



National Institutes of Health
Office of Data Science Strategy

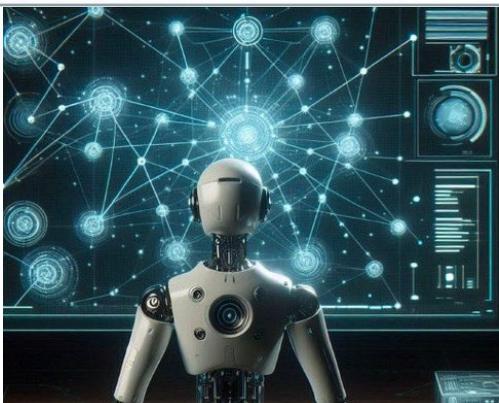
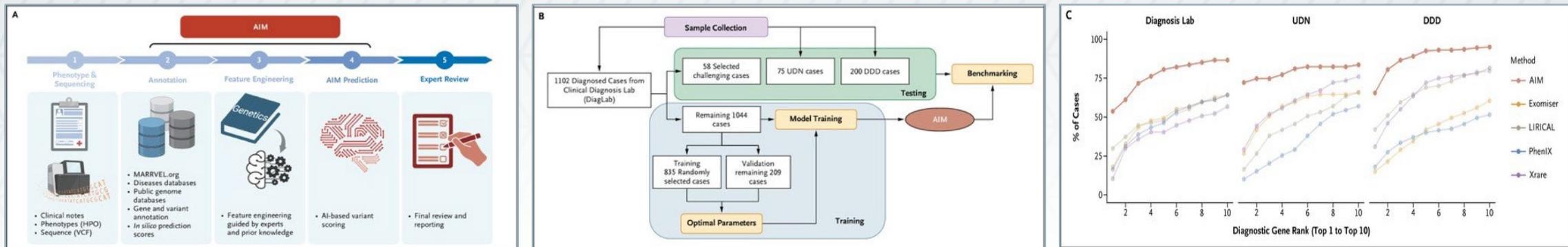
AI Activities, AI Education opportunities, and AI protection and security

Dr. Susan Gregurick,
Associate Director for Data Science, NIH
Wake Forest University

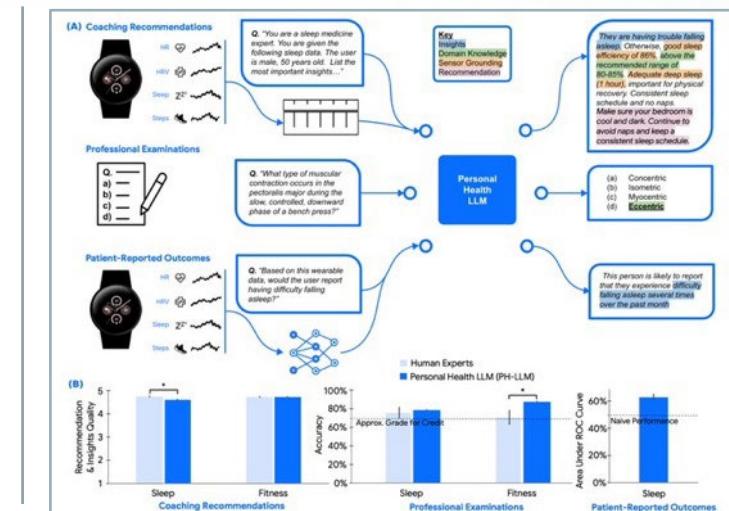
28 June 2024

What's happening right now

AI-MARRVEL – A Knowledge-Driven AI System for Diagnosing Mendelian Disorders | NEJM AI



GNN-RAG brings together the knowledge graph-processing abilities of graph neural networks and the language abilities of LLMs to unlock new applications.



Google just published a Personal Health Large Language Model

Accelerating Trustworthy AI



NATIONAL ARTIFICIAL INTELLIGENCE INITIATIVE
OVERSEEING AND IMPLEMENTING THE USE OF TRUSTWORTHY ARTIFICIAL INTELLIGENCE

ADVANCING TRUSTWORTHY AI

<https://www.ai.gov/strategic-pillars/advancing-trustworthy-ai/>



FEDERAL REGISTER
The Daily Journal of the United States Government

Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government

A Presidential Document by the Executive Office of the President on 12/08/2020

<https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government>



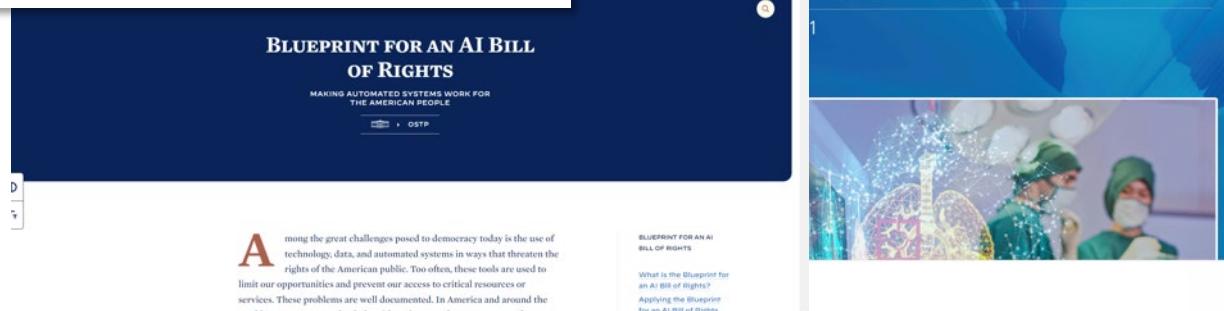
OCTOBER 30, 2023

Trustworthy AI (TAI) Playbook

HEALTH & HUMAN SERVICES

Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence

BRIEFING ROOM ▶ PRESIDENTIAL ACTIONS



BLUEPRINT FOR AN AI BILL OF RIGHTS
MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE
OSTP

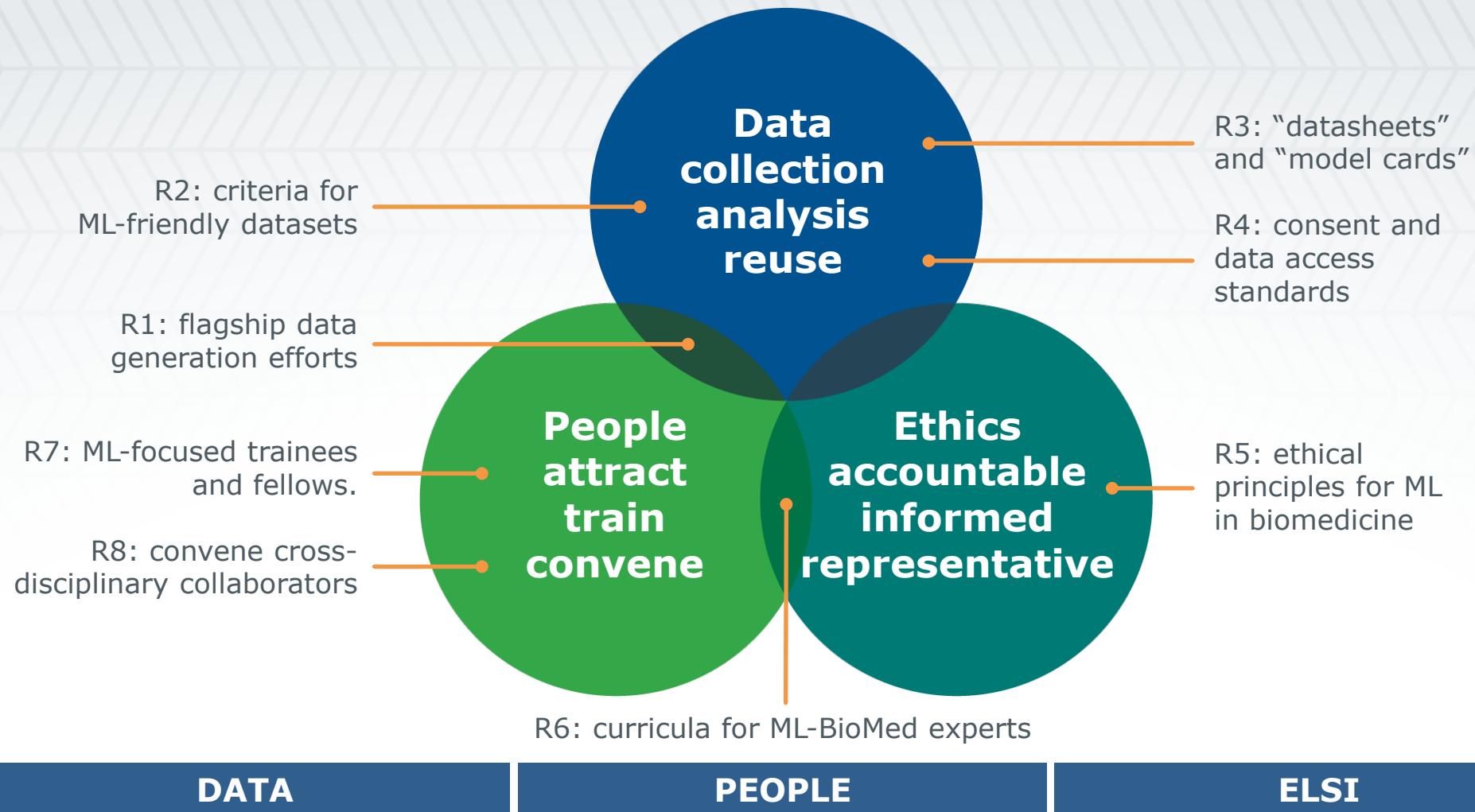
Among the great challenges posed to democracy today is the use of technology, data, and automated systems in ways that threaten the rights of the American public. Too often, these tools are used to limit our opportunities and prevent our access to critical resources or services. These problems are well documented. In America and around the world, systems supposed to help with patient care have proven unsafe,

<https://www.hhs.gov/sites/default/files/hhs-ai-strategy.pdf>
<https://www.hhs.gov/sites/default/files/hhs-trustworthy-ai-playbook.pdf>

Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence

- Promoting responsible innovation, competition, and collaboration
- Establish a program to identify and **attract top talent in AI**
- Launch a pilot program **implementing the National AI Research Resource (NAIRR)**.
- Support **2024 Leading Edge Acceleration Project** cooperative agreement awards to improve healthcare-data quality, support the responsible development of AI tools for clinical care, real-world-evidence programs, population health, public health, and related research.
- **Accelerate the National Institutes of Health Artificial Intelligence/Machine Learning Consortium to Advance Health Equity and Researcher Diversity (AIM-AHEAD) program** and showcasing current AIM-AHEAD activities in underserved communities.

Recommendations from 2019 ACD AI Working Group

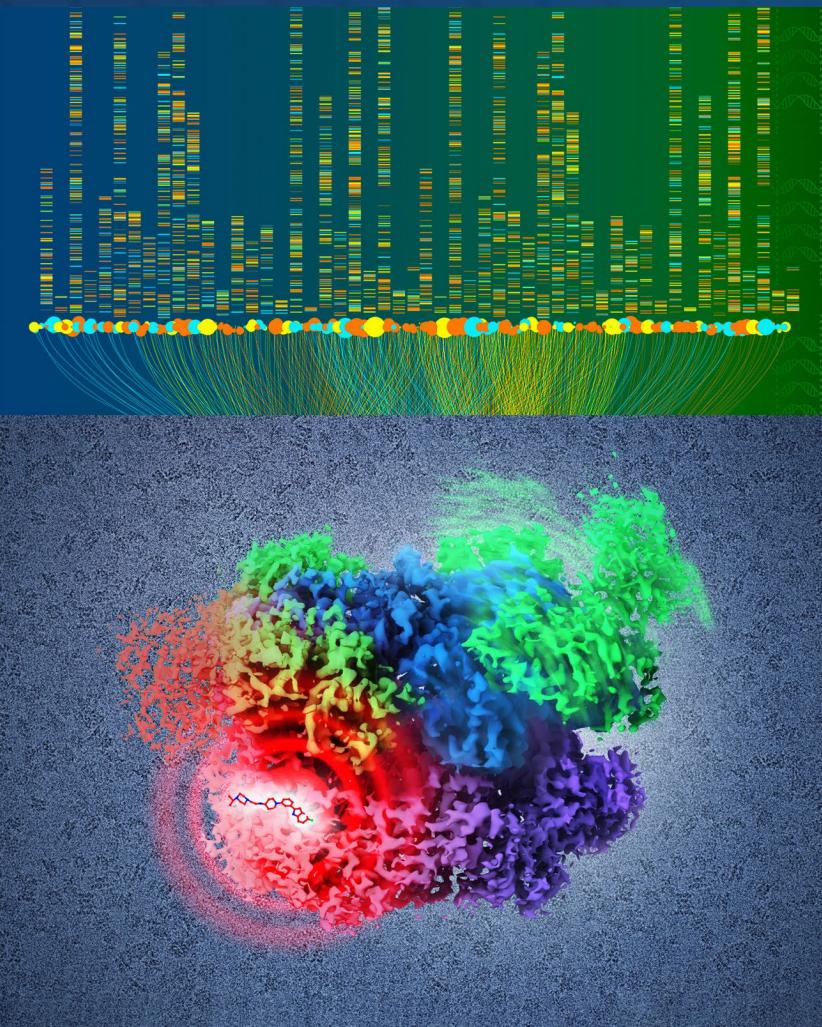


DATA

PEOPLE

ELSI

Progress is accelerated when advanced scientific methods, such as new data analytics, are applied to data that includes everyone, and when new discoveries are rapidly and equitably adopted in clinical care.



Examples of Progress

Bridge2AI to generate new “flagship” datasets and best practices for machine learning analysis.

\$300,000,000

AIM-AHEAD to enhance participation of underrepresented communities in AI/ML research.

\$250,000,000

ScHARe to test AI bias mitigation strategies and to advance health disparities research.

\$200,000,000

DEMONSTRATE to guide healthcare providers and systems in safe opioid prescribing.

\$150,000,000

CARD to extract insights on disease risk and protective factors from large networks of data

\$100,000,000

Improved operations in health through developing AI computation tools to improve screening

\$50,000,000

\$0

AI Total Budget

2019 2020 2021 2022

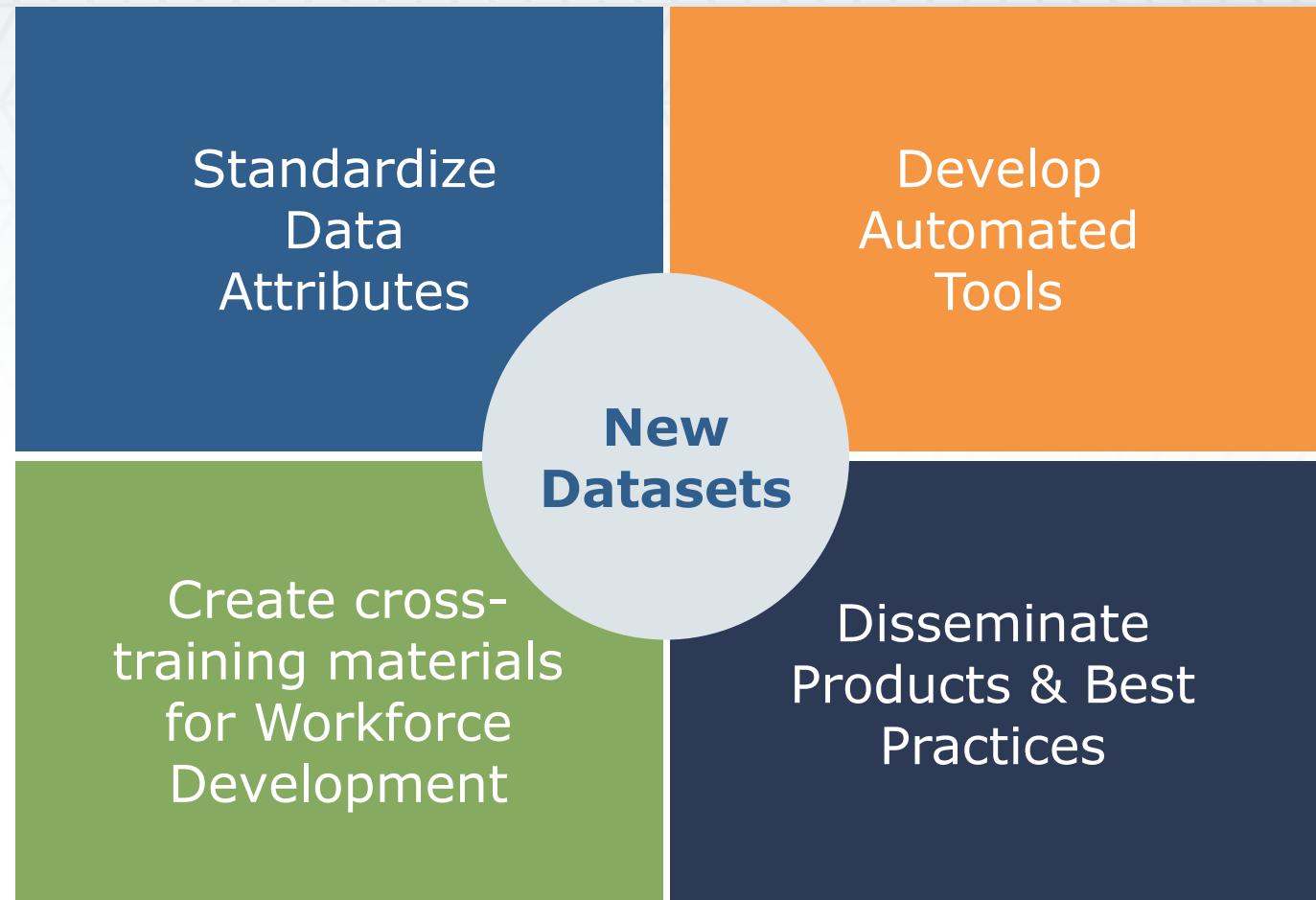
Challenges

- **Long development times** and low success rates in developing new treatments
- **Lack of clinical and healthcare related** data, with careful annotations, that are representative of the full diversity of the American people
- **Lack of data in lived experiences**, historical/cultural contexts such as social determinants of health (SODH)
- **Need for validated ethical and trustworthy AI algorithms** in health settings
- **Lack of diversity in AI workforce** and challenges in disseminating AI technologies to diverse populations

Bridge2AI Program Goals

Determining how to:

- Use biomedical and behavioral research grand challenges to generate **flagship data sets**
- **Prepare** AI/ML-friendly data
- Emphasize **ethical** best practices
- Promote **diverse teams**



Bridge2AI Milestones

The program established a public portal (<https://bridge2ai.org/>) for disseminating information regarding program activities and products.

Data

- The four data generation projects (DGPs) have started to release data
- External users worked with data in April 2024

People

- 5 internship programs to train researchers in both AI and biomedical research

Ethics

- Paper submitted (under review) which discusses the ethics considerations as part of data and AI modeling lifecycle

About the Artificial Intelligence/Machine Learning Consortium to Advance Health Equity and Researcher Diversity (AIM-AHEAD)



Goals

Enhance the participation and representation of researchers and communities currently underrepresented in the development of AI

Address health disparities and inequities using AI/ML

Improve the capabilities of this emerging technology

<https://aim-ahead.net/>

AIM-AHEAD Impact / Year 1 & 2 Awards

AIM-AHEAD supported over 274 awards to increase researcher diversity, address data & AI biases, engage underserved communities, and build institutional capacity



Training Programs

- Leadership Fellowship (50)
- Research Fellowship (47)



Community Engagement

- Hub Pilot Projects (35)



AI Health Equity Research

- Pilot Projects (21)
- Consortium Projects (21)



Institutional Capacity Building

- Program for Artificial Intelligence Readiness (15)
- Data and Infrastructure Capacity Building (13)



Joint training to increase researcher diversity in AI/ML by leveraging *All of Us* and N3C datasets, infrastructure, and training components.



**212 Applications
25 trainees**



National
COVID
Cohort
Collaborative

**120 Applications
50 trainees**

Impact

as of June 5, 2024

4,418+ TOTAL MEMBERS

2,519 MENTEES

1,170 MENTORS

1,306 INSTITUTIONS

**AIM-AHEAD Named
in White House
Executive Order**



National Institutes of Health
Office of Data Science Strategy

Examples of AIM-AHEAD Supported AI Studies

AIM-AHEAD-supported studies have appeared in high-impact journals, including Nature Communication, Scientific Report, Journal of Medical Internet Research AI, PLOS One, Journal of Clinical Oncology, Journal of Systemics, Cybernetics and Informatics, etc.

Journal of Systemics, Cybernetics and Informatics (2023) 21(2), 13-20
<https://doi.org/10.54808/JSCI.21.02.13>

Teaching Health Informatics in Middle School: Experience from an NIH AIM-AHEAD pilot

Gregory TARDIEU
Alexandria City Public Schools
Alexandria, VA 22314, USA

Senait TEKLE
Biomedical Informatics Center, The George Washington University
Washington, D.C. 20037, USA

Linda ZANIN
Alexandria City Public Schools
Alexandria, VA 22314, USA

JMIR AI | Published on 6.12.2023 in Vol 2 (2023) | Journal Information ▾ | Browse Journal ▾ | [Submit](#)

Preprints (earlier versions) of this paper are available at <https://preprints.jmir.org/preprint/52888>, first published September 18, 2023.



Developing Ethics and Equity Principles, Terms, and Engagement Tools to Advance Health Equity and Researcher Diversity in AI and Machine Learning: Modified Delphi Approach

Rachele Hendricks-Sturrup¹ ; Malaika Simmons¹ ; Shilo Anders² ;
Kammarauge Aneni³ ; Ellen Wright Clayton² ; Joseph Coco² ; Benjamin Collins² ;
Elizabeth Heitman⁴ ; Sajid Hussain⁵ ; Karuna Joshi⁶ ; Josh Lemieux⁷ ;
Laurie Lovett Novak² ; Daniel J Rubin⁸ ; Anil Shanker⁹ ; Talitha Washington¹⁰ ;
Gabriella Waters¹¹ ; Joyce Webb Harris² ; Rui Yin¹² ; Teresa Wagner¹³ ; Zhijun Yin² ;
Bradley Malin²

OPEN ACCESS | ORIGINAL REPORTS |  | January 10, 2023 |      

Derivation and Validation of a Clinical Risk Assessment Model for Cancer-Associated Thrombosis in Two Unique US Health Care Systems

Authors: Ang Li, MD, MS  , Jennifer La, PhD , Sarah B. May, MS , Danielle Guffey, MS, Wilson L. da Costa Jr, PhD , Christopher I. Amos, PhD , Raka Bandyo, MS, ... [SHOW ALL](#) ..., and Nathanael R. Fillmore, PhD  | [AUTHORS INFO & AFFILIATIONS](#)

Publication: Journal of Clinical Oncology • Volume 41, Number 16 • <https://doi.org/10.1200/JCO.22.01542>

AIM-AHEAD

Research and Fellowship

Opportunities

IN AI/ML AND HEALTH EQUITY

The AIM-AHEAD Clinicians Leading Ingenuity IN Al Quality (CLINAQ) Fellowship Program is a one-year fellowship (September 16, 2024 - August 29, 2025) that seeks to operate in parallel with clinical practice to empower clinicians in the field of Artificial Intelligence/Machine Learning (AI/ML).

Application Due Date: June 30, 2024

**APPLY
TODAY**

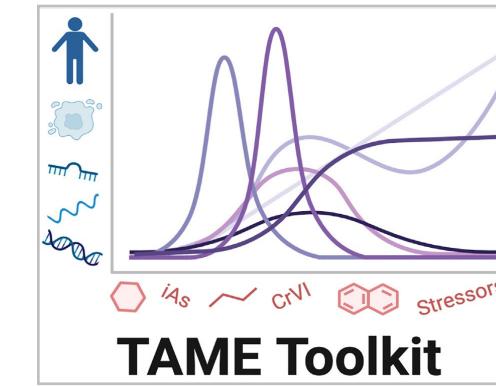
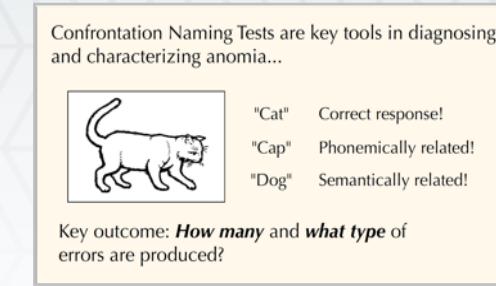


aim-ahead.net

NIH collaborative activities in AI



- NIH supplemental awards have resulted in new AI-Ready data sets and new training opportunities to develop the skills and competencies
- NIH takes an “ethics-first” approach by requiring that AI-Ready data be guided by a concern for human and clinical impact with attention to ethical, legal, and social implications of AI/ML



Towards Automatic Transcription of Post-Stroke Disordered Speech (NIDCD)

Stephen Bedrick, Oregon Health & Science University

Cancer Research Workforce Development in FAIR Artificial Intelligence and Machine Learning (NCI)

Douglas Cress, H. Lee Moffitt Cancer Center and Research Institute

The UNC inTelligence And Machine IEarning (TAME) Training Program (NIEHS)

Ilona Jaspers, The University of North Carolina at Chapel Hill

Supported activities to create AI-ready Data and Algorithms

- Data cleaning
- Improving data
- Removing spurious artifacts
- Adoption of ontologies
- Removing or characterizing biases
- Discovering and identifying imbalances
- Addressing specific challenges
- Developing and sharing documentation,
- Preparation of social determinants of health (SDOH)
- Preparing data

Multimodal AI

*Embed in context of ethical,
trustworthy AI practices
and assessment*

EXPECTED OUTPUTS:

- New systems-level biomedical research using multimodal AI technologies
- Elucidation of the unique opportunities, risks, and challenges for applying multi-modal AI methods
- Identification of considerations for the appropriate use of multimodal AI, relative to other methodologies



National Institutes of Health
Office of Data Science Strategy

Research Opportunity Announcement! **Apply to the Advancing Health Research Through Ethical, Multimodal AI Initiative**



**INFORMATIONAL
WEBINAR:**
April 19, 2024, 2-3pm EDT

**LETTER OF INTENT
(OPTIONAL) DUE:**
April 29, 2024

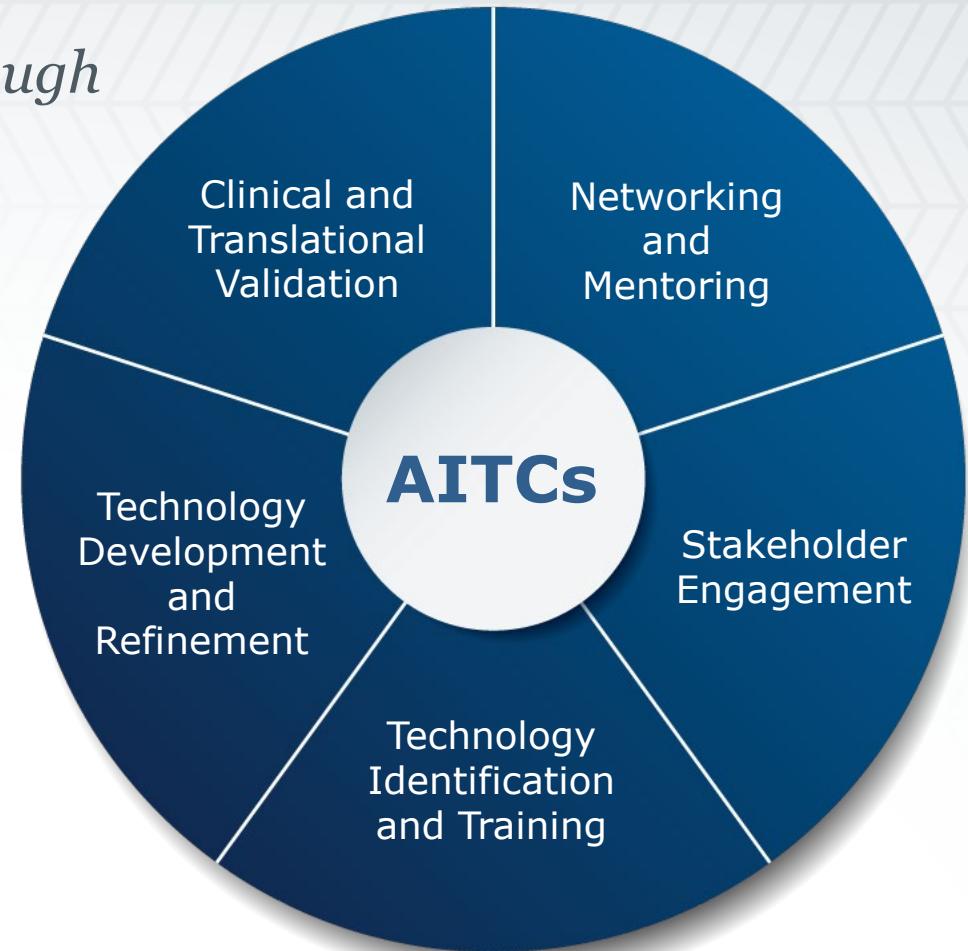
**PROPOSALS
DUE:**
May 16, 2024



Artificial Intelligence and Technology Collaboratories (AITC)

Helping Americans live longer, healthier lives through the application of AI and emerging technologies.

- AITCs focus on **advancing technology in age-related areas of need**, including:
 - Preventing cognitive decline
 - Caregiver support
 - Aging in place
 - End-of-life care
- AITCs use **AI/technology approaches**, including:
 - Generative AI
 - Robotics
 - Virtual reality
 - Wearable sensors



AITC PROJECT: Conversational AI to Detect Dementia in the Home

NEED:

Enhance early detection of dementia

OPPORTUNITY:

Develop conversational AI to predict cognitive status of older adults at home

APPROACH:

Administer clinical cognitive assessments in a HIPAA-secure environment using voice assistant technology



Agency Collaborations

- Machine Learning and Artificial Intelligence NSTC Subcommittee
- National AI Research Resource (NAIRR)
- Health and Human Services AI Task Force
- NCI-DOE Collaboration for Advanced AI to end Cancer

Impacts of NCI-DOE collaboration:

- **30 AI/ML publicly-available resources**
- **134 publications** since 2016
- **50+ public and private organizations** participating in innovation challenges



AGENCY COLLABORATIONS: National AI Research Resource (NAIRR)

NATIONAL AI RESEARCH RESOURCE: a shared research infrastructure facilitating access to compute, software, datasets, models, training and user support for researchers and students.

OBJECTIVE: to strengthen and democratize the U.S. AI Innovation ecosystem in a way that protects privacy, civil rights, and civil liberties.

GOALS:



Spur
innovation



Increase the
diversity of
talent in AI



Improve U.S.
capacity for
AI R&D



Advance
trustworthy
AI

NSF | NIH | DOE | NASA | NOAA

Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem

*An Implementation Plan for a
National Artificial Intelligence Research Resource*



NIH Contributions to the NAIRR Pilot



DATA: NIH contributes large amounts of curated, interoperable, AI-Ready Data



SECURITY: NIH brings different data together in secure, privacy-preserving ways to protect patient safety in AI



TRUSTWORTHY AND ETHICAL AI: NIH adopts ethical practices and addresses bias, diversity, and transparency to build trustworthy AI



BROADENING PARTICIPATION: NIH cloud-based platforms and programs lower the barriers to data science and grow capacity in under-served and -resourced communities

NIH Contributions to NAIRR Pilot

NIH Data and Computational Infrastructure Ecosystem

| | |
|-----------------------|---|
| Governance | <ul style="list-style-type: none">• Experience developing and overseeing federated interoperability |
| NAIRR Open | <ul style="list-style-type: none">• Integration of <u>ImmPort</u> datasets into the NAIRR• Integration of Health Equity Action Network (HEAN) datasets and <u>ScHARE</u> analysis tools into the NAIRR |
| NAIRR Secure* | <ul style="list-style-type: none">• Integration of the Medical Imaging and Data Resource Center (<u>MIDRC</u>) and National COVID Cohort Collaborative (<u>N3C</u>) into NAIRR Secure |
| Software Stack | <ul style="list-style-type: none">• Coordinate with NSF and DOE a NAIRR software stack community workshop |
| Classroom | <ul style="list-style-type: none">• NIH Cloudbase and other platform tools leveraged in NAIRR |
| Outreach | <ul style="list-style-type: none">• Leverage NIH networks to attract diverse users and data |



AI Ready Datasets for COVID and Clinical Modeling

National COVID Cohort Collaborative (N3C)



Data Diversity, Representation, and Harmonization are foundations for robust, trustworthy AI

COVID+
CASES
8.9m

Patients
22.5m

Rows of Data
32.7b

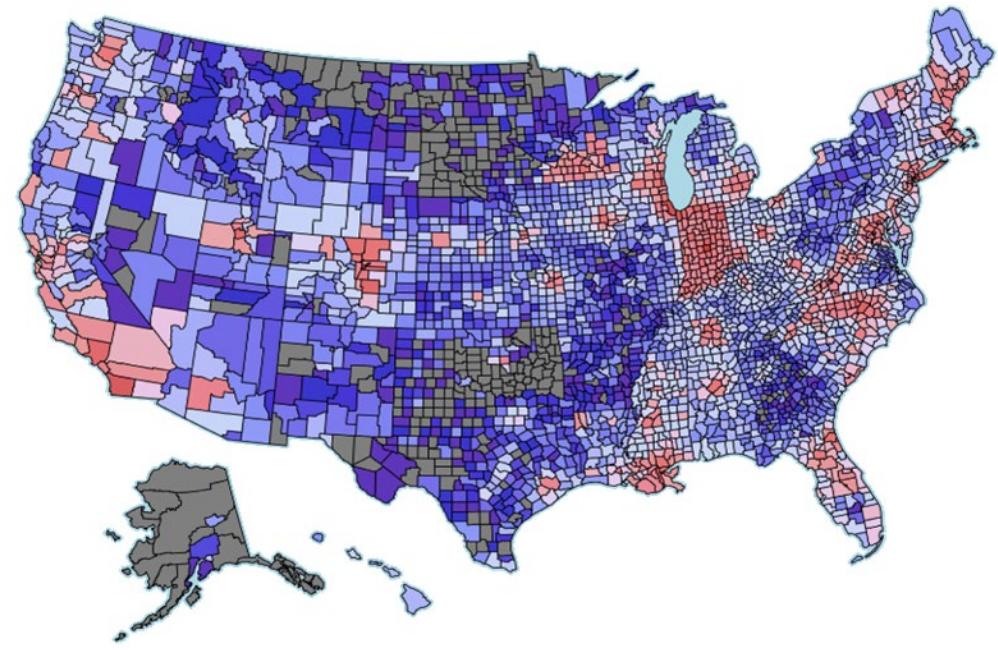
Contributors
Health Systems
84

Institutions
Using Data
391

Active
Investigators
>3900

Research
Studies
553

Citations and
H-Index
3890 / 29



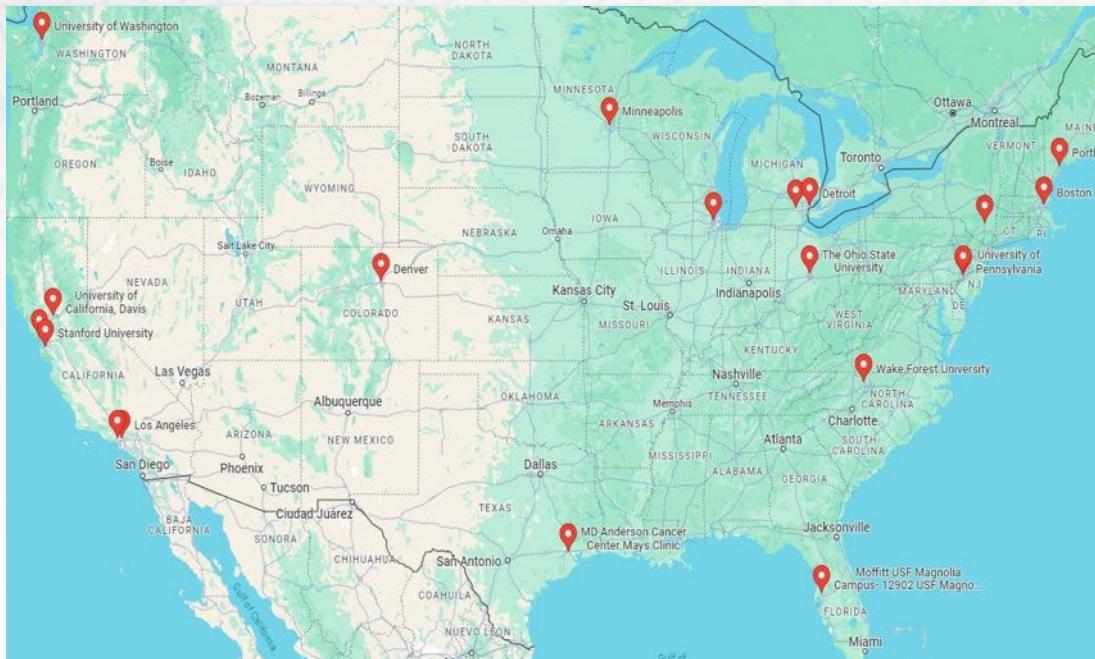
- Geographics: 50/50 States >92% of all US Counties in USA
- Representative of US population
- Source: Community, Academic, FQHCs
- Patient Mix: Inpatient ~20%, Outpatient ED ~80%
- Longitudinal Data: 1/1/2018 to Present
- <https://covid.cd2h.org/>



AI Ready Datasets for Medical Imaging AI Applications

Medical Imaging and Data Resource Center (MIDRC)

Information dense medical imaging data has many applications in AI algorithm development



Map of MIDRC Investigators

A collaboration of ~20 institutions and > 100 investigators from academia, community practices, FDA, and others

<https://www.midrc.org/>

309,270
Imaging studies
ingested

254

Total data downloads
this month

58
Publications

847
Registered Users

176,961
Imaging studies
released to the public

73,695
Cases

140+
Presentations

100+
Investigators

132,309
Imaging studies
undergoing quality and
harmonization

14.39 TB
Total size published

31
Algorithms

632
Collaborating
institutions



National Institute of
Biomedical Imaging
and Bioengineering



NATIONAL CANCER INSTITUTE
Center for Biomedical Informatics
& Information Technology

Medical Image De-Identification Benchmark (MIDI-B)

Challenge

The MIDI-B Challenge invites developers to evaluate their DICOM image de-identification tools, emphasizing automated methods that maintain research utility.

By utilizing diverse medical imaging data from the National Cancer Institute's Cancer Imaging Archive, participants can contribute to the MIDI-B Challenge report as co-authors.

Registration Ends: August 15, 2024

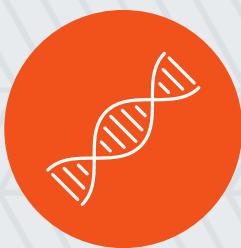
REGISTER



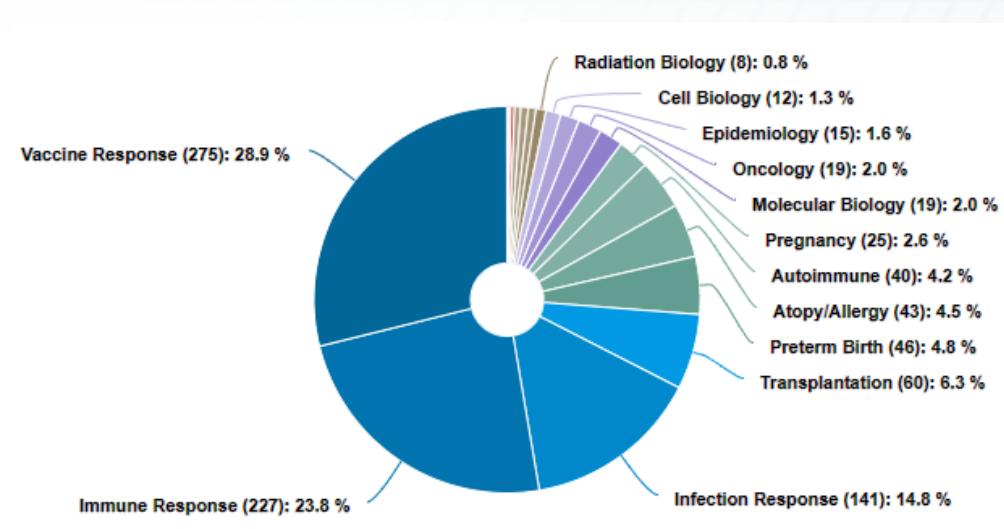
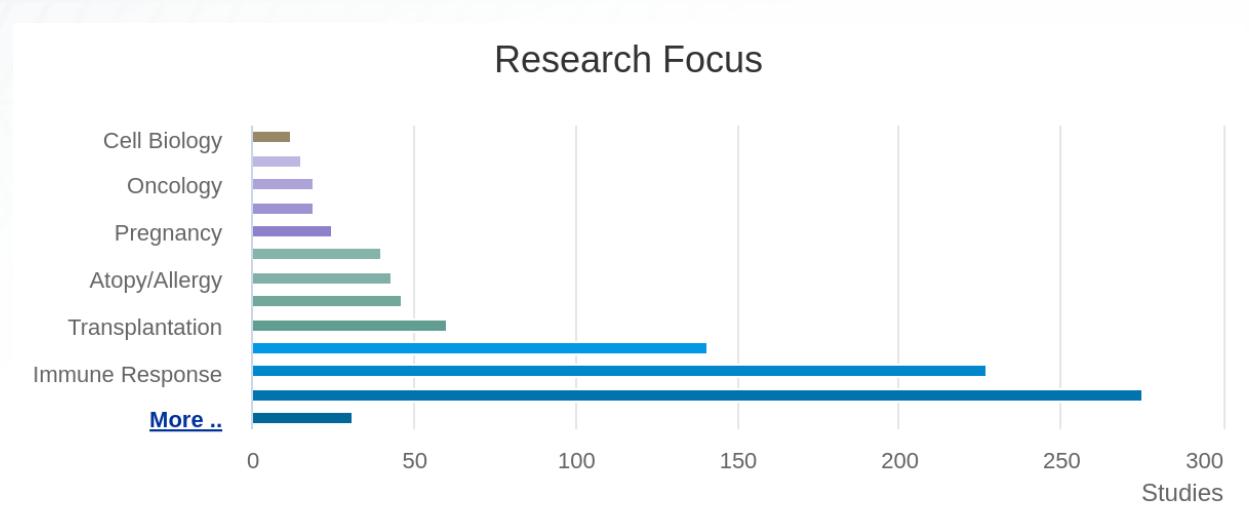
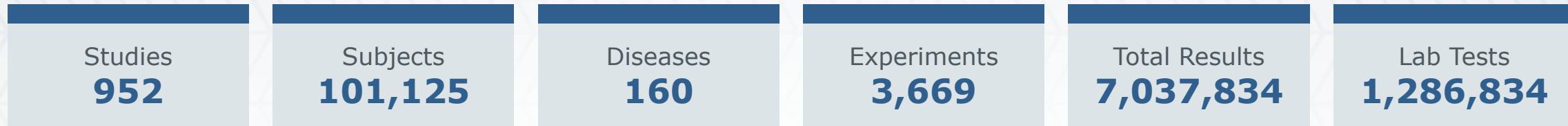
[datascience.
cancer.gov](https://datascience.cancer.gov)

AI Ready Datasets for Immunology Research

ImmPort



Immunology data supports AI development and data reuse in a broad range of applications across test and study types



AI Ready Datasets for Social Determinants of Health (SDOH) ScHARe Health Equity Action Network (HEAN)

Social Determinants of Health brings the lived experience to AI algorithms and meets communities where they are.



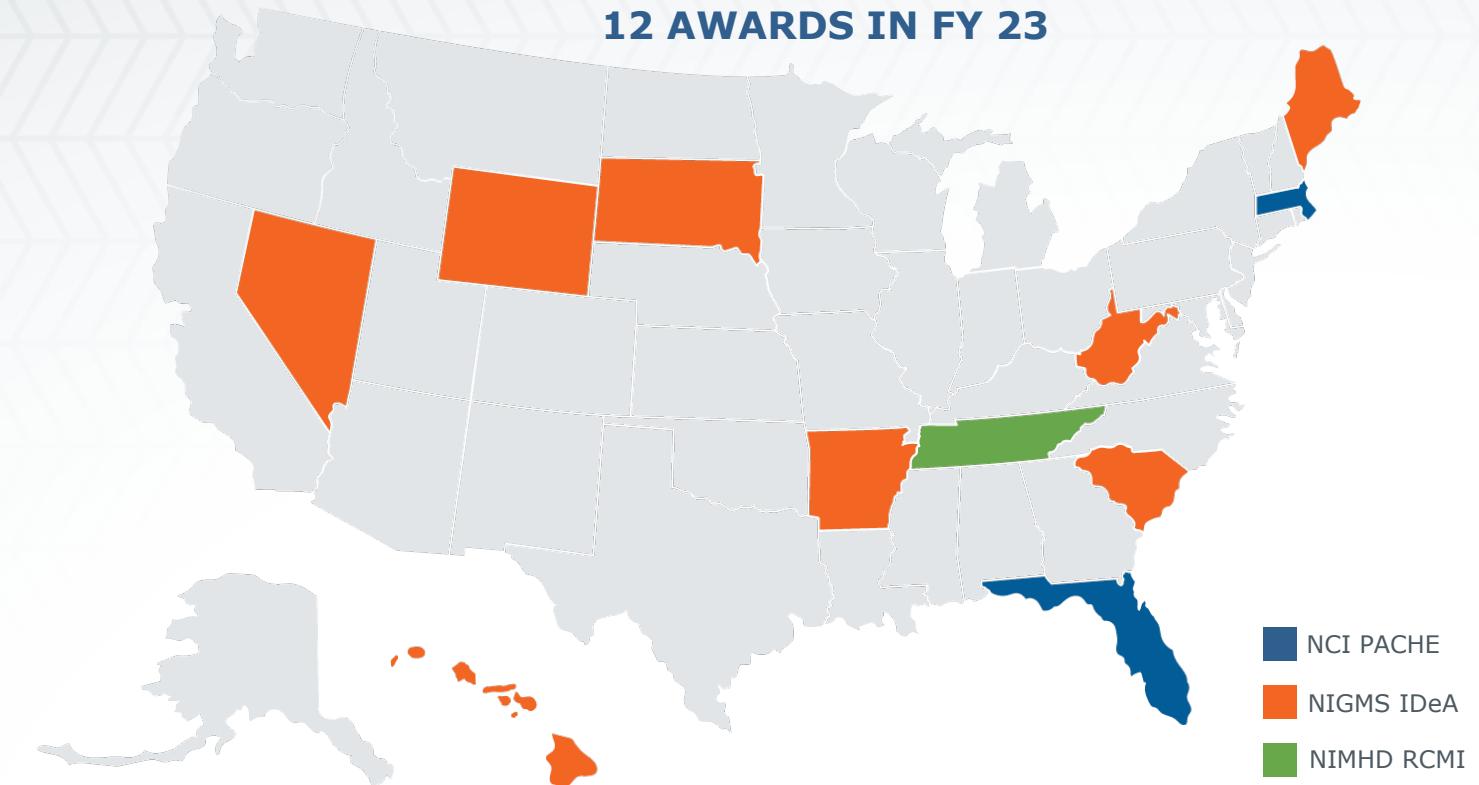
<https://www.nimhd.nih.gov/resources/schare/>



ODSS Supports Awards to Enhance Institutional Data Science Capacity

Particularly in institutions serving medically underserved populations and underrepresented students.

12 AWARDS IN FY 23



GOAL

To enhance institutional data science capacity by growing human capital, expanding institutional infrastructure, and building partnerships.

ACTION

- Published NOT-OD-23-123 in collaboration with:
- NIMHD Research Centers in Minority Institutions (RCMI) program;
- NIGMS Institutional Development Award (IDeA) program; and
- NCI Partnerships to Advance Cancer Health Equities (PACHE) program.

OUTCOME

- 19 applications and 12 awards
- Initiated PI meeting series
- Next receipt date: April 1, 2024

Support for AI-Enabled research

- Validation of Digital Health and Artificial Intelligence/Machine Learning Tools for Improved Assessment in Biomedical and Behavioral Research (NOT-CA-24-031)
- Harnessing Artificial Intelligence and Polypharmacology to Discover Pharmacotherapeutics for Substance Use Disorders (R41/R42 Clinical Trials Not Allowed) (RFA-DA-25-053 and RFA-DA-25-054 for R43/R44)
- Advancing Data Science Research in HIV: Responding to a Dynamic, Complex, and Evolving HIV Epidemic with Artificial Intelligence/Machine Learning (NOT-MH-23-350)
- Smart Health and Biomedical Research in the Era of Artificial Intelligence and Advanced Data Science (Partnership with NSF, NOT-OD-23-165)
- Explainable Artificial Intelligence for Decoding and Modulating Neural Circuit Activity Linked to Behavior (NOT-MH-23-110)



NATIONAL ACADEMY OF MEDICINE

ARTIFICIAL INTELLIGENCE IN HEALTH,
HEALTH CARE, AND BIOMEDICAL SCIENCE:

Discussion Draft

AI Code of Conduct Principles and Commitments

A new #NAMPerspectives outlines a framework for stakeholders from various disciplines and industries to align and harness the benefits of AI, while also safeguarding against potential harms. Through responsible implementation, the commentary argues that we can foster accurate, safe, reliable, and ethical AI advancements in health, health care, and biomedical science.

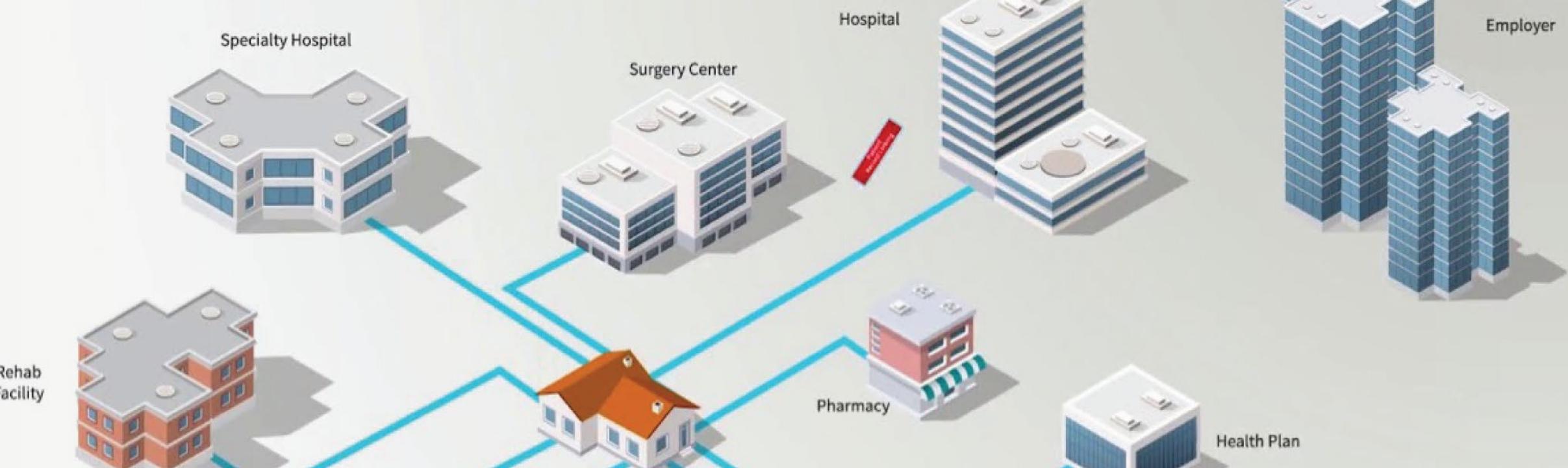
**READ
MORE**



nam.edu

AI at NIH

- AI and machine learning methods are not new to NIH
- We lack data resources at a scale necessary to optimally use AI/ML methods to improve health
- We must respectfully engage people to earn trust
- Diversity is critical to avoid harm to some populations
- Current needs:
 - Dramatic increase in data collection from the clinical care environment, including all populations
 - NIH resource to facilitate optimal and especially ethical use of AI for health



- Facilitate learning health system initiatives by achieving better data to assess health outcomes that matter to individual people and to society overall
- Eliminate costly data formatting and collection redundancies that create silos
- Reduce clinical care site burden for data submission
- Increase data quality and speed time to data access for use in agency decision-making
- Deliver data from the clinical care environment with diversity sufficient to allow use of AI methods



**Thank you
for your time and
attention today**

Are there any questions?