Detecting and Addressing
BIAS
in Data, Humans, and Institutions

Stony Brook University

Mission: to seed a generation of researchers trained to identify and mitigate biases that arise when data-centric methods are applied to real-world problems
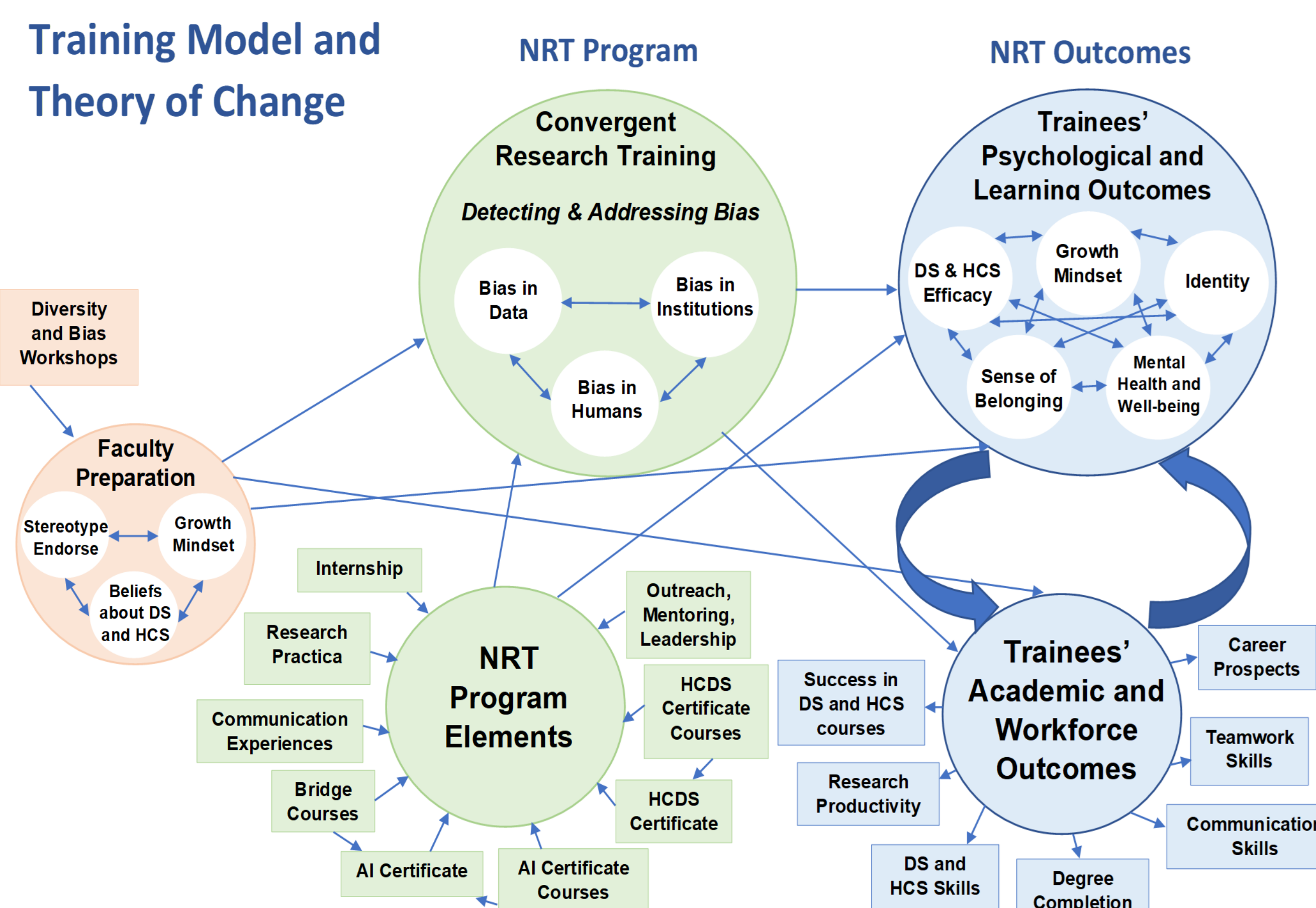
Introduction

Data science and AI are powerful tools for generating new knowledge, fueling innovation, and dealing with society's most pressing problems. However, "big data" and machine learning tools can perpetuate biases that advantage some people, and disadvantage others. This training project (NSF 2125295) bridges perspectives from the human-centered sciences with those from the data sciences in support of convergent research projects.

This project includes faculty and Ph.D. students from the data sciences (DS), as well as from the human-centered sciences (HCS). Participating Ph.D. programs include **Africana Studies, Applied Mathematics & Statistics, Cognitive Science, Computer Science, Economics, Linguistics, Political Science, Psychology, Neurobiology & Behavior, and Sociology.**
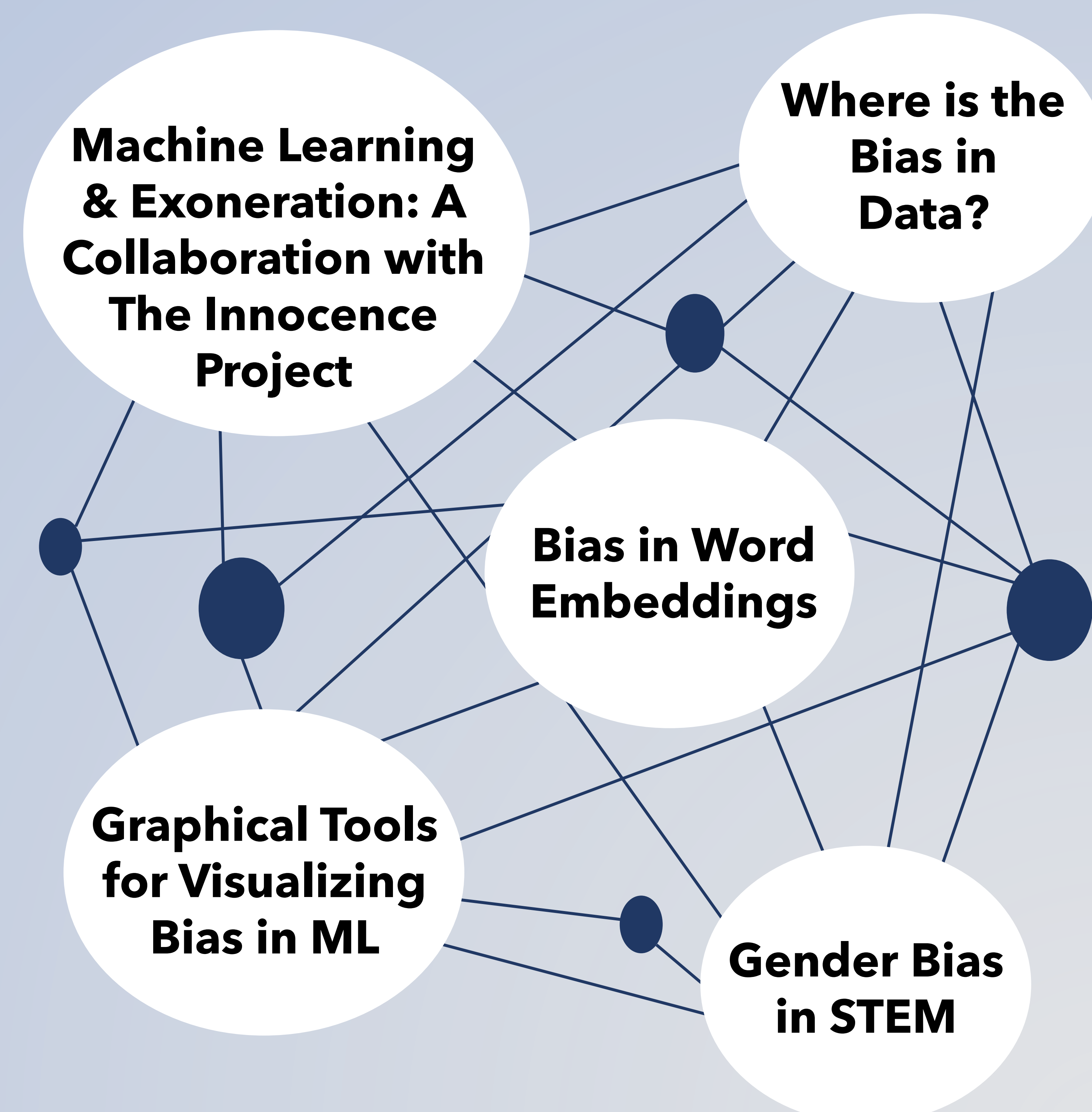
Y1 Achievements

✓ Recruited first cohort of 11 trainees from 7 PhD programs
✓ Launched Fall 2022 Research Practicum
✓ Created Bridge Mentoring Structure (DS students paired with HCS faculty and vice versa)
✓ Began approval process for 2 graduate certificates

2022-2023 Cohort: Rosa Bermejo, Kalina Kostyszyn, Carl Wiedemann, James May, John Murzaku, Veronica Oelerich, Sekine Ozturk, Amie Paige, Karin Hasegawa, Pei-Hsun Hsieh, Medhini Urs

Leadership Team: Susan E. Brennan (PI), C.R. Ramakrishnan, Wei Zhu, Bonita London, Jeffrey Heinz
Project Coordinator: Kristen Kalb-DellaRatta
Project Evaluation: Catherine Good, Elevate Learning, LLC

Features of the Traineeship Model

Convergent Research Practica
Data Science Internships
Bridge Courses

AI Focus   **HCDS Focus**

DS/CS Core Courses
34 Funded Trainees
34 Non-funded Trainess
Travel Funds

Research Practicum, Fall 2022
Sampling of Topics Underway:

- **Machine Learning & Exoneration: A Collaboration with The Innocence Project**
- **Where is the Bias in Data?**
- **Bias in Word Embeddings**
- **Graphical Tools for Visualizing Bias in ML**
- **Gender Bias in STEM**

Training Model and Theory of Change

NRT Program — NRT Outcomes

Convergent Research Training — *Detecting & Addressing Bias*: Bias in Data, Bias in Institutions, Bias in Humans

Diversity and Bias Workshops

Faculty Preparation — Stereotype Endorse, Growth Mindset, Beliefs about DS and HCS

NRT Program Elements — Internship, Research Practica, Communication Experiences, Bridge Courses, AI Certificate, AI Certificate Courses, Outreach, Mentoring, Leadership, HCDS Certificate Courses, HCDS Certificate

Trainees' Psychological and Learning Outcomes — DS & HCS Efficacy, Growth Mindset, Identity, Sense of Belonging, Mental Health and Well-being

Trainees' Academic and Workforce Outcomes — Success in DS and HCS courses, Research Productivity, DS and HCS Skills, Degree Completion, Career Prospects, Teamwork Skills, Communication Skills