

Artificial Intelligence (AI) Readiness at UCSF: Current State and Future Directions

Implementation and Evaluation of AI in Real-World Clinical Settings Seminar Series

January 12, 2024

Overview of Seminar Series

- **Implementation and Evaluation of AI in Real-World Clinical Settings**
- Website: <https://ai.ucsf.edu/seminar>
- Co-sponsors:
 - Department of Epidemiology and Biostatistics
 - Division of Clinical Informatics and Digital Transformation
 - Bakar Computational Health Sciences Institute
 - UCSF-UC Berkeley Computational Precision Health
 - UCSF K scholars Program
- Upcoming talks:
 - Jan – UCSF Panel on AI Readiness
 - "Feb" – Maryellen Giger, University of Chicago (ML in radiology)
 - March – Nigam Shah, Stanford University (Clinical informatics, LLMs)
 - And more to come!



Sara Murray, MD, MAS

Chief Health AI Officer

Associate CMIO, Inpatient Care

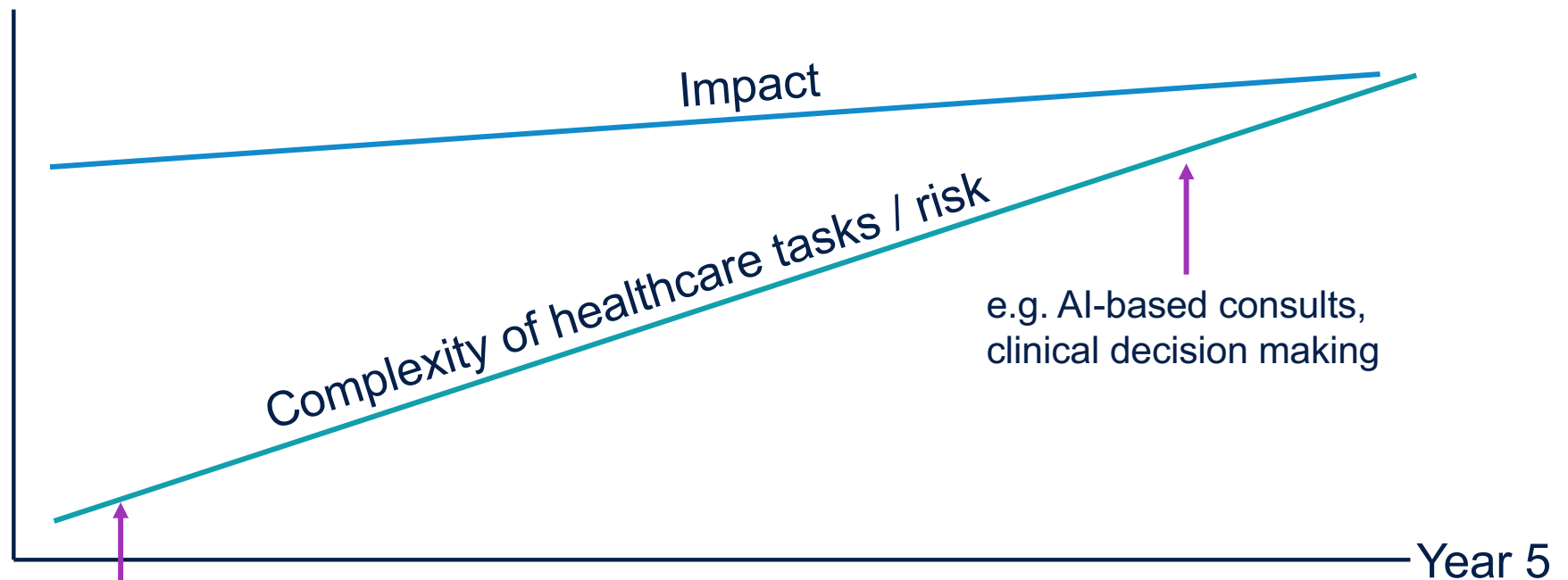
Associate Professor of Clinical Medicine, DoC-IT

Vision for the Use of AI at UCSF Health

*UCSF Health will be the national leader for the use of **trustworthy AI at scale** to address the most pressing problems in healthcare.*

- 1. AI at scale:** Enterprise-wide AI implementations with wide engagement of clinical champions
- 2. Ensuring trustworthy AI:**
 - Establish processes to ensure the best possible solutions are selected for problems that truly warrant an AI solution
 - Establish systems for longitudinal AI and impact monitoring

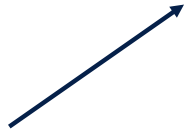
5 Year Horizon of AI at Scale at UCSF Health: High Impact, Increasing Complexity



e.g. "keyboard liberation,"
automation to support clinicians
working at top of license

e.g. AI-based consults,
clinical decision making

A Brief History of AI in the Health System



ACUTE CARE	POP HEALTH	OPERATIONS
Deterioration Index Early Detection of Sepsis Unplanned Readmission In-Hospital Fall Risk ICU Mortality (Benchmarking) ICU Length of Stay (Benchmarking) Risk of ICU Readmission or Mortality <i>Surgical Site Infection</i> * <i>Pediatric Deterioration Index</i> * <i>Acute Kidney Injury</i> * <i>Unplanned Readmission (CHF)</i> *	ED Visit or Hospitalization: <ul style="list-style-type: none">• General• Pediatric• Asthma Hospitalization for Heart Failure Negative Outcomes of Type 2 Diabetes Risk of Hypertension Risk of Myocardial Infarction Opioid Abuse Risk <i>Pediatric Asthma Exacerbation</i> * <i>End of Life Care Index</i> *	No-Show Appointments Remaining Length of Stay ED Likelihood of Admission OR Block Utilization
		SOFTWARE (included with Epic) In Basket Recipient Suggestion Patient Matching <i>Cogito Catalog Recommendation</i> * <i>Chart Review Quick Summaries</i> * <i>Orders Search Improvements</i> *

* Indicates models available in future Epic releases

Many of the Initial Epic Models Failed to Deliver

Original Investigation

June 21, 2021

External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients

Andrew Wong, MD¹; Erkin Otles, MEng^{2,3}; John P. Donnelly, PhD⁴; et al

[» Author Affiliations](#) | [Article Information](#)

JAMA Intern Med. 2021;181(8):1065-1070. doi:10.1001/jamainternmed.2021.2626

FREE

HEALTH AFFAIRS FOREFRONT

RELATED TOPICS:

ELECTRONIC MEDICAL RECORDS | ETHICS | CLINICS | PATIENT HARM | REGULATION | SYSTEMS OF CARE
| HEALTH DISPARITIES | HEALTH CARE PROVIDERS | HEALTH EQUITY | POPULATIONS

Discrimination By Artificial Intelligence In A Commercial Electronic Health Record—A Case Study

[Sara G. Murray](#), [Robert M. Wachter](#), [Russell J. Cucina](#)

JANUARY 31, 2020

10.1377/forefront.20200128.626576

Double Dice Roll Outperforms a Built-In Model for Predicting Remaining Length of Stay: Lessons Learned from a Prospective Evaluation

Hossein Soleimani; Sara G. Murray

A Brief History of AI in the Health System



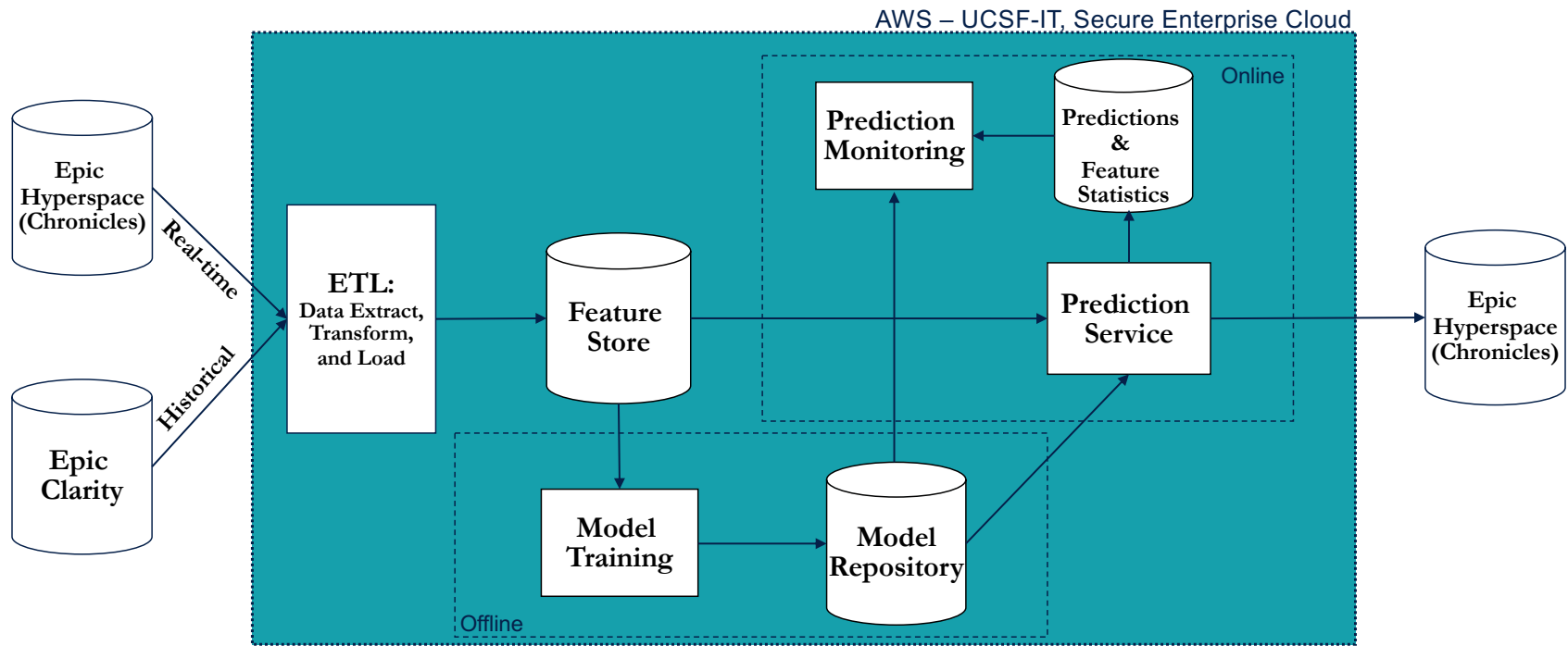
ACUTE CARE	POP HEALTH	OPERATIONS
Deterioration Index Early Detection of Sepsis Unplanned Readmission In-Hospital Fall Risk ICU Mortality (Benchmarking) ICU Length of Stay (Benchmarking) Risk of ICU Readmission or Mortality Surgical Site Infection * Pediatric Deterioration Index * Acute Kidney Injury * Unplanned Readmission (CHF) *	ED Visit or Hospitalization: <ul style="list-style-type: none"> • General • Pediatric • Asthma Hospitalization for Heart Failure Negative Outcomes of Type 2 Diabetes Risk of Hypertension Risk of Myocardial Infarction Opioid Abuse Risk Pediatric Asthma Exacerbation * End of Life Care Index *	No-Show Appointments Remaining Length of Stay ED Likelihood of Admission OR Block Utilization
SOFTWARE (included with Epic)		
In Basket Recipient Suggestion Patient Matching Cogito Catalog Recommendation * Chart Review Quick Summaries * Orders Search Improvements *		

* Indicates models available in future Epic releases

Very Limited

Deployment of research or custom health models
 Ex. Vital sign de-escalation, delirium, post-op LOS

Health IT Platform for Advanced Computing (HIPAC)



Increasing Focus on AI Fairness in Healthcare

Attorney General Bonta Launches Inquiry into Racial and Ethnic Bias in Healthcare Algorithms

Press Release / Attorney General Bonta Launches Inquiry into Racial and Ethn...

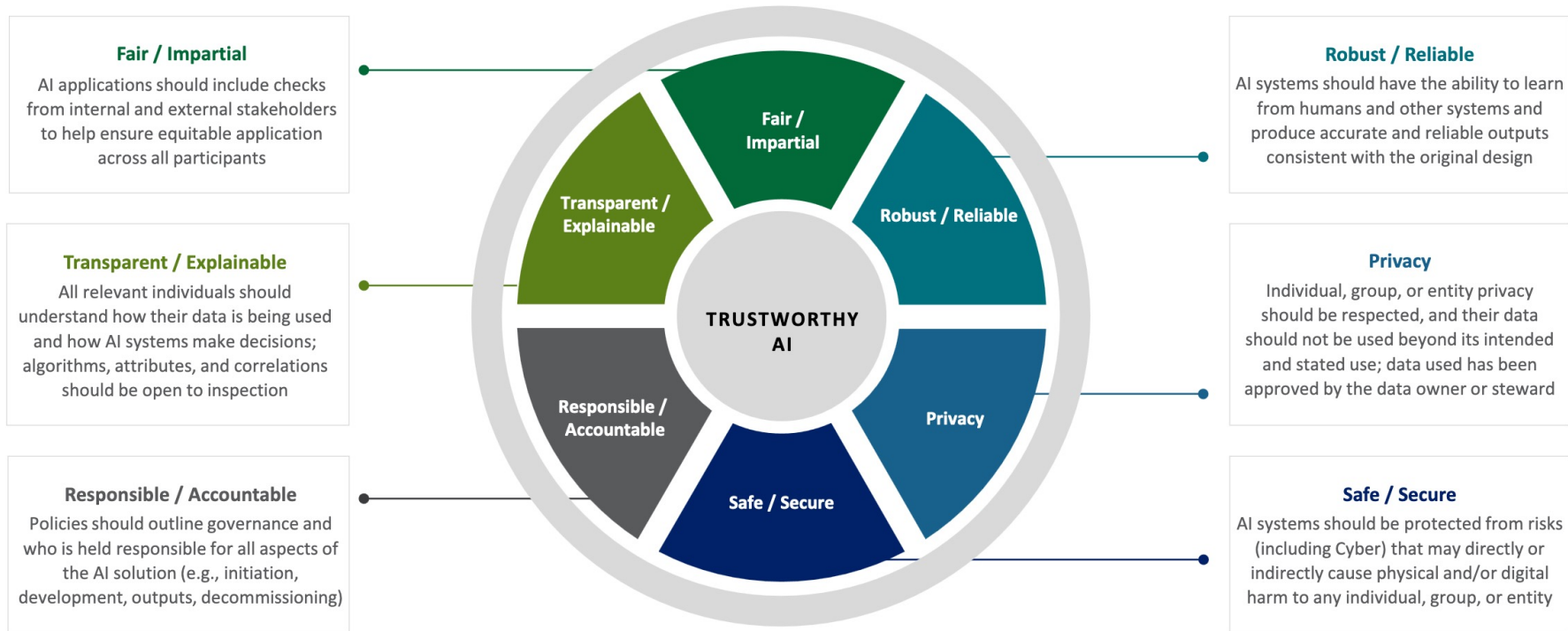


Wednesday, August 31, 2022

Contact: (916) 210-6000, agpressooffice@doj.ca.gov

Sends letters to 30 hospital CEOs across the state requesting information regarding the use of commercial healthcare decision-making tools

UCSF Health Adopted the HHS Trustworthy AI Guidelines



TAI principles are not mutually exclusive, and tradeoffs often exist when applying them.

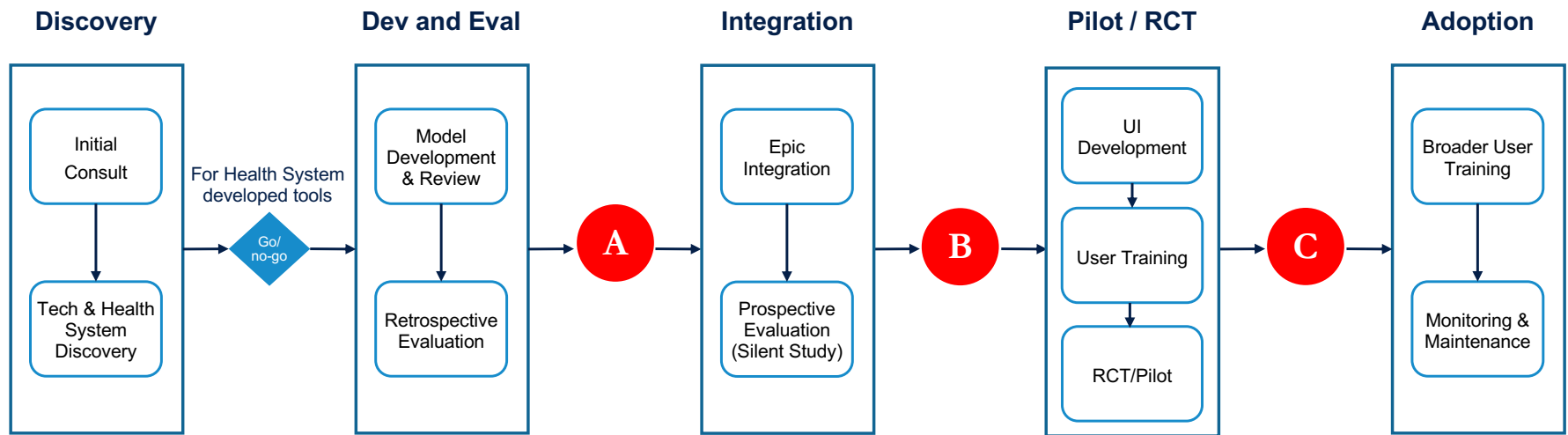
AI Governance Committee for UCSF Health

To guide UCSF in the evaluation and deployment of trustworthy and ethical Artificial Intelligence (AI) and Machine Learning (ML) tools in support of improved care delivery and health system operations, and ensure that all models that touch healthcare delivery are deployed in a responsible fashion

Unified Governance Regardless of AI Origin:

- Health System Development / Partnership
- Research / Academic Community
- Procured from Vendors

AI Governance at UCSF Health



Challenges of Effective Governance

- Backlog of embedded AI tools needing review, most were acquired in a decentralized way (divisions, departments)
- Due to extensive time required for review, if we're going to do this right, the oversight committee warrants increased resourcing (similar to an IRB)
- Expansion of ethical framework as we evaluate generative AI tools

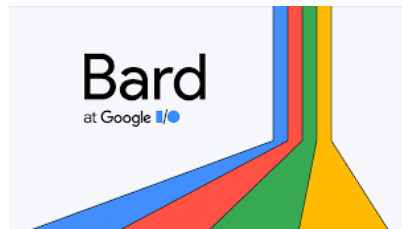
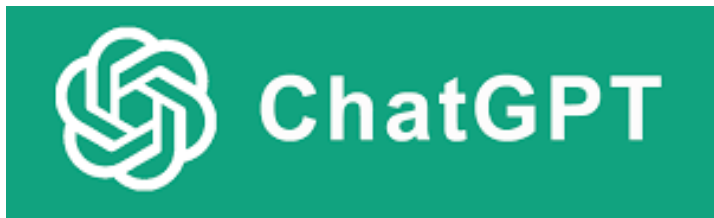


GORDON AND BETTY
MOORE
FOUNDATION



UNIVERSITY OF CALIFORNIA Office of the President

A Paradigm Shift for AI in Healthcare



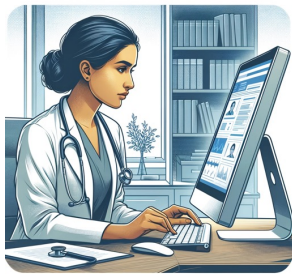
Engine Data Source Large Language Model

Azure OpenAI OpenAI Web GPT-3.5 Turbo (16k)

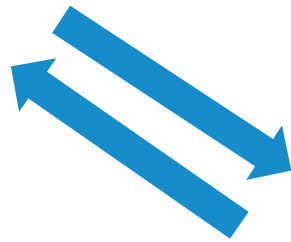
 Greetings Sara! I'm Versa, how may I assist you today?



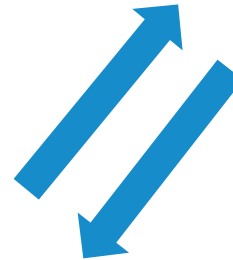
HIPAC + Versa in the health system



EHR



HIPAC



- 3 instances of GPT-4
- designed to support other LLMs

Initial Principles for Selecting Generative AI Tools

- High-impact potential over 1-2 years
- Adds value across many clinical settings
- Lower-risk
 - Generative AI tools remain largely experimental
 - Mistakes should be easily detected and corrected by humans
- Flexible and modifiable in response to feedback
- Measurable Key Performance Indicators



Mark Pletcher, MD MPH

Chair, Dept of Epidemiology and Biostatistics

Director, CTSI Informatics and Research Innovation Program

Need for clinical evaluation

- HIPAC prospective algorithm evaluation is first step
 - Does the AI predict like it should? As well as in retrospective analysis?
 - Does it predict differently for different sets of patients?
- But still need clinical evaluation
 - What happens when you deliver an AI algorithm's results to clinicians?
 - Will it result in benefit or harm? Will clinicians use it? Will it make their lives better? What will happen to costs?
- Evaluation after AI implementation is critical!

Need for clinical evaluation

- CTSI Informatics Program specializes in modifying APeX to improve healthcare delivery
 - APeX-Enabled Research (AER) Program
 - AI Pilots, Learning Health System Demonstration Projects
- We run RCTs of EHR interventions to make sure they are improving outcomes
 - LHS Oversight Committee
 - Unintended consequences are the rule, not the exception
 - Demonstration projects have supported evaluation and A/B testing, and *often result in de-implementation*

Need for clinical evaluation

- AI can generate amazing predictions and content...but:
 - How do you design delivery of that content to clinicians?
 - What do you nudge clinicians to do?
 - Will clinicians do what the AI recommends?
 - Will bias in the algorithm lead to worsening inequities?

Need for clinical evaluation

○ My recommendations:

- 1) Equity-focused approach to AI implementation
 - We cannot get the bias out of the AI.
 - But we can design the way we USE the results of an algorithm to deliberately counteract AI bias
- 2) Use modeling studies to design for equity
 - Microsimulation modeling of our health system population
- 3) Use research methods to evaluate impact
 - Define outcomes we care about and measure them over time
 - Interrupted time series analyses
 - Randomized trials



Ida Sim, MD, PhD

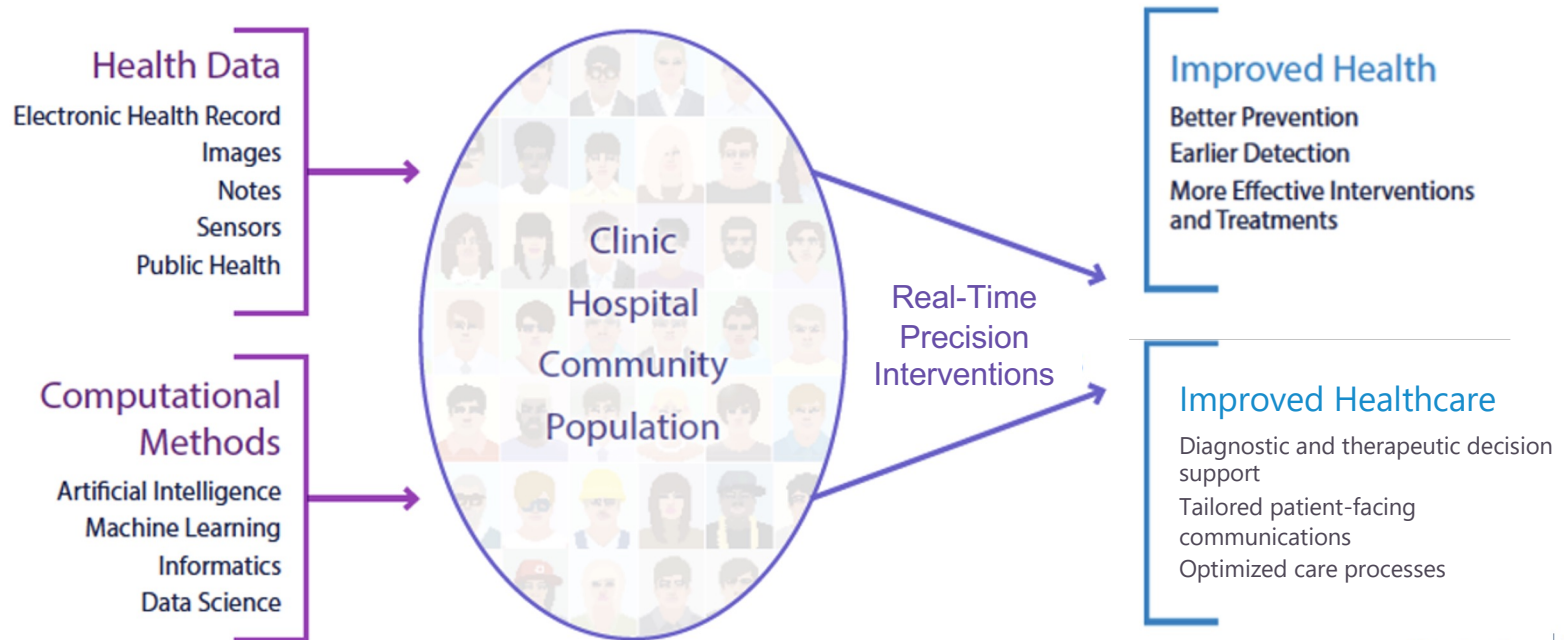
Professor of Medicine and Computational Precision Health

Co-Director, UCSF UC Berkeley Joint Program in
Computational Precision Health

UCSF Chief Research Informatics Officer

What is Computational Precision Health?

A new discipline at the intersection of machine learning, statistics, medicine, and population health



Unique Bi-Campus Department-Equivalent

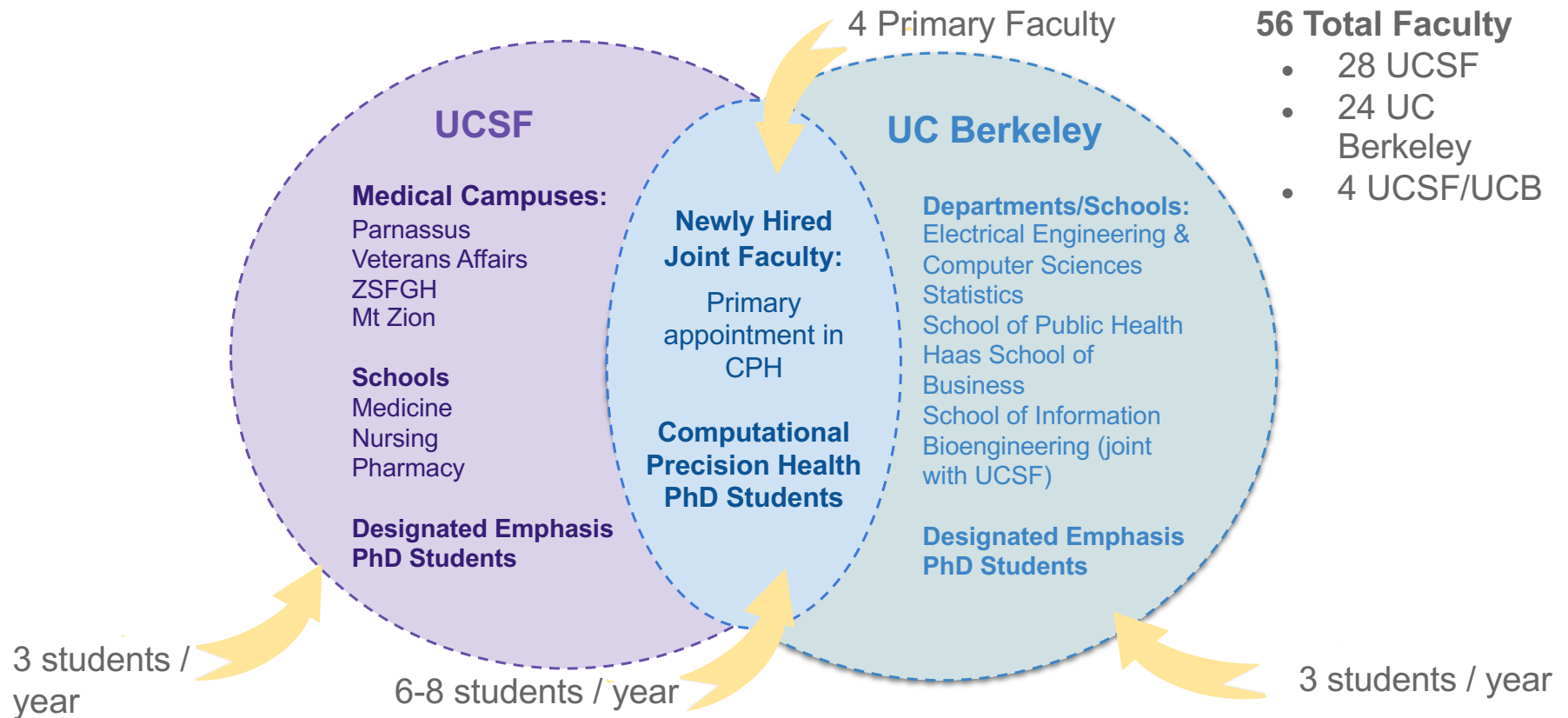
Singular entity spanning UCSF and UC Berkeley with one interwoven agenda, building a new interdisciplinary culture



**Clinical and population health sciences,
clinical care**



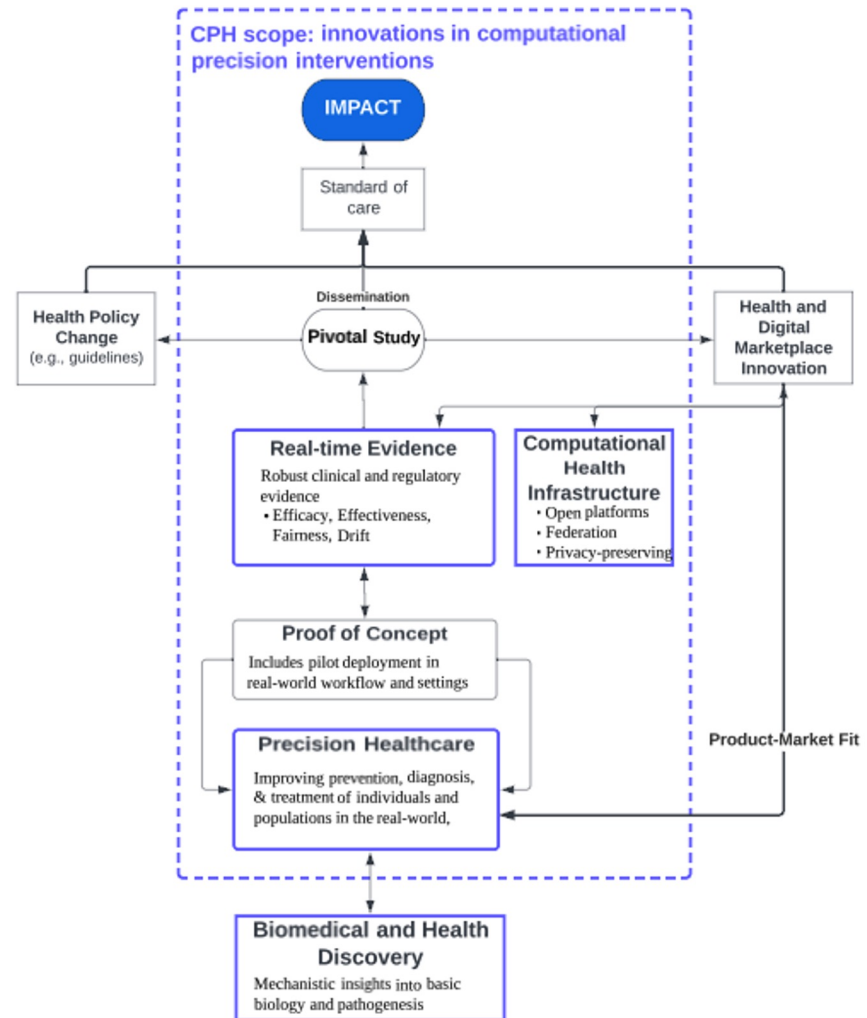
**Computer science, statistics,
public health**




UCSF Health | UC Health | Partner care settings

Berkeley | UCSF

CPH Innovation to Impact Pipeline



UCSF



Please use this form to register your interest in using Versa.

Versa is UCSF's secure generative AI platform. It is now available in limited release featuring a chat user-interface and API gateway.


Please indicate which tool(s) you would like to request access to (you can choose both):

- Versa chat - the web user-interface for chatting with generative-AI tools.
- Versa API - the programmatic access tool for interacting with generative-AI model via API



Engine: Azure OpenAI | Data Source: OpenAI Web | Large Language Model: GPT-3.5 Turbo (16k) | Sample Questions: Tell me about UCSF



 Greetings Ida! I'm Versa, how may I assist you today?

Enter your question



Capabilities

- ✓ Comply with HIPAA
- ✓ Keep prompts and data within UCSF's secure enterprise cloud
- ✓ Support multiple LLM providers, with OpenAI being the first (AWS Bedrock and Google Bard pending): GPT-3.5 Turbo and GPT-4 available, Llama2 pending
- ✓ Serve all mission areas
- ✓ Integrate with our enterprise data warehouse
- ✓ Two products **Versa API** and **Versa Chat** to cater to programmers and non-programmers, respectively.
 - ✓ RAG Q1 2024
 - ✓ Fine-tuning Q2 2024

Versa Chat Vectorization Example

University of California San Francisco Logout

UCSF Versa

Engine: Azure OpenAI | Data Source: OpenAI Web | Large Language Model: GPT-3.5 Turbo (16k)

Greetings Ki! I'm Versa, how may I assist you today?

How do I get pet insurance for my dog?

Getting pet insurance for your dog is a responsible decision to help protect your furry friend's health and provide financial assistance in case of unexpected medical expenses. Here are the general steps to obtain pet insurance:

1. Research different pet insurance providers: Look for reputable insurance companies that offer pet insurance in your area. Compare their coverage options, premiums, deductibles, and customer reviews to find the best fit for your needs.
2. Understand the coverage options: Read the policy details carefully to understand what's covered and what's not. Most pet insurance plans cover accidents and illnesses, but

Answer about pet insurance according to publicly available information

University of California San Francisco Logout

UCSF Versa

Engine: Azure OpenAI | Data Source: UCSF Intranet: HR (Prototype) | Large Language Model: GPT-3.5 Turbo (16k)

Greetings Ki! I'm Versa, how may I assist you today?

How do I get pet insurance for my dog?

To get pet insurance for your dog, you can enroll in coverage through Nationwide. You can visit the Nationwide website at petinsurance.com/uc to learn more and enroll online. Alternatively, you can call Nationwide at 877-738-7874 to speak to a representative and enroll over the phone. Once enrolled, your coverage will be effective approximately 14 days after your application is approved.

References are as below:
<https://llmstudiodev.azurewebsites.net/documents/HR/welcome-kit.pdf>

Same question but different answer because it is in the context of our HR content such as benefits and leave of absence

Answer is sourced from the new hire welcome kit, which is loaded into Versa Chat via vectorization

Versa users exploring use cases across themes

Information Extraction

- Extracting radiological findings from radiology reports and cancer staging from notes.
- Validating against data that has been annotated through manual chart reviews.
- Identifying adverse drug reactions from clinical notes.

Knowledge Management

- Simplifying nursing training materials into short, customized summaries for different job roles and situations.
- Making HR policies and guidelines easily accessible to employees through chat, saving them from long reading sessions.
- Loading verified medical literature into Versa and leveraging the repository as a trusted source for finding new knowledge.

Writing and Coding Assistance

- Generating content for research proposals, workshops, and interviews.
- Synthesizing scientific articles, assisting with literature reviews, and helping with presentation outlines and writing.
- Crafting easy-to-understand summaries for various audiences, such as patients, providers, and researchers.
- Providing commands for statistical analyses and assisting with coding tasks.

Responsible Use of LLMs

Coming February 2024!

- **New ai.ucsf.edu**
- **Versa Education Modules**
 - **For Versa ChatGPT users:**
 - Overview of LLMs
 - Basics of Prompt Engineering
 - Using Versa
 - **For Versa API users:**
 - Pointers to technical documentation and code
 - Best practices for LLM developers
 - Starting February, Versa users will have to complete SumTotal modules to gain Versa access
- **Active conversations with IRB on guidance**

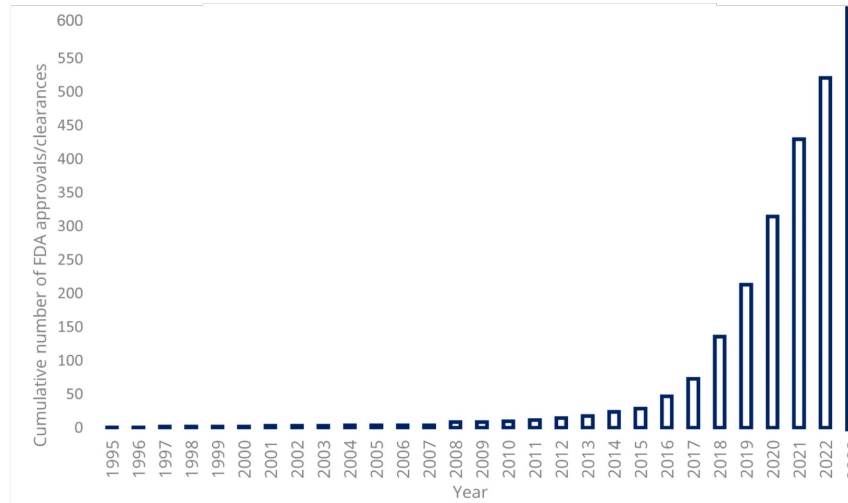


Jean Feng, PhD, MS

Assistant Professor

Department of Epidemiology and Biostatistics

FDA Approval of AI algorithms



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

Figure 1: SaMD IMDRF risk categorization

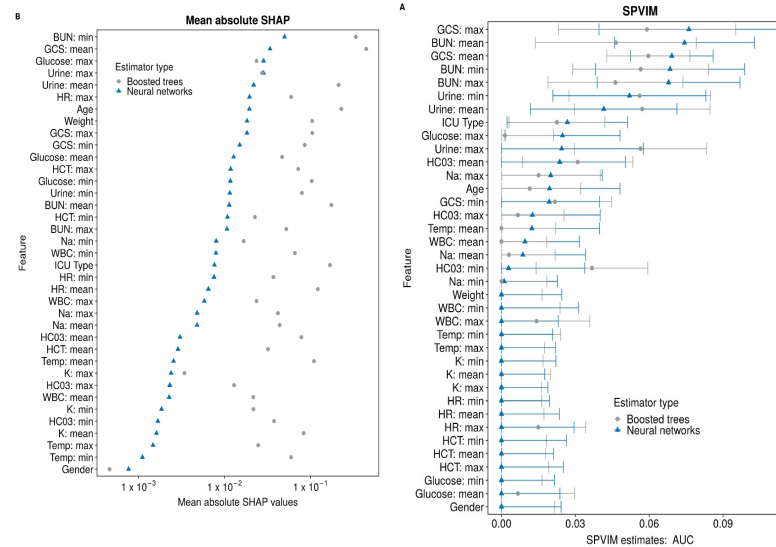
Source: <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-ai-enabled-medical-devices>

AI algorithms present new methodological challenges

- **Interpretability:** AI algorithms are opaque. How can we know they are doing something reasonable?
- **ML is dynamic:** How can we anticipate the effects that AI algorithms will have on a system and detect potential feedback loops?
- **Fairness:** AI algorithms tend to perform well for the majority group. How can we detect subpopulations that an AI algorithm performs poorly for?
- **Causality:** How can we develop AI algorithms that go beyond learning correlations and can be interpreted causally?
- **Generalizability:** AI algorithms rely on patterns in the data from a given context, in a given time period. How can we check that an AI algorithm will generalize? How can we monitor AI algorithms to determine when they are outdated and need updating?
- **Evaluation for Generative AI:** How can we evaluate the performance of AI algorithms that can perform a wide variety of tasks?
- And more!

AI algorithms present new methodological challenges

- **Interpretability:** AI algorithms are opaque. How can we know they are doing something reasonable?



Williamson and Feng, ICML 2020

The Mythos of Model Interpretability

IN MACHINE LEARNING, THE CONCEPT OF INTERPRETABILITY IS BOTH IMPORTANT AND SLIPPERY.

ZACHARY C. LIPTON



AI algorithms present new methodological challenges

- **Interpretability:** AI algorithms are opaque. How can we know they are doing something reasonable?
- **ML is dynamic:** How can we anticipate the effects that AI algorithms will have on a system and detect potential feedback loops?

Runaway Feedback Loops in Predictive Policing*

Danielle Ensign
University of Utah

Sorelle A. Friedler
Haverford College

Scott Neville
University of Utah

Carlos Scheidegger
University of Arizona

Suresh Venkatasubramanian†
University of Utah

DANIPHYE@GMAIL.COM

SORELLE@CS.HAVERFORD.EDU

DROP.SCOTT.N@GMAIL.COM

CSCHIED@CSCHIED.NET

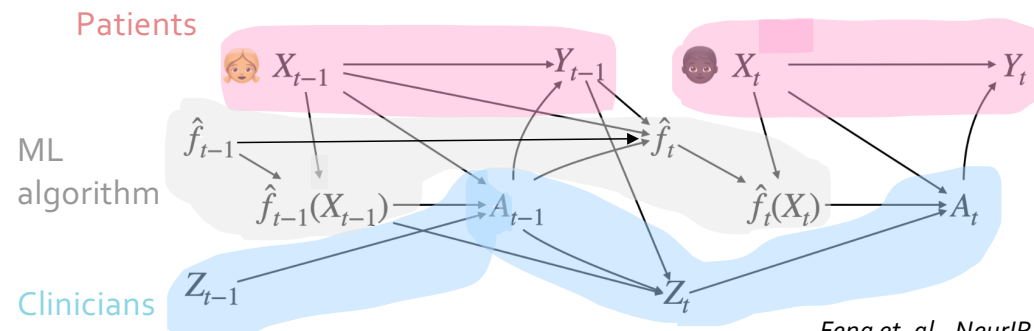
SURESH@CS.UTAH.EDU

How Algorithmic Confounding in Recommendation Systems Increases Homogeneity and Decreases Utility

Allison J.B. Chaney
Princeton University
Department of Computer Science
achaney@princeton.edu

Brandon M. Stewart
Princeton University
Department of Sociology
bms4@princeton.edu

Barbara E. Engelhardt
Princeton University
Department of Computer Science
bee@princeton.edu



Feng et. al., NeurIPS Regulatable AI Workshop 2023

AI algorithms present new methodological challenges

- **Interpretability:** AI algorithms are opaque. How can we know they are doing something reasonable?
- **ML is dynamic:** How can we anticipate the effects that AI algorithms will have on a system and detect potential feedback loops?
- **Fairness:** AI algorithms tend to perform well for the majority group. How can we detect subpopulations that an AI algorithm performs poorly for?

Subgroup analysis for clinical trials



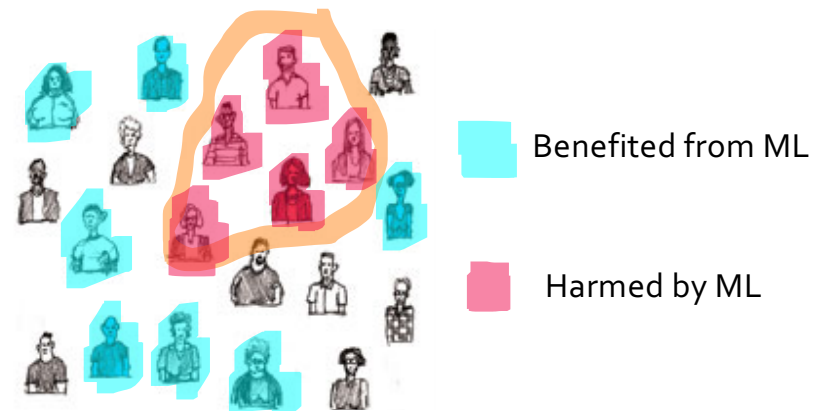
AI algorithms present new methodological challenges

- **Interpretability:** AI algorithms are opaque. How can we know they are doing something reasonable?
- **ML is dynamic:** How can we anticipate the effects that AI algorithms will have on a system and detect potential feedback loops?
- **Fairness:** AI algorithms tend to perform well for the majority group. How can we detect subpopulations that an AI algorithm performs poorly for?

Subgroup analysis for clinical trials



Subgroup analysis for ML algorithms



AI algorithms present new methodological challenges

- **Interpretability:** AI algorithms are opaque. How can we know they are doing something reasonable?
- **ML is dynamic:** How can we anticipate the effects that AI algorithms will have on a system and detect potential feedback loops?
- **Fairness:** AI algorithms tend to perform well for the majority group. How can we detect subpopulations that an AI algorithm performs poorly for?
- **Causality:** How can we develop AI algorithms that go beyond learning correlations and can be interpreted causally?
- **Generalizability:** AI algorithms rely on patterns in the data from a given context, in a given time period. How can we check that an AI algorithm will generalize? How can we monitor AI algorithms to determine when they are outdated and need updating?
- **Evaluation for Generative AI:** How can we evaluate the performance of AI algorithms that can perform a wide variety of tasks?
- And more!

Building effective AI solutions requires cross-disciplinary collaboration

- **Picking the right clinical problem:** Given the capabilities of ML, what clinical problems can it solve?
- **Extracting, cleaning, and understanding the data:** What limitations exist in the data (e.g. confounding, missing data, measurement error)?
- **Training and selecting a model:** Different ML algorithms make different tradeoffs between accuracy, interpretability, fairness, speed, cost, and more. What are the most appropriate tradeoffs?
- **Model deployment and evaluation:** How can we deploy the model in a way that is transparent, reliable, and accessible to all stakeholders?



- Healthcare providers
- Clinical informaticists
- ML developers
- Statisticians
- Patients
- ...

AI algorithms present new regulatory challenges

2019 Discussion Paper



Questions / Feedback on GMLP:

- What additional considerations exist for GMLP?
- How can FDA support development of GMLP?
- How do manufacturers and software developers incorporate GMLP in their organization?

154 comments in the Docket

2023 Draft Guidance

Contains Nonbinding Recommendations

Draft – Not for Implementation

Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence/Machine Learning (AI/ML)-Enabled Device Software Functions

Draft Guidance for Industry and Food and Drug Administration Staff

DRAFT GUIDANCE

This draft guidance document is being distributed for comment purposes only.

Document issued on April 3, 2023.

You should submit comments and suggestions regarding this draft document within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to <https://www.regulations.gov>. Submit written comments in the Dockets Management Staff, Food and Drug Administration, 5630 Fishers Lane, Room 1061, (HFA-305), Rockville, MD 20852. Identify all comments with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions about this document regarding CDREH-regulated devices, contact cdrehe@hqs.fda.hhs.gov. For questions about this document regarding CDREH-regulated devices, contact cdrehe@hqs.fda.hhs.gov. For questions about this document regarding CDREH-regulated products, contact drugs@hqs.fda.hhs.gov. For questions about this document regarding combination products, contact the Office of Combination Products at combination@fda.gov.

FDA U.S. FOOD & DRUG ADMINISTRATION

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Devices and Radiological Health
Center for Biologics Evaluation and Research
Center for Drug Evaluation and Research
Office of Combination Products in the Office of the Commissioner

32 comments in the Docket

