



Artificial Intelligence/**M**achine Learning Consortium
to **A**dvance **H**ealth **E**quity **A**nd Researcher **D**iversity

AIM-AHEAD *All of Us* Training Program

Cohort II

Informational Webinar

October 31, 2024, 1:00pm Central

Welcome to AIM-AHEAD



Introduction

The Artificial Intelligence/Machine Learning Consortium to Advance Health Equity and Research Diversity (AIM-AHEAD) program was established by the National Institutes of Health (NIH).

Purpose

The purpose of AIM-AHEAD is to enhance diversity in the field of artificial intelligence and machine learning (AI/ML), with emphasis on reducing health disparities and promoting health equity. This will be achieved by engaging in a fair, equitable, and transparent process of building a consortium of AI/ML partners to promote health equity and an inclusive and diverse workforce.

Consortium Building

Many communities have untapped potential to contribute new expertise, data, recruitment strategies, and cutting-edge science to the AI/ML field. The AIM-AHEAD Coordinating Center (A-CC) was created to increase participation and engagement through mutually beneficial partnerships, stakeholder engagement, and outreach to advance health equity.

The AIM-AHEAD Coordinating Center



Introduction

The A-CC consists of four cores, focused on various initiatives to achieve AIM-AHEAD's mission. The cores include institutions and organizations that have a mission to serve underrepresented or underserved groups impacted by health disparities.

Leadership Core

Lead, recruit, and coordinate the AIM-AHEAD Consortium

Data Science Training Core

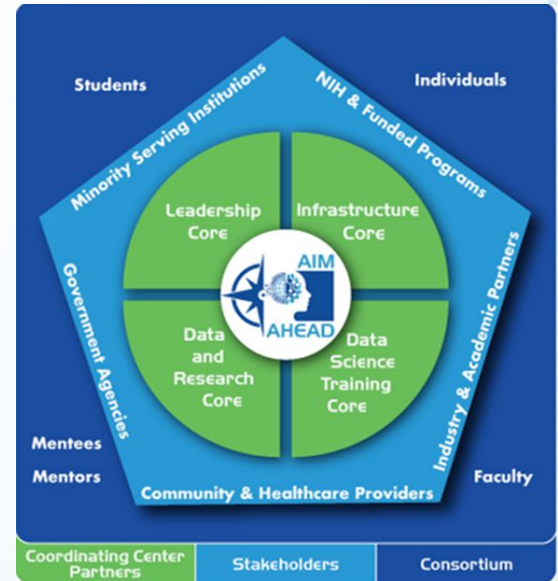
Assess, develop, and implement data science training curriculum

Data and Research Core

Address research priorities and needs to form an inclusive basis for AI/ML

Infrastructure Core

Assess data, computing, and software infrastructure to facilitate AI/ML and health disparities research



NIH AIM-AHEAD Leadership Team



Samson Gebreab
*Program Lead, AIM-AHEAD
Office of Data Science Strategy,
NIH*



Eva Lancaster
*Program Director
Office of Data Science Strategy,
NIH*



Emir Khatipov
*Program Director
Office of Data Science Strategy,
NIH*



Christian Evans
*Program Specialist, AIM-AHEAD
Office of Data Science Strategy,
NIH*

Program Directors



**Robert T. Mallet, PhD,
Co-Director**

Regents Professor, Department of
Physiology and Anatomy
University of Texas Health Science
Center, Fort Worth, TX



**Toufeeq A. Syed, PhD,
Co-Director**

Associate Professor and Assistant Dean of
Education Informatics
McWilliams School of Biomedical Informatics
University of Texas Health Science Center,
Houston, TX



**Legand L. Burge III, PhD,
Co-Director**

Professor of Computer Science
Executive Director, Howard West Initiative
Howard University, Washington, DC

Program Purpose



Purpose

The central goal of the AIM-AHEAD *All of Us* Training Program is to increase researcher diversity in AI/ML by training individuals from diverse backgrounds who are committed to gaining proficiency in AI/ML data analysis and applying their expertise to benefit communities underrepresented in biomedical research.

Program Partnership



Partnership

The AIM-AHEAD consortium (Data Science Training Core and Communications Hub), *All of Us*, and RTI International are partnering to offer AIM-AHEAD stakeholders, trainees, mentees, and consortium partners a training opportunity designed to increase researcher diversity in AI/ML by leveraging the *All of Us* data and infrastructure (Researcher Workbench).



Researcher Workbench



All of Us
RESEARCHER WORKBENCH

The Researcher Workbench is a cloud-based platform where registered researchers access Registered and Controlled Tier data. It provides tools for data analysis, storage, and collaboration, allowing high-powered queries using R or Python within the integrated Jupyter Notebook environment.

Data Now Available in the Researcher Workbench



413,350+
Survey Responses



337,500+
Physical Measurements



312,900+
Genotyping Arrays



287,000+
Electronic Health Records



245,350+
Whole Genome Sequences



15,600+
Fitbit Records



1,000+
Long-Read Sequences

Training Overview



Trainees will learn to use tools available in the Researcher Workbench to access data within the *All of Us* database.

Trainees will complete courses and receive support on using R, Python, and Jupyter Notebook. Additionally, trainees will complete data use case cases to support model development for *All of Us* data subsets. Data use case training will include:

Validating models

**Merging/validating data
across *All of Us* sources**

**Considering biases that may
be present and detected or
missed by the model**

Building a supervised model

**Splitting data into subsets for
model training and testing**

Program Trainee Objectives



Objective 1

Analyze

The trainee will apply R, Python, and/or Jupyter Notebook to analyze *All of Us* datasets from diverse and underrepresented communities.



Objective 2

Hypothesize

The trainee will formulate hypotheses testable by applying AI/ML and advanced data analyses to *All of Us* data.



Objective 3

Present

The trainee will present their project at the AIM-AHEAD Annual Meeting 2025.



Outcome

After completing advanced training in coding, model development, hypothesis testing, and data analysis, trainees will be equipped to apply AI/ML approaches to analyze complex datasets. They will join a committed community of professionals dedicated to extending AI/ML benefits to underrepresented communities in biomedical research.

Program Benefits



Stipend

An \$8,000 stipend upon successful completion of trainee milestones

A \$2,000 allowance to attend the AIM-AHEAD Annual Meeting 2025



Support

Support and guidance from an experienced AIM-AHEAD mentor

Support from the AIM-AHEAD Data Science Training Core

Direct 1:1 guidance, virtual office hours, AIM-AHEAD HelpDesk support, and concierge services

Support from RTI Workbench and ML Coaches



Training

Training on:

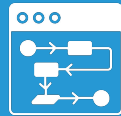
- Use and applications of R, Python, and Jupyter Notebook
- Hypothesis development for testing by analysis of *All of Us* data
- Data analysis using *All of Us* Researcher Workbench
- Model building, data merging, and validation across *All of Us* sources
- Data splitting methods for model training vs. testing
- Detecting and addressing biases in model development

AIM-AHEAD Mentorship Process



Each trainee will be matched with a mentor who will provide ongoing support throughout the training program. Mentors are matched with mentees using the Connect Platform. Mentorship matches are made using:

AIM-AHEAD CONNECT



AI Algorithm



Administrative
Matching



Mentor Pool
Search

Applicant Eligibility



Institution

Higher Education Institutions

Public, Private, HSIs, HBCUs, TCUs, AANAPISI, or NAH Serving Institutions

Non-Academic Organizations

Nonprofits with or without 501(c)(3) status, Tribally derived institutions, or For-Profit Businesses

Data Use and Registration Agreement (DURA)

Must hold an active [DURA](#) or obtain one within a month of award



Applicant

Citizenship

Must be a U.S. Citizen, Permanent Resident, or Non-Citizen U.S. National

Education

Post-baccalaureate and graduate students, early-career investigators, or employees with a Bachelor's degree in a related field

Skills & Experience

Prior programming experience and statistics knowledge

*Experience in R/Python coding, data management, and coursework in statistics is **strongly recommended.***

Application Process



Applications must be submitted between October 18, 2024 and November 18, 2024

Note: Please use Chrome, Firefox, or Edge browsers.



Familiarize yourself with the program requirements outlined in the call for applications



Create an account on [AIM-AHEAD Connect](#) and register as a “mentee/learner”



Gather all of the required application materials



Submit application for review using the [InfoReady](#) platform



25 trainees will be selected

Application Requirements



Submission Deadline: November 18, 2024 by 11:59 PM EST

- ✓ **Profile Information:** Name, organization, department, position, research area, and contact.

- ✓ **Letters of Support:** A supervisor's letter confirming training time and contact info is required, along with one faculty recommendation attesting to the applicant's skills and readiness for advanced data analytics.

- ✓ **Transcripts:** Official or photocopy of undergraduate and graduate (if applicable).

- ✓ **NIH Biosketch:** Max 5 pages.

- ✓ **Statement of Rationale:** Max 900 words—goals, research question, coding plan, relevant experience, and long-term objectives.

Program Timeline



Funding Cycle

2024-2025



Release Date

October 18, 2024



Application Deadline

November 18, 2024 by 11:59 PM EST



Notice of Award

January 6, 2025



Program Start Date

January 15, 2025



AIM-AHEAD Annual Meeting 2025

July 2025



Program Length

8 month program



Application Resources

(Items linked)

- **CFA Link** (QR code on last slide)
- **InfoReady**
- **AIM-AHEAD Connect**
- **NIH biosketch sample**

Data Use and Registration Agreement (DURA)

(Items linked)

- **List of institutions with active DURAs**
- **DURA Request Form**

NOTE: If a DURA is not currently held by your institution, one must be obtained within 30 days of the program start date in order to remain in the program.



Assessing Research Topic Viability for *All of Us*

(Items linked)

All of Us Data Repository: Comprehensive details on the entirety of the *All of Us* data repository

All of Us Data Dictionaries: What data fields are available?

All of Us Data Browser: What survey data, health conditions, and other data types are available?

Research Projects: conducted using the *All of Us* data

Cohort I Metrics



PROGRAM REFLECTIONS

"I ENJOYED THE HANDS-ON INTERACTION WITH THE NOTEBOOKS AND THE CHALLENGES AT THE END, WHICH ENCOURAGED ME TO PUT WHAT I REVIEWED INTO PRACTICE."

"THE FLOW OF THE PYTHON COURSE WAS VERY WELL STRUCTURED, AND THE JUPYTER NOTEBOOK MADE FOLLOWING THE VIDEOS EASY."

"THE COURSE WAS CONCISE AND EACH STEP WELL EXPLAINED."

"THE EXAMPLES GIVEN TO DEMONSTRATE R WERE AWESOME."

KEY OUTCOMES



95%

CONFIDENT IN EXPLAINING AI/ML CONCEPTS TO OTHERS
(19/20 TRAINEES)



95%

PROFICIENT IN DRAWING INSIGHT FROM COMPLEX DATA SETS
(19/20 TRAINEES)



70%

CAPABLE OF USING AI/ML TO UNDERSTAND COMMUNITY HEALTH DISPARITIES
(14/20 TRAINEES)



82%*

CONFIDENT IN REAL-WORLD APPLICATION
(47/57 RESPONSES)

*AVERAGE BASED ON 3 COURSE EVALUATIONS WITH 57 TOTAL TRAINEE RESPONSES



100% COURSE COMPLETION



96% PROGRAM GOALS COMPLETED ON THE CONNECT PLATFORM

83%

ATTENDED 2024 AIM-AHEAD ANNUAL MEETING

74%

PRESENTED AT 2024 AIHES POSTER SESSION

+110

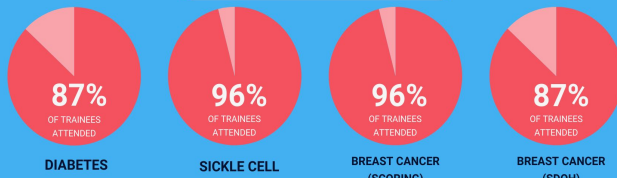
SUCCESSFUL MENTOR ENGAGEMENTS



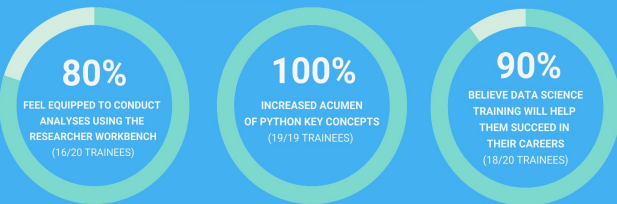
TRAINEE MILESTONE COMPLETION



DATA USE CASE TRAININGS



ACHIEVEMENT METRICS



This report provides a comprehensive snapshot of key metrics from the AIM-AHEAD *All of Us* Training Program for Cohort 1, comprised of 23 trainees. It highlights participation, feedback, task completion, and overall progress toward the program's goals. (Some metrics may vary based on trainee evaluation response totals; variations have been indicated below).

Questions?



Please feel free to
submit a Help Desk
Ticket
Training Program
Help Desk:

[LINK](#)

Use the QR code
above to access
the AIM-AHEAD *All
of Us* Training
Program Call for
Applications