This project is based on work supported by NSF Grant #2125295

Stony Brook University

Detecting and Addressing BIAS in Data, Humans, and Institutions

About Us

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 project (NSF #2125295) bridges perspectives from the human-centered sciences with those from the data sciences to train PhD students from 8 different STEM disciplines to collaborate on high-impact convergent research.

Top row (left to right): Sekine Ozturk (Psychology, Clinical), Srivardhan Jangili (political Science), Tina Behzad (Computer Science), Ignacio Urbina (Political Science), Weiling LI (Psychology, Cognitive), Veronica Oelerich (Political Science).
Second row (left to right): Karin Hasegawa (Applied Mathematics & Statistics), Rosa Bermejo (Psychology, Social & Health), Peter Zeng (Computer Science), Pei-Hsun Hsieh (Political Science), MacKenzie Johnson (Psychology, Cognitive), Medhini Urs (Psychology, Cognitive).
Third row (left to right): Zhengxiang (Jack) Wang (Linguistics), Ritik Raina (Psychology, Cognitive), James May (Psychology, Cognitive), John Murzaku (Computer Science), Kalina Kostyszyn (Linguistics), Kiera Gross (Computer Science).
Fourth row (left to right): Carl Wiedemann (Psychology, Social & Health), AJ Castle (Women's, Gender, and Sexuality Studies), Dasha Likhacheva (Psychology, Social & Health), Dana Golden (Economics), Gilvir Gill (Computer Science), Evan West (Computer Science).
Bottom row (left to right): Brett Indelicato (Economics), Adil Soubki (Computer Science), Alexandra Anthonioz (Psychology, Social & Health), Amit Kumar Das (Computer Science), Amie Paige (Psychology, Cognitive), Benjamin Hechtman (Applied Mathematics & Statistics).

Top row (Leadership Team): Susan E. Brennan (PI), C.R. Ramakrishnan (Co-PI), Wei Zhu (Co-PI)
Bottom Row: Atefeh Hamedi (Project Coordinator), Catherine Good (Project Evaluator).

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

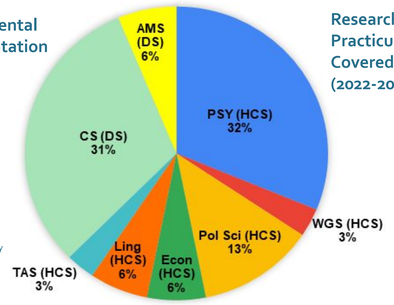
Convergent Research Practicum
Optional Data Science Internships
Cross-Disciplinary Courses
Cross-Disciplinary Mentoring

Data Science (DS)

Human-Centered Science (HCS)

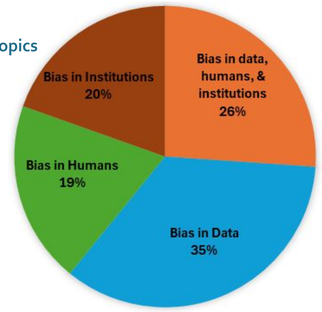
DS Courses: Python, Machine Learning
HCS Research & Methods Courses
34 NRT-Funded Trainee-Years
1/2 of Trainees Have Been Funded
Travel Awards for All

Departmental Representation

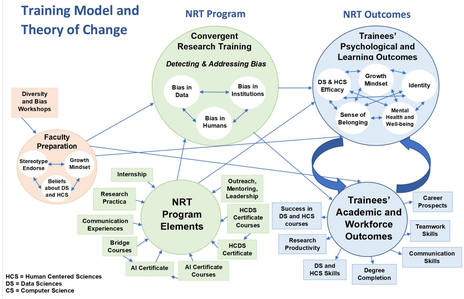


AMS = Applied Math & Statistics
CS = Computer Science
WGS = Women's, Gender, & Sexuality Studies
TAS= Technology, AI, and Society
Econ = Economics
Ling = Linguistics
Poli Sci = Political Science
Psy = Psychology

Research Practicum Topics Covered (2022-2025)



Training Model and Theory of Change



HCS = Human Centered Sciences
DS = Data Sciences
CS = Computer Science

Participating Faculty

Trainee Research Highlights

Our project, now it its 5th funded year, has been successful at retaining trainees (whether domestic or international) throughout their doctoral training; most continue participating in the Research Practicum, held weekly during each semester. Here, we focus on two convergent research projects led by talented international students from Computer Science, Linguistics, and Psychology.

Bias NRT Trainees Weiling Li (Cognitive Science) and Jack (Zhengxiang) Wang (Linguistics) presented their collaborative study "LVLMs are Bad at Overhearing Human Referential Communication" at the Conference on Empirical Methods in Natural Language Processing (EMNLP 2025).

LVLMs are Bad at Overhearing Human Referential Communication

Zhengxiang Wang, Weiling Li, Panagiotis Kaliosis, Owen Rambow, Susan E. Brennan

Abstract

During conversation, speakers collaborate on spontaneous referring expressions, which they can then re-use in subsequent conversation with the same partner. Understanding such referring expressions is an important ability for an embodied agent so that it can carry out tasks in the real world. This requires integrating and understanding language, vision, and conversational interaction. We study the capabilities of seven state-of-the-art Large Vision Language Models (LVLMs) as overhearers to a corpus of spontaneous conversations between pairs of human discourse participants engaged in a collaborative object-matching task. We find that such a task remains challenging for current LVLMs, which fail to show a consistent performance improvement as they overhear more conversations from the same discourse participants repeating the same task for multiple rounds. We release our corpus and code for reproducibility and to facilitate future research.

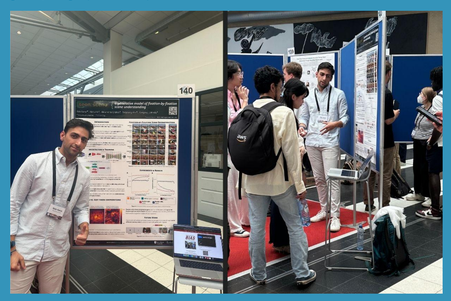
Wang, Z., Li, W., Kaliosis, P., Rambow, O., & Brennan, S. E. (2025, November). LVLMs are Bad at Overhearing Human Referential Communication. In Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (pp. 16769-16793).

Seen2Scene: a generative model of fixation-by-fixation scene understanding

Ritik Raina, Abe Leite, Alexandros Graikos, Seoyoung Ahn, Gregory J. Zelinsky

Abstract

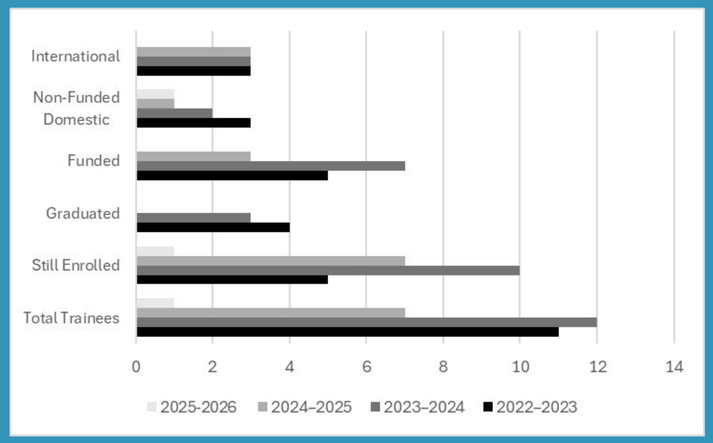
Human scene understanding dynamically evolves over the course of sequential viewing fixations from a gist-level understanding to a more detailed comprehension of the scene. Each fixation provides rich visual information about objects and their spatial relationships. To model this incremental process, we introduce Seen2Scene, a framework for modeling human scene understanding by controlling the inputs used to generate a visual hypothesis of the scene. Seen2Scene uses a self-supervised encoder to extract features from fixated scene regions, which guide a pre-trained text-to-image latent diffusion model through a modular adapter framework. As fixations accumulate, the model iteratively refines its visual hypotheses, filling in unseen areas with contextually plausible content. We evaluated Seen2Scene on COCOFreeView using two experimental conditions: fixation only conditioning to isolate the contribution of foveal information, and fixation+gist conditioning to examine how non-fixated scene information integrates with foveal details. Results show that initial fixations drive the greatest gains in semantic and perceptual fidelity and that the fixation+gist condition reached high-fidelity scene understanding with the fewest fixations, thus demonstrating the importance of integrating peripheral gist information with visual details collected foveally.



Bias-NRT Trainee Ritik Raina (Cognitive Science) presented a poster titled "Seen2Scene: a generative model of fixation-by-fixation scene understanding" at the 8th Annual Conference on Cognitive Computational Neuroscience.

Raina, R., Leite, A., Graikos, A., Ahn, S., & Zelinsky, G. J. (2025). Seen2Scene: A generative model of fixation-by-fixation scene understanding. Conference on Cognitive Computational Neuroscience (CCN).

Cohorts of Trainees 2022-2026



Sustainability and Impacts on the Workforce in Human-Centered Data Science

Our BIAS NRT project will be sustainable into the future via two core elements:

1. Our **Human-Centered Data Science (HCDS) Advanced Graduate Certificate,** approved in 2023 by the State University of New York (SUNY) and the New York State Education Department (NYSED), requires funded trainees and encourages non-funded trainees to "cross the bridge" by mastering advanced graduate coursework in both computational and human-centered data sciences, as well as to complete IRB Training. As of Spring 2026, 7 of 31 trainees have completed the certificate and 75% are still enrolled.

2. The weekly **Research Practicum** and the HCDS certificate will be integrated into a new Ph.D. program within Stony Brook's newly created *Dept. of Technology, AI, and Society*, created with funding from NY Governor Kathy Hochul.

Current and former trainees have recently interned or are currently employed at AAAS/U.S. Dept. of Transportation, U.S. Dept. of Health & Human Services, Northwell Health, CPCS (an Innocence Organization), Meta, Google, Amazon, Adobe, U. Penn, Oxford U., and Florida Southern College.

Evaluation: Impactful Program Components

Most Positively Rated Programmatic Components:

- Perceived Social Support
- Convergent Research Practicum
- Human-Centered Data Science Certificate

Trainees Rating the Research Practicum More Positively Have These Psychosocial Outcomes:

- Stronger *sense of belonging* in HCDS
- Higher ratings of HCDS Certificate as more useful
- More positive attitudes towards DS and HCDS

Research Practicum's Weekly Exit Ticket Feedback

- High satisfaction with intellectual quality and relevance
- Praise for breadth of topics
- Strong alignment of topics with program's focus
- High quality and accessibility of invited speakers
- RP contributes to community building
- RP Fosters professional development

What We Have Learned

- The culture of the program, the HCDS Certificate, and the Research Practicum are cornerstones of our success.
- Encouragement and support for mentoring is needed; adding peer mentoring enhanced mentorship quality.
- The Research Practicum intentionally provides opportunities for engagement and active collaboration.
- Trainee-led discussions build shared purpose, belonging, and identity (trainees led the entire practicum in Semester #7).

Learn More!

Poster created and presented by Atefeh Hamedi, Project Coordinator