Data Science in the next 5 years

- Improve Capabilities to Sustain the NIH Policy for Data Management and Sharing

- Develop Programs to Enhance Human Derived Data for Research

- Provide New Opportunities in Software, Computational Methods, and Artificial Intelligence

- Support for a Federated Biomedical Research Data Infrastructure

- Strengthen a Broad Community in Data Science

Goal 1

Capabilities to Sustain the NIH Data Management and Sharing Policy

Challenges

- Need for the generation of FAIR Data in a manner that will foster greater sharing and the integration of scientific results
- Need for cost effective strategies for sustainable, secure, and accessible biomedical data repositories and knowledgebases

Objectives to Address Challenges

1) Support the biomedical community to manage and share data

2) Enhance FAIR data and greater data harmonization

3) Strengthen NIH's data repository and knowledgebase ecosystem

Goal 2

Enhance Human Derived Data for Research

Challenges

- Need for acquisition and protection of data obtained from electronic health records, and other real-world data, that preserves privacy and enhances participant consent
- challenges in data quality, privacy and confidentiality, policy, regulatory, and ethical issues associated with healthcare and administrative data
- need to better understand the ethical, legal, and social implications of data linkage

Objectives to Address Challenges

1) Improve access to and use of clinical and real-world data
2) Adopt health IT standards for research
3) Enhance the adoption of social and environmental determinants of health for health equity

Goal 3

New Opportunities in Software, Computational Methods, and AI

Challenges

- Emergence of innovations in trustable artificial intelligence (AI) approaches that reduces bias and risks
- Multi-dimensional data integration remains a significant challenge for biomedical and behavioral research

Objectives to Address Challenges

1) New opportunities to improve biomedical AI analysis
2) Develop cutting edge software technologies
3) Support FAIR software sustainability

NIH Office of Data Science Strategy

Accelerating Trustworthy AI

NATIONAL ARTIFICIAL INTELLIGENCE INITIATIVE
OVERSEEING AND IMPLE...

ADVANCING TRUSTWORTHY AI

Trustworthy AI (TAI) Playbook
U.S. DEPARTMENT OF HEALTH & HUMAN SERVICES

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

OCTOBER 30, 2023

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

🏛 ▸ BRIEFING ROOM ▸ PRESIDENTIAL ACTIONS

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

FEDERAL REGI
NATIONAL ARCHIVES — The Daily Journal of the United States

🏠 ▪ Sections ▾ | ⊙ Browse ▾ | ᐩ Search ▾ | 🔖 Reader Aid...

**Promoting the Use of Trustwor...
Federal Government**

A Presidential Document by the Executive Office of the Pre...

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

FDA **U.S. FOOD & DRUG ADMINISTRATION**
CENTER FOR DEVICES & RADIOLOGICAL HEALTH

**Artificial Intelligence/Machine Learning (AI/ML)-Based
...s a Medical Device (SaMD) Action Plan**

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

A mong 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,

BLUEPRINT FOR AN AI BILL OF RIGHTS

What is the Blueprint for an AI Bill of Rights?
Applying the Blueprint for an AI Bill of Rights

NIH Office of Data Science Strategy

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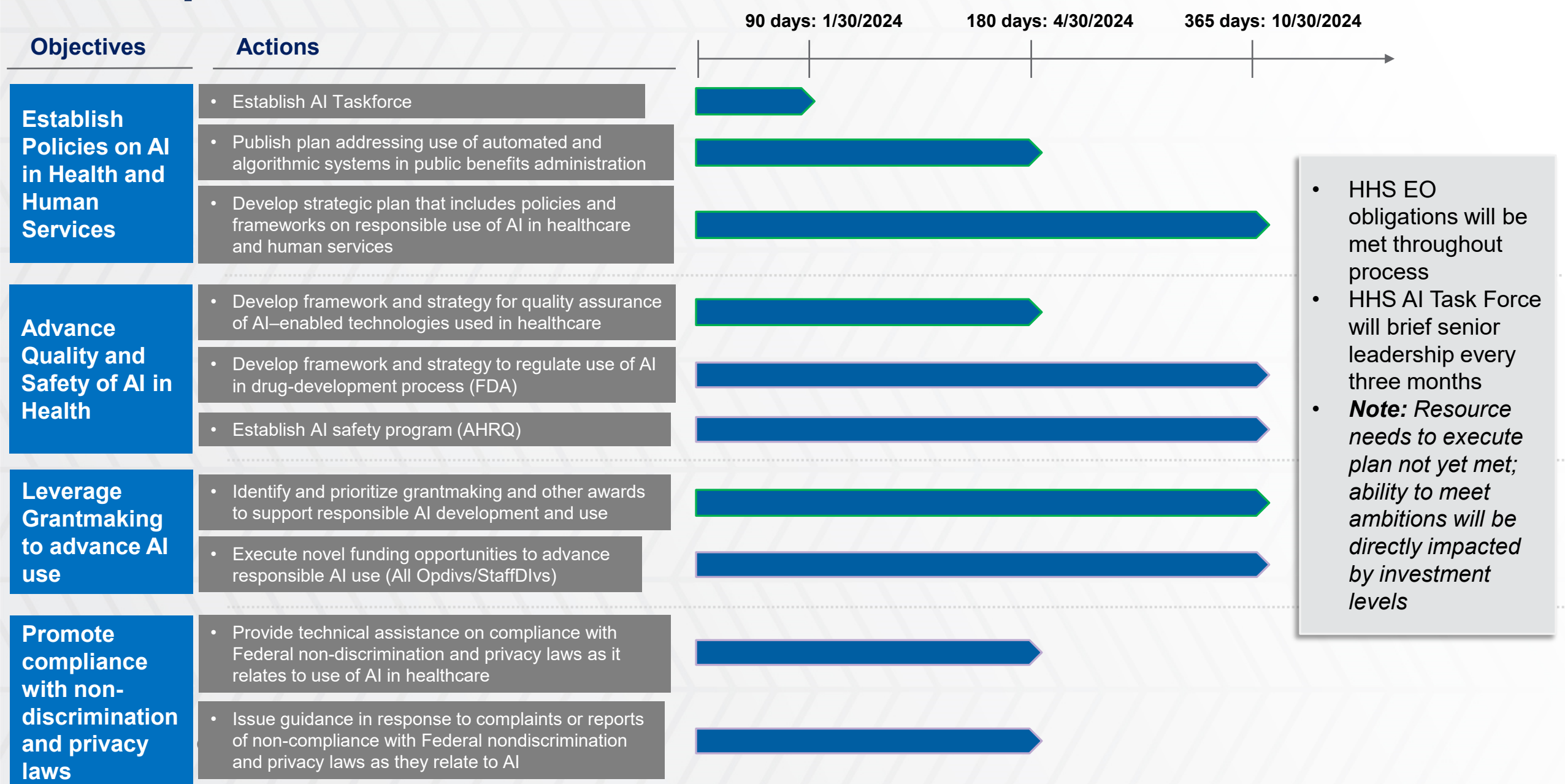
Objectives

Establish Policies on AI in Health and Human Services	**Advance Quality and Safety of AI in Health**	**Leverage Grantmaking to advance AI use**	**Promote compliance with non-discrimination and privacy laws**

NIH Office of Data Science Strategy

Examples of Progress

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

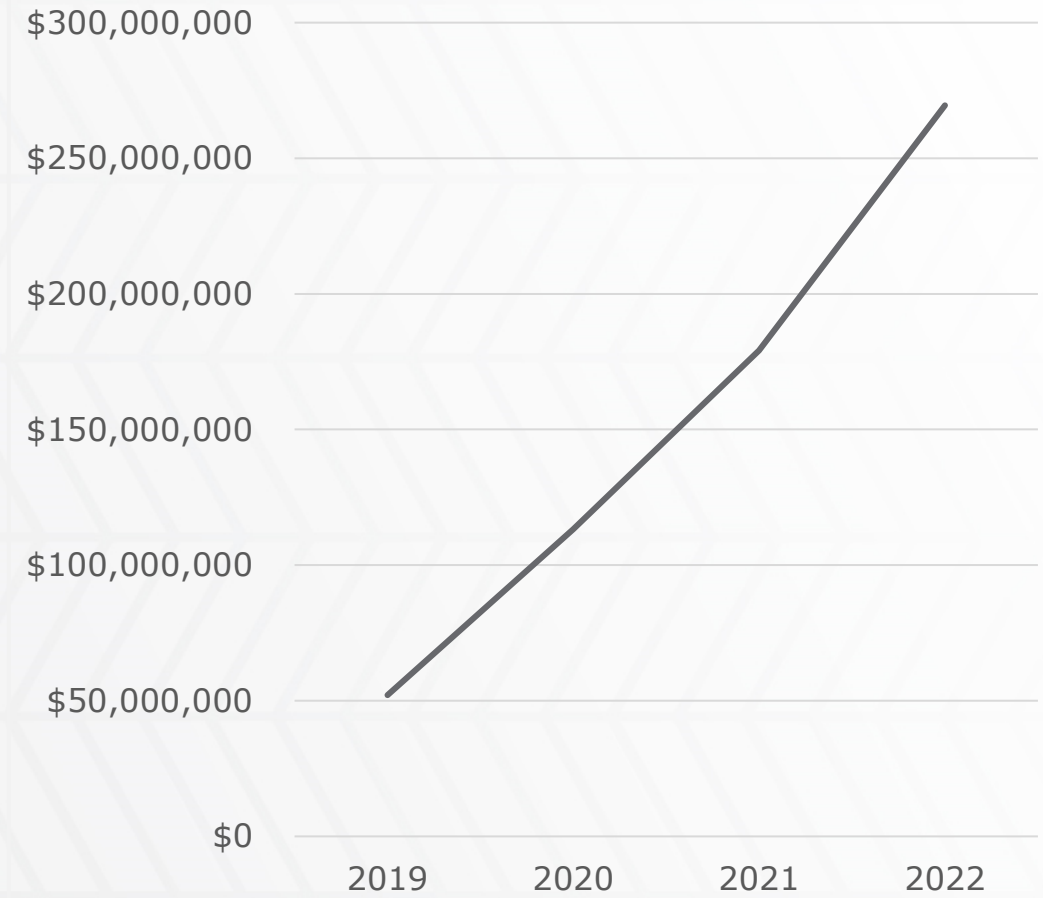**AIM-AHEAD** to enhance participation of underrepresented communities in AI/ML research.

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

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

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

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

AI Total Budget

$300,000,000

$250,000,000

$200,000,000

$150,000,000

$100,000,000

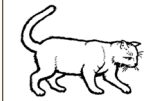$50,000,000

$0

2019   2020   2021   2022

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

Confrontation Naming Tests are key tools in diagnosing and characterizing anomia...

"Cat"   Correct response!
"Cap"   Phonemically related!
"Dog"   Semantically related!

Key outcome: **How many** and **what type** of errors are produced?

ICADS
Integrated Program in Cancer and Data Science

Clinical and Research Data Made AI FAIR
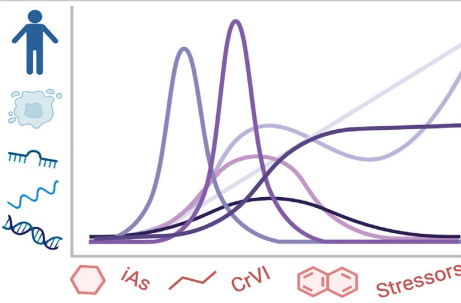
- Findable
- Accessible
- Interoperable
- Reusable

- Coursework via USF
- Workshops via RET
- Public Videos via YouTube

WORKFORCE
Members Impacted
- Ph.D Students
- Postdoctoral Fellows
- Moffitt Faculty and Staff
- Potential Clinical Trainees
- Researchers world-wide
- Clinicians world-wide

Machine and Deep Learning Algorithm

Clinical Decision Making → Improvement in Health and Patient Outcome

iAs   CrVI   Stressors

TAME Toolkit

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 lEarning (TAME) Training Program (NIEHS)
Ilona Jaspers, The University of North Carolina at Chapel Hill

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

An Implementation Plan for a National Artificial Intelligence Research Resource

January 2023

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:

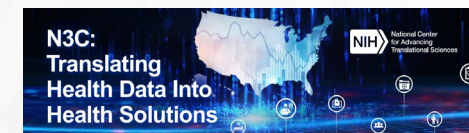| Spur **innovation** | Increase the **diversity** of talent in AI | Improve U.S. **capacity** for AI R&D | Advance **trustworthy AI** |

NSF | NIH | DOE | NASA | NOAA

NIH Contributions to NAIRR Pilot

Governance	• Experience developing and overseeing federated interoperability
NAIRR Open	• Integration of *ImmPort* datasets into the NAIRR • Integration of Health Equity Action Network (HEAN) datasets and ScHARE analysis tools into the NAIRR
NAIRR Secure*	• Integration of the Medical Imaging and Data Resource Center (MIDRC) and National COVID Cohort Collaborative (N3C) into NAIRR Secure
Software Stack	• Coordinate with NSF and DOE a NAIRR software stack community workshop
Classroom	• NIH Cloudlab and other platform tools leveraged in NAIRR
Outreach	• Leverage NIH networks to attract diverse users and data

* NIH and DOE jointly lead NAIRR Secure

NIH Data and Computational Infrastructure Ecosystem

NIH⟩ *STRIDES Initiative*
STRIDES / CloudLab

IMMPORT Private Data
Your site for managing ImmPort data uploads

ScHARe

N3C: Translating Health Data Into Health Solutions — National Center for Advancing Translational Sciences

MIDRC
MEDICAL IMAGING AND DATA RESOURCE CENTER.

Goal 4

Support for a Federated Biomedical Research Data Infrastructure

Challenges

- Creation of opportunities for exploration of new technologies and computing paradigms for biomedical research

Objectives to Address Challenge

1. Develop, test, validate, and implement ways to integrate NIH data and infrastructure
2. Ensure a robust and connected data resource ecosystem that includes collaborative data management platforms, curation, analysis, and sharing of data and metadata
3. Develop new capabilities for data search and discovery

Goal 5

Strengthen a Broader Community in Data Science

Challenges

- Develop and nurture data science talent from a diverse array of scientific interests and institutions

Objectives to Address Challenge

1. Increase training opportunities in data science
2. Develop and advance initiatives to expand the data science workforce
3. Broaden and champion capacity building and community engagement efforts
4. Enhancing data science collaboration within the NIH Intramural Research Program

RFI for NIH Strategic Plan for Data Science

- Read and submit your comment on the draft NIH Strategic Plan for Data Science, 2023-2028

- **The NIH seeks comments on any of the following topics**:
  - The appropriateness of the goals of the plan, the strategies and implementation tactics proposed to achieve them; including potential benefits, drawbacks or challenges
  - Opportunities for NIH to partner to achieve these goals
  - Emerging research needs and opportunities that should be added to the plan.
  - Any other topic the respondent feels is relevant for NIH to consider in developing this strategic plan.

- **Last day to submit: March 15, 2024**

https://bit.ly/3vc4MTq

Thank you for your time and attention