Artificial Intelligence in Drug & Biological Product Development Hybrid Public Workshop

Tuesday, October 7, 2025
9:00 a.m. – 5:00 p.m.

The National Press Club
Washington, DC

FDA U.S. FOOD & DRUG ADMINISTRATION

CLINICAL TRIALS **TRANSFORMATION** INITIATIVE

Welcome and Opening Remarks

Morgan Hanger

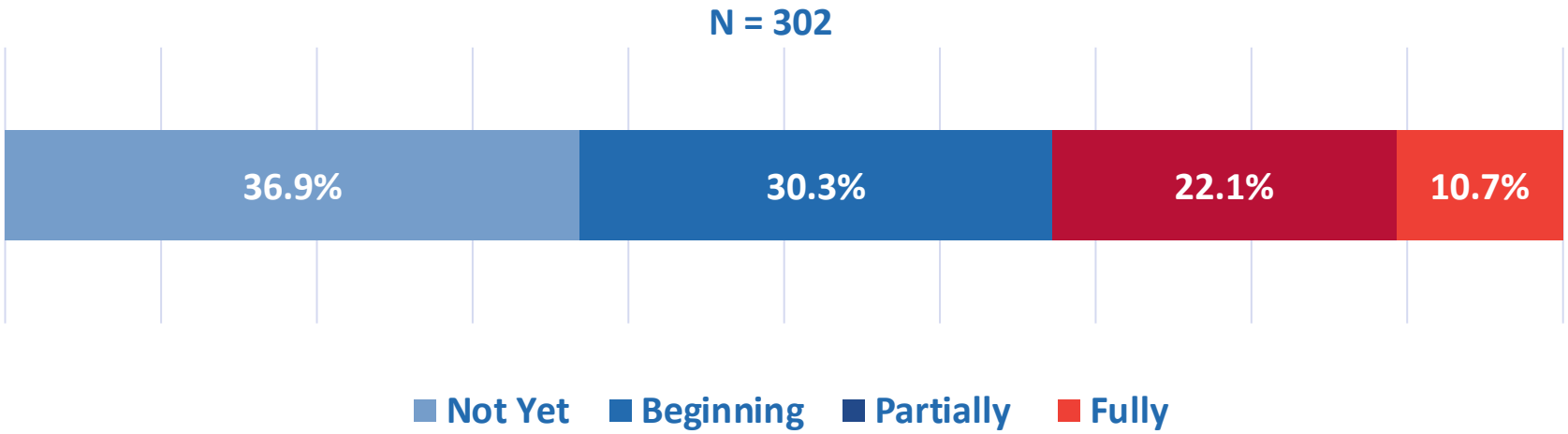
Executive Director
Clinical Trials Transformation Initiative

FDA U.S. FOOD & DRUG ADMINISTRATION

CLINICAL TRIALS TRANSFORMATION INITIATIVE

Hybrid Public Workshop on Artificial Intelligence in Drug & Biological Product Development

Implementation of AI/ML across 36 design and planning, execution, and regulatory submission activities



Lamberti MJ, et al. The Adoption and Use of Artificial Intelligence and Machine Learning in Clinical Development. Therapeutic Innovation & Regulatory Science (2025) 59:1074–1086

Challenges to Implementation

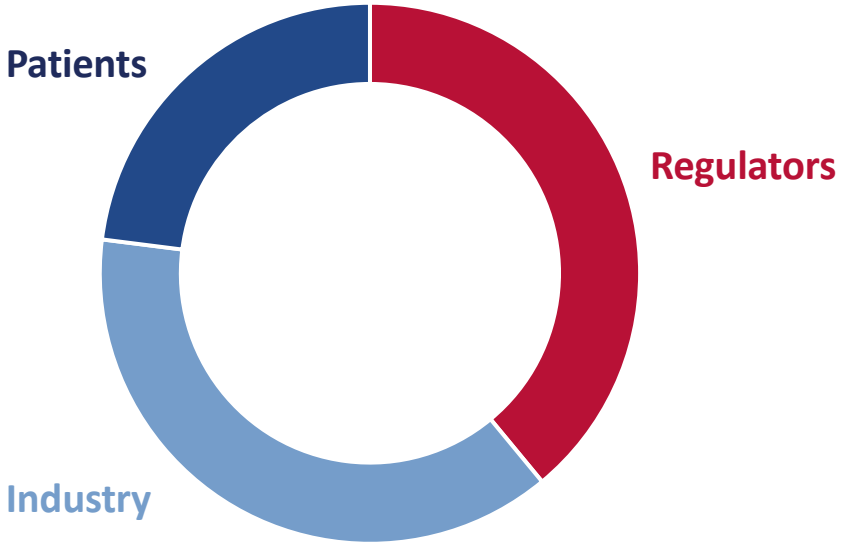
CSDD Survey, N=221(2024)	
Data quality	41%
Trust in AI/ML generated output	37%
Intellectual Property (IP) and other legal concerns related to data sharing	29%
Governance of ethical and privacy concerns	28%

Source: Lamberti MJ, et al. The Adoption and Use of Artificial Intelligence and Machine Learning in Clinical Development. Therapeutic Innovation & Regulatory Science (2025) 59:1074–1086

CTTI Survey, N=219 (2025)	
Data quality/availability	56%
Technical expertise	56%
Regulatory compliance	53%
Ethical concerns	48%
Resistance to adoption	47%

Source: Unpublished CTTI survey, closed June 2025. Convenience sample of organizations in the clinical trials enterprise including but not limited to pharma/biotech, CROs, AROs, tech vendors, patient groups, regulators, health systems, trialists.

Who is Resistant to AI/ML Adoption?



Source: Unpublished CTTI survey, June 2025. N=218

U.S. FOOD & DRUG ADMINISTRATION

CLINICAL TRIALS TRANSFORMATION INITIATIVE

- The views expressed in this presentation do not necessarily represent the policies of the FDA
- Mentions are not endorsements
- Disclosures: None

Agenda

Time (EDT)	Content
9:00 a.m.	Welcome and Opening Remarks
9:05 a.m.	Keynote Speakers
9:15 a.m.	FDA Update: Considerations for the Use of Artificial Intelligence
9:30 a.m.	Session 1: Where Are We Now?
10:40 a.m.	Break
10:55 a.m.	Session 2: Data Quality, Reliability, Representativeness, and Access in AI-Driven Drug Development
12:05 p.m.	Lunch
1:35 p.m.	Session 3: Model Performance, Explainability, Transparency, and Interpretability in AI-Driven Drug Development
2:45 p.m.	Break
3:00 p.m.	Session 4: Navigating the Future of AI in Drug Development
4:15 p.m.	Discussion
4:45 p.m.	Concluding Remarks
5:00 p.m.	Adjourn

Questions throughout the day?

PLEASE SUBMIT HERE:

Keynote Address

Shantanu Nundy

Advisor on Artificial Intelligence (Contractor), Office of the Commissioner

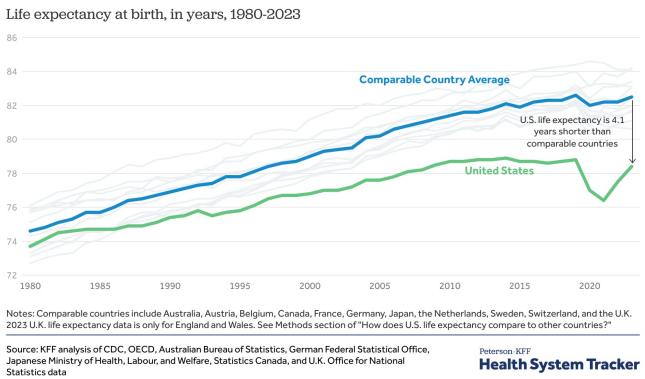
U.S. Food & Drug Administration

Accelerating AI in Drug Development

Shantanu Nundy, MD
Advisor on Artificial Intelligence (contractor)
Office of the Commissioner, FDA
October 7, 2025

16,000,000 birthdays lost in the U.S.

- 100M Americans lack regular access to care
- 75M live in health professional shortage areas
- Medical error 3rd leading causes of death
- 95% of rare diseases lack FDA-approved treatments
- Life expectancy ~4 years behind peers



Accelerating AI in Drug Development

- Real-world data and evidence
- Reduce animal testing and model toxicity
- AI-enabled endpoints and biomarkers
- Clinical trial design
- Clinical trial enrollment

Accelerating AI in Drug Development

- **Real-world data and evidence**
- Reduce animal testing and model toxicity
- AI-enabled endpoints and biomarkers
- Clinical trial design
- Clinical trial enrollment

AI for RWD and RWE

Incidence of Remission in Adults With Type 2 Diabetes: The Diabetes & Aging Study

Andrew J. Karter,[1] Shantanu Nundy,[2,3] Melissa M. Parker,[1] Howard H. Moffet,[1] and Elbert S. Huang[4,5,6]

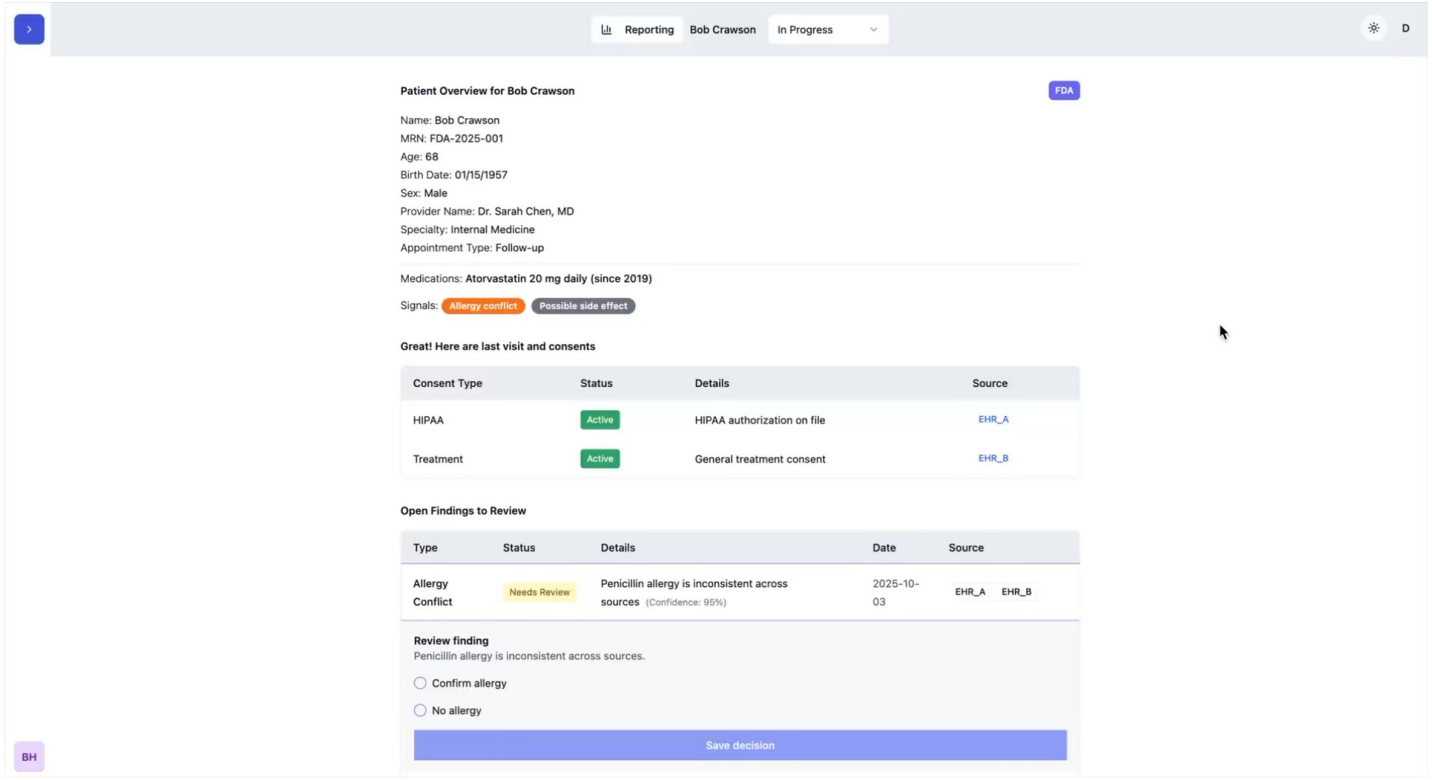
Diabetes Care 2014;37:3188–3195 | DOI: 10.2337/dc14-0874

Table 3—Incidence rates of remission for the full cohort and stratified by time since diagnosis

	Total person-years at risk*	Incident events (*n*)	7-Year cumulative incidence (95% CI)†	Incidence rate per 1,000 person-years (95% CI)
Any remission				
All	586,725	1,761	1.60% (1.53–1.68)	3.00 (2.86–3.14)
Time since diagnosis <2 years (*n* = 18,451)	88,473	776	4.55% (4.25–4.88)	8.77 (8.15–9.39)
Time since diagnosis 2–3 years (*n* = 18,127)	89,526	424	2.54% (2.31–2.79)	4.74 (4.29–5.19)
Time since diagnosis 4–5 years (*n* = 15,122)	75,304	228	1.67% (1.46–1.89)	3.03 (2.63–3.42)
Time since diagnosis 6–9 years (*n* = 20,270)	102,270	152	0.82% (0.70–0.96)	1.49 (1.25–1.72)
Time since diagnosis ≥10 years (*n* = 30,326)	147,333	98	0.37% (0.30–0.45)	0.67 (0.53–0.80)
Baseline diabetes therapy, no medication (*n* = 13,502)	62,009	1,538	12.33% (11.76–12.93)	24.80 (23.56–26.04)
Baseline diabetes therapy, OHA only (*n* = 84,968)	419,376	212	0.28% (0.24–0.32)	0.51 (0.44–0.57)
Baseline diabetes therapy, insulin (*n* = 22,625)	105,340	11	0.05% (0.03–0.10)	0.10 (0.04–0.17)
Partial remission	587,341	1,615	1.47% (1.40–1.54)	2.75 (2.62–2.88)
Complete remission	593,216	140	0.14% (0.12–0.16)	0.24 (0.20–0.28)
Prolonged remission	170,356	6	0.007% (0.003–0.02)	0.035 (0.007–0.063)

OHA, oral hypoglycemic agent. *Defined as the total number of person-years 12 months after cohort inception for partial and complete remission, and after 60 months for prolonged remission. †Defined as 100% minus the cumulative survival probability calculated using the Kaplan-Meier method.

AI for RWD and RWE



Accelerating AI in Drug Development

- Real-world data and evidence
- **Reduce animal testing and model toxicity**
- AI-enabled endpoints and biomarkers
- Clinical trial design
- Clinical trial enrollment

AI to Reduce Animal Testing & Model Toxicity

Regulatory Toxicology and Pharmacology 138 (2023) 105329

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

Check for updates

Re-evaluating the need for chronic toxicity studies with therapeutic monoclonal antibodies, using a weight of evidence approach

Hsiao-Tzu Chien [a,s,1,2], Helen Prior [b,1,2], Laura Andrews [c], Leon van Aerts [a,1], Annick Cauvin [d], David O. Clarke [e,1], Kaushik Datta [f], Maggie Dempster [g,1], Noel Dybdal [h,1], Wendy Freebern [i], Lolke de Haan [j,1], Danuta Herzyk [k,1], Adam Hey [l], Thomas Kissner [m], Sven Kronenberg [n,1], Michael W. Leach [o,1], Donna Lee [h], Katrin Schutte [p], Fiona Sewell [b,1], Kevin Trouba [i,1], Peter Ulrich [q], Lucinda Weir [r,1,3], Peter van Meer [a,1,3,*]

Accelerating AI in Drug Development

- Real-world data and evidence
- Reduce animal testing and model toxicity
- **AI-enabled endpoints and biomarkers**
- Clinical trial design
- Clinical trial enrollment

AI for New Endpoints & Biomarkers

nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › articles › article

Article | Open access | Published: 16 July 2025

Detecting structural heart disease from electrocardiograms using AI

Timothy J. Poterucha, Linyuan Jing, Ramon Pimentel Ricart, Michael Adjei-Mosi, Joshua Finer, Dustin Hartzel, Christopher Kelsey, Aaron Long, Daniel Rocha, Jeffrey A. Ruhl, David vanMaanen, Marc A. Probst, Brock Daniels, Shalmali D. Joshi, Olivier Tastet, Denis Corbin, Robert Avram, Joshua P. Barrios, Geoffrey H. Tison, I-Min Chiu, David Ouyang, Alexander Volodarskiy, Michelle Castillo, Francisco A. Roedan Oliver, ... Pierre Elias ✉ + Show authors

Nature **644**, 221–230 (2025) | Cite this article

Accelerating AI in Drug Development

- Real-world data and evidence
- Reduce animal testing and model toxicity
- AI-enabled endpoints and biomarkers
- **Clinical trial design**
- Clinical trial enrollment

AI for Clinical Trial Design

Accelerating AI in Drug Development

- Real-world data and evidence
- Reduce animal testing and model toxicity
- AI-enabled endpoints and biomarkers
- Clinical trial design
- **Clinical trial enrollment**

AI for Clinical Trial Enrollment

AI Will Move Drug Discovery at the Speed of Trust

VIEWPOINT

TRUST IN HEALTH CARE

Promoting Trust Between Patients and Physicians in the Era of Artificial Intelligence

Shantanu Nundy, MD, MBA
The World Bank, Washington, DC; and George Washington University Milken Institute School of Public Health, Washington, DC.

Tara Montgomery, BA
Civic Health Partners, Brooklyn, New York.

Robert M. Wachter, MD
Department of Medicine, University of California, San Francisco.

Advances in technology are increasingly changing relationships between consumers and providers of products and services in every business. The democratization of recognized expertise that accompanies the use of information technology can be a positive force for improving access, cost, and equity, but also can challenge the role and status of traditional experts, including marginalizing them. Consumers no longer need a professional taxi driver to tell them the fastest way to the airport and can book flights without the help of a travel agent.

While the internet allows individuals to access tremendous amounts of information in many domains, they often still need to partner with credentialed experts who offer additional knowledge and experience and access to certain services that remain under their control.

Health care is one such example. Even patients

physicians and systems, such as for clinical decision support and system strengthening, physician assessment and training, quality improvement, clinical documentation, and nonclinical tasks, such as scheduling and notifications; (2) use of health care AI by patients including triage, diagnosis, and self-management; and (3) data for health care AI involving the routine use of patient data to develop, validate, and fine-tune health care AI as well as to personalize the output of health care AI.

Each of these applications has the potential to enable and disable the 3 components of trust: competency, motive, and transparency.

Competency

Competency reflects both the extent to which physicians are perceived to have clinical mastery and patients' knowledge and self-efficacy of their own health.

Keynote Address

M. Khair ElZarrad

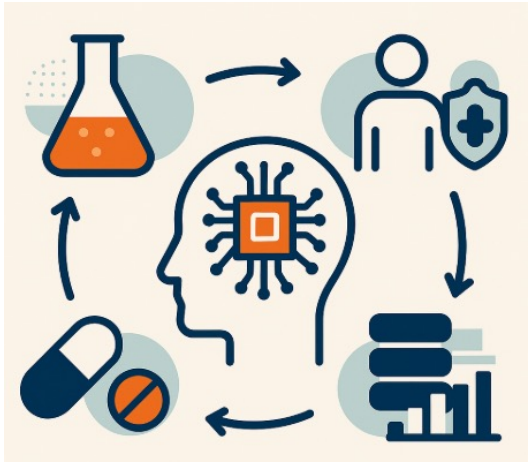
Director, Office of Medical Policy (OMP), Center for Drug Evaluation and Research (CDER)

U.S. Food & Drug Administration

U.S. FOOD & DRUG ADMINISTRATION

Predictability & Consistency

- Risk-based approach
- Increased engagements
- Transparency
- Data governance
- Context of use



Challenges & Opportunities

- Coordinated FDA approach
  - Multidisciplinary review
  - Rapid engagement
- Global harmonization
  - Intersection of privacy and confidentiality - regional laws & regulations
- Upskilling
- Responsive infrastructure

Balancing a rapidly evolving area with providing responsive regulatory input

U.S. FOOD & DRUG ADMINISTRATION

CLINICAL TRIALS TRANSFORMATION INITIATIVE

Update: *Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products*

Gabriel Innes
Assistant Director,
Data Science and Artificial Intelligence Policy, OMP, CDER
U.S. Food & Drug Administration

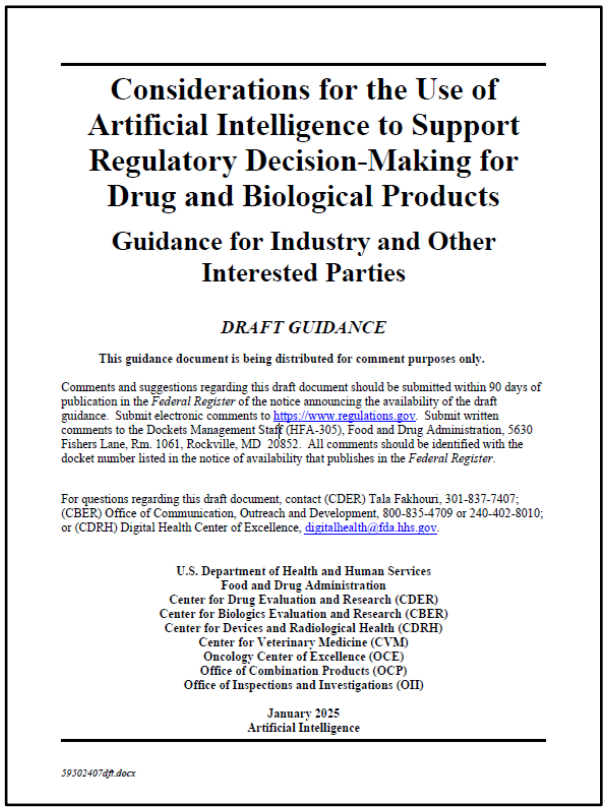
Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products: *UPDATES*

Gabriel Innes, VMD, PhD

October 7, 2025

Some slides adapted from Dr. Tala Fakhouri

AI Guidance – Purpose and Scope



Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products

Guidance for Industry and Other Interested Parties

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted 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 to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document, contact (CDER) Tala Fakhouri, 301-837-7407; (CBER) Office of Communication, Outreach and Development, 800-835-4709 or 240-402-8010; or (CDRH) Digital Health Center of Excellence, digitalhealth@fda.hhs.gov.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)
Center for Veterinary Medicine (CVM)
Oncology Center of Excellence (OCE)
Office of Combination Products (OCP)
Office of Inspections and Investigations (OII)

January 2025
Artificial Intelligence

5930240?dft.docx

Purpose: Providing a risk-based credibility assessment framework to establish the credibility of an AI model for a particular context of use (COU)

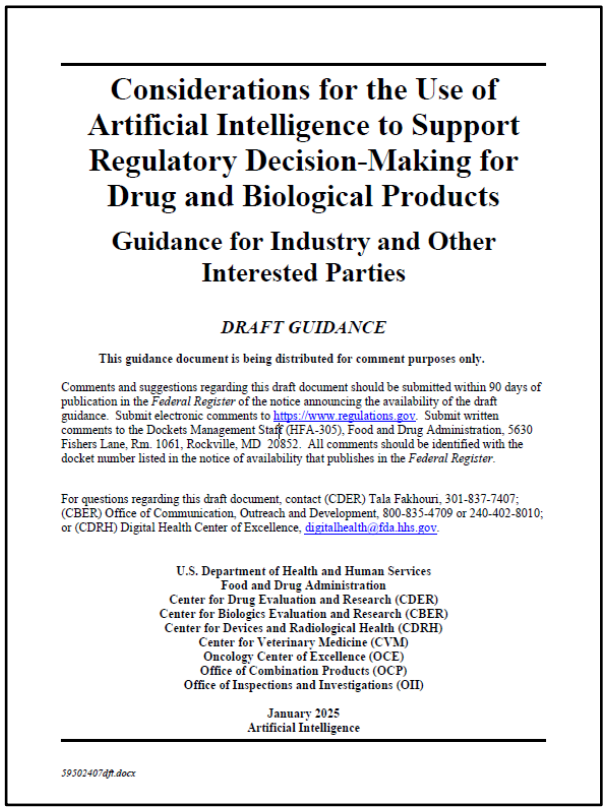
In Scope: Use of AI to produce information or data intended to support regulatory decision-making regarding safety, effectiveness, or quality for drugs and biological products

***Not* in Scope:**

- **Drug discovery***
- **Operational efficiencies (unrelated to patient safety drug quality, or reliability of results from a non-clinical or clinical study), such as**
  - **Internal workflows**
  - **Resource allocation**
  - **Drafting/writing a regulatory submission**

*except in very specific situations

AI Guidance – Key Points



Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products

Guidance for Industry and Other Interested Parties

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted 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 to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document, contact (CDER) Tala Fakhouri, 301-837-7407; (CBER) Office of Communication, Outreach and Development, 800-835-4709 or 240-402-8010; or (CDRH) Digital Health Center of Excellence, digitalhealth@fda.hhs.gov.

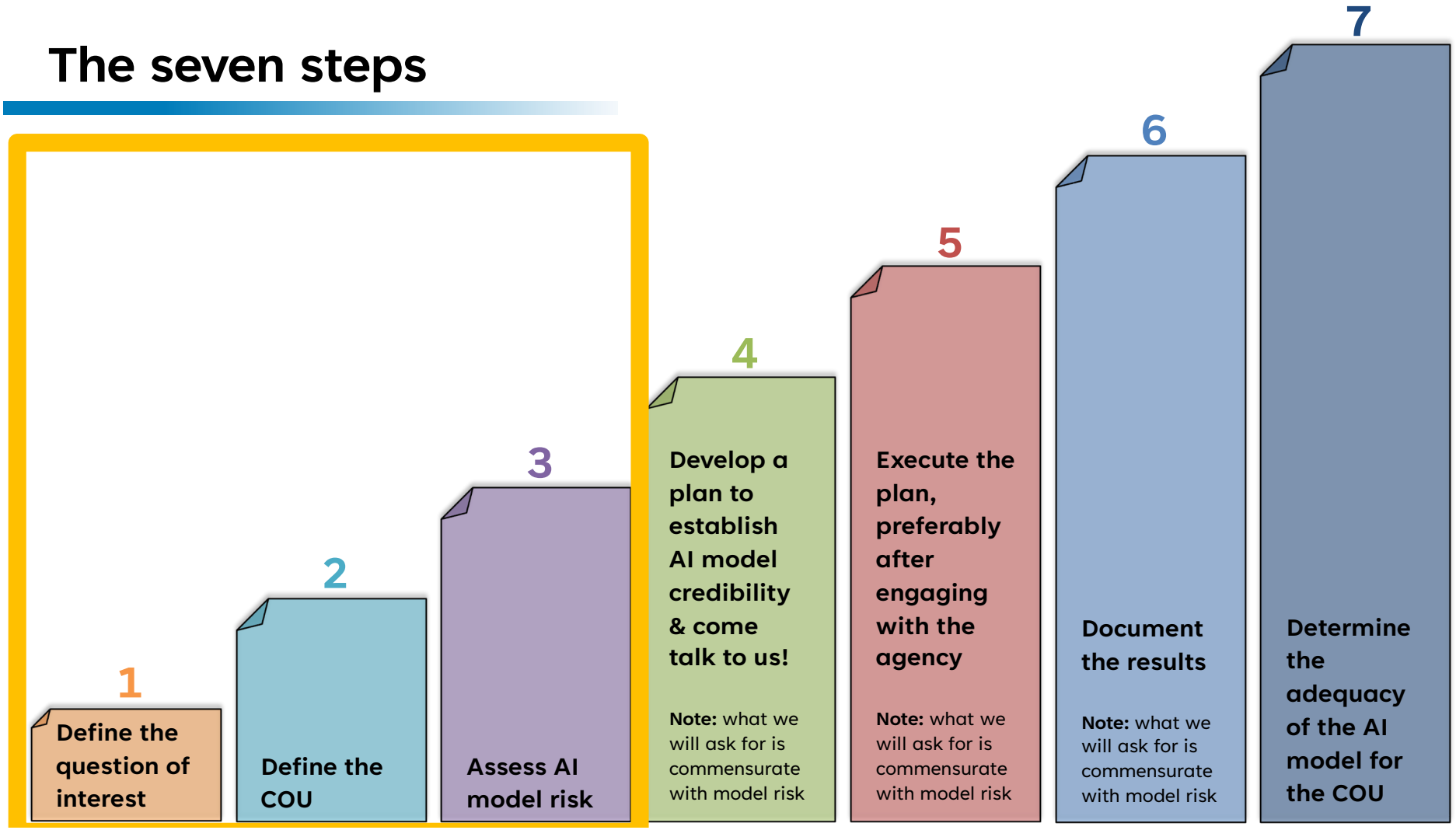
U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)
Center for Veterinary Medicine (CVM)
Oncology Center of Excellence (OCE)
Office of Combination Products (OCP)
Office of Inspections and Investigations (OII)

January 2025
Artificial Intelligence

5930240?dft.docx

- Risk-based credibility assessment framework consisting of seven steps to establish and evaluate the credibility of AI model outputs for specific COU, with activities commensurate to AI model risk
- Life cycle maintenance requirements for AI models used in contexts where performance may change over time or across deployment environments, particularly in pharmaceutical manufacturing
- Early engagement with FDA is strongly encouraged to set expectations regarding appropriate credibility assessment activities based on model risk and COU
- Comprehensive documentation requirements including credibility assessment plans and reports that describe model development, training data, evaluation processes, and performance metrics tailored to the specific COU

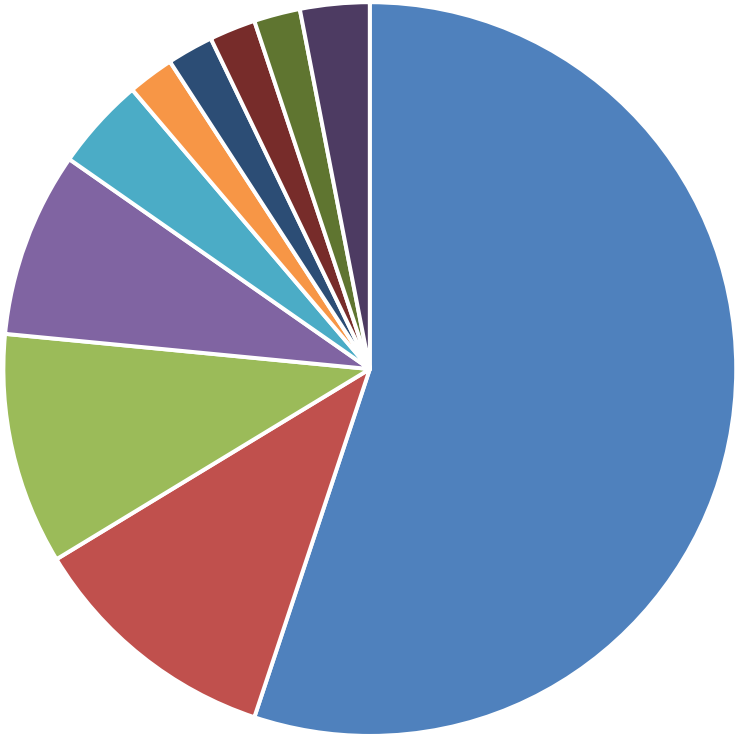
The seven steps



Two Examples Provided in Draft Guidance

	Drug A	Drug B
Step 1: Question of Interest	"Which participants can be considered low risk [of having a life-threatening drug-related adverse reaction] and do not need inpatient monitoring after dosing?"	"Do vials of Drug B meet established fill volume specifications?"
Step 2: COU	"The output from the AI model will be used to stratify participants into low- versus high-risk groups for the potentially life-threatening adverse reaction to Drug A (the AI model's role). In this context, the sponsor is proposing that *only* the AI model will be used to determine whether the participant is considered low risk and whether they will need inpatient or outpatient monitoring after dosing (the AI model's scope)."	"An AI-based model will be used to analyze data obtained from visual images of the vials to determine if a deviation in volume has occurred (the AI model's role). However, as part of release testing, independent verification of the fill volume is performed on a representative sample for each batch. Therefore, the AI-based model will *not* be the sole determinant for the release of product (the AI model's scope)."
Step 3: Model Risk	High	Medium
Model Influence	High: AI model is sole determinant	Low: fill volume measured by a manufacturer on representative sample of each batch
Decision Consequence	High: potential mortality	High: incorrect volume measurements would have a high impact on product quality

Commentors (n = 98)



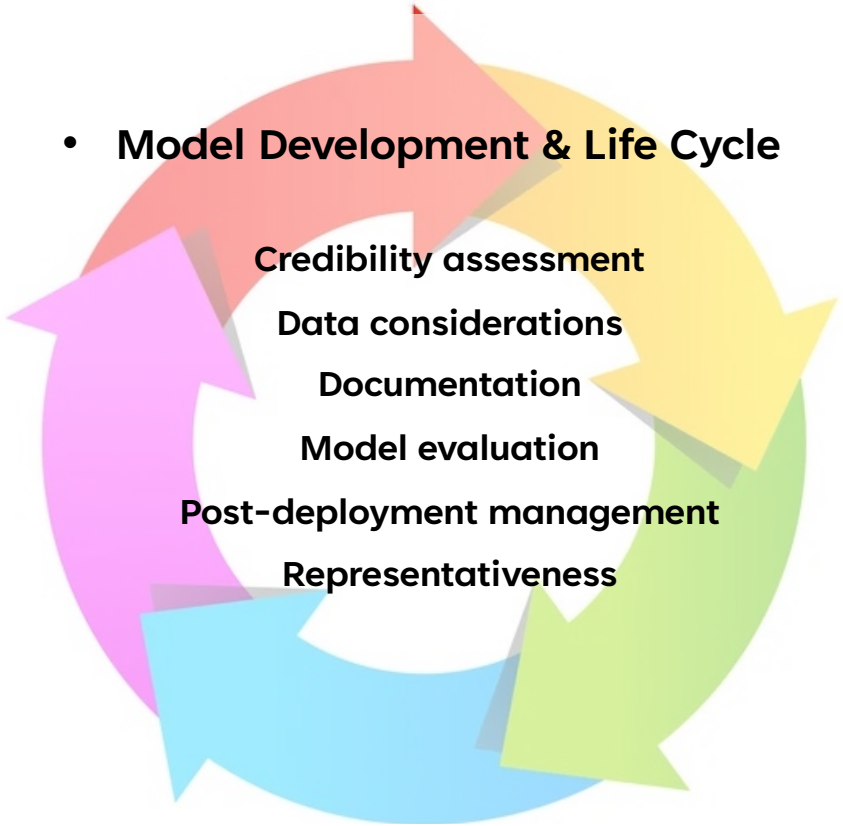
- Regulated Industry (Including Trade Associations) (n = 54)
- Private Industry (Tech) (n = 11)
- Patient and Consumer Advocacy and Other Non-Profit (n = 10)
- Scientific and Academic Experts (n = 8)
- International Organization (n = 4)
- Healthcare Professionals (Physicians, Hospitals) (n = 2)
- Legal/Regulatory Consulting (n = 2)
- Other Government Agencies and Legislators (n = 2)
- Private Citizens or Individuals (n = 2)
- Unknown (n = 3)

Comments by Section (n = 1,455)

Section	No. of Comments
General Comments	706
I. INTRODUCTION	57
II. SCOPE	54
III. BACKGROUND	63
IV. CONSIDERATIONS FOR AI USE IN THE DRUG PRODUCT LIFE CYCLE	3
A. A Risk-Based Credibility Assessment Framework	472
B. Special Consideration: Life Cycle Maintenance of the Credibility of AI Model Outputs in Certain COUs	73
C. Early Engagement	27

Large Themes of Public Comments (non-exhaustive)

- Guidance Clarity
  - Scope
  - Terminology
  - Use Case Examples
- Guiding Principles for Good AI Practices in Drug and Biological Product Development
- Regulatory Engagement & Processes



Opportunity to Continue Shared Learnings

THANK YOU!

U.S. FOOD & DRUG ADMINISTRATION

CLINICAL TRIALS TRANSFORMATION INITIATIVE

Session 1: Where Are We Now?

Moderator:
M. Khair ElZarrad
Director, OMP, CDER
U.S. Food & Drug Administration

Dana Lewis
Independent Researcher, Developer, and Founder
OpenAPS

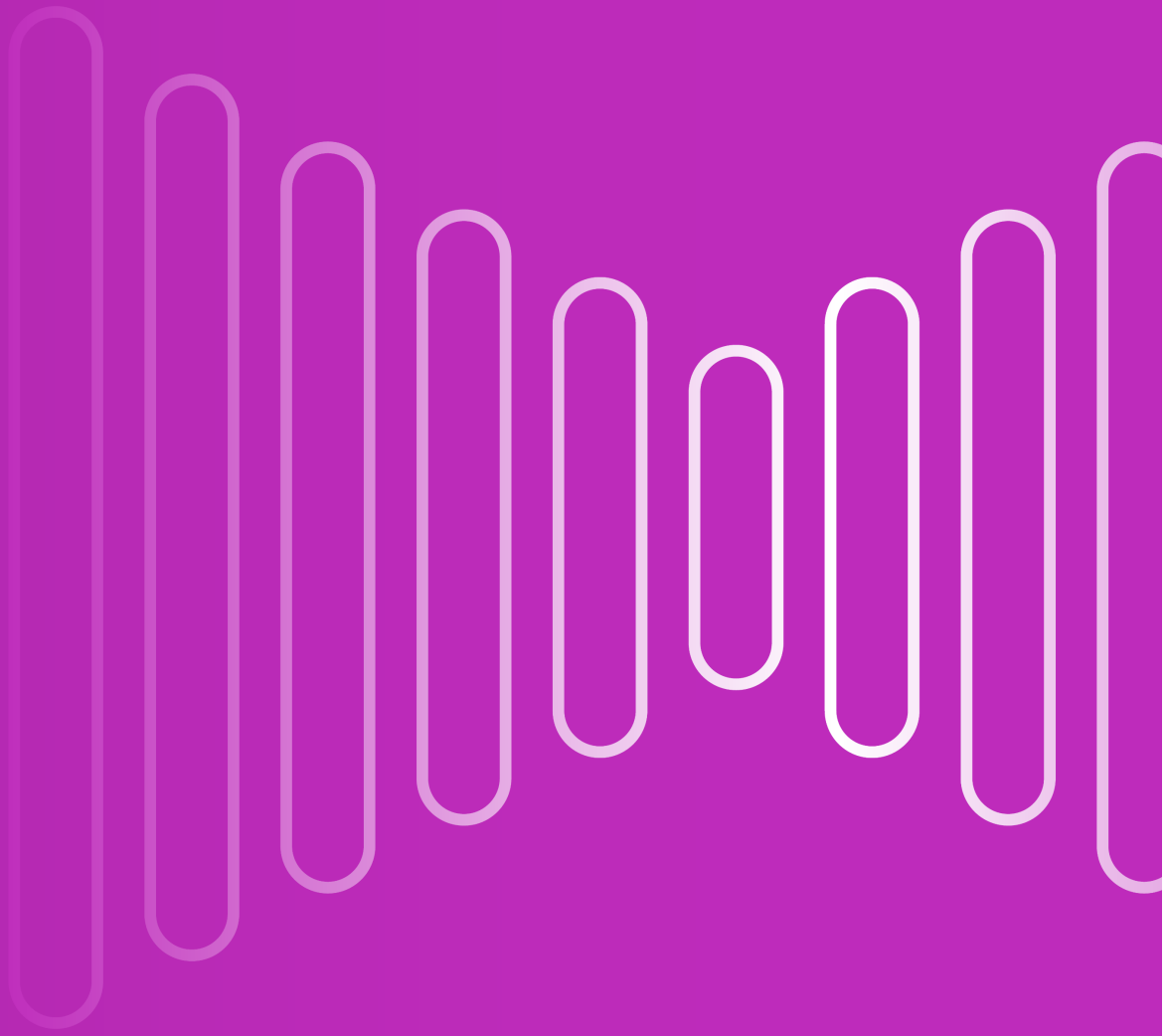
Greg Meyers
Executive Vice President and Chief Digital and Technology Officer
Bristol Myers Squibb

Thomas Osborne
Chief Medical Officer
Microsoft Federal

AI in Drug & Biological Development

Greg Meyers
Chief Digital and Technology Officer
Bristol Myers Squibb

Oct 7, 2025

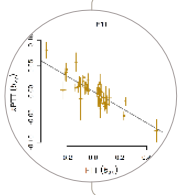


Areas of Opportunity We See With AI in Product Development

Bristol Myers Squibb®

Causal human biology

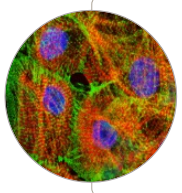
Problem statement: Causal inference from large human datasets is increasingly complex
AI/ML solution: Three efforts to inform on causal human biology based on target perturbation



In vivo genetics-based causal inference

Data (N>1M)

Mutation → drug
Target
Disease

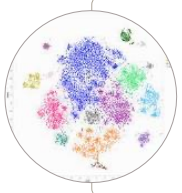


In vitro platforms & phenotypic screens

Tissue/cells
Patients

Screen for targets & drugs

In vitro model



In silico perturbation & generative AI models

>30M cells

Foundation cell model

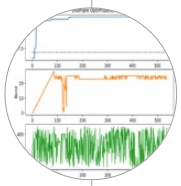
Predict beneficial target

Bristol Myers Squibb®

Matching modality to mechanism

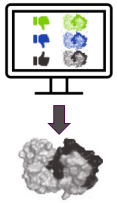
Problem statement: Molecular design is slowed by the need to translate large, complex data sets into actionable hypotheses

AI/ML solution: Accelerated virtual screening and ML-guided MPO with dynamic profiling

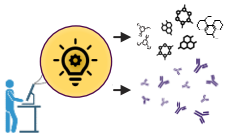


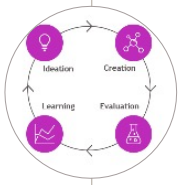
Early Discovery: Identifying hits that can become leads through virtual screening and design

Virtual Screening



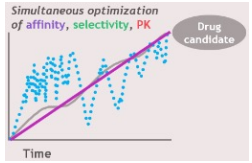
Design



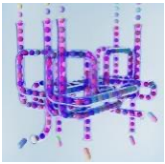


Late Discovery: Delivering high quality development candidates through multi-parameter optimization and dynamic profiling

Multi-parameter Optimization



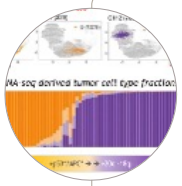
Dynamic Profiling



Bristol Myers Squibb®

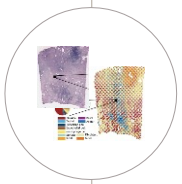
Path to clinical proof-of-concept

Problem statement: Accelerate PoC leveraging increasingly complex disease, clinical and translational data
AI/ML solution: Identify disease segments, actionable biomarkers and dose projection, early



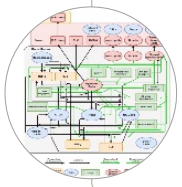
Disease & patient segmentation

- High dimensional, multi-modal, clinical and translational data integrated to develop biologically relevant patient segments
- Assets mapped to biologically relevant disease segment



Clinically relevant BMX/DX

- Defined asset specific patient population
- Clinically actionable classifier & BMX/DX strategy
- Opportunity to screen/select patients for speed to clinical PoC

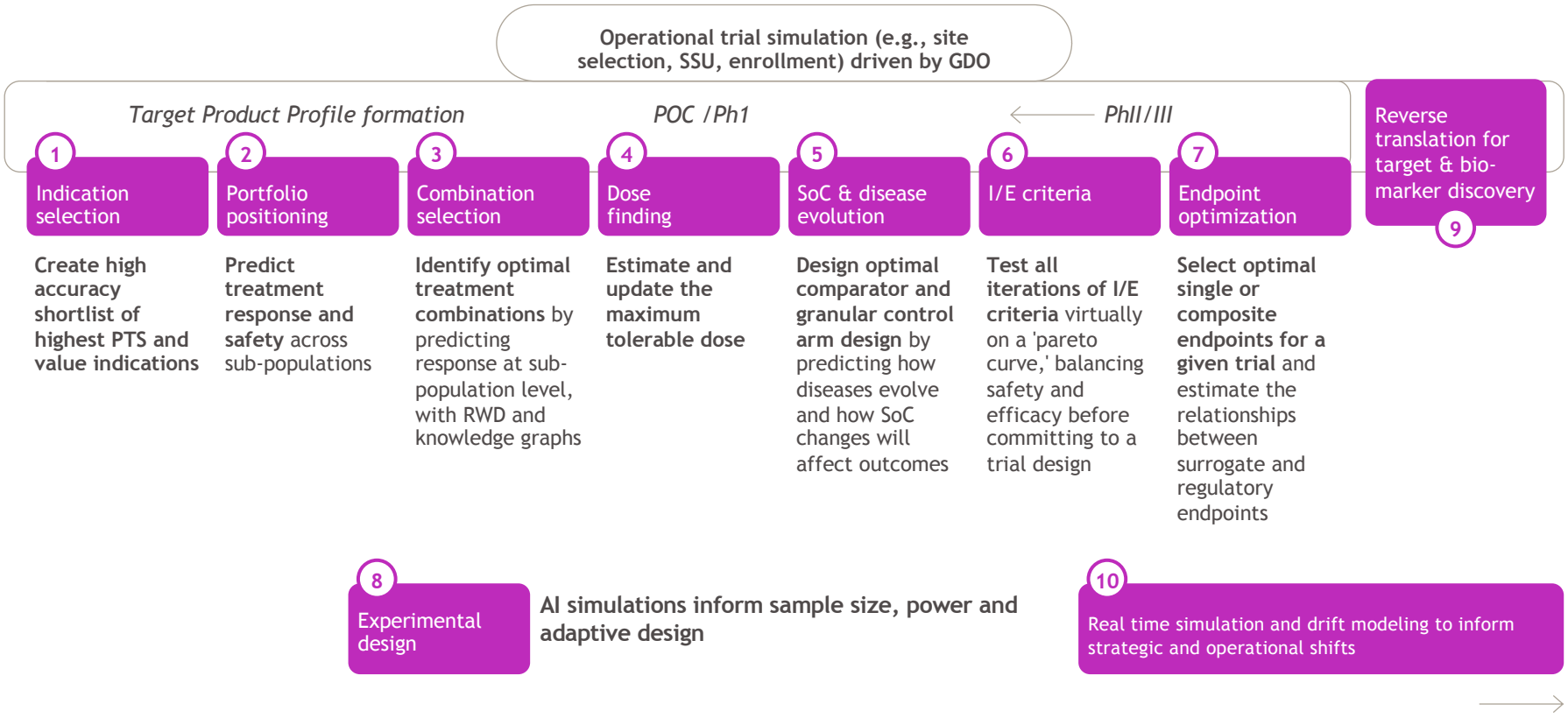


Mechanistic modeling dose-to-patient segment

- Disease, modality, mechanism based QSP models
- AI/ML calibrated models integrating clinical BMX data
- Modality, mechanism and dose projection aligned to proposed patient segment

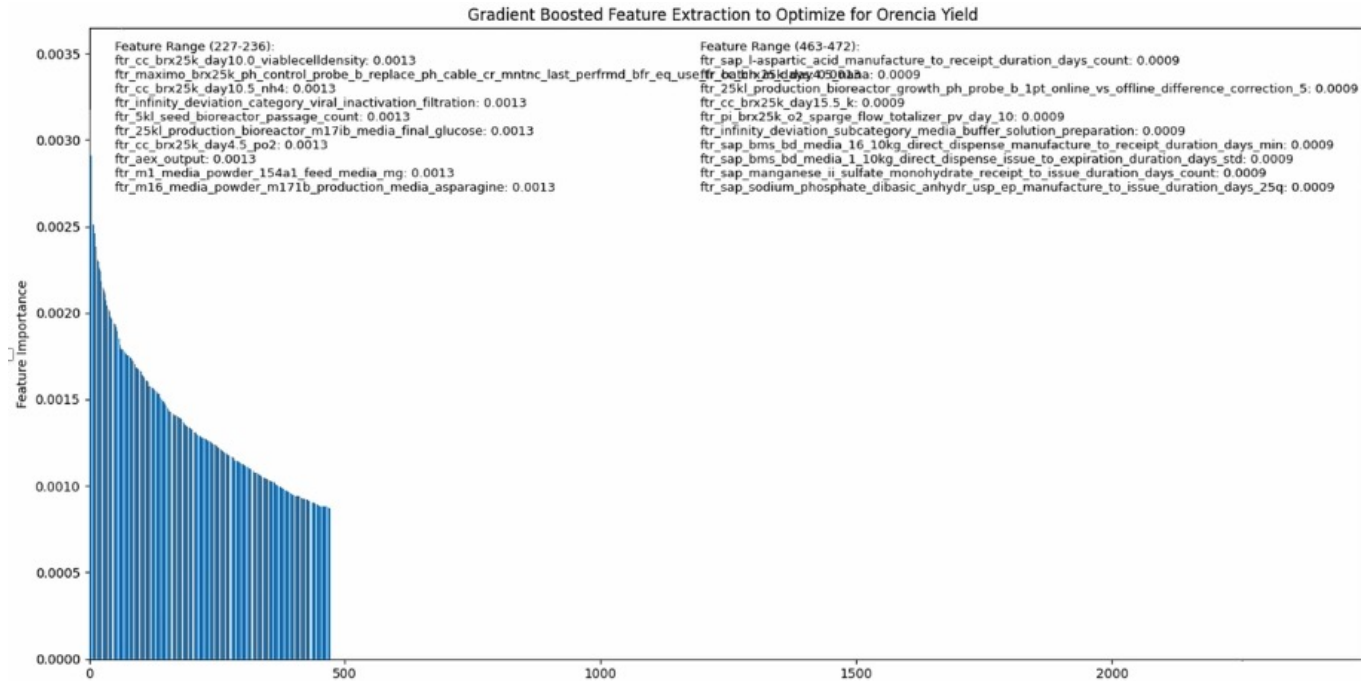
Bristol Myers Squibb®

AI is Useful at Every Stage of Clinical Development



Bristol Myers Squibb®

AI-Based Parameter Optimization Within Manufacturing



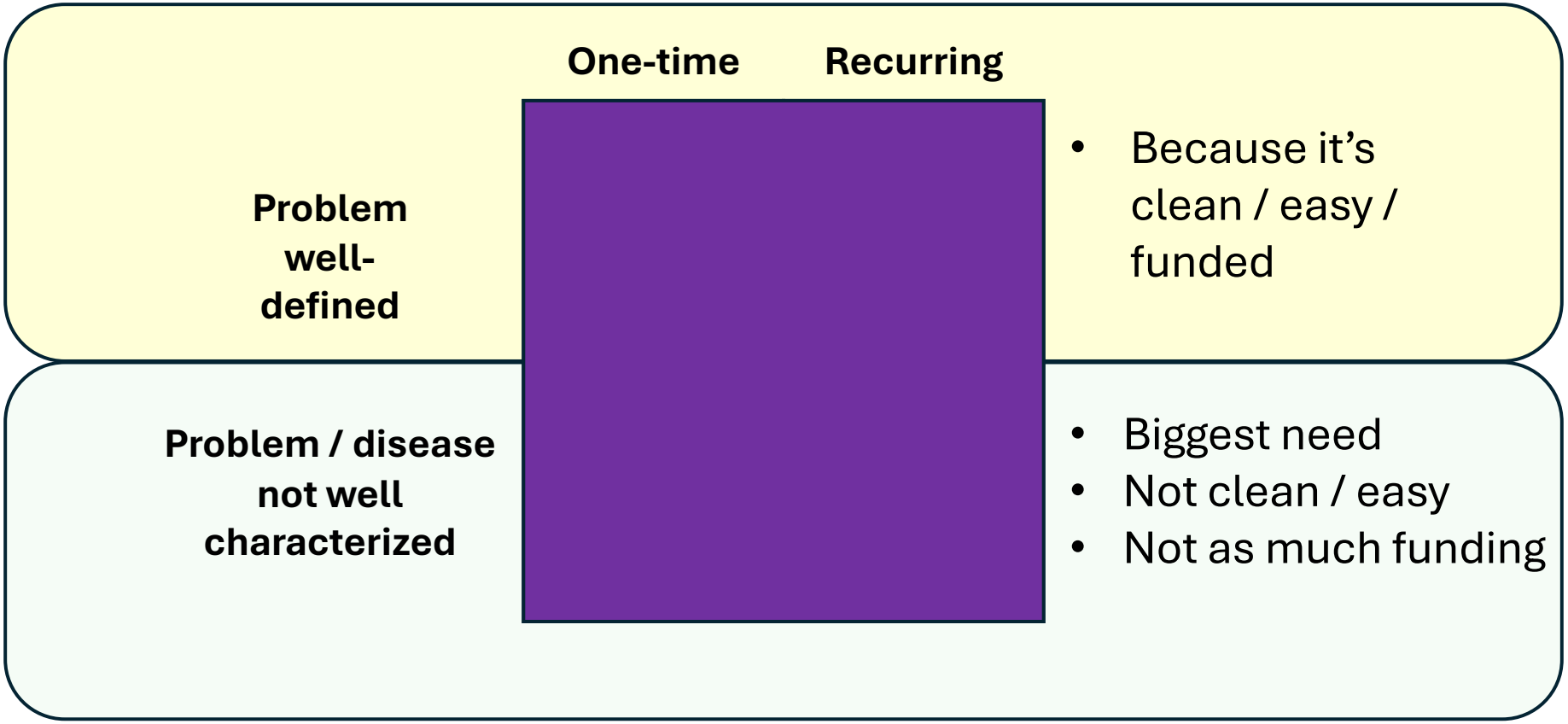
Bristol Myers Squibb®

The Knowable Unknown

(It’s time to gather more real-world evidence and use AI to accelerate drug development)

@DanaMLewis

What are we even talking about when we talk about using AI?



@DanaMLewis

A Study on LLMs & Chart Notes

1 chart note

+

11 questions

»»

4 prompt series

1. Standard
2. Randomized
3. Persona
4. Randomized Persona

»»

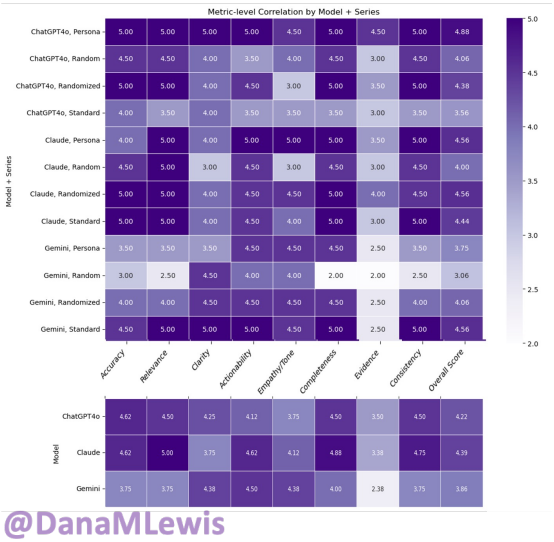
3 LLMs (May-June 2024)

1. Claude 3 Opus
2. ChatGPT 4o
3. Gemini 1.5

»»

Raters:

- blinded to model
- Scoring rubric

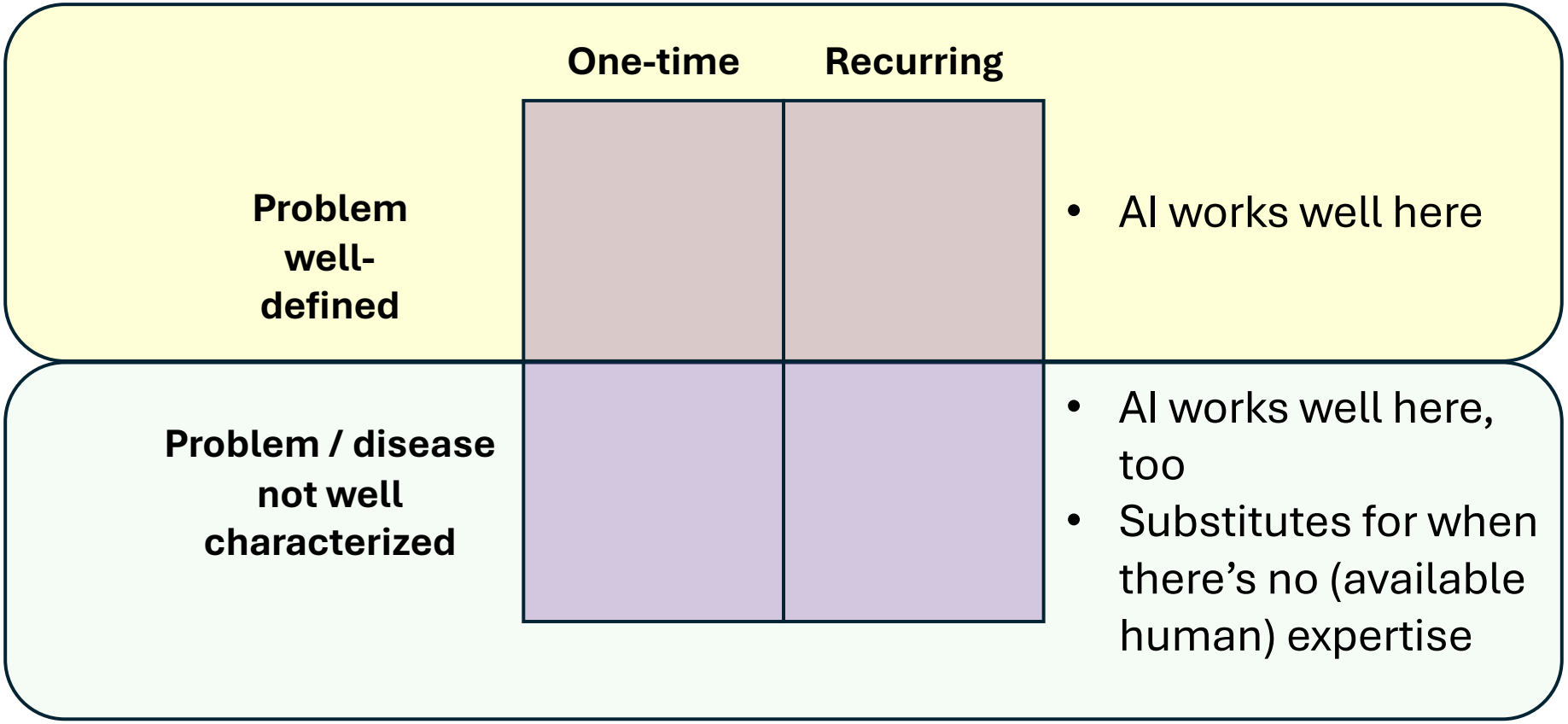


@DanaMLewis

- This research underscores **the potential of LLMs to augment patient understanding** of clinical notes but also highlights the importance of **prompt design** and **model selection**.
- Incorporating a "**persona**" into the **prompt** (e.g., "You are an oncologist…") enhanced **all** LLMs' performance.

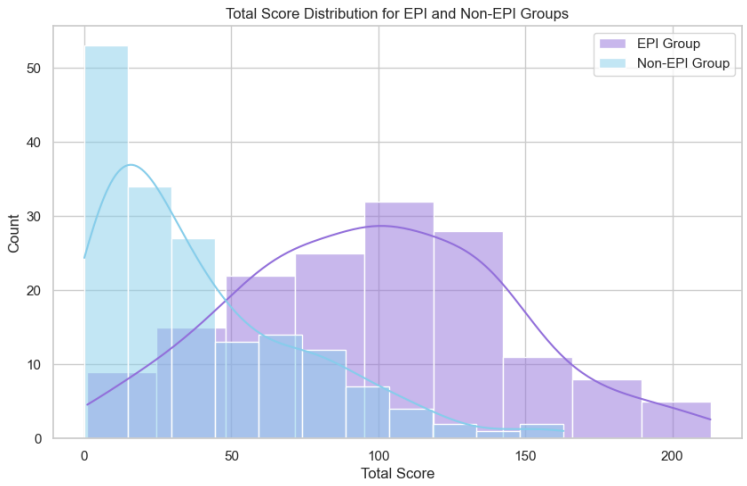
A Proof-of-Concept Study for Patient Use of Open Notes with Large Language Models.
JAMIA Open. 2025. Salmi, Lewis, et al. DOI: 10.1093/jamiaopen/ooaf021

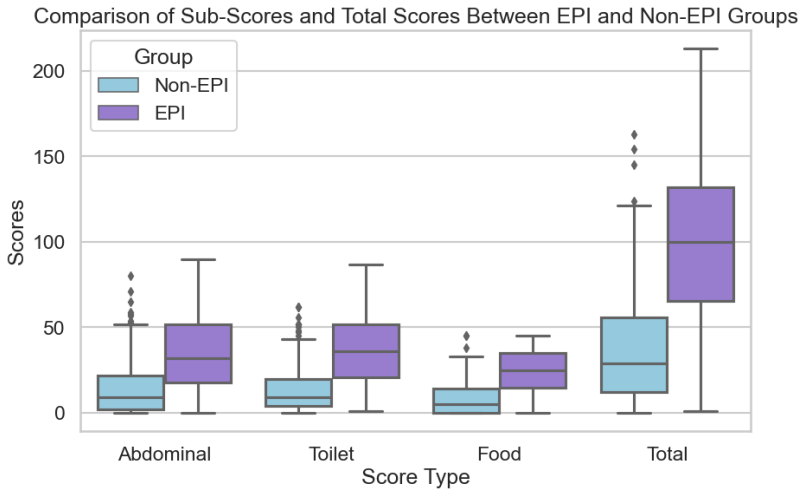
AI is better than nothing (when there's no available human)



@DanaMLewis

How to titrate when titration hasn't been studied?



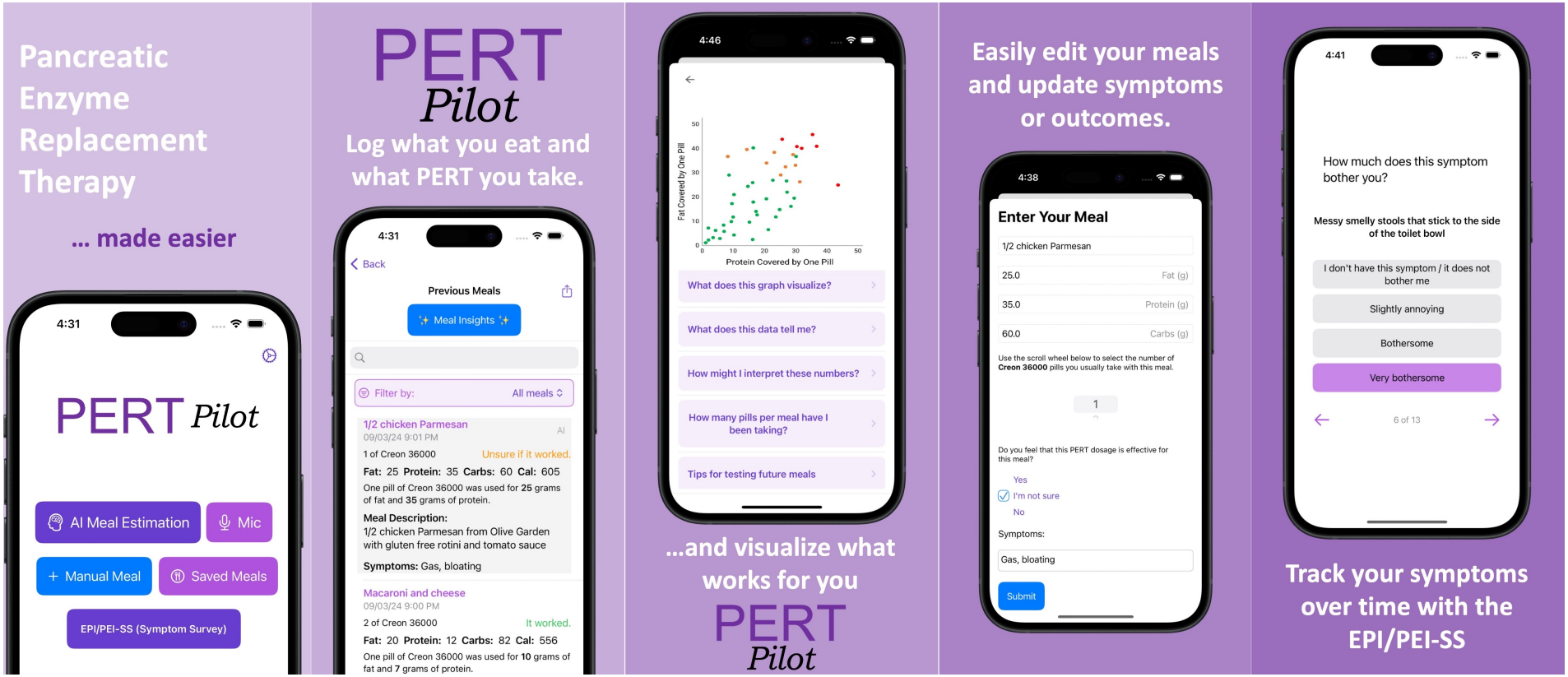


(n=155 with EPI, n=169 without EPI)

- **EPI/PEI-SS demonstrated 0.86 AUC to discriminate between EPI / not**
- **Pilot data suggests can be used to track titration outcomes**

"Development of Novel Symptom Score to Assist in Screening for Exocrine Pancreatic Insufficiency". Epidemiologia. 2025. DOI: 10.3390/epidemiologia6030048

@DanaMLewis



- Personal use
- Research use

@DanaMLewis

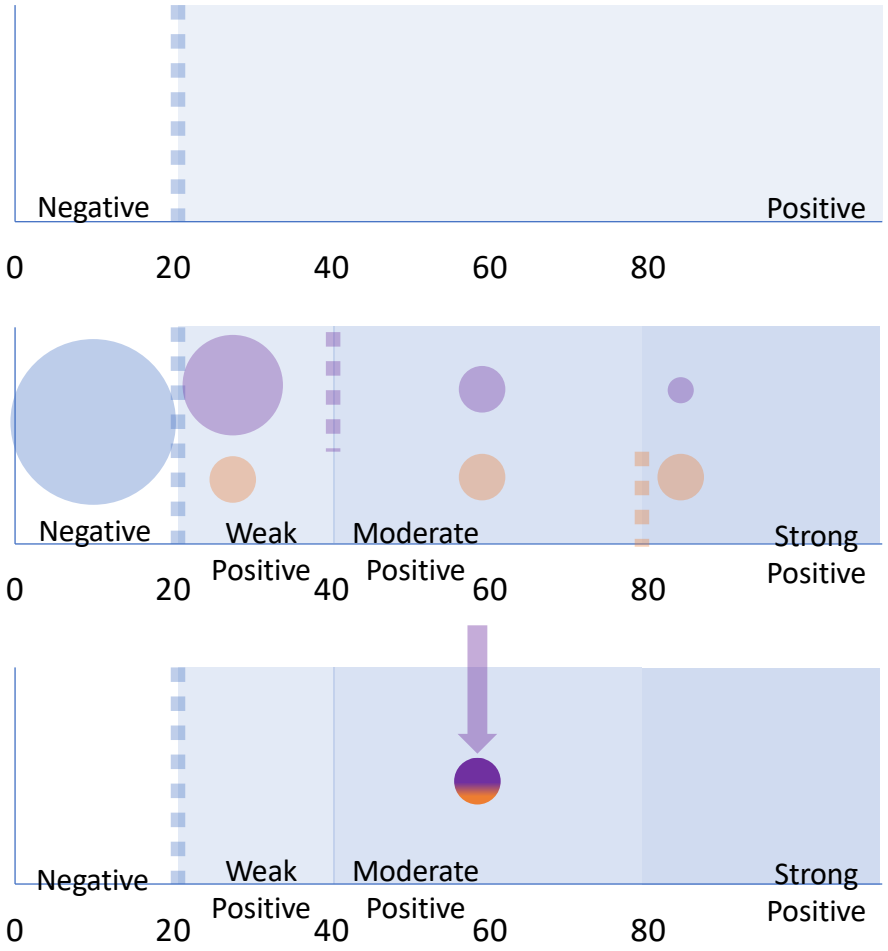
A to Z

(But most of this is not feeding back into the drug development pipeline)

- **A**nalyzing data
- **B**uilding apps to track data
- **C**urating data to show HCPs
- **D**eciding what is useful to track
- **E**vidence finding & appraisal
- **F**iguring out what was said
- **G**oal creation and evaluation
- **H**ome triage (when to escalate care)
- **I**nsurance navigation
- **J**ournaling (structured or unstructured)
- **K**nowledge seeking
- **L**anguage and literacy
- **M**edication comprehension
- **N**utrition analysis and ideas
- **O**rganizing and transforming data
- **P**reparing for visits
- **Q**uick questions and comparisons
- **R**esearch access (deep research, etc.)
- **S**hared decision-making evaluation
- **T**ranslating language and/or concepts
- **U**ncertainty assessments
- **V**oice capture
- **W**orkflow automation (transforming data)
- **X-**ray & other imaging explanations
- **Y**early and other timeframe analyses
- **Z**-scores & normalizing data

@DanaMLewis

Problem area: data reporting resolution hampers everyone



"Using a cutoff of 20, around 35% are positive."

We don't know if the distribution is thin-tailed or heavy-tailed.

Clinicians & AI are left to instinct and bias to contextualize results.

@DanaMLewis | Dana@OpenAPS.org

This means:

- If we can't diagnose and/or characterize the population, this influences recruitment and population identification for drug development and trials
- Limits interpretability of study/clinical trial data
- Possibly studies the wrong thing for the wrong audience
- Limits downstream translation into clinical practice

Let's think about:

- Regulators:
  - Make sure policies don't penalize for RWE
  - Make easier pathways for repurposing
- Companies & researchers:
  - Record & retain & report more granular data
  - Make open datasets…and use pre-prints
  - **Actually make data available** ~~on request~~
  - Create pathways to draw in real-world data
  - Existing drugs still have a lot of room for improvement (titration, repurposing, targeting, etc.)

@DanaMLewis | Dana@OpenAPS.org

There is a lot of unknown

that we can know now – or soon –

by using AI.

@DanaMLewis | Dana@OpenAPS.org

We can’t fix everything (yet),

but we should do what we can -

now

@DanaMLewis | Dana@OpenAPS.org

AI in Biological Product Development

An FDA Hybrid Public Workshop

October 7, 2025
National Press Club, Washington, D.C

Thomas Osborne, MD
Chief Medical Officer: Microsoft (Federal)
Clinical Professor *(*adjunct): Stanford

How is Biomedical Knowledge Generated? & How **AI** Can Help?

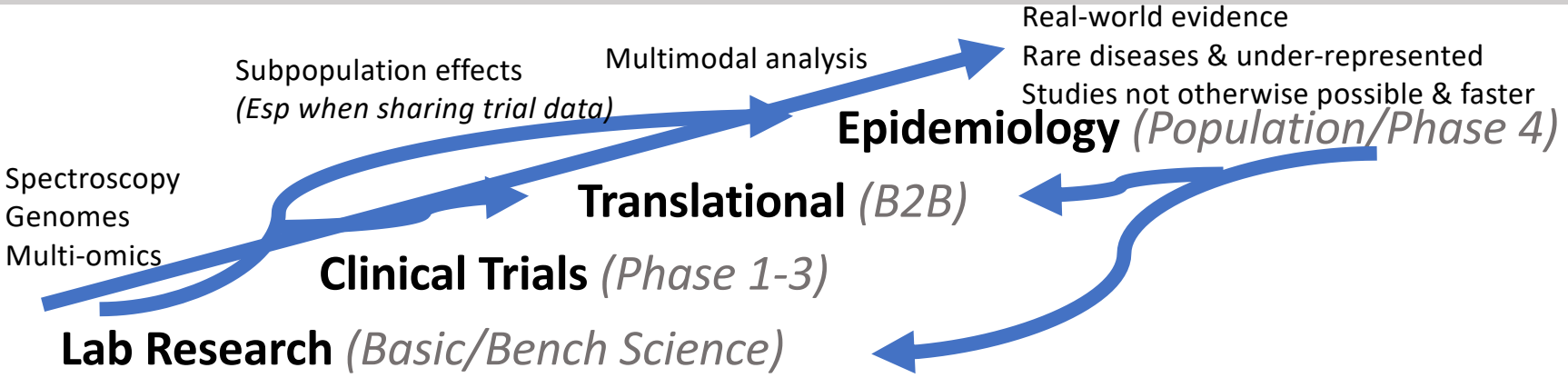
Uncover patterns humans didn't hypothesize

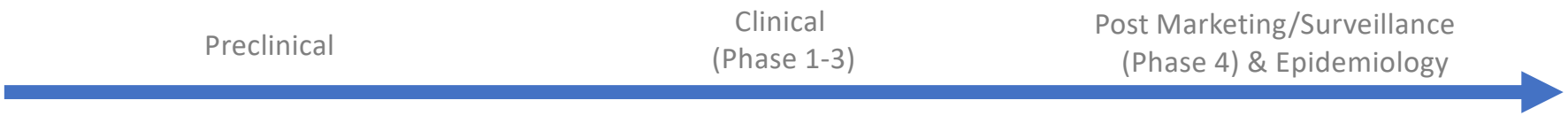
ML

- *Complex pattern finding*
- *Generate new hypotheses*
- *Virtual experiments*
- *Modeling/simulations*
- *Predictions*
- *Informed Rx discovery*

LLM

- *Unstructured to structured*
- *Summarizing/synthesizing data*
- *Language of nature/biology (bio-transformers)*
- *Talk with data*
- *Informed Rx discovery*



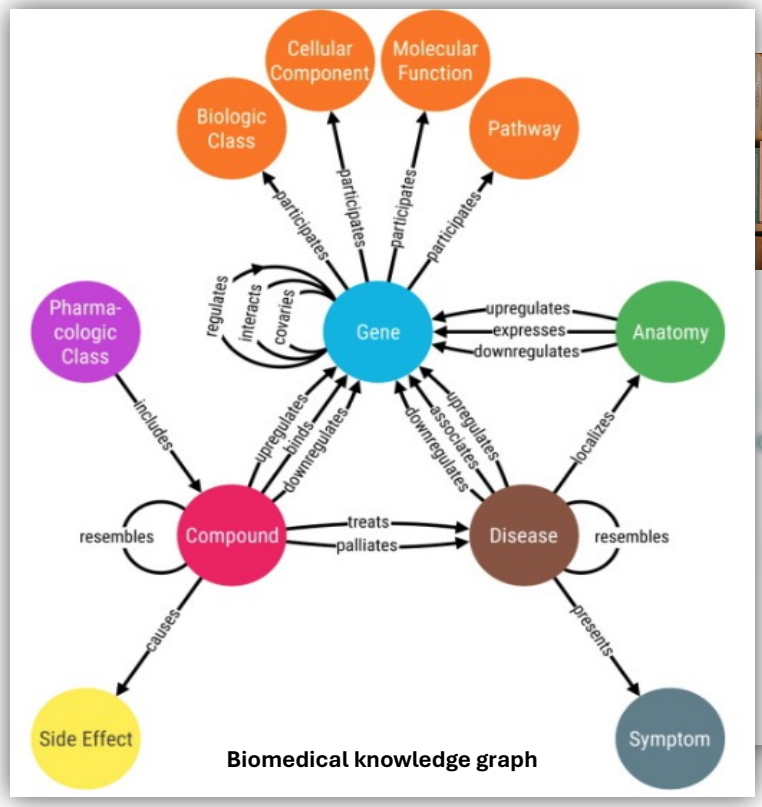

Preclinical
Clinical
(Phase 1-3)
Post Marketing/Surveillance
(Phase 4) & Epidemiology

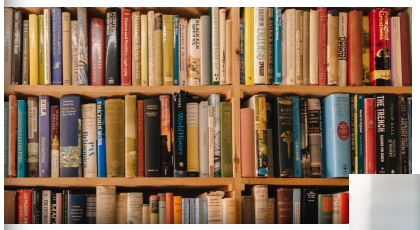
Preclinical

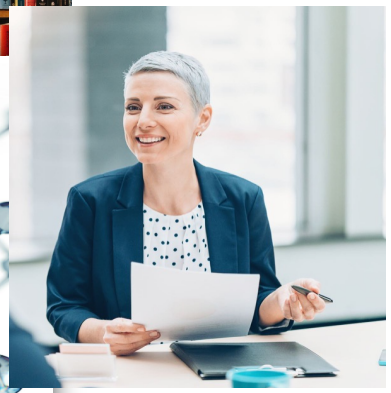
Clinical
(Phase 1-3)

Post Marketing/Surveillance
(Phase 4) & Epidemiology

Map disease pathways and find new therapeutic targets



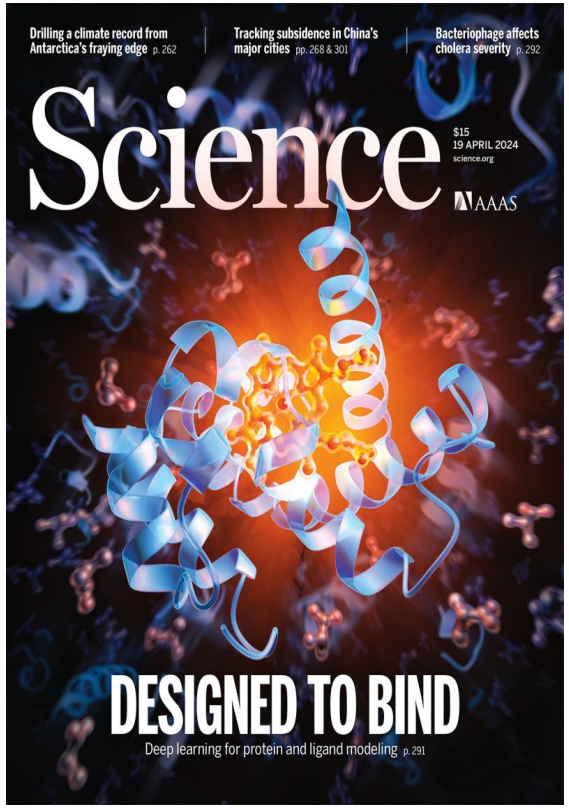




multi-omics datasets

Nicholson DN, Greene CS. Constructing knowledge graphs and their biomedical applications. Computational and structural biotechnology journal. 2020 Jan 1;18:1414-28.

Preclinical → Clinical (Phase 1-3) → Post Marketing/Surveillance (Phase 4) & Epidemiology



- Reveal hidden target **binding sites**
- ***Virtual screening*** */prediction of millions of molecules that are likely to bind to targets*
- Structure-based **drug design**
- Structure–activity relationships to predict a compound's **ADMET** *properties:*
  - *(Absorption, Distribution, Metabolism, Excretion, Toxicity)*

Preclinical

Clinical (Phase 1-3)

Post Marketing/Surveillance (Phase 4) & Epidemiology

arXiv > cs > arXiv:2509.06602

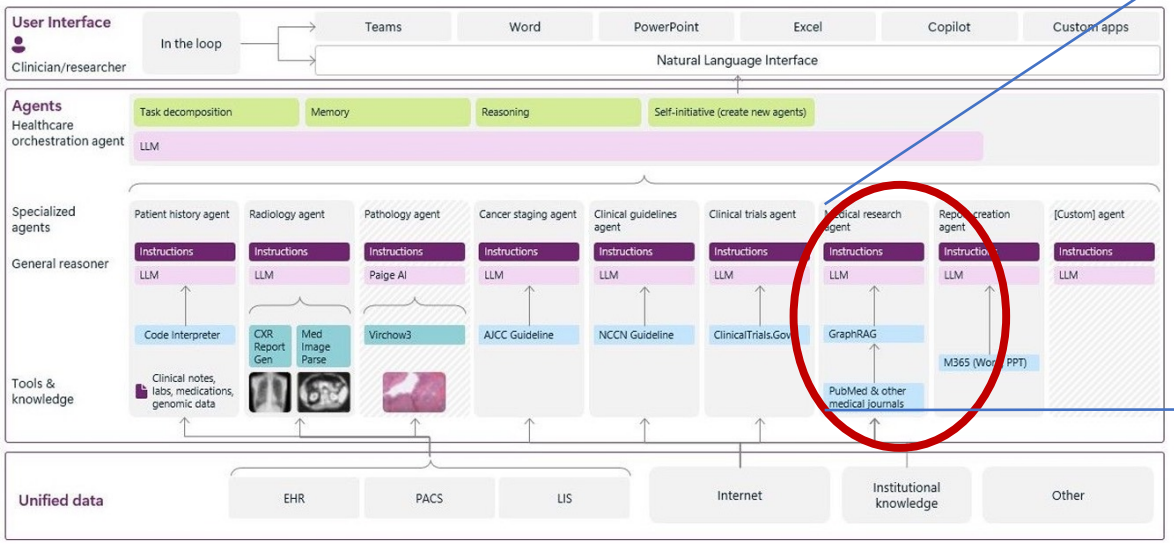
Search... Help | Ad

Computer Science > Machine Learning

[Submitted on 8 Sep 2025 (v1), last revised 11 Sep 2025 (this version, v2)]

Demo: Healthcare Agent Orchestrator (HAO) for Patient Summarization in Molecular Tumor Boards

Matthias Blondeel, Noel Codella, Sam Preston, Hao Qiu, Leonardo Schettini, Frank Tuan, Wen-wai Yim, Smitha Saligrama, Mert Öz, Shrey Jain, Matthew P. Lungren, Thomas Osborne



Preclinical

Clinical (Phase 1-3)

Post Marketing/Surveillance (Phase 4) & Epidemiology

EHR mining to identify patients who match trial criteria

Preclinical

Clinical (Phase 1-3)

Post Marketing/Surveillance (Phase 4) & Epidemiology

Adaptive trial designs *(supported by AI algorithms)*

Monitor incoming data & adjust the trial (e.g., drop a futile arm or recalibrate dosage)

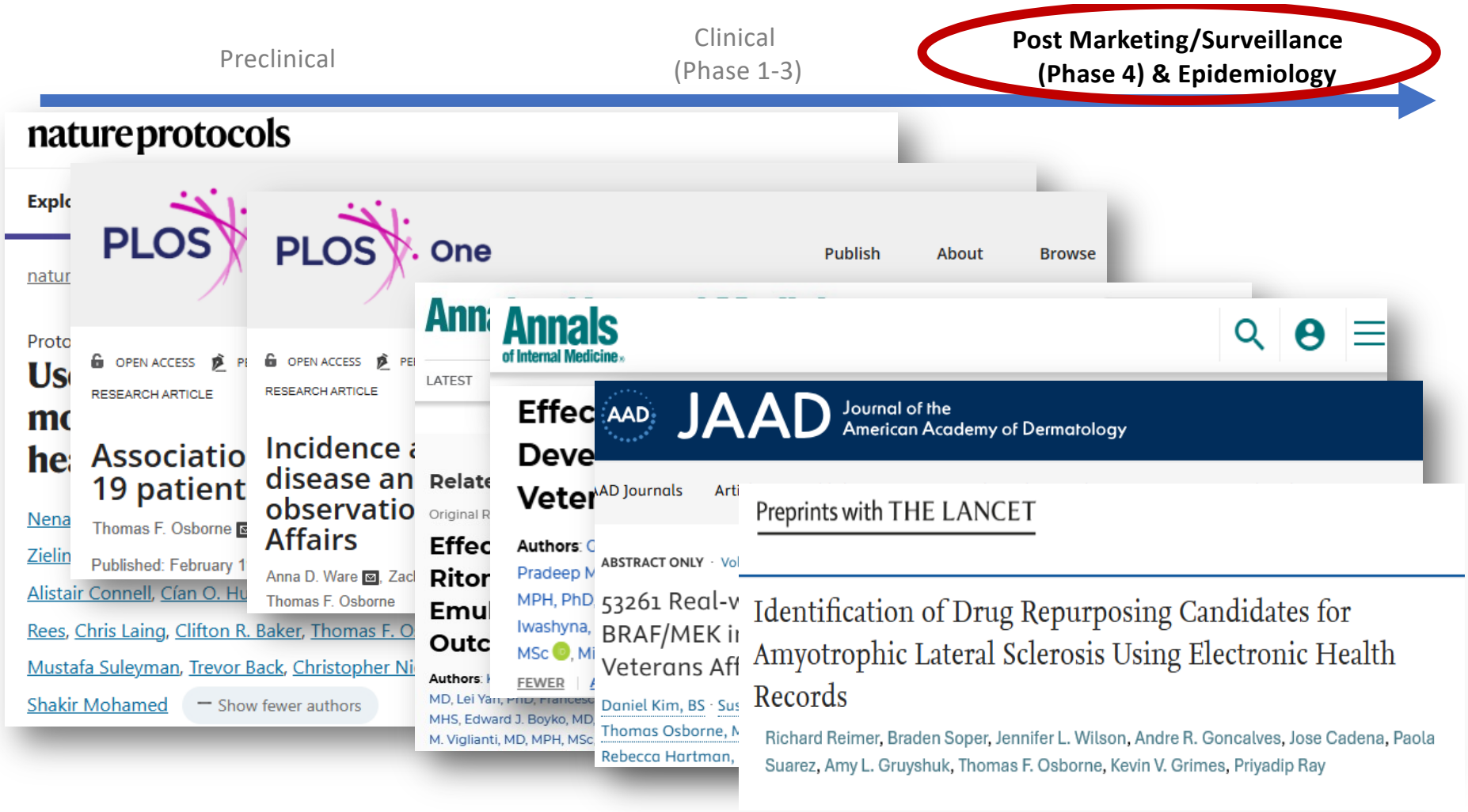
Preclinical

Clinical (Phase 1-3)

Post Marketing/Surveillance (Phase 4) & Epidemiology

Digital Twin Control Arms

Example:
Participants may all receive the experimental drug.
AI generates a **matched "digital twin"** outcome for each patient as if they had gotten placebo.


Preclinical
Clinical
(Phase 1-3)
Post Marketing/Surveillance
(Phase 4) & Epidemiology
nature protocols
PLOS One
Publish
About
Browse
OPEN ACCESS
RESEARCH ARTICLE
Thomas F. Osborne
Anna D. Ware
Annals of Internal Medicine
JAAD
Journal of the American Academy of Dermatology
Preprints with THE LANCET
Identification of Drug Repurposing Candidates for Amyotrophic Lateral Sclerosis Using Electronic Health Records
Richard Reimer, Braden Soper, Jennifer L. Wilson, Andre R. Goncalves, Jose Cadena, Paola Suarez, Amy L. Gruyshuk, Thomas F. Osborne, Kevin V. Grimes, Priyadip Ray

Thank you!

Questions?

PLEASE SUBMIT HERE:

BREAK

We will reconvene at 10:55 a.m.

Session 2: Data Quality, Reliability, Representativeness, and Access in AI-Driven Drug Development

Moderator:
Hussein Ezzeldin
Associate Director, Advanced Technologies, OBPV, CBER
U.S. Food & Drug Administration

Moderator:
Lanyan Fang
Supervisory Pharmacologist, OGD
U.S. Food & Drug Administration

Wesley Anderson,
Quantitative Medicine Scientist
The Critical Path Institute

Michelle Longmire
Co-Founder and Chief Executive Officer
Medable, Inc.

Sheraz Khan,
Senior Director, Generative AI
Pfizer

CRITICAL PATH INSTITUTE 20 YEARS OF IMPACT

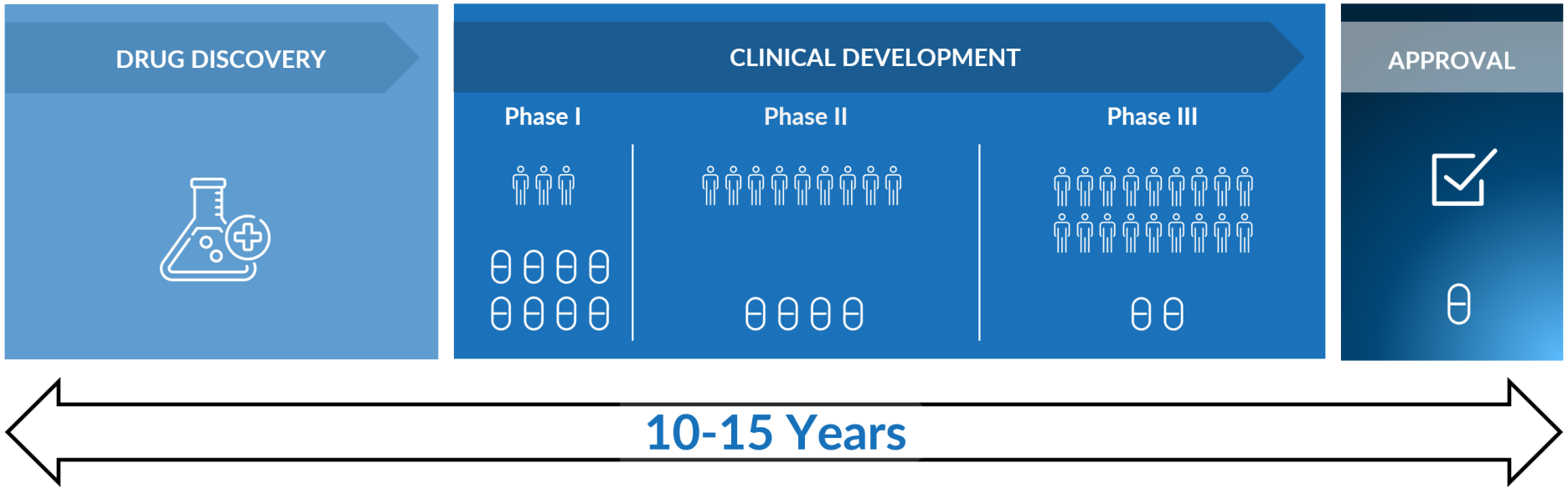
Building Reliable AI for Drug Development: Insights from The Critical Path Institute

Wes Anderson, Ph.D.

Scientist, Quantitative Medicine

The Problem

Lab discovery to regulatory-approved therapy is slow and costly with a high failure rate*



*90% drugs in clinical development ultimately fail

The Solution

CRITICAL PATH INSTITUTE 20 YEARS

Forge diverse collaborations to identify hurdles

- Industry & research scientists
- Regulatory agencies
- Patient and patient advocacy groups
- Industry leaders
- Donors and philanthropic community

Craft tools and solutions

Achieve regulatory success

FDA

- 8 Qualifications Decisions
- 11 Letters of Support
- 1 Fit for Purpose endorsement

EMA

- 9 Qualification Opinions
- 10 Letters of Support

PMDA

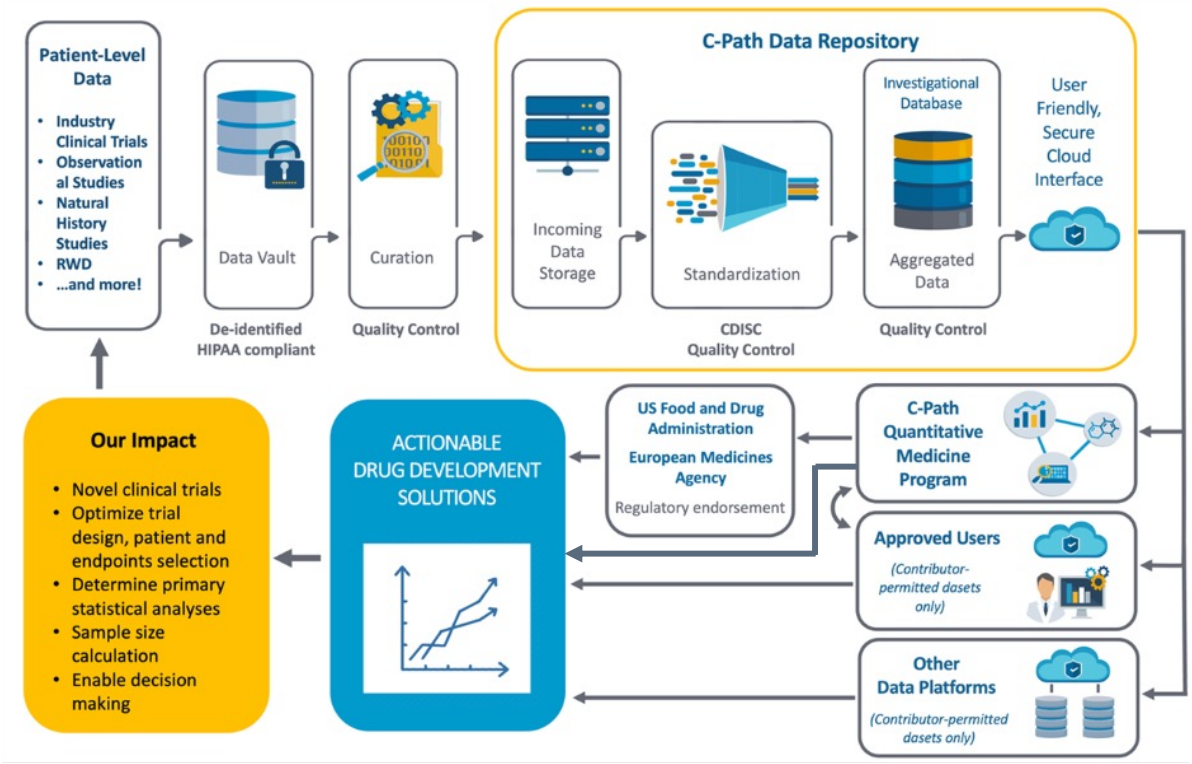
- 1 Qualification Decision

Clinical Impact

The Data Pillars of Responsible AI-Driven Drug Development

Pillar	Why it Matters in AI
Data Quality	Garbage in, Garbage out – bad inputs degrade model performance
Data Reliability	Accurate, complete, and consistent data ensure proper performance
Data Representativeness	Prevent bias, promote fairness, and ensure reliability
Data Access	Without usable, shareable data, model training is constrained

Quantitative Medicine

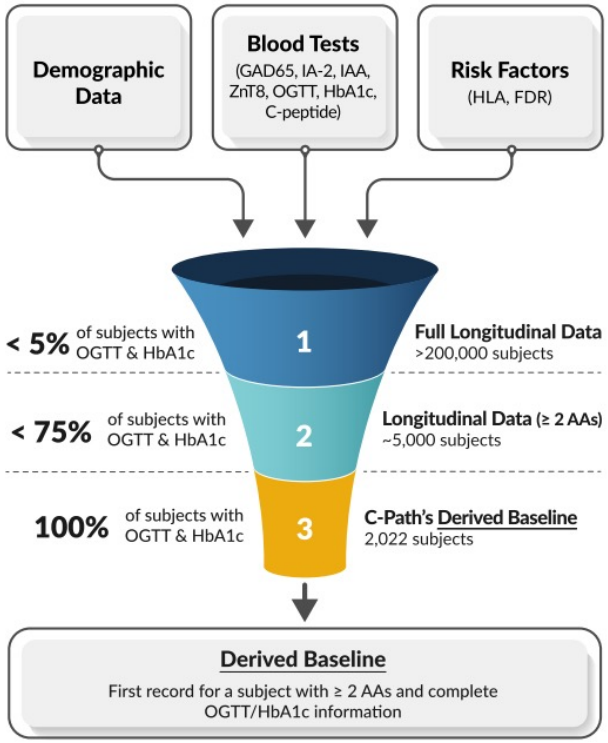


Use Case Example: T1DCTE Tool

Purpose: Developed to support clinical trial enrichment strategies using islet autoantibodies (AAs) and clinical features to optimize prevention trial design in T1D.

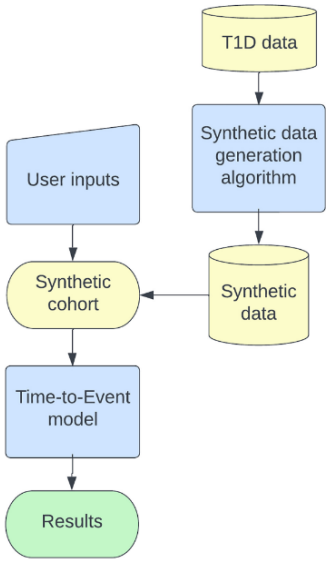
Uses a **time-to-event** predictive model for T1D onset

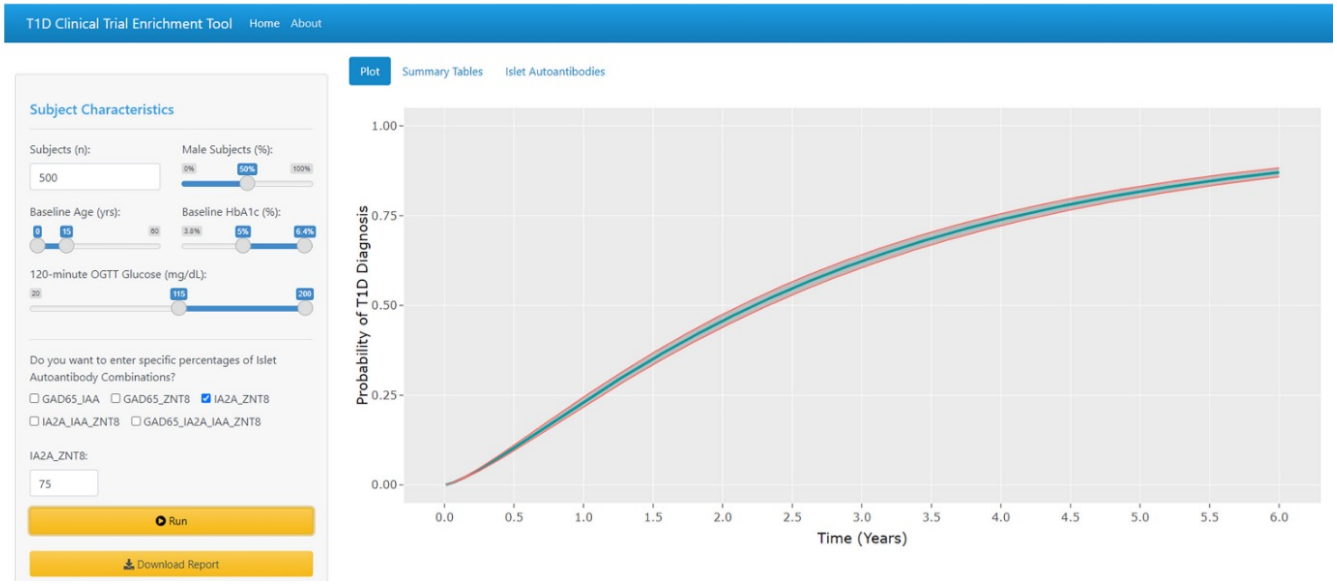
Model estimates the **time-varying probability of T1D diagnosis**



Use Case Example: T1DCTE Tool (cont) – Synthetic Data Generation

- To increase accessibility of the tooling, patient privacy must be considered





- > 2 islet autoantibodies now a qualified biomarker for clinical trial enrichment in T1D

Link to the tool: https://t1d-cte.c-path.org/

Pauley M, Henscheid N, David SE, Karpen SR, Romero K, Podichetty JT; Type 1 Diabetes Consortium (T1DC). T1dCteGui: A User-Friendly Clinical Trial Enrichment Tool to Optimize T1D Prevention Studies by Leveraging AI/ML Based Synthetic Patient Population. Clin Pharmacol Ther. 2023 Sep;114(3):704-711. doi: 10.1002/cpt.2976. Epub 2023 Jun 30. PMID: 37326252.

The Clinical Impact

Indication	C-Path's Solution	Clinical Impact
Alzheimer's disease	2 Clinical Trial Simulation (CTS) Tools, 2 biomarkers	**First disease-modifying drugs**
Tuberculosis	Multiple quantitative tools	**First new drug and drug regimen**
Polycystic Kidney Disease	CTS Tool and model-based imaging biomarker	**First disease-modifying drug**
Type 1 Diabetes	Model-based biomarkers	**First prevention drug**
Irritable Bowel Syndrome with constipation	Diary for symptoms	**Label expansion for symptomatic drug**
Friedrich's ataxia	Data & Analytics Platform to support generation of external controls	**First disease-modifying drug**
Duchenne Muscular Dystrophy	5 disease progression models, 3 CTS tools, 1 biomarker	**First non-steroidal treatment for all variants**
Kidney Transplantation	Composite biomarker endpoint	**Accelerated treatment options with improved drug tools and trial strategies for faster clinical development**
Parkinson's disease	3 CTS tools, 1 biomarker, multiple DHT solutions	
Huntington's disease	Staging system, 3 disease progression models	

In closing,

The real foundation of the future of drug development isn't just algorithms. It's data, and not just any data:

Data that is clean, complete, and curated.

Data that reflects real-world diversity.

Data that is trusted, shared, and governed.

The proper data infrastructure makes good models possible through data that is fit-for-purpose, regulatory-grade, and designed to serve science, not skew it.


CRITICAL PATH
INSTITUTE
Advancing Drug Development.
Improving Lives. Together.
c-path.org

Using Data to address bottlenecks in drug development

Rebuilding Trial Execution with Data and Agentic AI.

The White Space: Where Data Gaps Slow Drug Development

CRAs spend 31 hrs/month on administrative burden alone*

Protocol amendments add 4–6 months and cut enrollment by >30%*

Study start-up averages 8 months from site ID to activation*

Every day of delay costs sponsors $600K–$8M*

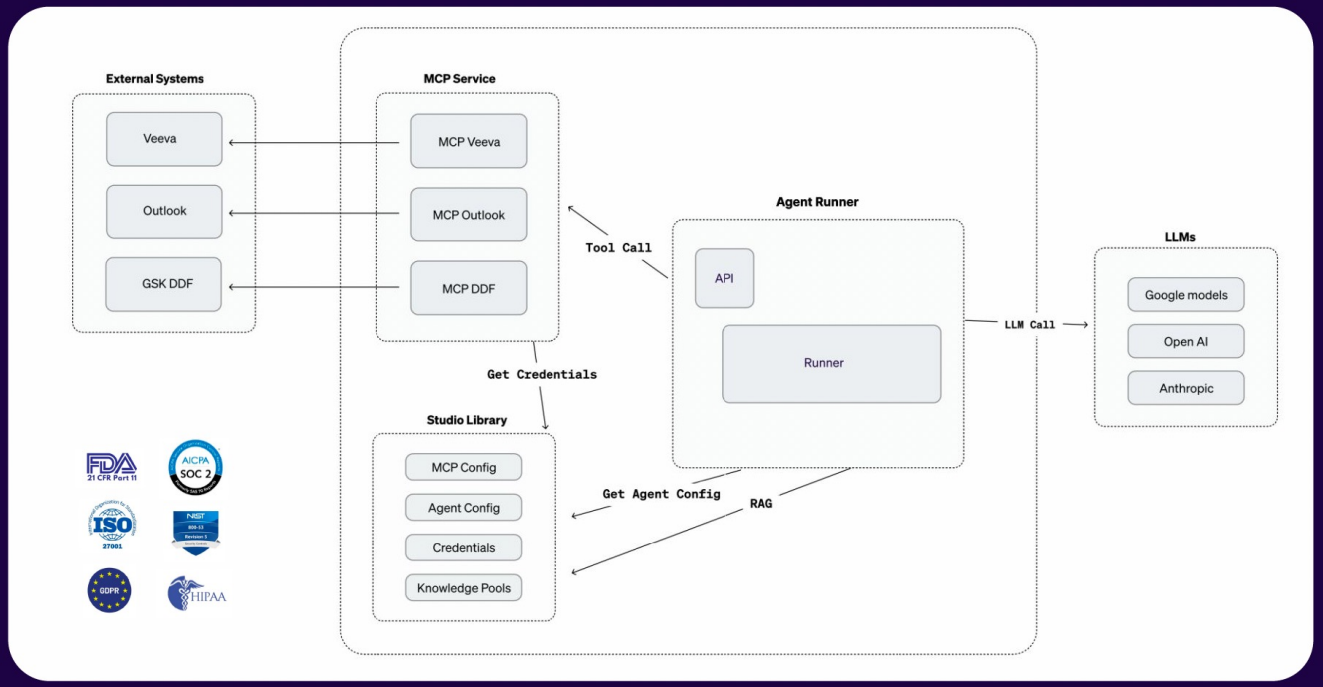
Tufts Center for the Study of Drug Development (CSDD) Impact Reports (various)

CHALLENGE

We’ve reached the limit of human-only clinical development

Agents for life sciences

Manual & Disconnected → Agentic & Human-in-the-Loop



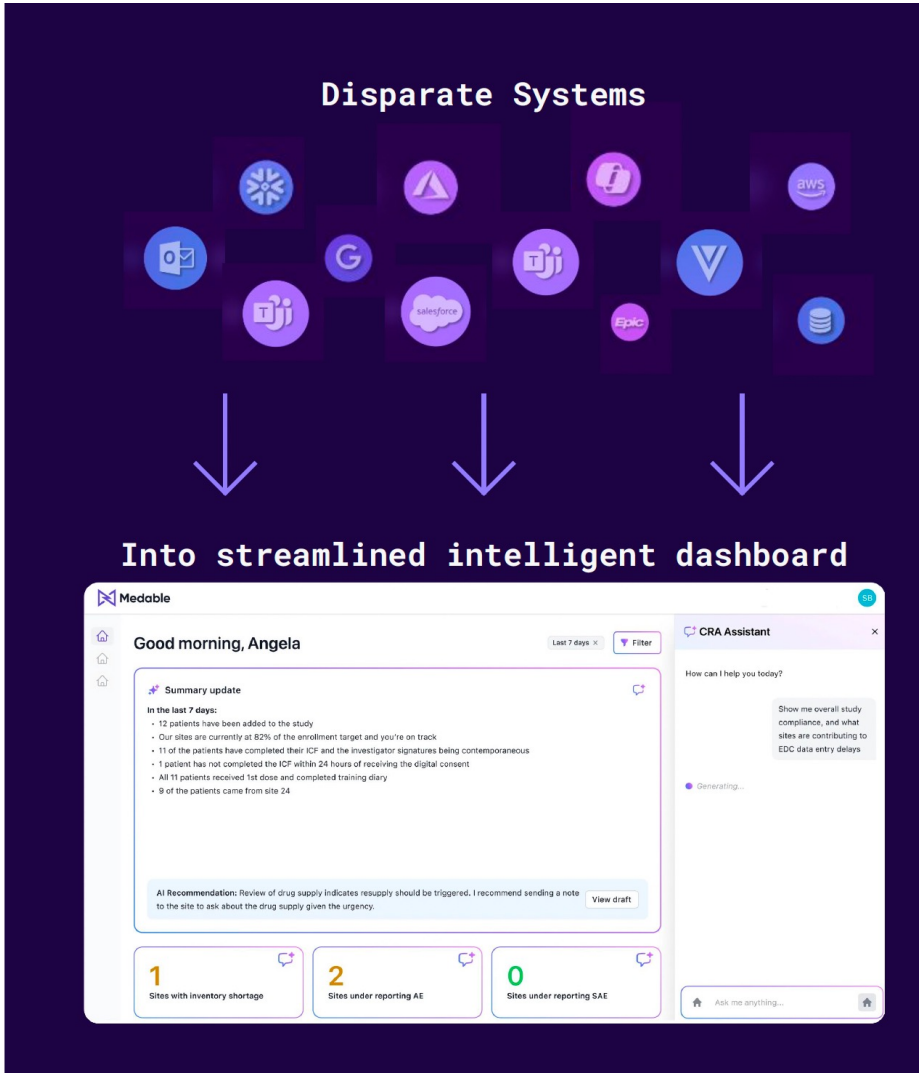
Connecting Systems in Clinical Research

Example: the CRA Agent

Purpose:
Assists Clinical Research Associates in preparing for and conducting site visits.

How it works:
→Pulls data from CTMS, EDC, RBQM systems
→Identifies site-specific risks
→Suggests follow-up actions
→Pushes them to downstream systems.

Value:
Saves CRAs time, enhances site oversight, increases quality signal detection.



What works and what's next

What works

Seamless data integration &
Automated risk detection

What's next

CAPTCHA paradox: Agents blocked by human verification systems. API access through MCPs solves this problem

Dependence on the LLM's reasoning quality: Agents can only reason as well as their model;

Shifting mindsets from deterministic to agentic systems: Agentic models mirror human variability. Regulations built for deterministic software now face the challenge of governing probabilistic, reasoning-based systems while still ensuring patient safety.

Session 2: Data Quality, Reliability, Representativeness, and Access in AI-Driven Drug Development

Building the Foundation: Shared Foundation Model for Digital Endpoints in Drug Development

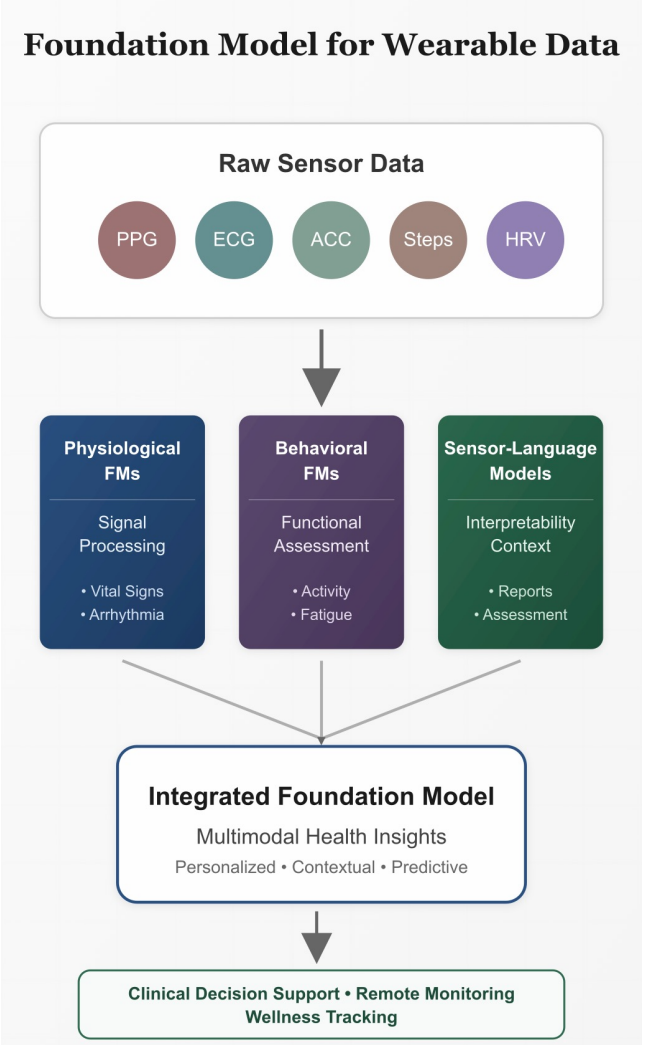
Sheraz Khan, Ph.D.
David Isom, B.Sc.

Pfizer Inc.

Artificial Intelligence in Drug & Biological Product Development

FDA/CTTI Workshop

October 7, 2025



Task-Specific Narrow AI approaches fail to address the complexity and diversity of real-world digital health data

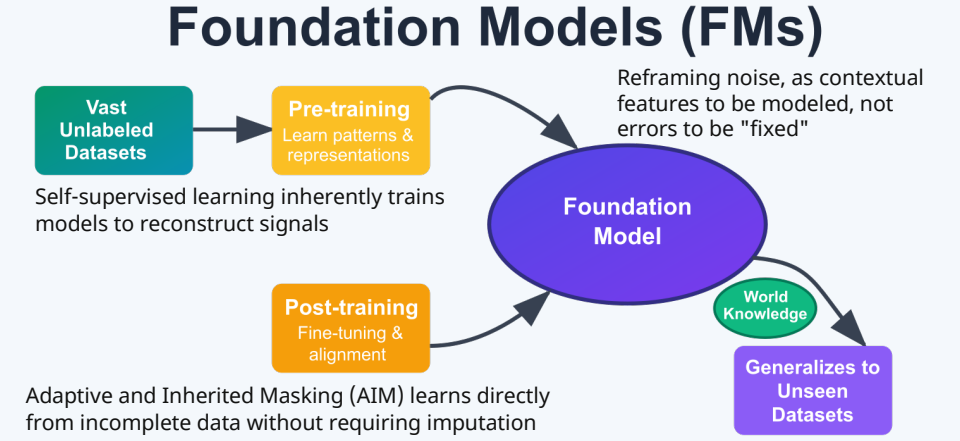
Limitations of Narrow AI

Task-Specific and Brittle

Poor generalization across diverse populations, devices, and "in the wild" conditions

Scalability Challenges & Poor Generalization

Heavy reliance on manual feature engineering and task specific large labeled datasets

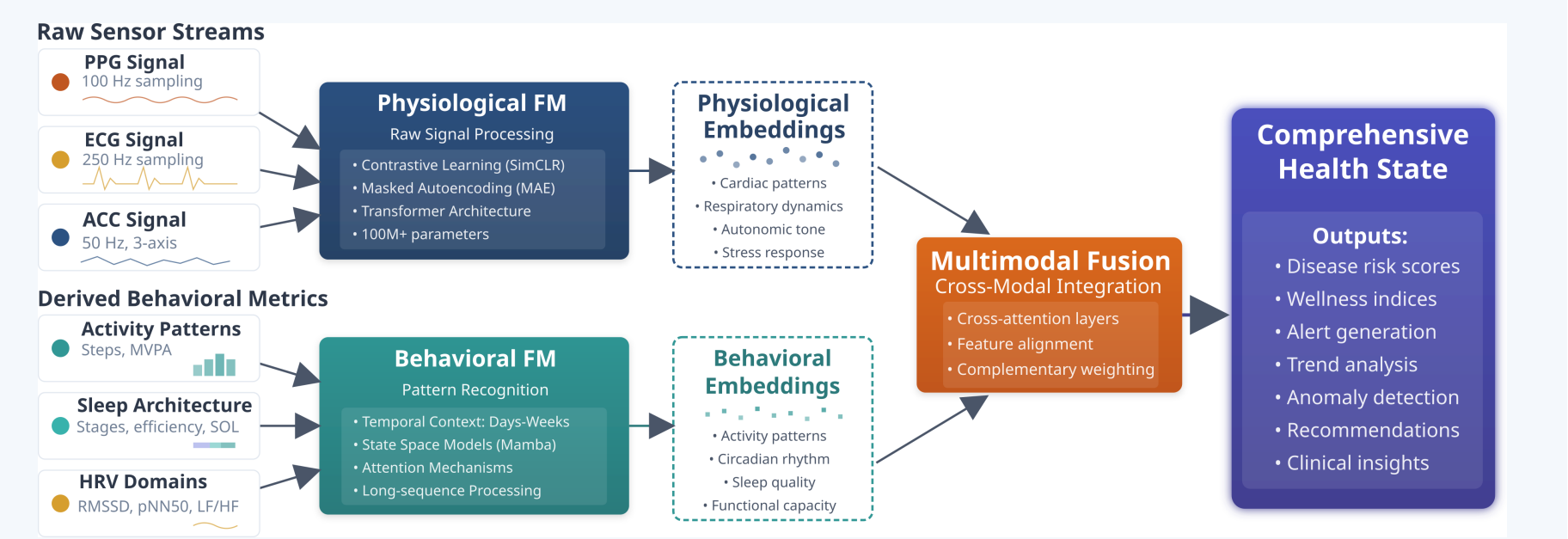


The Spectrum of FMs for Digital Measures

Type	Data Source	Primary Function	Example Applications
Physiological FMs	Raw Sensor Data (PPG, ECG, ACC)	Signal processing and feature extraction	Vital sign monitoring, arrhythmia detection
Behavioral FMs	Derived Metrics (Steps, HRV, Sleep)	Functional capacity assessment	Activity patterns, fatigue monitoring
Sensor-Language Models	Numerical + Semantic Data	Interpretability and context	Personal coach, health assessment

Lee et at., 2025, Erturk et al., 2025, Khasentino et al., 2025, Cosentino, et al., 2024

Foundation models enable unprecedented scale, adaptability, and efficiency in drug development



KEY ADVANTAGES

Robustness & Generalizability
Captures conserved patterns across populations, devices & trial sites

Enhanced Sensitivity
Uncovers subtle temporal dynamics: circadian, motor, cardiovascular

Efficient Adaptation
Rapid fine-tuning with minimal labeled data requirements

CLINICAL TRIAL IMPACT

Patient Stratification
Digital phenotyping identifies responder subgroups

Efficacy Endpoints
Continuous real-world function measures

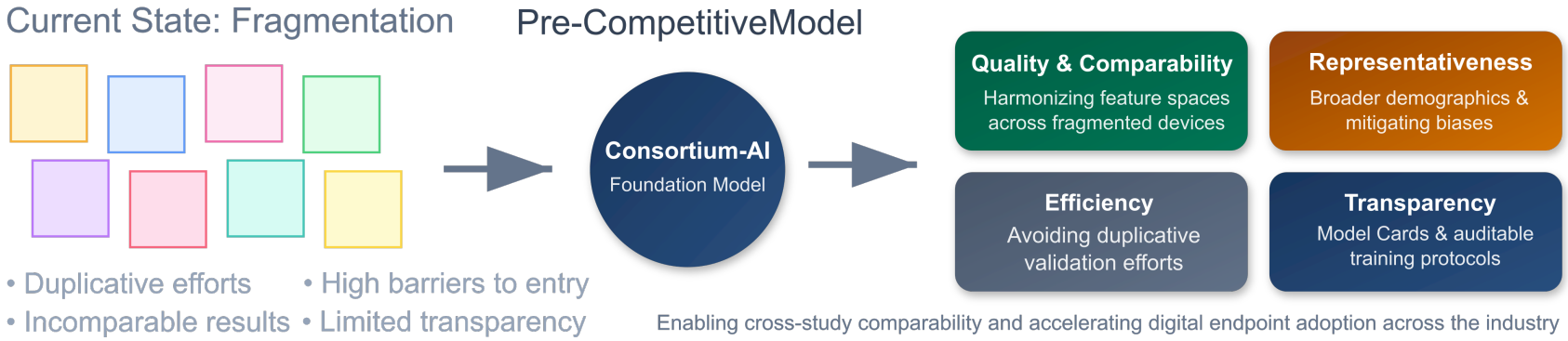
Safety Monitoring
Real-time physiological perturbation detection

Novel Biomarkers
Discovery beyond predefined features

Lee et at., 2025, Erturk et al., 2025, Khasentino et al., 2025, Cosentino, et al., 2024

A pre-competitive shared responsibility, foundation model ecosystem addresses critical gaps in quality, equity, and efficiency

Pre-Competitive Foundation Model Ecosystem: Bridging Critical Gaps in Quality, Equity, and Efficiency



Enabling cross-study comparability and accelerating digital endpoint adoption across the industry

Shared Responsibility framework for Foundation Models

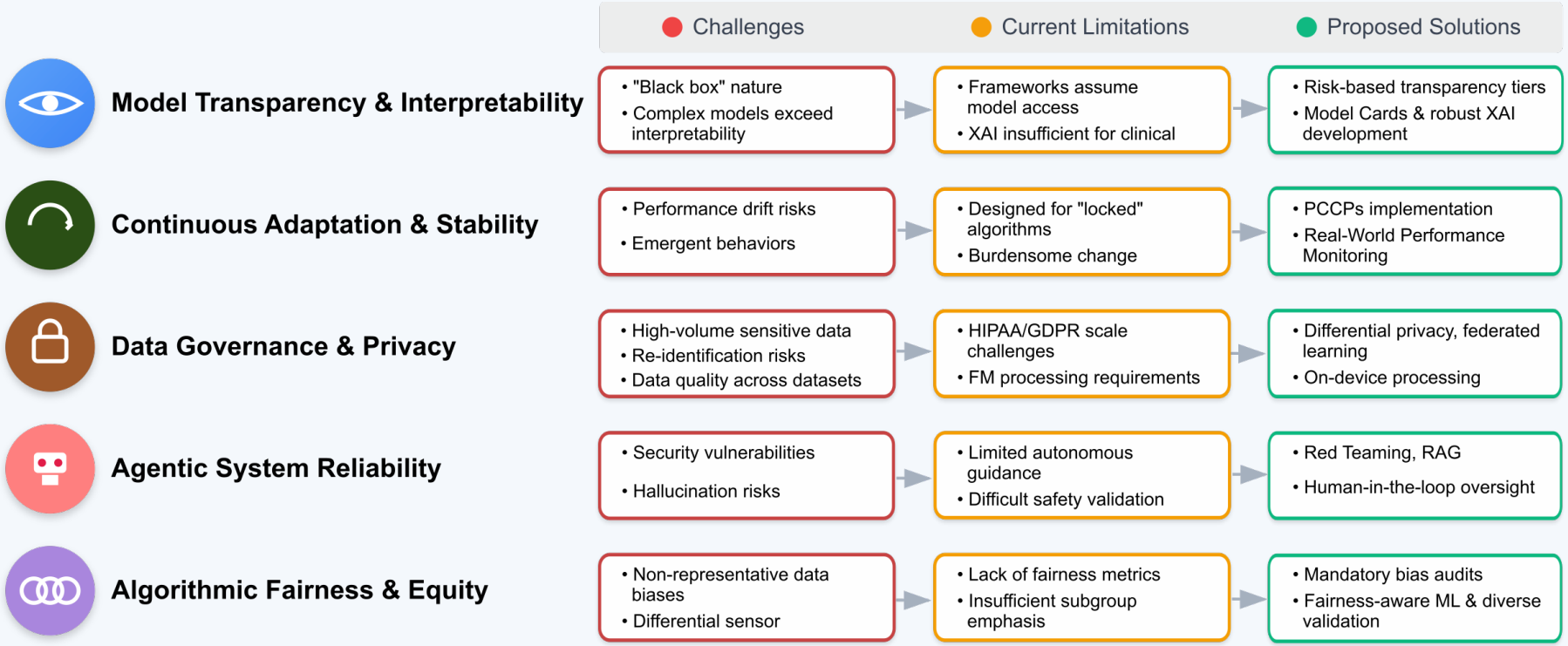
FM Developer (Model Steward)

- Ensuring robustness of base model
- Comprehensive documentation
- Fairness and bias mitigation
- Transparency in training methodology

Application Developer (Sponsor)

- Validating fine-tuned model for specific clinical CoU
- Context-specific performance evaluation
- Regulatory submission for specific use case
- Post-market surveillance

The path forward, regulatory and governance challenges for foundation models in drug development



Industry collaboration in the next 12 months could establish the foundation for digital health innovation

Key Takeaways

Transformative Potential

FMs offer a paradigm shift for developing objective, sensitive digital biomarkers

Shared Benefit

A pre-competitive foundation model raises the floor on reliability and equity

Regulatory Alignment

Collaborative approach streamlines regulatory pathways and validation

Collaboration Model

Consortium & Governance Structure

Establish cross-industry, academic, and regulatory working groups

Data Contributions Framework

Share privacy-preserving frameworks for sharing diverse, de-identified sensor data

Open Evaluation, benchmarks & Endpoints

Public benchmark endpoints and standardized documentation templates

What Next

Join the Steering Group

Trusted partnership mediated governance and standardization working group for wearable FMs

Nominate Datasets

Contribute diverse datasets and propose benchmark endpoints for evaluation

Initiate Pilot Project

Participate in shared FM benchmarking focused on fairness and generalizability

Questions?

PLEASE SUBMIT HERE:

LUNCH

We will reconvene at 1:35 p.m.

Session 3: Model Performance, Explainability, Transparency, and Interpretability in AI-Driven Drug Development

Moderator:
Hao Zhu
Director, Division of Pharmacometrics, OTS/OCP/DPM
U.S. Food & Drug Administration

Moderator:
Nicole Mahoney
Executive Director, Regulatory Policy, and Global Regulatory Policy Lead, Data and Digital Technologies
Novartis

Prasanna Rao
Chief Products and Innovation Officer
Saama

Andrea Downing
President and Co-Founder,
The Light Collective

Oanh Dang
Regulatory Science and Applied Research Lead, Office of Surveillance and Epidemiology, CDER
U.S. Food & Drug Administration

FDA/CTTI AI meeting

AI model performance, explainability, and interpretability across the drug development life cycles

Bridging Performance, Interpretability & Explainability for Regulatory trust

Performance	Interpretability	Explainability
• How well the model works measured usually with accuracy, efficiency & speed	• How well humans understand why it works	• How well the model (or agent) can communicate its decision in a way stakeholders trust

Performance earns attention, explainability earns trust.

AI Use cases

Planning phase

- Site selection
- Study builder
- Protocol drafting
- Data Review plan generation
- Metadata mapping

Trial conduct phase

- Data Management
  - Anomaly detection
  - Query Management
- Monitoring
  - RBQM
  - Data surveillance

Submission phase

- Analysis & Reporting
  - SDTM generation
  - Tables Listing and figures analysis
- CSR generation

Clinical Development

Approach to Responsible AI

Operationalizing Core Principles Through Quality & Development Standards

AI Development & Lifecycle Management

- Lifecycle Management
- Security and Privacy by Design
  - Security and Architecture Reviews
  - Regulatory, Compliance, and Data Privacy Reviews
  - Secure Coding Practices

Verification, Validation & Monitoring

- Model Verification
  - Fairness
  - Performance
  - Security
- Model Validation
  - Valid & Reliable
  - Explainable
  - Interpretable
- Monitoring

Acceptable Use

- AI Tool Evaluation
  - Vendor Management
  - Tool Evaluation
  - Access Management & Reviews
  - Human in the Loop philosophy
- Training and Best Practices
- Implementation & Monitoring

AI Governance Framework

Explainability and Intrepretability

Moving from “Black box” to “Glass box”

	Transparency	Traceability
 **Explainable AI**	• Shows which variables or inputs most influenced a prediction • Example: site selection model highlighting patient density and investigator track record	• Focused on internal logic and reproducibility • Example: anomaly detection flags linked to actual source data anomaly to explain
 **Interpretable AI**	• Converts model reasoning into language that regulators, clinicians, and patients can understand • Example: “black box to glass box” explaining the “thinking,” “planning” and “action” of a reasoning model	• Links model outputs back to underlying datasets and decisions • Example: CSR generation where each sentence maps to trial data tables

Implementation of Explainable AI

Integrated into our Model Development Life Cycle

Planning & Design

Define the model's **intended use**

Justify the **model choice**

Identify specific **input and output data**

Embed an **interpretable validation** approach

Establish **human oversight** from the start

Development

Prioritize a **human-in-the-loop** approach

Document model decisions and their rationale

Balance explainability with **intellectual property** protection

Validation

Validate model **accuracy, robustness,** and **safety**

Provide a **Confidence Score** for each prediction

Align with **FDA guidance** on Good AI/ML Validation Practices

Use **statistical items** and citations to prove output validity

Monitoring & Maintenance

Continuously monitor for **drift** and **hallucinations**

Trigger **retraining** when performance degrades

Provide **transparency** and **communication to user** via user manuals

Ensure accountability with an **AI Governance Framework Policy**

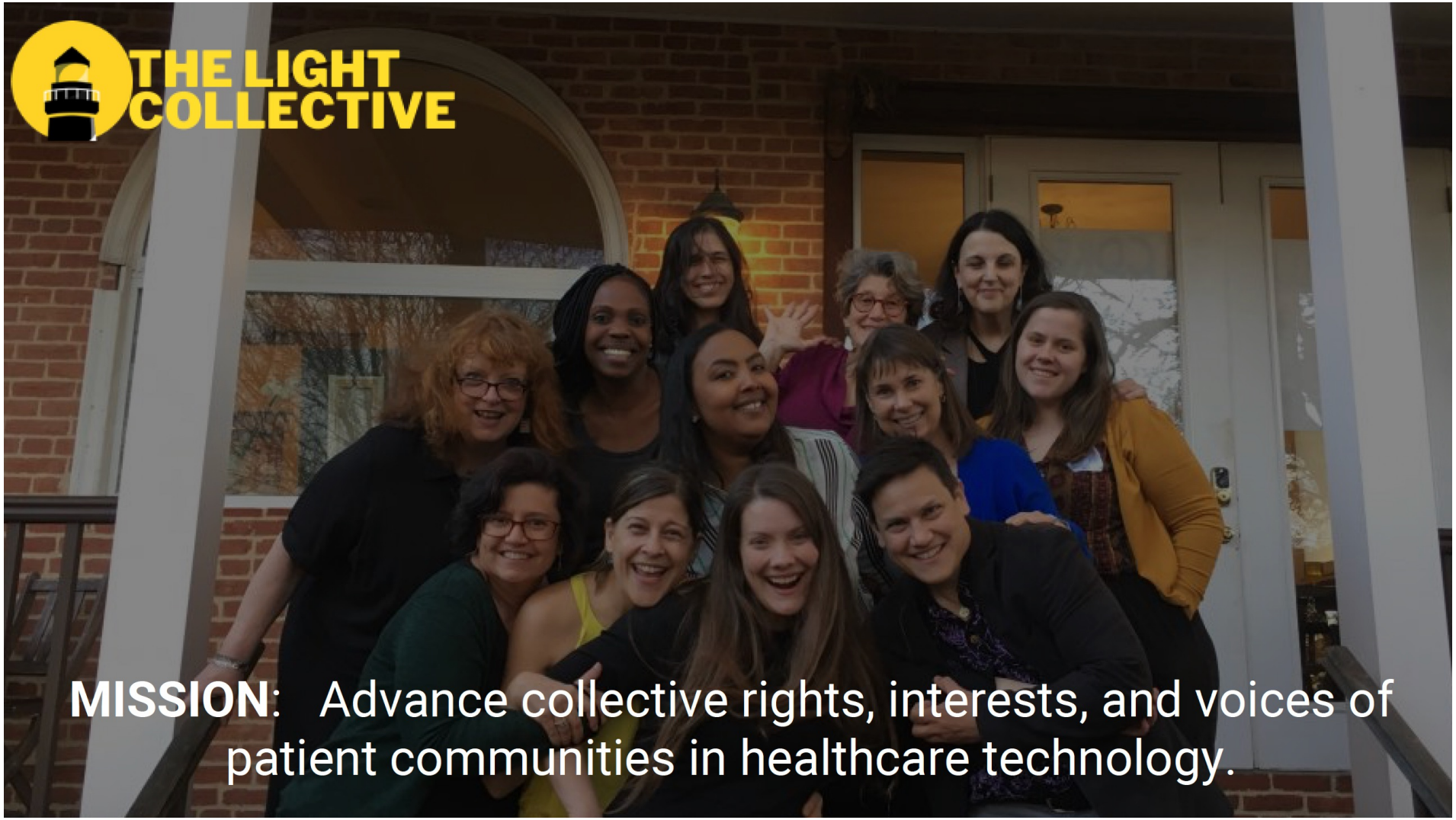
Thank you!

THE LIGHT COLLECTIVE

Patients In The Drug Development AI Loop

Performance, Explainability, Transparency, and Interpretability For Patients As The Ultimate Stakeholders

Andrea Downing
www.lightcollective.org


THE LIGHT COLLECTIVE
MISSION: Advance collective rights, interests, and voices of patient communities in healthcare technology.

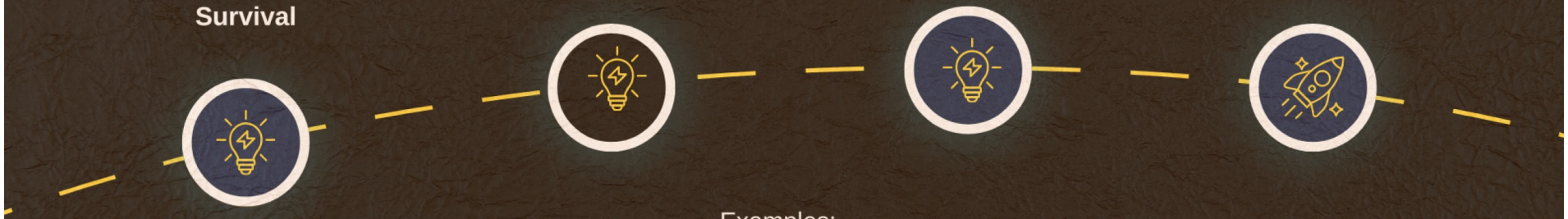
Patient Partnered Innovation: Historical Perspective

1. Built Community-Led Networks for Survival

2. Became Experts in the Science

3. Transformed Clinical Trials & Policy

5. Targeted Leverage



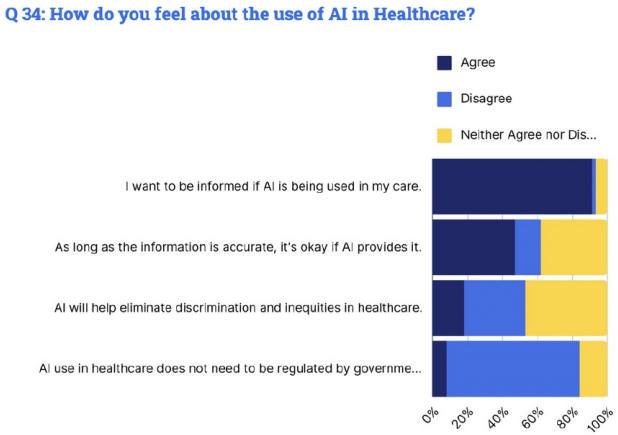
Examples:

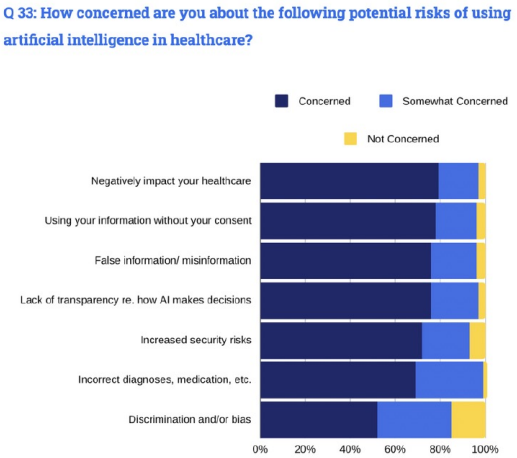
HIV Community → AZT Got FDA Approval
Type 1 Diabetes → Continuous Glucose Monitoring
Cystic Fibrosis Community ---> Kalydeco
Breast Cancer Community → Herceptin
Long Covid Community → Off Purpose Targeted Therapies

How might we design and govern AI to serve patients as the ultimate stakeholders?

91% of patients in our network wanted to be informed if AI is being used in their care.

.....But HOW do we do this?





Source: The Light Collective. Tangled in the Web 2024: Patient Safety & Privacy Concerns When Seeking Knowledge, Community, and Support Online.

www.lightcollective.org

THE LIGHT COLLECTIVE

Explaining, interpreting, or building transparency for patients at scale ….

If you let AI algorithms sit in an active system, it can drift to better or drift to worse — and you don't know unless you measure it.

Known risks of generative AI in healthcare — the 'known unknowns'

Lack of transparency: Specific to LLMs, there is little or nil transparency on pre-training data as developers rarely share this openly. Contention over copyright and intellectual property infringement has surfaced when developers use copyrighted data or paywalled documents in the pre-training process without seeking approval. Model distillation is a technique designed to enhance computational efficiency of LLMs by taking the outputs of a larger, more complex model to teach a smaller model to achieve a similar performance. Model or knowledge distillation using proprietary, closed sourced models may infringe on intellectual property laws. There are some models that use synthetic data for pre-training, which may be undetectable.

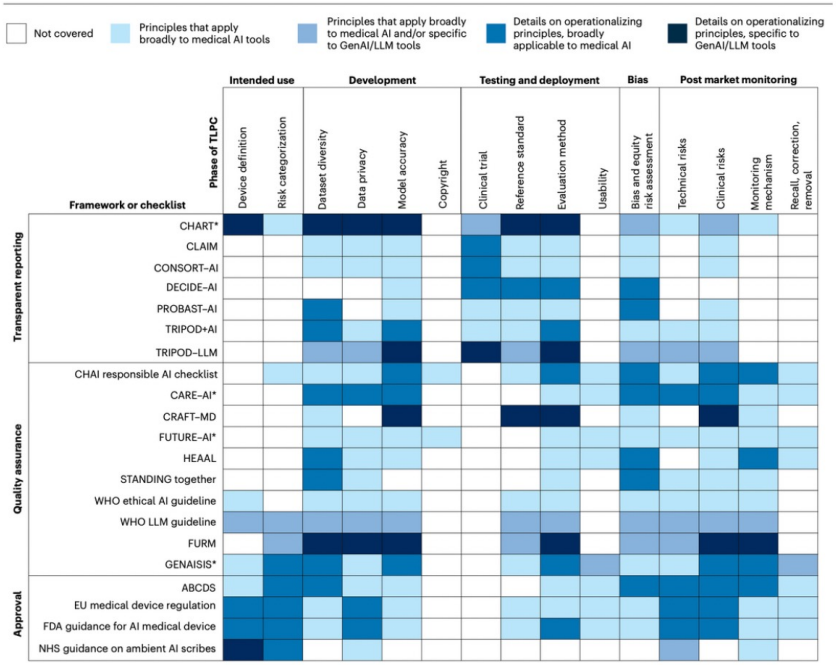
Lack of explainability: The process by which GenAI or LLMs derive an output is often opaque. The introduction of reasoning models such as OpenAI's o1, Claude 3.7, Gemini 2.5, DeepSeek R1 may mitigate this risk to some extent, by displaying the chain of thought, or reasoning process of LLMs in output derivation. However, reasoning coherence in complex medical contexts such as analyzing multi-modal data (imaging, laboratory results or patient history) for differential diagnosis remains unproven.

Hallucinations: Sometimes referred to as confabulations, hallucination is the generation of plausible sounding but incorrect or nonsensical answers, leading to misinformation in medical contexts.

Bias: Generative AI or LLMs may perpetuate existing biases present in their training data, resulting in unfair treatment recommendations or diagnostic inaccuracies for certain populations. Traditional techniques used to reduce AI model bias such as data resampling may not be applicable to pre-trained LLMs.

Data privacy: The use of patient information for LLM pre-training or fine-tuning without obtaining explicit informed consent contravene rights of data policies. De-identified or synthetic data generated using real patient data may run into the risk of re-identification.

Susceptibility to adversarial attacks: Both open-source and closed-source LLM systems are susceptible to adversarial attacks such as prompt injection, potentially compromising sensitive patient information and disrupting clinical operations.



Ong, J.C.L., Ning, Y., Collins, G.S. et al. International partnership for governing generative artificial intelligence models in medicine. Nat Med (2025). https://doi.org/10.1038/s41591-025-03787-4

PATIENT AI RIGHTS *Initiative*

GORDON AND BETTY MOORE FOUNDATION

THE LIGHT COLLECTIVE

1 Patient-Led Governance

Patients at risk should be central to creating AI healthcare rules and standards.

2 Independent Duty to Patients

AI in healthcare must uphold the fiduciary duty to prioritize patient care.

3 Transparency

Transparency in AI healthcare requires informing patients about data use, AI-driven guidance, and providing evidence of AI efficacy in care.

4 Self-Determination

Patients' right to self-determination means they should have the option to choose or refuse AI interventions in their care.

5 Identity Security and Privacy

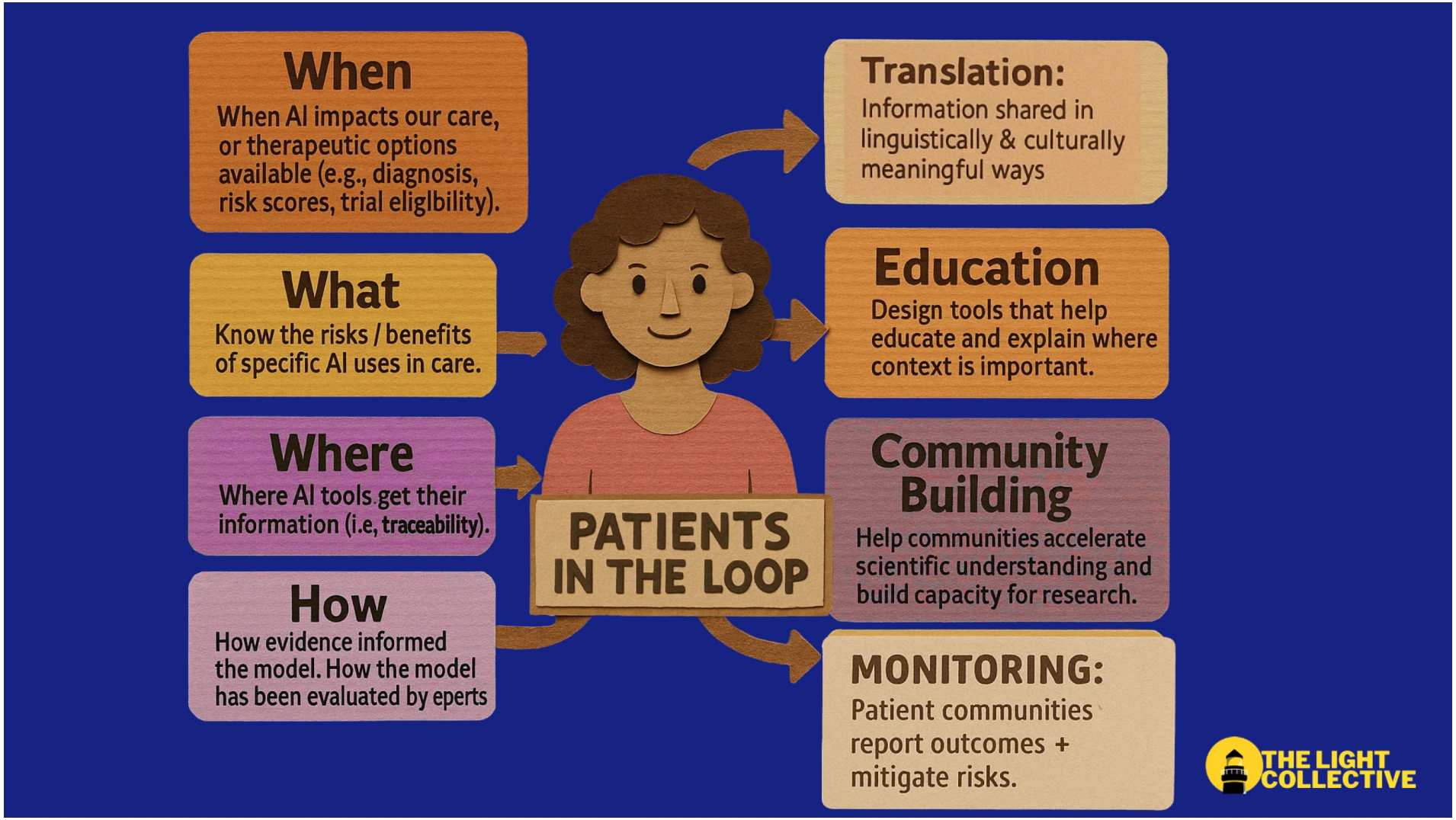
AI in health must ensure patient safety, privacy, and identity protection while respecting their choices.

6 Right of Action

Organizations must bear the risk for AI in healthcare and provide legal remedies if AI causes harm

4 Shared Benefit

AI in healthcare must fairly include and benefit those who risk the most, particularly patients.


When
When AI impacts our care, or therapeutic options available (e.g., diagnosis, risk scores, trial eligibility).
What
Know the risks / benefits of specific AI uses in care.
Where
Where AI tools get their information (i.e, traceability).
How
How evidence informed the model. How the model has been evaluated by eperts
PATIENTS IN THE LOOP
Translation:
Information shared in linguistically & culturally meaningful ways
Education
Design tools that help educate and explain where context is important.
Community Building
Help communities accelerate scientific understanding and build capacity for research.
MONITORING:
Patient communities report outcomes + mitigate risks.
THE LIGHT COLLECTIVE

Bringing PAIR To The Health AI Community: Emerging AI + Policy Initiatives

FUTURE DIRECTIONS

PATIENT-GOVERNED AI OVERSIGHT COUNCIL

Pilot independent, patient-led body co-chairs with clinical experts

PROCUREMENT POLICY FOR INSTITUTIONS

Bake patient governannce and safety/quality criteria into RFPs and vendor onboarding

BUDGET

Dedicate 10% of Health AI budgets to patient governance infrastructure. Fund capacity building and training for patient advocates and communities to co-design

DEPLOY PATIENT-GOVERNED AI SANDBOXES

Create a test bed for Patient AI Rights Initiative's principles into practice. Co-produce consent tools, patient-led ovaluation metrics, and opt-out/appeal flows embedded in portals and care navigation

THE LIGHT COLLECTIVE

Acknowledgements

Jill Holdren, MPH
Grace Cordovano, PhD, BCPA
Dave deBronkart
Andrea Downing
Ysabel Duron, BA
Lesley Kailani Glenn, BS
Dana Lewis
Marlena Murphy, MA
Valencia Robinson, Ed.S
Liz Salmi, AS
Chethan Sarabu MD
Christine Von Raesfeld

Thank you to our sponsor:

GORDON AND BETTY
MOORE
FOUNDATION

Valencia Robinson | Tiah Tomlin-Harris | Christine Von Raesfeld | Mike Mitteiman

Ashley Dedmon | Estela Mata | Marlena Murphy | Ivis Febus-Sampayo

Andrea Downing | Erica Olenski | Whitman Corlew

Eric Perakslis PhD | Danny van Leeuwen | Sneha Dave | Joanne Hayashi

Marissa Thomas | Lyndsey Womack | Arianna Apodaca | Victoria Riley

Tambre Leighn | Catherine M. | Marquita Goodluck | Dr. Grace

Dr. Maya R Cummings

Liz Salmi | Ysabel Duron | Jill Holdren | Dave deBronkart

Chethan Sarabu, MD | Chisa Nosamiefan

David White | Alicia Staley | Lesley Glenn | Robin Dubin

FDA U.S. FOOD & DRUG ADMINISTRATION

User-Centric Design for AI in Pharmacovigilance (PV)

Oanh Dang, Pharm D, BCPS
Regulatory Science and Applied Research Lead
FDA/Center for Drug and Evaluation/Office of Surveillance and Epidemiology
CTTI 2025 Public Workshop: Artificial Intelligence in Drug and Biological Product Development
October 7, 2025

The opinions expressed in this lecture are those of the presenter, and do not necessarily represent the views of the US Food and Drug Administration or the US Government

User-Centric Design for Explainability, Interpretability, and Transparency for Enhancing Trust in AI for PV

*FDA Digital Health and Artificial Intelligence Glossary – Educational Resource | FDA

User-Centric Design

- Placing the PV safety reviewer end user at the center of how AI tools are designed ensures the system aligns with their workflow and decision-making processes

Explainability & Interpretability*

- Explainable AI answers the question "Why" an AI system made a particular decision*
- Interpretable AI refers to the meaning of AI systems' output in the context of their designed functional purposes*

Transparency, Model Performance, & Trust

- Methodology documentation and a user interface that enables users to validate or correct outputs fosters transparency and trust among PV end users, even when model performance is imperfect

Information Visualization Platform (InfoViP)

- InfoViP's Individual Case Safety Report (ICSR) Deduplication Pipeline and Assessability Classifier demonstrate how user-centric design can support AI trust and adoption

The Need for AI in Pharmacovigilance

Background → Challenge → Information Visualization Platform (InfoViP)

Pharmacovigilance

- All scientific and data gathering activities for detection, assessment and understanding of adverse events
- Individual case safety report (ICSRs) are an important data source for ongoing monitoring post-market drug safety surveillance to detect adverse event safety signals

The FDA Adverse Event Reporting System (FAERS)

- > 31.9 million ICSRs; receives >2 million reports annually
- Variable data quality
- Data contained in structured and unstructured formats

AI-enabled Decision Support Tool

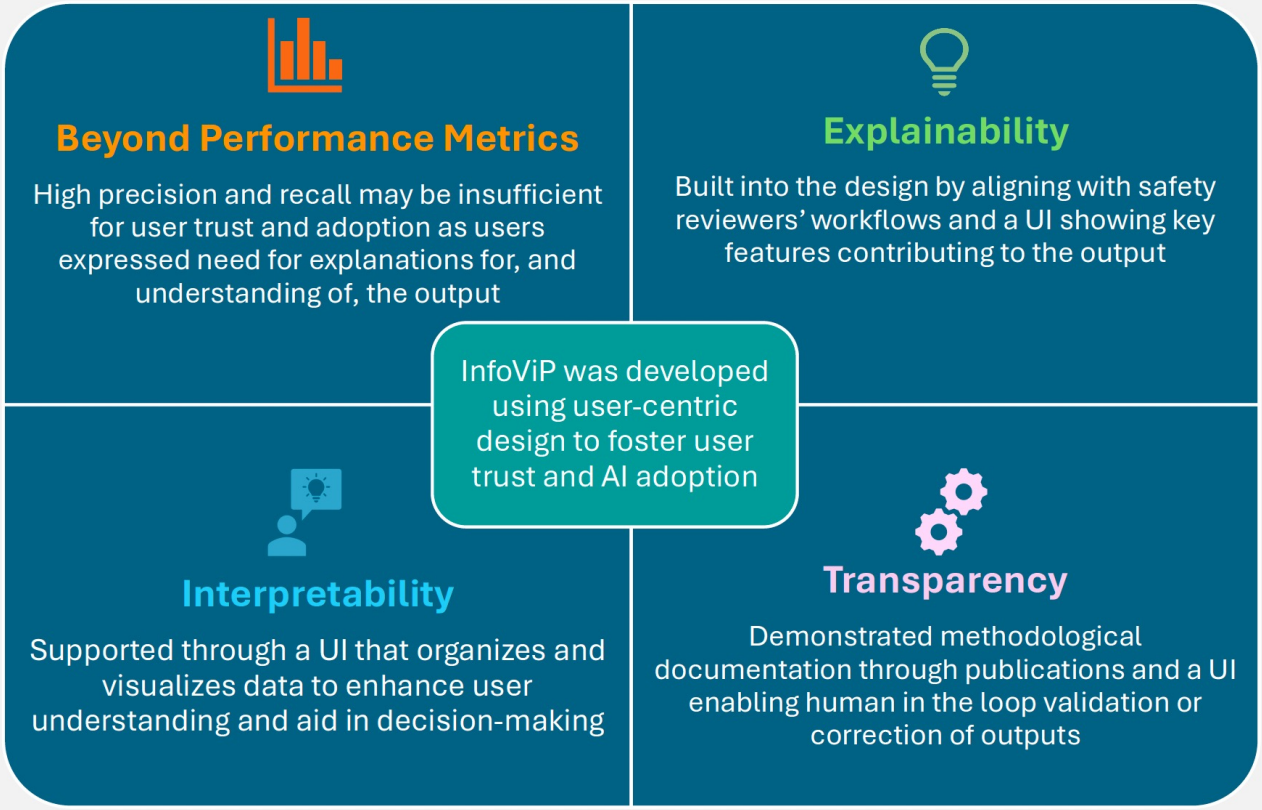
- Deduplication Pipeline for duplicate detection of ICSRs
- Assessability Classifier for low information or incomplete unassessable reports

User-Centric Design Approach

- Collaborative development with end users
- Iterative feedback integration
- Platform and user interface designed to enhance AI explainability, interpretability, and transparency to foster trust and user adoption

- Good Pharmacovigilance Practices and Pharmacoepidemiologic Assessment. March 2005.
- FDA's Adverse Event Reporting System (FAERS) Public Dashboard (accessed 9/28/2025)
- Spiker J, Kreimeyer K, Dang O, Boxwell D, Chan V, Cheng C, Gish P, Lardieri A, Wu E, De S, Naidoo J, Lehmann H, Rosner GL, Ball R, Botsis T. Information Visualization Platform for Postmarket Surveillance Decision Support. Drug Saf. 2020 Sep;43(9):905-915

User-Centric Design Enhances Explainability, Interpretability, and Transparency



InfoViP ICSR Deduplication Pipeline

Issue

- Duplicate ICSR records can skew or misrepresent actual drug safety signals
- Contribute to operational inefficiency

Applications of User-Centric Design

Figure 1. Alignment of InfoViP's multi-step deduplication pipeline with safety reviewers' manual workflow builds in explainability

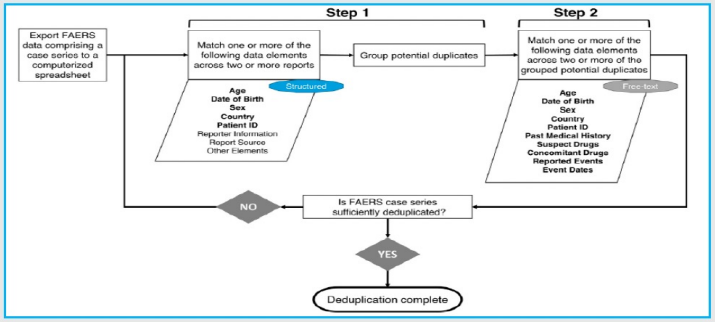
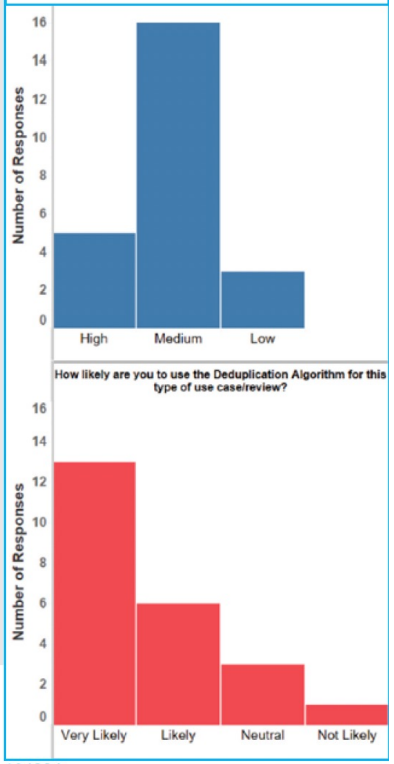


Figure 2. Representation of UI enabling visualization of NLP-derived evidence from ICSRs to support human in the loop verification of duplicate reports

	Sample Case 1	Sample Case 2
Age	(missing)	27
Date of Birth	(missing)	(missing)
Sex	(missing)	Female
Country	USA	USA
Product List	ALCOHOL:BUPROPION:HYDROCODONE BITARTRATE AND ACETAMINOPHEN:OLANZAPINE:SERTRALINE HYDROCHLORIDE(PS):TOPIRAMATE	ACETAMINOPHEN\HYDROCODONE BITARTRATE(PS):ALCOHOL:BENZODIAZEPINE:BUPROPION:OLANZAPINE:SERTRALINE HYDROCHLORIDE:TOPIRAMATE
Narrative	This case, manufacturer control number 2010US-31516 from UNITED STATES refers to a Female, 27 Years-old, who presented with withdrawal syndrome after ingesting olanzapine, unk., sertraline, unk., Hydrocodone and acetaminophen, unk, and Topiramate, unk, which was fatal. This case was reported in Literature. The route of administration of the suspect product was reported to be ingestion. Patient also took ethanol, unk, bupropion, unk, and benzodiazepine, unk, which were considered as co-suspect medications. On an unknown date, patient took the medications. It was reported that patient had prehospital cardiac and/or respiratory arrest and her hospital records were reviewed. The reason for the exposure of medications was reported as other-withdrawal, as per the literature. The chronicity of the medications was reported to be acute and the relative contribution to fatality was reported to be contributory. According to the authors, amongst all the suspect products, ethanol was the primary cause for the event. The event outcome was reported as Fatal at the time of this report. Case Outcome: Fatal	Literature report from the USA of OTHER WITHDRAWAL and CARDIAC AND/OR RESPIRATORY ARREST with HYDROCODONE/ACETAMINOPHEN therapy. On an unreported date, the patient ingested unknown amounts of HYDROCODONE/ACETAMINOPHEN, ETHANOL, BUPROPION, BENZODIAZEPINE, OLANZAPINE, SERTRALINE and TOPIRAMATE as a result of OTHER WITHDRAWAL. On an unreported date, the patient's hospital records were reviewed. The patient experienced a pre-hospital cardiac and/or respiratory arrest. On an unreported date, the patient died following the ingestion of HYDROCODONE/ACETAMINOPHEN, ETHANOL, BUPROPION, BENZODIAZEPINE, OLANZAPINE, SERTRALINE and TOPIRAMATE. The patient's chronicity for HYDROCODONE/ACETAMINOPHEN, ETHANOL, BUPROPION, BENZODIAZEPINE, OLANZAPINE, SERTRALINE and TOPIRAMATE therapies was reported as acute exposure. The reporters believed that the relative contribution to the fatality was contributory due to the HYDROCODONE/ACETAMINOPHEN, ETHANOL, BUPROPION, BENZODIAZEPINE, OLANZAPINE, SERTRALINE and TOPIRAMATE. ETHANOL, BUPROPION, BENZODIAZEPINE, OLANZAPINE, SERTRALINE and TOPIRAMATE were also considered suspect.

Figure 3. Majority of users indicated 'medium and high' confidence in and were 'very likely' or 'likely' to use the deduplication output



- Kreimeyer K, Menschik D, Winiecki S, Paul W, Barash F, Woo EJ, Alimchandani M, Arya D, Zinderman C, Forshee R, Botsis T. Using Probabilistic Record Linkage of Structured and Unstructured Data to Identify Duplicate Cases in Spontaneous Adverse Event Reporting Systems. Drug Saf. 2017 Jul;40(7):571-582.
- Kreimeyer, Kory & Dang, Oanh & Spiker, Jonathan & Gish, Paula & Weintraub, Jessica & Wu, Eileen & Ball, Robert & Botsis, Taxiarchis. (2022). Increased Confidence in Deduplication of Drug Safety Reports with Natural Language Processing of Narratives at the US Food and Drug Administration. Frontiers in Drug Safety and Regulation. 2. 918897. 10.3389/fdsfr.2022.918897.
- Kreimeyer K, Spiker J, Dang O, De S, Ball R, Botsis T. Deduplicating the FDA adverse event reporting system with a novel application of network-based grouping. J Biomed Inform. 2025 May;165:104824.

InfoViP ICSR Assessability Classifier

Issue

- Incomplete or low quality ICSRs
- Operational inefficiency

Applications of User-Centric Design

Figure 1. Top contributing NLP-derived features displayed in UI to enhance understanding of output and support human in the loop verification

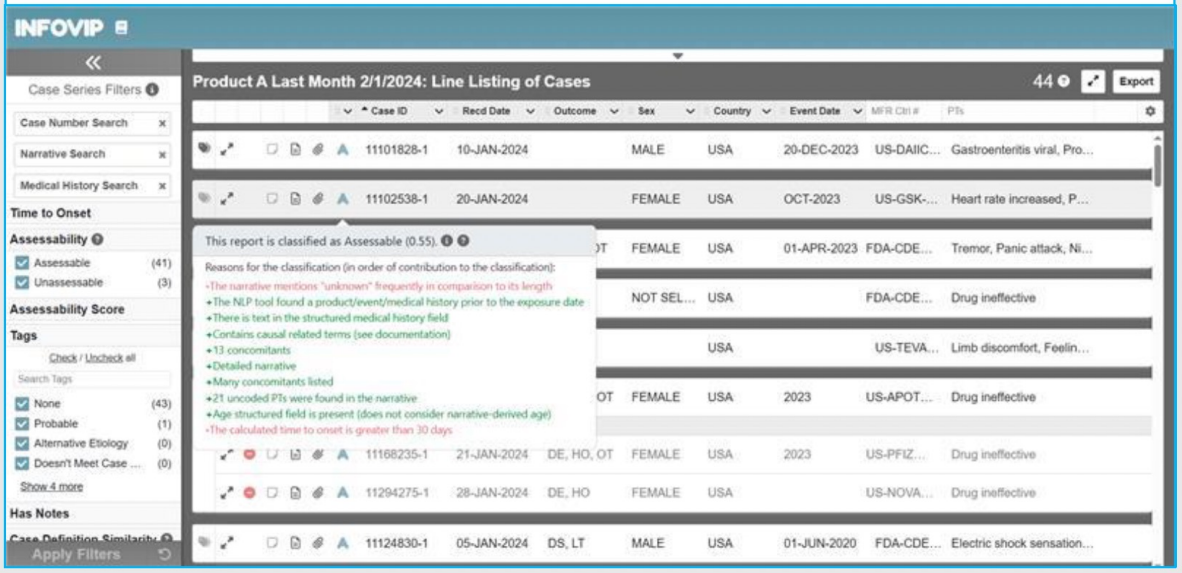


Figure 2. Safety reviewers can choose to view, explore, and analyze all classification features, including NLP-derived evidence, through an export function enabled through an interactive UI

FAERS Case #	Assessability	Assessability Score	Number of Adverse Events Found Only in Text	Age Structured Value Present	Primary Diagnosis in Text	Narrative Medical History (Pre-morbidity)	Reporter is Medical Expert	From Literature Source	Many Concomitant Products	Length of Narrative	Number of Concomitant Products	Contains Causal-Related Text Term	Coded With CDER/OSE Designated Medical Eve	Onset Within 30 Days	Report Type	Primary Suspect Stop Date Value Present	Structured Medical History Value Present	Symptoms Before Exposure	Amount of Text Content With Timestamp	Frequency of Unknowns in Text
xx	Unassessable	0.2	(-) 1	(-) No	(+) A diagnosis was found in the narrative (generally with keywords "pDx" or "diagnosis")	(+) The NLP tool detected at least one medical history features within the narrative	(+) Yes	(-) No	(-) No	(-) Detailed narrative	(-) 1	(-) Does not contain causal related terms (see documentation)	(-) No	(-) The calculated time to onset is greater than 30 days	(-) The report type is PERIODIC	(-) No	(+) There is text in the structured medical history field	(-) The NLP tool did not find a product/event/ medical history prior to the exposure date	(-) 0% of the NLP extracted features have an associated timestamp	(-) The narrative mentions "unknown" frequently in comparison to its length

- Kreimeyer K, Dang O, Spiker J, Muñoz MA, Rosner G, Ball R, Botsis T. Feature engineering and machine learning for causality assessment in pharmacovigilance: Lessons learned from application to the FDA Adverse Event Reporting System. Comput Biol Med. 2021 Aug;135:104517

Extending User-Centric Design to Large Language Models (LLMs)

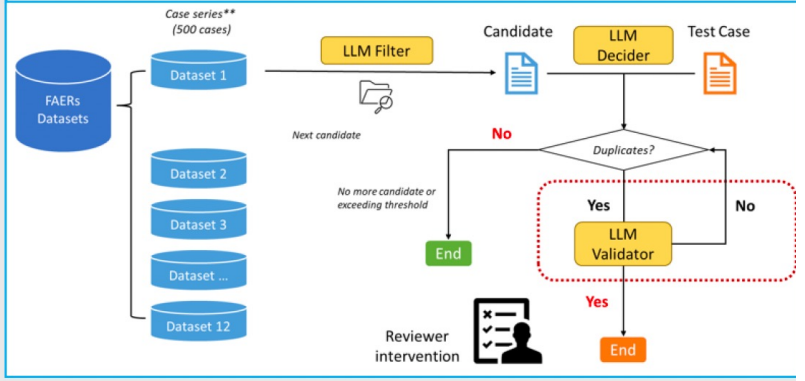
A key challenge with LLMs is limited user trust due to the "black box" nature of its outputs

Exploratory Application of User-Centric Design

- Modular LLM-powered Deduplication Pipeline*
  - Aligns with manual ICSR deduplication workflow; follows logical stepwise approach
  - Applies structured sequential prompts to guide LLM through each stage of deduplication process (Figure 1)
    - LLM Filter: Extracts key features from ICSRs, filters ICRs likely to be duplicates
    - LLM Decider: Uses chain-of-thought prompt for ICSR duplicate classification
    - LLM Validator: Provides reasons and explanations for its decision

- LLM for ICSR Assessability
  - Developed and refined expert-driven prompt rules to define assessability based on user feedback
  - LLM prompts include instructions to generate evidence behind outputs to enhance explainability

Figure 1.* Modular LLM-powered Deduplication Pipeline



*Wu L., Xu J, Dang O, Ball R poster at 4th Global Summit on Regulatory Science Annual Conference Sept. 18-19, Little Rock, AR

Summary

Building Trust Through User-Centric Design for AI in PV

- Performance metrics alone may not be enough as accuracy doesn't guarantee user trust or adoption
  - To bridge the gap, user-centric approaches, including thoughtfully designed UI, can complement and enhance explainability, interpretability, and transparency
- In our explorations, we observe that user-centric design approaches might be extended to "black box" LLMs to foster trust
- Collaborative development with end users, supported by iterative feedback integration, is essential for building trust

Acknowledgements

- Robert Ball, MD, MPH, Scm
- Taxiarchis Botsis, PhD, MSc, MPS
- Jonathan Spiker BS
- Kory Kreimeyer MS
- Joshua Xu, PhD
- Leihong Wu, PhD
- Monica Munoz, PharmD, PhD
- Ellen Pinnow, PhD
- Division of Pharmacovigilance Safety Reviewers

Questions?

PLEASE SUBMIT HERE:

BREAK

We will reconvene at 3:00 p.m.

U.S. FOOD & DRUG ADMINISTRATION

CLINICAL TRIALS TRANSFORMATION INITIATIVE

Session 4: Navigating the Future of AI in Drug Development

Moderator:
Gabriel Innes
Assistant Director for Data Science and AI Policy, OMP, CDER
U.S. Food & Drug Administration

Moderator:
Rebecca Nebel
Senior Director, Science and Regulatory Advocacy
PhRMA

Jon Walsh
Founder & Chief Scientific Officer
Unlearn.AI

Ryan Hoshi
Director of Regulatory Policy & Intelligence
AbbVie

Jessilyn Dunn
Assistant Professor of Biomedical Engineering and Biostatistics & Bioinformatics
Duke University

UNLEARN

Using AI for the hard problems

Jonathan Walsh

Artificial Intelligence in Drug & Biological Product Development FDA / CTTI workshop

October 2025

Artificial intelligence can meaningfully contribute to how we answer fundamental scientific questions

	Drug development	→	Can we find and identify drugs that are safe and effective to treat a particular disease?
	Physics	→	Can we test theories of the universe through observation and planned experiments?
	Climate	→	Can we predict how our climate will change and measure our impacts on it?

Currently, the key processes it impacts are data analysis and modeling & simulation, with eyes towards discovery

	Drug development	→	Computational chemistry for compound properties and discovery; model disease course; simulate trial outcomes and drug effects; discern signals from complex data sources (e.g., imaging)
	Physics	→	Characterize complex data from experiments; use efficient models to simulate computationally expensive processes; model systems without clear first principles behavior
	Climate	→	Use efficient models to simulate computationally expensive processes; characterize data from satellite imaging; model systems beyond where first principles can predict behavior

A key transition is converting human-centric systems to model-centric ones

Human-centric

- Value proposition of models unclear
- Interpretability prized
- Simplicity/explainability over performance

Human in the loop

- Initial proof points for model value
- Interpretability valued
- Focus on system pieces with impact

Model-centric

- **“Just use a model”**
- **Interpretability mostly irrelevant**
- **Best practices rule to ensure quality**

Example: covariates for use in efficacy analyses of randomized trials to add statistical power

	Standard covariates (e.g., key demographics, disease features, baseline severity)	→	High mistrust, lack of clear value, regulatory concern
	Proof points from model-derived covariates (retrospective analyses, broader technical interest, positive regulatory feedback)	→	Engagement from savvy or high need teams, growing awareness
	Clear case studies, common code and models, low barriers to adoption	→	De-facto expectation for use, standardization high

These things will happen – how do we help them succeed*?

**Success = positive impact with broad adoption*

Set the rules of engagement

Regulators

- ✓ Define the framework to assess and build AI-based applications
- ✓ Define clear boundaries for “not yet” or “work with us to evaluate”
- ❑ Build broader collaboration pathways to define case studies and best practices
- ❑ Publish case studies and push the industry forward

Invest in Computation

Sponsors

- ❑ Build data resources as the foundation for modeling
- ❑ Build/buy analytics and modeling capabilities
- ❑ Identify clear use cases to move from human-centric to human in the loop or to model-centric
- ❑ Find out what blocks adoption and solve it

Investment in data resources can be a driver for change

Easy initial investment – build harmonized data resources in key areas, especially where internal data (e.g., past studies) exists

Use data for improvements in decision making, e.g., trial planning and simulations, better understanding of indications and populations

Work on processes so that stakeholders can get the answers they want from data quickly – in hours, not weeks

Leverage data and models where needed to answer problems that arise in development. Sure-fire way to build proof points for teams to understand the value of data resources and activities on top of them

Multiple proof points in individual systems can roll up into greater capabilities and centered activities – don’t do “build it and they will come”

UNLEARN

Thank you.

AI and DHTs for Disease Detection & Monitoring

Jessilyn Dunn, PhD

October 7, 2025

Artificial Intelligence in Biological Product Development

US FDA & CTTI Workshop

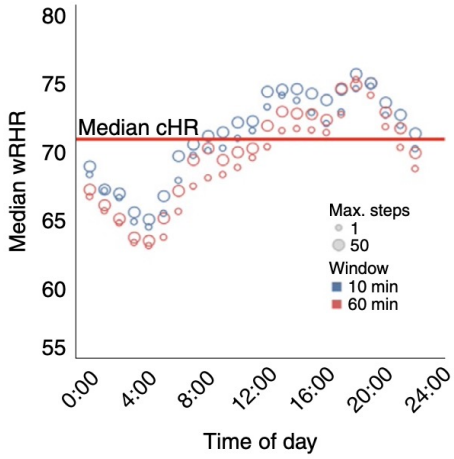
Duke UNIVERSITY | BIG IDEAs LAB

Utility of Data from DHTs

Continuous biosignal measurement can now be seamlessly integrated into daily life

Wearables capture more consistent biosignals than in-clinic measurements

- Biosignals change over the course of the day
- Normative values are insufficient– no one is average



Dunn, Kidzinski et al. *Nature Medicine*, 2021
https://news.mit.edu/2020/sensors-monitor-vital-signs-0423

Duke | BIG IDEAS LAB

DHT Data Deluge

- Wearables generate massive amounts of data
- How can we use that data to make actionable insights?
- Solution: Apply AI to develop Digital Biomarkers

Digital data becomes a Digital Biomarker when a relationship to a health outcome is established

Target Outcomes $= f($ **Measurements** $)$

- Clinical measurements
- Survey data
- Illness events

- Heart rate
- Skin Temperature
- Skin Conductance
- Accelerometry
- Location

Duke | BIG IDEAS LAB

Digital biomarkers as proxies for clinically-relevant measurements

npj | Digital Medicine

www.nature.com/npjdigitalmed

ARTICLE OPEN

Check for updates

Engineering digital biomarkers of interstitial glucose from noninvasive smartwatches

Brinnae Bent[1], Peter J. Cho[1], Maria Henriquez[2], April Wittmann[3], Connie Thacker[3], Mark Feinglos[3], Matthew J. Crowley[3] and Jessilyn P. Dunn[1,4]✉

Prediabetes affects one in three people and has a 10% annual conversion rat
interventions. Management of glycemic health is essential to prevent progressi
commercially-available and noninvasive method for monitoring glycemic healt
a critical need for innovative, practical strategies to improve monitoring and m
dataset of 25,000 simultaneous interstitial glucose and noninvasive wearable
feasibility of using noninvasive and widely accessible methods, including sma
continuously detect personalized glucose deviations and to predict the exact i
and 87% accuracy, respectively. We also establish methods for designing varia
from noninvasive wearables toward interstitial glucose prediction.

npj Digital Medicine (2021)4:89 ; https://doi.org/10.1038/s41746-021-00465-w

Open access Original research

BMJ Open Diabetes Research & Care

Non-invasive wearables for remote monitoring of HbA1c and glucose variability: proof of concept

Brinnae Bent,[1] Peter J Cho,[1] April Wittmann,[2] Connie Thacker,[2] Srikanth Muppidi,[3] Michael Snyder,[3] Matthew J Crowley,[2] Mark Feinglos,[2] Jessilyn P Dunn [1,4]

alence continues to grow and
liagnostic gap in one-third
s pre-diabetes. Innovative,
ve monitoring of glycemic
ed. In this proof-of-concept
nship between non-invasive
trics and demonstrate the
ive wearables to estimate
hemoglobin A1c (HbA1c) and

ıods We recorded over 25 000
ıuous glucose monitor (CGM)
ın wearable (skin temperature,
rate, and accelerometry
s in 16 participants with
re-diabetes (HbA1c 5.2–6.4).
able to develop machine
bA1c recorded on day 0 and
from the CGM. We tested
ıodel on a retrospective,
10 additional participants and
iM-based HbA1c estimation

of data from 26 participants
models of glucose variability
sing non-invasive wearables,

Significance of this study

What is already known about this subject?

- Several studies report that wearables can non-invasively capture metrics reflecting autonomic nervous system activity, which is a demonstrated correlate to glycemic health.

What are the new findings?

- Glycemic variability metrics can be estimated with high accuracy using non-invasive wearables.
- Non-invasive wearables can be used to predict hemoglobin A1c (HbA1c) with similar accuracy to a continuous glucose monitor.

How might these results change the focus of research or clinical practice?

- Our findings from this proof-of-concept study suggest that wearables could potentially be used as part of a strategy to remotely monitor diabetes and detect undiagnosed pre-diabetes. Because wearables are so prevalent in the general population, leveraging these ubiquitous devices for purposes including glycemic monitoring and pre-diabetes detection and monitoring could represent a major advance in clinical pre-diabetes care.

nature medicine ARTICLES

https://doi.org/10.1038/s41591-021-01339-0

Check for updates

Wearable sensors enable personalized predictions of clinical laboratory measurements

Jessilyn Dunn[1,2,3,4,5,10]✉, Lukasz Kidzinski[4,10], Ryan Runge[1,4], Daniel Witt[2,3], Jennifer L. Hicks[4], Sophia Miryam Schüssler-Fiorenza Rose[1,5,6], Xiao Li[1,7], Amir Bahmani[1], Scott L. Delp[4,8], Trevor Hastie[9]✉ and Michael P. Snyder[1,5]✉

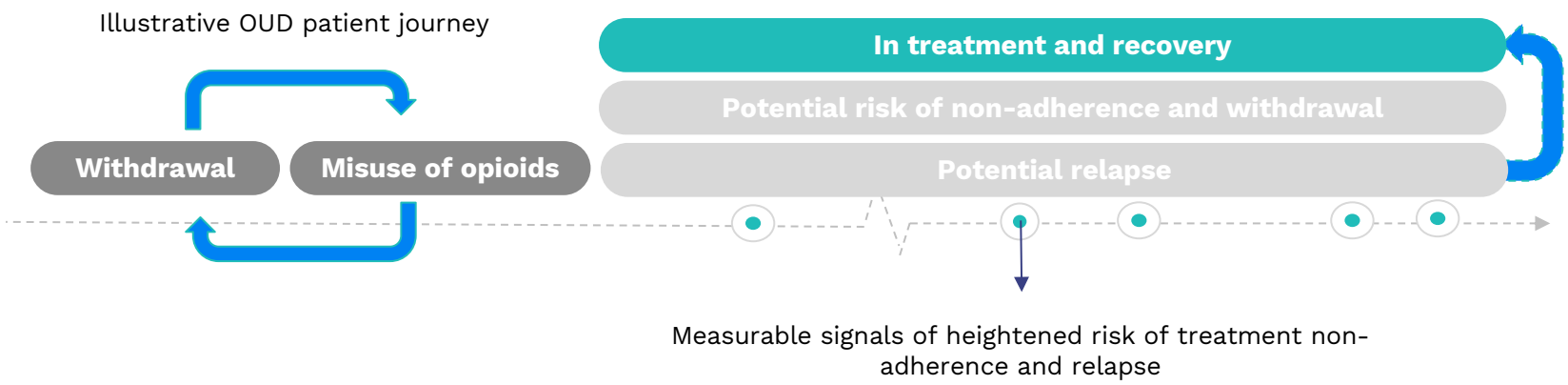
Vital signs, including heart rate and body temperature, are useful in detecting or monitoring medical conditions, but are typically measured in the clinic and require follow-up laboratory testing for more definitive diagnoses. Here we examined whether vital signs as measured by consumer wearable devices (that is, continuously monitored heart rate, body temperature, electrodermal activity and movement) can predict clinical laboratory test results using machine learning models, including random forest and Lasso models. Our results demonstrate that vital sign data collected from wearables give a more consistent and precise depiction of resting heart rate than do measurements taken in the clinic. Vital sign data collected from wearables can also predict several clinical laboratory measurements with lower prediction error than predictions made using clinically obtained vital sign measurements. The length of time over which vital signs are monitored and the proximity of the monitoring period to the date of prediction play a critical role in the performance of the machine learning models. These results demonstrate the value of commercial wearable devices for continuous and longitudinal assessment of physiological measurements that today can be measured only with clinical laboratory tests.

Duke | BIG IDEAS LAB

We have used noninvasive wearables data to…

- Predict HbA1c
- Predict interstitial glucose values
- Detect personalized interstitial glucose excursions
- Estimate glycemic variability metrics
- Explain the variance of key clinical labs (hematologic, immune, etc.)
- Detect infection
- …& much more

Developing A Digitally-derived Tool To Support The Prevention Of Relapse In Opioid Use Disorder



"The vast majority of our patients return to use. It is an expected part of the illness." (clinician)

Duke | BIG IDEAS LAB

DATACC by DiME

Triangle CERSI

Ongoing Pilot Study:
DHTs to Support the Prevention of Relapse in Opioid Use Disorder

*Evaluate the **feasibility, acceptability, and data quality of using sensor-based digital health technologies (sDHTs) and electronic patient-reported outcomes (ePROs)** for real-time monitoring of opioid relapse among individuals with OUD.*

Study Timeline

1. **Patients receiving MOUD will be recruited through virtual and in-person visits with their provider.**
2. **If the patient agrees to participate, they will read and sign the e-consent and complete an screening/eligibility form.**
3. **Participants will review online materials about study participation and complete initial ePRO surveys**
4. **A Fitbit and Oura ring will be sent to the address indicated**
5. **Participants will be asked to complete the first set of ePRO surveys and wear the Fitbit and Oura for a week to collect baseline data**
   Data from both Fitbit and Oura for 6 of 7 days will be needed for the study to continue.
6. **For 3 months the participants will:**
   1. Complete daily ePRO surveys through their smartphone
   2. Wear both the Fitbit and Oura ring daily, making sure they are charged and synced to their smartphone.
7. **During the study, investigators will monitor for adherence and unexpected events. Upon data collection completion, researchers will analyze the data and disseminate findings.**

Oura Ring Gen 4 and Fitbit Charge 6

Duke | BIG IDEAS LAB

DATAcc by DiME

Triangle CERSI

Current Challenges In Digital Biomarker Discovery

- Device/Data Validity & Accuracy
- Interpretability & FAIR principles
- The “Data Deluge”
- Regulatory oversight of tools
- “ELSI”, Privacy, Security

Examples of Biosignal AI Algorithm Underperformance

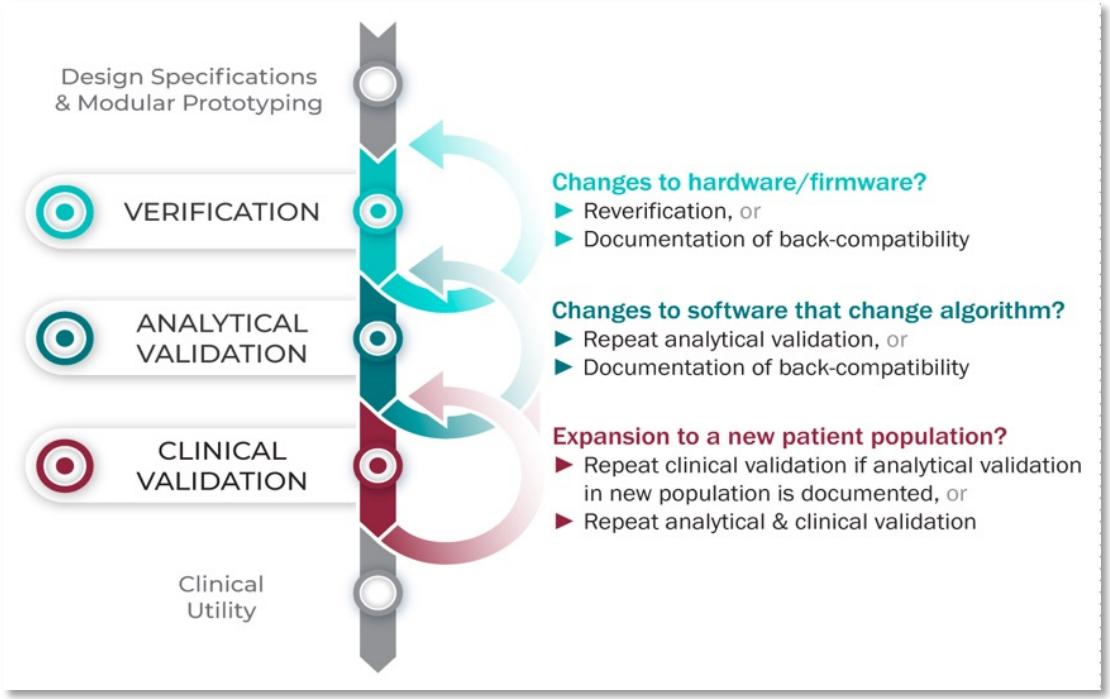
Biosignal	Algorithm purpose	Circumstance of underperformance	Reason for underperformance	Impact of underperformance
Pulse Ox	Detect low SpO2	Dark skin tones[33]	Melanin absorbs and hinders sensing of light	Missed hypoxemia in dark skin[26,27]
Contactless Body Temp	Detect fever	Light[28] or dark[29] skin; females[34]; High temp	Core temp dynamic range/fluctuations; thermoregulation[35,36]	Miss fevers or flu[28,29]
Auscultatory BP	Measure BP	Obesity	Poor cuff fit	BP overestimation (up to 10 mmHg)[30,31]
Physical Motion	Detect falls	Elderly in nursing homes	Falls often supported (furniture; guardrails)	Missed falls in elderly in nursing homes[37,38]
Physical Motion	Detect crashes	Winter athletes	Skiing has similar motion profile to crashes	False alarms to emergency responders[32]

Wearables Software Updates Impact Data They Output

Duke | BIG IDEAS LAB

Collins, T. *et al.* Version Reporting and Assessment Approaches for New and Updated Activity and Heart Rate Monitors. *Sensors* **19**, 1705 (2019).

Modular Evaluation of Digital Measures



Duke | BIG IDEAS LAB

Goldsack et al, *npj Digital Medicine*, 2020

DiME

The Digital Biomarker Discovery Pipeline (DBDP) provides open source ***data***, ***code***, ***algorithms***, and ***educational resources*** to make discovering digital biomarkers more accessible and establish best practices for the field.

DBDP.org

Bent et al. *JCTS*, 2020.

Duke | BIG IDEAS LAB

Duke | BIG IDEAs Lab

NIH National Institute of Diabetes and Digestive and Kidney Diseases

R01 DK133531

BARDA Biomedical Advanced Research and Development Authority

POWERED BY DRIVe

NSF U.S. National Science Foundation

CAREER #2339669

Lab Members

Mahmud Islam, PhD
Karnika Singh, MS
Stephanie Sullivan, MS
Kaileigh Moertl, MS
Leeor Hershcovich
Yihang Jiang
Anita Silver
Hayoung Jeong
Bill Chen
Lauren Lederer
Jerry Yang

FDA Triangle CERSI

U01FD007857

FDA Subject Matter Experts

Roxane Modares
Annie Saha
Gioia Guerrieri
Robert Wright
Jisun Yi

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Christin Daniels
Jennifer Cavender
Paul Watkins
Susan Halabi
Robert Mentz
Ehsan Samei

DATACC by DiME

North Carolina Central University

Janice Collins McNeil

Candice Taguibao
Jen Goldsack
Dan Karlin

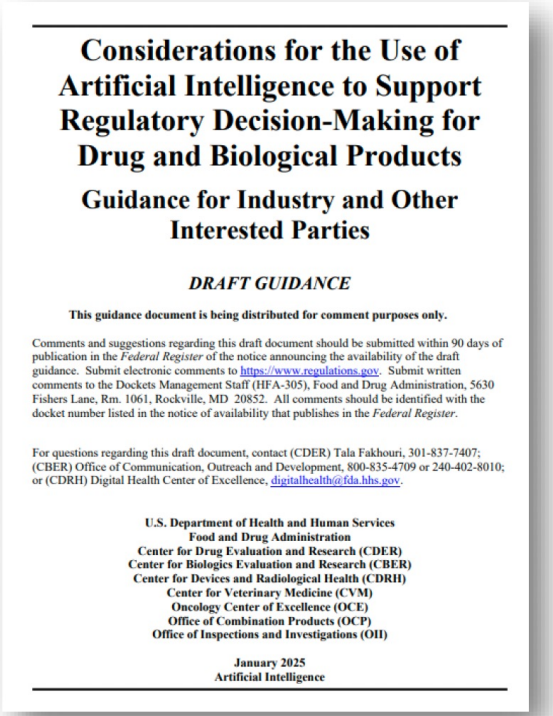
Navigating the Future: Industry Perspectives and Principles for Responsible AI in Drug Development

Ryan Hoshi, PhD, MBA
Director, Regulatory Policy & Intelligence
AbbVie
ryan.hoshi@abbvie.com

abbvie

FDA Draft Guidance: "Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products"

- Provides recommendations to industry on the use of artificial intelligence (AI) to produce information or data intended to support regulatory decision-making regarding the safety, effectiveness, or quality for drug and biological products.
- Strongly support the FDA's ongoing efforts to provide guidance and regulatory clarity to sponsors on how to establish credibility of AI model outputs used in regulatory decision-making.



Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products

Guidance for Industry and Other Interested Parties

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted 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 to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

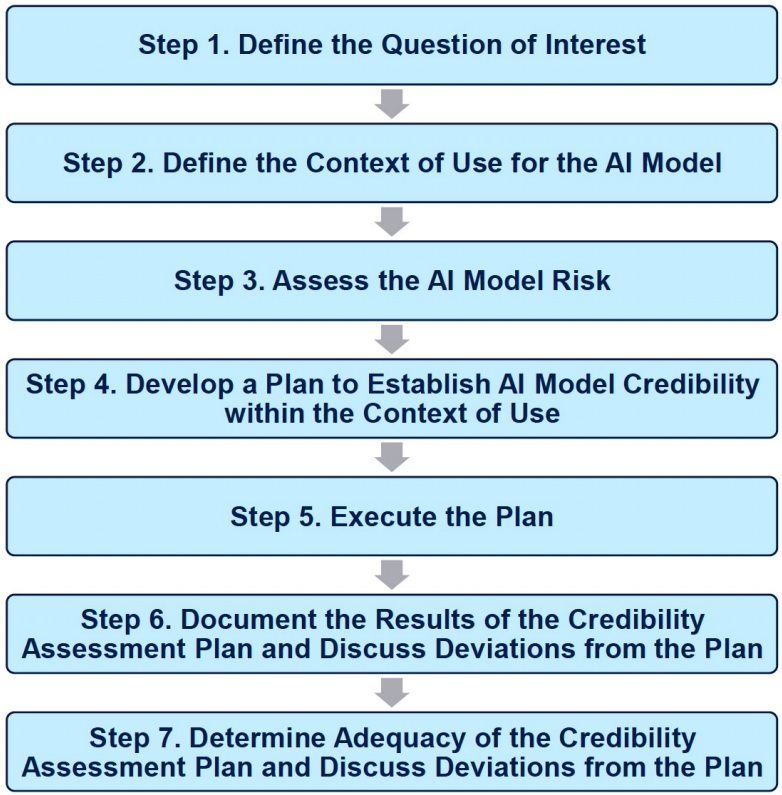
For questions regarding this draft document, contact (CDER) Tala Fakhouri, 301-837-7407; (CBER) Office of Communication, Outreach and Development, 800-835-4709 or 240-402-8010; or (CDRH) Digital Health Center of Excellence, digitalhealth@fda.hhs.gov.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)
Center for Veterinary Medicine (CVM)
Oncology Center of Excellence (OCE)
Office of Combination Products (OCP)
Office of Inspections and Investigations (OII)

January 2025
Artificial Intelligence

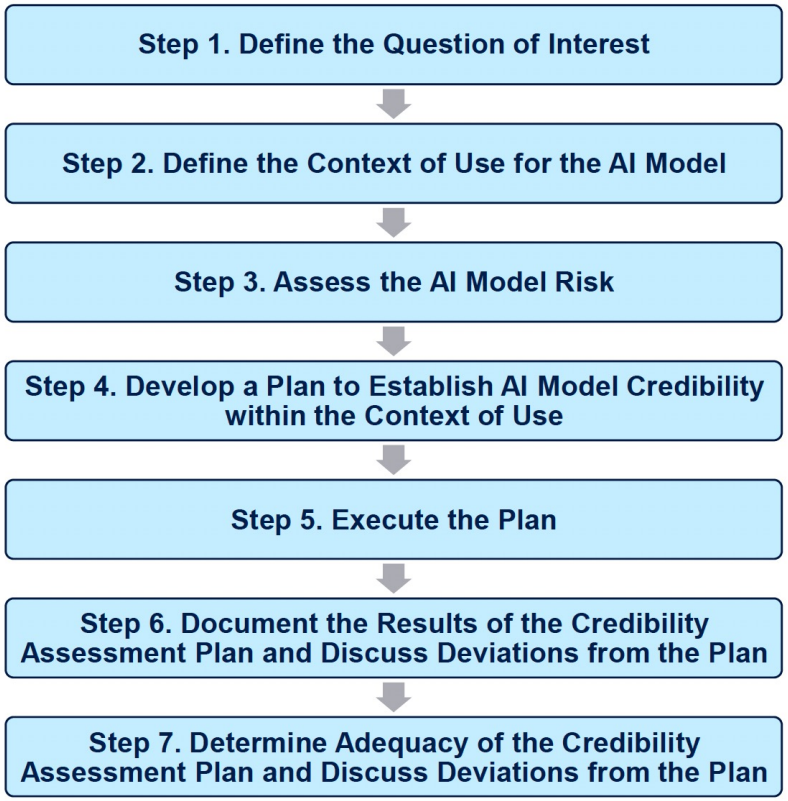
US FDA Draft Guidance. (2025). "Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products."

Risk-Based Credibility Assessment Plan (CAP) Framework



Special Consideration:
Life Cycle Maintenance

Risk-Based Credibility Assessment Plan (CAP) Framework



Questions to consider:

- What type of documentation needs to be provided in the meeting package or regulatory submission?
- Are sponsors expected to request FDA feedback for all AI use cases?
- Does FDA have the resources it needs to provide timely feedback to sponsors?

Special Consideration: Life Cycle Maintenance

Industry Perspectives: Key Considerations

Scope

- Support FDA's exclusion of "AI models (1) in drug discovery or (2) when used for operational efficiencies…that do not impact patient safety, drug quality, or the reliability of results from a nonclinical or clinical study."
- Clearly defined exemptions for AI use cases that are outside FDA's authority and other low-risk scenarios where the credibility assessment framework does not apply.

CAP Framework

- Provide flexibility in model development by leveraging elements of the Predetermined Change Control Plan[1] (PCCP) developed for AI-enabled medical devices.
- Appreciate examples described in the guidance, but recommend elaborating the clinical development and commercial manufacturing examples across all framework steps.

Harmonization

- Encourage global alignment and multistakeholder engagement through international organizations like ICH, IMDRF, and PIC/S.
- Support FDA's Digital Health and Artificial Intelligence Glossary[2] as an important resource but encourage formal alignment and harmonization of key AI terms and definitions referenced in FDA guidance and international collaborations.

Third-Party Models

- Third-party or pre-trained models may limit access to proprietary information such as the underlying model architecture, features, parameters, and training/tuning data.
- We recommend that the FDA clarify expectations for these scenarios such as recommendations regarding the use of master files to support the sharing of confidential information.

1. US FDA Guidance. (2025). "Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions."
2. FDA Digital Health and Artificial Intelligence Glossary – Educational Resource

Industry Perspectives: Key Considerations

Templates

- Documentation templates for regulatory submissions can help standardize content, streamline review, and enhance consistency.
- Submission template for risk-based credibility assessment framework can help align submission content with regulatory expectations. For example, FDA developed a template for integrating real-world data (RWD) and real-world evidence (RWE) in submissions.[1]

Documentation

- The draft guidance indicates that the credibility assessment report may be submitted as part of a regulatory filing, meeting package, or provided to the FDA upon request.
- Clarify documentation and submission expectations (meeting package vs. regulatory filing) for low-risk and high-risk models, including when model modifications are made.

GMLPs

- Good Machine Learning Practice Principles (GMLPs)[2], such as those developed through IMDRF include important concepts related to representative datasets, training, testing, robustness, and performance monitoring.
- Recommend collaborating with stakeholders to develop core GMLP principles to address use of AI in drug development.

Model Acceptance

- FDA's AI-enabled devices database[3] provides important insights on regulatory acceptance of emerging technologies.
- Encourage FDA to provide appropriate, publicly shareable information on acceptance of AI models in drug development: stage of drug product life cycle, therapeutic area, and types of model applications (e.g., dose selection/optimization, endpoint assessment, synthetic controls, pharmacovigilance).

abbvie

1. US FDA Guidance. (2022). "Submitting documents using real-world data and real-world evidence to FDA for drug and biological products."
2. IMDRF/AIML WG/N88. (2025). "Good machine learning practice for medical device development: Guiding principles."
3. FDA CDRH Artificial Intelligence-Enabled Medical Devices List

Principles for the Use of Artificial Intelligence in Regulatory Activities

- **FDA Modernization**: Support FDA's ongoing efforts to use technology, like the GenAI, to modernize regulatory operations and help FDA staff work more efficiently.
- **Principles:** Medical Device Manufacturers Association (MDMA) developed a set of principles guiding the use of AI in FDA premarket activities. These principles provide important considerations for ensuring responsible use of AI in medical product reviews.

Use of AI tools: a two-tier approach

Short-term

Recommend FDA use AI to assist with basic administrative tasks: meeting planning, meeting minutes, tracking due dates, and document preparation.

Long-term

Explore use of AI to assist reviewers in assessment and decision making. However, this advancement requires further stakeholder engagement to ensure a transparent, validated and secure process.

Opportunities: Automate screening of regulatory submissions. Ensure consistency in FDA information requests and deficiencies.

Principles for the Use of Artificial Intelligence in Regulatory Activities

1. Transparency of Use and Assessment of Benefit

- Ensure transparent disclosure of AI tools used by the FDA, including scope of use, and limitations, along with human and technical safeguards to protect data rights and mitigate potential risks.
- Assess current metrics and documentation to monitor performance, and identify potential gaps in current regulation, guidance, or statute for FDA's use of AI for regulatory purposes.

2. AI for Efficiency and Assistance: Human Oversight

- FDA's Mission is critical for ensuring patient safety and public health. FDA staff maintain integral decision-making roles and responsibilities.
- Encourage the continued exploration of AI applications that support innovation by enhancing the regulatory process.

Principles for the Use of Artificial Intelligence in Regulatory Activities

3. Training and Procedures to Guide Use of AI in Regulatory Review

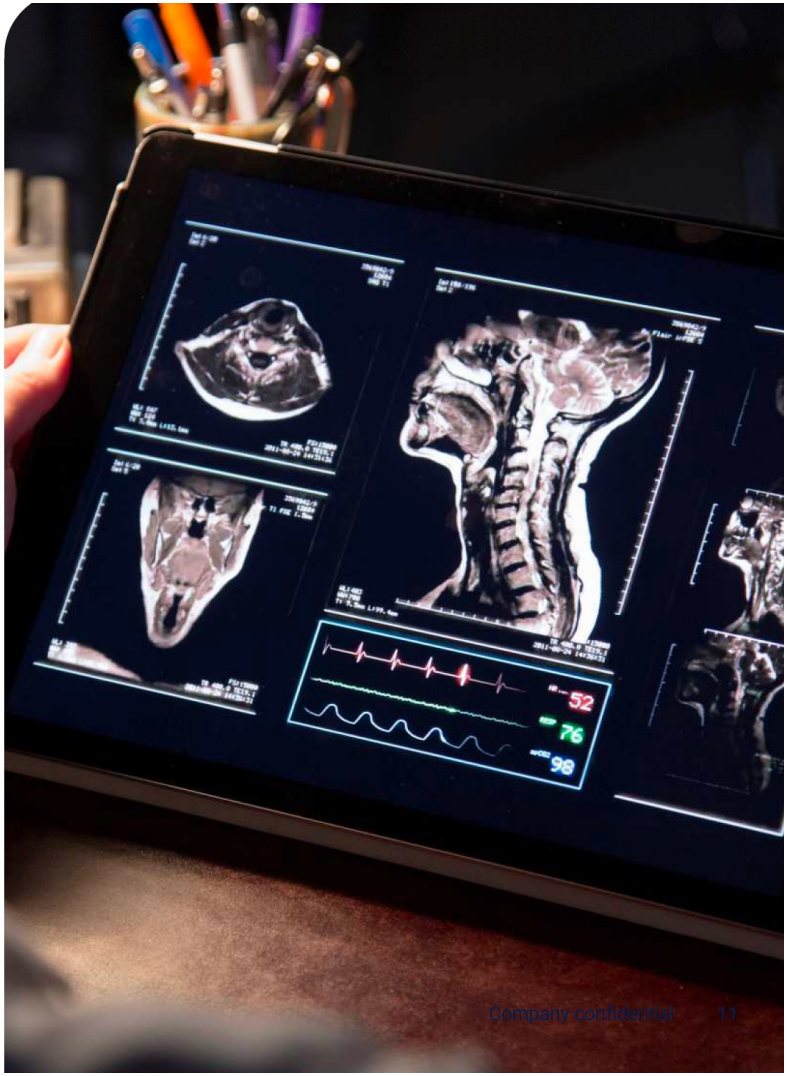
- Recommend FDA adopt and publish principles regarding its use of AI to assist regulatory review.
- Implement FDA staff training to understand AI tool limitations and how to effectively use and interpret AI outputs. Establish procedures for using AI tools to ensure consistency across FDA staff and offices.
- The agency should disclose if Generative AI is used for FDA responses or review letters.

4. Confidentiality and Security of Information

- FDA regulatory submissions contain confidential data and proprietary information that must be handled properly and consistent with applicable law.
- Information security should be maintained at the highest level to ensure protection of manufacturer intellectual property.
- FDA should clearly demonstrate how it will protect proprietary and confidential information and provide assurances that AI tools, associated data, and libraries will not be disclosed.
- Patient information must be protected.

Conclusion

- Regulatory clarity and flexibility enables sponsors to innovate and take actionable steps to implement FDA guidance.
- Clearly defined exemptions for AI use cases that are outside FDA's authority and flexibility for low-risk scenarios where the credibility assessment framework may not apply.
- Encourage global alignment through international organizations like ICH, IMDRF, and PIC/S.
- AI tools have the potential to enhance regulatory review activities, and by fostering transparency and ongoing stakeholder dialogue, we can help ensure these modernization efforts deliver meaningful benefits for patients.



Thank you!

Questions?

PLEASE SUBMIT HERE:

Discussion

Moderator:
Anindita Saha
Associate Director for Data Science and Artificial Intelligence Policy (acting), OMP, CDER
U.S. Food & Drug Administration

Qi Liu
Associate Director for Innovation & Partnership, Office of Clinical Pharmacology, Office of Translational Sciences, CDER
U.S. Food & Drug Administration

Shantanu Nundy
Advisor on Artificial Intelligence (Contractor), Office of the Commissioner
U.S. Food & Drug Administration

U.S. FOOD & DRUG ADMINISTRATION

CLINICAL TRIALS TRANSFORMATION INITIATIVE

Concluding Remarks

Qi Liu
Associate Director for Innovation & Partnership,
Office of Clinical Pharmacology, Office of
Translational Sciences, CDER
U.S. Food & Drug Administration

U.S. FOOD & DRUG ADMINISTRATION

CLINICAL TRIALS TRANSFORMATION INITIATIVE

THANK YOU!

Please take the post-workshop survey here:

Note: We will also email a copy for your convenience.