on. **Note:** *If Participants will not be notified, strong justification is provided.*
☐	***Un-Blackboxing:*** Confirm protocol describes how the model(s) function; the process and role of the model's output in final decision-making are explained ***and*** (if consent is required) comprehensible to the participants (e.g., is the "black box" addressed?).
☐	**Data Source:** Protocol describes method and sources of data collection (Example: Application Programming Interface (API); scraping (automated programs to collect data, faces, voices, etc. from a website in a methodical way, including URLs) to provide access to the data of an application or operating system)
☐	**Data Disposition:** Confirm application describes what will happen to the data when this specific project is complete. **Example:** Will the model continue using the data for future training? Will the model be shared? With whom?
	2. Explainability (Human interpretability): *Confirm protocol is written so researcher can examine the input features that were most important in making the decisions it made.*
☐	Describes how they are using the best available interpretability technology.
☐	Confirm commitment to updating model as technology improves.
☐	**Training and Monitoring:** Application describes continuous training/iteration and monitoring of model (to account for data change, or model drift over time)**. Note:** ***(i)*** Model training should be done with prospective data collection. ***(ii)*** If no re-training, protocol should explain why.
B) Justice*: No group bears the burden of testing (or being the test of) new technologies while other groups reap the rewards*	
☐	**Representativeness:** Confirm the diversity in the **data source** meets the needs of the study design and procedures (including recruitment) to ensure equitable selection. Consider race, skin tone, gender, disability, etc.
☐	**Minimize Disparities:** Protocol describes how algorithmic decisions do not create discriminatory or unjust impacts, such as health disparities, when comparing data across different demographics or affected communities and individuals. **Example:** Technology generalizable to groups outside those the model was trained on; Ensuring external validation and model re-calibration prior to implementing in real-world or clinical workflow.
☐	**Secondary Participants/Incidental Participant:** Describes what features of data will be used in the final model. ***Example:*** *a project focuses on broader populations (group) characteristics or environment, but to do so, individual measurable properties and/or characteristics of a phenomenon being observed contain potential PII/PHI such as age, gender, height, weight, gait, voice or facial recognition, etc.). Project collects data on each individual so that the AI can learn how to single out "noise" or "silence" outside data.*
C) Beneficence: Do no harm; minimize harm; maximize benefit. To adequately assess the risk-to-benefit ratio ***in uncertain and non-transparent AI***, and confirm the risks of participation do not outweigh the potential benefits of participating in the study, consider the following:	
☐	Describes who will directly benefit from this technology. Describes how findings and general knowledge benefit the populations of which the data originates. **Note:** If the benefit is limited to a specific population or setting, justification is required.

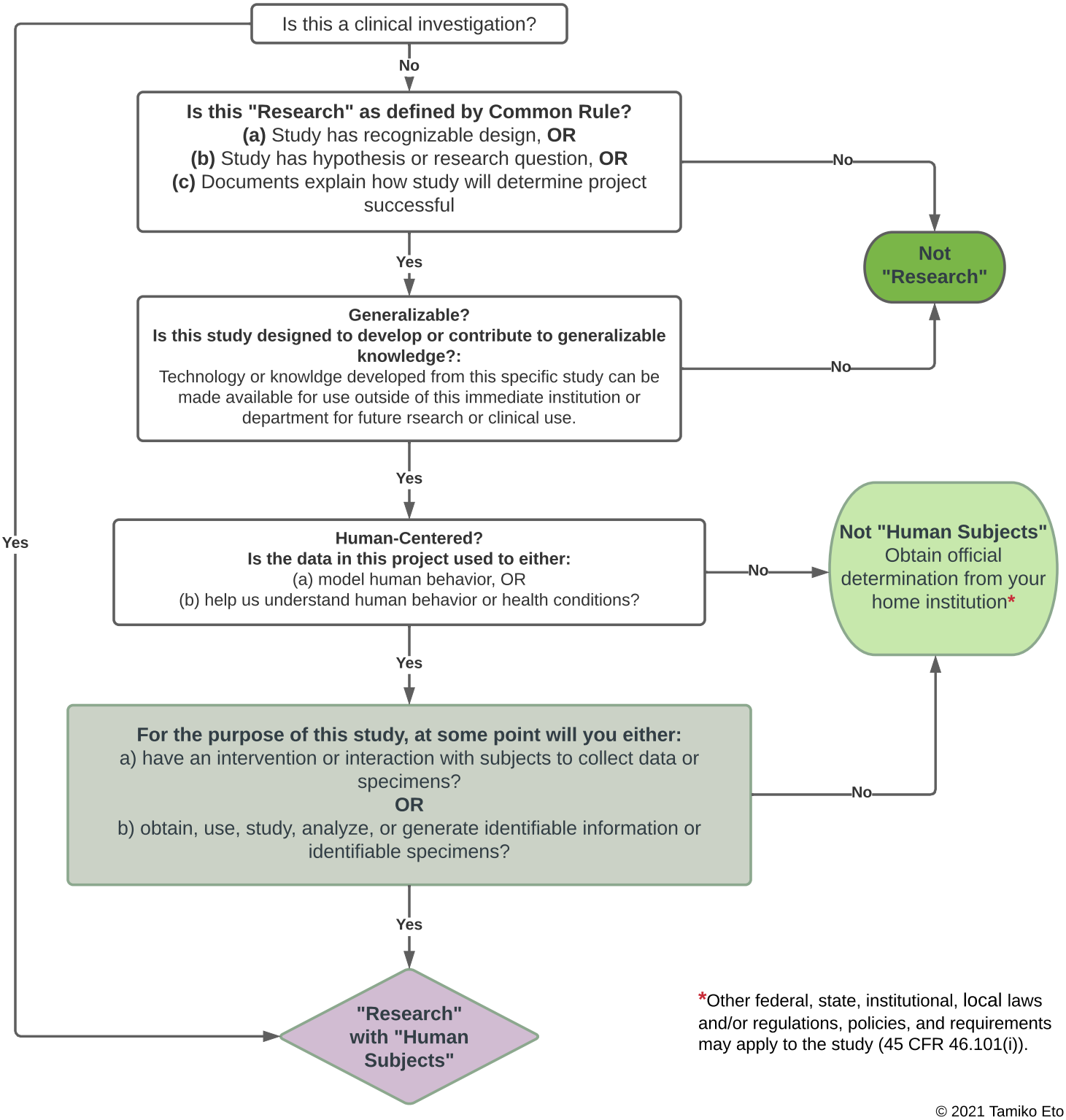
Artificial Intelligence Human Subjects Research (AI HSR)
IRB Reviewer Checklist

(C)(1) Monitoring Plan / Risk Mitigation: Confirm plan for monitoring how the AI is being used is clearly described.	
☐	***What could go wrong?*** Describes what possible mistakes it could make, be abused, or cause harm to others (e.g., nefarious use, dual use, incrimination of illegal activities, bias in algorithm, etc.)
☐	Describes possible risk(s) if any action or output is acted on autonomously, especially if such action might affect a human's health or wellbeing.
☐	Describes adequate controls in place for preventing abuse during the research, and after the research is complete.
☐	Describes iteration requirements and plans for continuous monitoring and evaluation of the data (retraining model); if not needed, PI must explain why. **Example:** the real-world environment doesn't change.

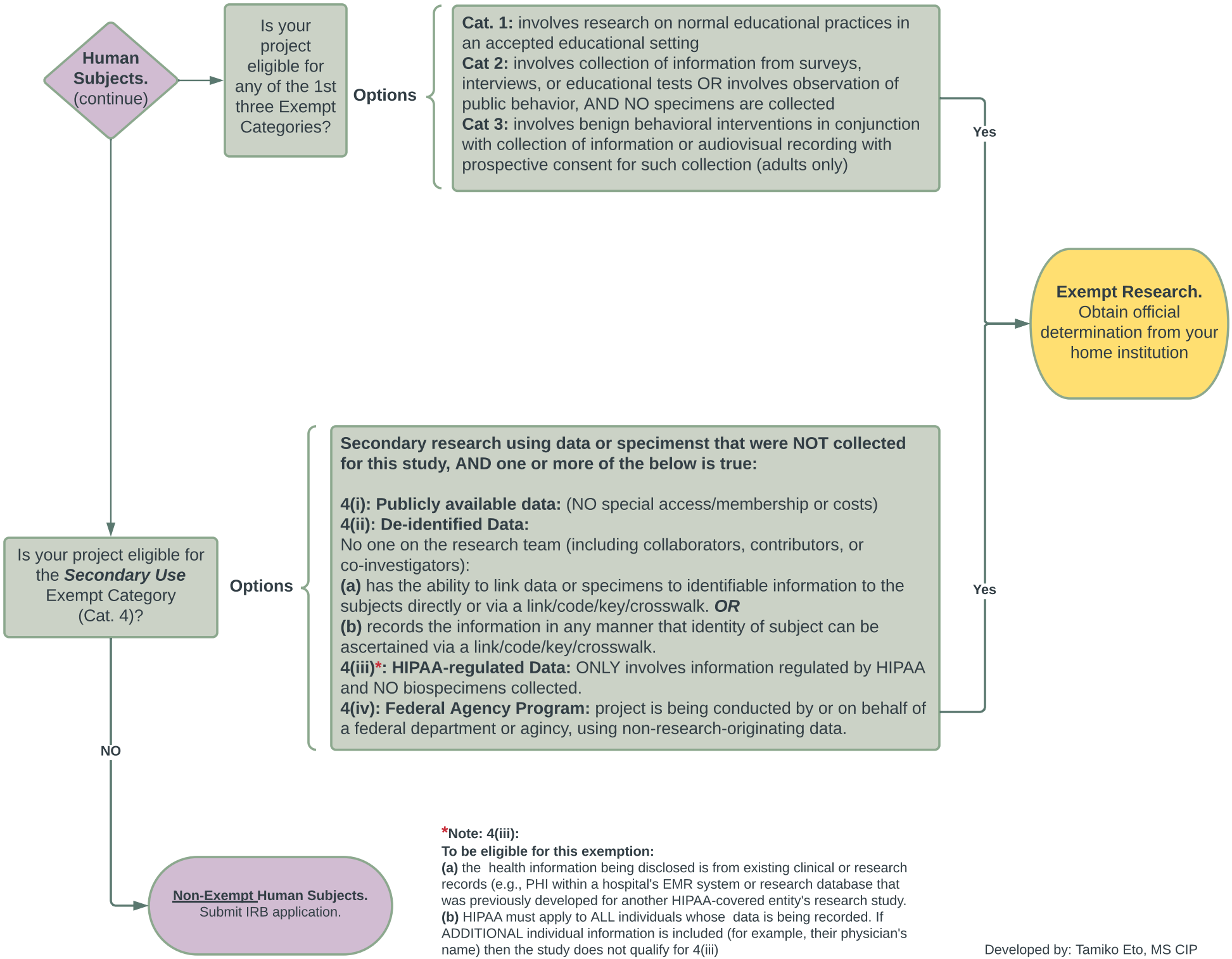
VII. Privacy & Confidentiality (45 CFR 46.111(a)(7); 21 CFR 56.111(a)(7))			
(A) Privacy: **AI-specific concerns about data use:** To what extent do the subjects have control over the circumstance around sharing oneself (and/or their data/information) with others?			
☐			**Privacy Limitations addressed?** Consent (if required), ***and*** application clearly explain limitations of privacy and confidentiality (e.g., due to utilization of external vendor services such as Google, Amazon, etc.)
☐			**Privacy Concerns addressed?** PI and IRB should consider if the subject would want this information kept private. Would they be surprised or unhappy if they found out you were using it?
(B) Data Collection & Maintenance			
☐	☐	☐	**3rd Party Data Collection or Storage?** Data use and Terms of Use/Service (ToU/ToS) requirements of third-party sources such as Facebook, Instagram, Twitter, dating websites, YouTube, LinkedIn, other social media websites, etc. have been reviewed by PI and provided to IRB for review.
☐	☐	☐	**Merging Datasets:** Consent (if applicable) ***and*** IRB application describe **(i)** if (and how) participant's data will be combined with other datasets, **(ii)** the possibility of re-identification and/or obtaining additional information, **(iii)** why this information is needed, and (**(iv)** name of additional data source(s).
☐			**Data Minimization:** Justification for each datapoint is included: only includes the bare minimum necessary in order to meet the study's purpose (absolutely necessary, and that the study goals could not practicably be achieved without that specific data).
(C) AI-specific Confidentiality Considerations: Does the researcher's plan include specific considerations for future data usage in iterative training models.			
☐			Consent (if applicable) ***and*** application describe how participant's audio/visual/biometric (voice, finger, facial, retina scans, etc.) data is used, stored (coded, transposed, etc.), shared, destroyed/not destroyed, de-identified/not de-identified, etc. ***during*** and ***after*** this specific project ends?
☐			Describes any reasonably foreseeable purposes in which participant data may be used in the future, how it will be shared, with whom it will be shared, how long it will be stored, when it will be destroyed.
☐	☐	☐	***Biometric datapoints used to determine eligibility?*** i.e., for, or access to a program, service, or opportunity, consent form (if applicable). Confirm IRB application describes those.

VIII. Misc. Considerations		
☐	☐	**Future Modifications Considerations:** Can the protocol be designed broad enough so that model changes can fit within the approved scope of the study? ***Example 1:*** *Allowing modifications to algorithm/device so long as the general procedures and design of study are not altered, and risks do not increase.*
☐	☐	***Accountability:*** Confirm protocol describes how technology is designed and implemented in publicly accountable ways, such as an obligation to report; explains and justifies specific decisions, mitigates negative impacts and potential harms.

Aritificial Intelligence Human Subjects Research
(AI HSR) Determination Decision Tree
(to be used for AI/ML HSR Determinations)



*Other federal, state, institutional, local laws and/or regulations, policies, and requirements may apply to the study (45 CFR 46.101(i)).



***Note: 4(iii):**
To be eligible for this exemption:
(a) the health information being disclosed is from existing clinical or research records (e.g., PHI within a hospital's EMR system or research database that was previously developed for another HIPAA-covered entity's research study.
(b) HIPAA must apply to ALL individuals whose data is being recorded. If ADDITIONAL individual information is included (for example, their physician's name) then the study does not qualify for 4(iii)
(c) data cannot be shared with a non-HIPAA covered entity

Developed by: Tamiko Eto, MS CIP
Version 1: 12/15/2021

Databases of Approved or Cleared AI

Database of Existing Classifications (by type of software, diagnostic, etc.)
https://www.fda.gov/medical-devices/device-software-functions-including-mobile-medical-applications/examples-device-software-functions-fda-regulates

Medical AI Evaluation Device Database: 141 FDA approved AI and How Each Device Was Evaluated
https://ericwu09.github.io/medical-ai-evaluation/

The Medical Futurist (TMF): Database of FDA-approved AI-based Algorithms
https://medicalfuturist.com/fda-approved-ai-based-algorithms/

AI Central. Data Science Institute. American College of Radiology. Detailed Database of FDA-cleared AI Medical Products
https://aicentral.acrdsi.org/

STAT Database: AI Tools (Excel Sheet)
https://www.statnews.com/wp-content/uploads/2021/02/STAT_FDA_cleared-_AI_tools.xlsx

Governmental and non-governmental database of 222 devices approved in US and Europe
Supplementary Appendix to Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (PDF-856KB)

Regulatory Support Documents

De-identification How-To:
https://fpf.org/wp-content/uploads/2017/06/FPF_Visual-Guide-to-Practical-Data-DeID.pdf

AI and data Protection Risk Mitigation and Management Toolkit
https://ico.org.uk/about-the-ico/ico-and-stakeholder-consultations/ai-and-data-protection-risk-mitigation-and-management-toolkit

Medical Device Determinations:
https://www.fda.gov/medical-devices/classify-your-medical-device/how-determine-if-your-product-medical-device

Frequently Asked Questions About Medical Devices information Sheet:
https://www.fda.gov/files/about%20fda/published/Frequently-Asked-Questions-About-Medical-Devices---Information-Sheet.pdf
SaMD:
https://www.fda.gov/medical-devices/digital-health-center-excellence/software-medical-device-samd

Device Software Functions Including Mobile Medical Applications
https://www.fda.gov/medical-devices/digital-health-center-excellence/device-software-functions-including-mobile-medical-applications

Examples of Software Function for Which FDA Will Exercise Enforcement Discretion (won't enforce requirements under FD&C Act)
https://www.fda.gov/medical-devices/device-software-functions-including-mobile-medical-applications/examples-software-functions-which-fda-will-exercise-enforcement-discretion

FDA Medical Device Data Systems (hardware or software that transfers, stores, converts formats, displays data)
https://www.fda.gov/regulatory-information/search-fda-guidance-documents/medical-device-data-systems-medical-image-storage-devices-and-medical-image-communications-devices

Examples of Mobile Apps that are NOT Medical Devices
https://www.fda.gov/medical-devices/device-software-functions-including-mobile-medical-applications/examples-mobile-apps-are-not-medical-devices

Examples of Device Software Functions FDA Regulates
https://www.fda.gov/medical-devices/device-software-functions-including-mobile-medical-applications/examples-device-software-functions-fda-regulates

Clinical Decision Support Software
https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-decision-support-software

Regulatory Support Documents, Cont'd

FDA Guidance and FAQs for 510(k)s, & When to Submit a 510(k) for Change to an Existing Device
https://www.fda.gov/media/116418/download & https://www.fda.gov/media/99812/download

FDA Guidance on Deciding When to Submit a 510(k) for a Software Change to an Existing Device
https://www.fda.gov/media/99785/download

Premarket Submissions for Device Software Functions (including SaMD and SiMD) (replaces 2005 guidance)
https://www.fda.gov/media/153781/download

Artificial Intelligence and Machine Learning in SaMD
https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device

Explaining Explainable Machine Learning (XML) – What Research Review Boards Need to Know
https://www.eventscribe.net/2021/AER-SBER21/fsPopup.asp?efp=Rk9IR1FPWFoxNDU4NA&PresentationID=922676&rnd=0.139438&mode=presinfo

For further information please visit the following FDA websites:
https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=11&showFR=1
https://www.fda.gov/regulatory-information/search-fda-guidance-documents/part-11-electronic-records-electronic-signatures-scope-and-application

Proposed and Current Policies

FDA AI/ML-Based SaMD Action Plan
https://www.fda.gov/media/145022/download

Policy for Mobile Medical Applications
https://www.fda.gov/media/80958/download

Artificial Intelligence Act (AIA)
https://artificialintelligenceact.eu/the-act/

If you think you have software that may fall under the FD&C Act, email: digitalhealth@fda.hhs.gov or DeviceDetermination@fda.hhs.gov

AI/ML Relevant Publications

- American College of Radiology and Radiology Society of North America. (2020. Letter to the FDA: Docket No. FDA-2019-N-5592. Public Workshop – Evolving Role of Artificial Intelligence in Radiological Imaging. Comments of the American College of Radiology. Accessed 3 Apr 2022. https://www.acr.org/-/media/ACR/NOINDEX/Advocacy/acr_rsna_comments_fda-ai-evolvingrole-ws_6-30-2020.pdf
- Dilmegani, C. (2022). 4 Reasons for Artificial Intelligence (AI) Project Failure in 2022. AI Multiple. Accessed 10 Apr 2022. https://research.aimultiple.com/ai-fail/
- Joseph, S. (2021). Artificial Intelligence Myth Vs Reality: Where Do Healthcare Experts Think We Stand? Accessed 3 Apr 2022. https://www.forbes.com/sites/sethjoseph/2021/09/30/artificial-intelligence-myth-vs-reality-where-do-healthcare-experts-think-we-stand/?sh=61097b0c66ba
- Kent, J. (2021). FDA Evaluations of Medical AI Devices Show Limitations. Health Analytics newsletter. Accessed 3 Apr 2022. https://healthitanalytics.com/news/fda-evaluations-of-medical-ai-devices-show-limitations
- Mandl, K. D., Szolovits, P. & Kohane, I. (2001). Public standards and patients' control: how to keep electronic medical records accessible but private. *BMJ* **322**, 283–287.
- Martinez, C., Jonker, E. (2020). A Practical Path Toward Genetic Privacy in the United States. Future of Privacy Forum. Privacy Analytics.
- Ross, Casey. (2021). Epic's AI algorithms, shielded from scrutiny by a corporate firewall, are delivering inaccurate information on seriously ill patients. Accessed 3 Apr 2022. https://www.statnews.com/2021/07/26/epic-hospital-algorithms-sepsis-investigation/
- Mueller, S. T., Hoffman, R. R., Clancey, W., Emrey, A., & Klein, G. (2019). Explanation in human-AI systems: A literature meta-review, synopsis of key ideas and publications, and bibliography for explainable AI, p. 90.
- Popescu, D.M., Shade, J.K., Lai, C. et al. (2022). Arrhythmic sudden death survival prediction using deep learning analysis of scarring in the heart. Nat Cardiovasc Res 1, 334–343.
- Strickland, Eliza. (2019). How IBM Watson Overpromised and Underdelivered on AI Healthcare. 2 Apr. Spectrum IEEE. Accessed 3 Apr 2022. https://spectrum.ieee.org/how-ibm-watson-overpromised-and-underdelivered-on-ai-health-care
- Varghese, J. et al., (2020. Artificial Intelligence in Medicine: Chances and Challenges for Wide Clinical Adoption.. Visceral Medicine. Oct 12. 36:443-449.
- Wong, A. et al., (2021). External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. JAMA Intern Med. Aug. 1. 181(8):1065-1070.
- Wu, E. et al., (2021). How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. Nature Medicine. Vol 27. 576-585.