

Today, computer scientists are asked to confront some of society's most pressing challenges, from reducing healthcare inequity to ensuring AI safety. My goal as an educator is to **develop responsible computer scientists** with the technical skills to build problem-solving technology and the critical perspective to ensure it benefits humanity.

TEACHING EXPERIENCE

I apply my teaching philosophy across courses, from large introductory classes to advanced seminars. As a **guest lecturer on AI Interpretability** in **MIT 6.C35: Interactive Data Visualization and Society** and **Brown CSCI 1032: Introduction to Sociotechnical Systems and HCI**, I paired the technical details of interpretability (e.g., monosemantic features) with discussions of their implications (e.g., human confirmation bias). Framed around a real case where a state-of-the-art AI model used meaningless features to diagnose cancer, I guided students in hypothesizing ways to identify and communicate its flaws. Through this process, students not only learned core interpretability concepts but also practiced critical reflection. This was exemplified by one student who, during our discussion, derived a novel interpretability method for communicating a language model's internal concepts to human stakeholders by generating maximally activating text sequences.

Developing responsible computer scientists also means helping students **connect computational thinking to real-world problems** early in their education. Over **two semesters as a graduate teaching assistant (TA)** for **MIT 6.1010 Fundamentals of Programming**, I led over 100 undergraduate TAs and developed teaching materials for 400+ students, most of whom were learning to program for the first time. To make programming concepts intuitive, I created assignments and recitations that connected each week's topic to a relatable problem, ranging from learning data structures by solving circuits to learning graphs by implementing autocomplete. By the end of the course, students with no prior coding experience were implementing complex program interpreters and fully functional games.

Finally, to ensure we can solve complex problems across domains, I aim to train a diverse community of technologists. Research in learning science shows that students learn most effectively when they are actively engaged and feel a sense of belonging [4, 5]. Thus, when leading weekly *Fundamentals of Programming* recitations for 30 students, I implemented strategies to create an **interactive and supportive learning environment**. I learned each student's name and structured recitations as collaborative workshops, where we live coded, diagrammed, and reasoned through problems together. Over time, students who were initially hesitant to speak up became active participants. As one student reflected, **"Her diagrams and explanations were very useful. I wish she had taught all of the classes."**

TEACHING PLANS

Given my prior experience, I am equipped and excited to teach **foundational computer science courses**, as well as courses on **human-computer interaction, visualization, and responsible AI**. In addition, I am interested in developing courses at the intersection of these fields, such as:

Human-Centered AI: This course explores how human-computer interaction (HCI) principles can guide the design of AI systems. To understand the impact of AI design choices, students will explore technical topics (e.g., data collection, optimization functions, model alignment) alongside societal considerations (e.g., fairness, governance, plurality). Through a semester-long group project, students will partner with local organizations to gain practical experience building AI solutions that meet real-world needs.

AI Interpretability: A course on current theories and methods for understanding AI models from a computational and human perspective. Students will investigate how models represent concepts and perform computation, as well as how people interpret and communicate model behavior. Through

seminal readings, weekly hands-on assignments, and a final project, students will analyze model reasoning and explore the implications of interpretability for AI safety and human understanding.

MENTORING EXPERIENCE

As with teaching, my goal as a mentor is to help students become skilled and responsible researchers. During my PhD, **I have mentored five undergraduate, master's, and junior PhD students** on projects spanning HCI and AI. Instead of assigning side experiments for my own papers, I guide students in leading publishable research projects. This approach gives students ownership over the entire research pipeline, and as a result, **three of my students have first-authored publications** [1, 2, 3] and **three have gone on to top PhD programs** at MIT and Harvard.

I mentored Hyemin (Helen) Bang on her MIT MEng thesis. While she had little prior research experience, with my mentorship, she led two independent projects. I initially advised her on hypothesis formulation and experiment design, and over time, handed off more responsibility for interpreting the results and defining the next steps. Her first project extended my alignment metrics to automatically reveal misalignments (e.g., spurious correlation and bias) in computer vision models, resulting in an ECCV workshop paper [1]. Her second project studies the mechanisms behind why LLMs produce misleading chart captions (in preparation for ICML 2026). She is now a PhD student at MIT.

I also use mentorship to bridge research communities. I mentored Zoe De Simone during her interdisciplinary MIT master's thesis, combining my HCI expertise with her faculty advisor's AI background. I advised Zoe on an interactive interface to inspect and edit generative image models' worldviews, and guided her through user studies that addressed sensitive issues, including how models reflect cultural identity. With my mentorship, Zoe submitted her system to an HCI conference [2] and is now an MIT PhD student. Our collaboration has continued, producing research at the intersection of AI and HCI, and I am excited to continue collaborative mentorship in my faculty role.

COMMUNITY BUILDING

As a first-generation female scientist, it was my professors who helped me feel a sense of belonging in academia. Paying this forward, I am excited to use my faculty position to build vibrant academic communities that support their members, students, and the broader public.

Mentorship is critical to student success [6], so I invest in inclusive, evidence-based practices, recently completing *MIT's Research Mentorship Certificate*. All five students I have mentored come from diverse backgrounds, and three are thriving PhD students (two of whom had not considered graduate school until I encouraged them to apply). These experiences affirm that **my mentoring approach attracts and retains diverse and valuable members of our research communities**, and I look forward to expanding my mentorship skills and reach as a professor.

I also aim to create communities that bring together researchers across disciplines and expand participation in research. To connect the visualization and AI communities, I have **organized the VISxAI workshop at IEEE VIS for the past five years**. Since these complementary fields rarely share the same conference venues, VISxAI has become a catalyst for interdisciplinary collaboration, broadening the conference's intellectual scope and connecting researchers who might not otherwise interact. By inviting interactive articles rather than traditional papers, **the workshop intentionally lowers the barrier to participation**, attracting many first-time authors, including undergraduates and engineers. As a result, VISxAI is one of the conference's longest-running workshops and a blueprint for the community design I plan to continue throughout my career.

References

Underlined authors are my mentees.

- [1] Explanation Alignment: Quantifying the Correctness of Model Reasoning At Scale. Hyemin Bang, Angie Boggust, and Arvind Satyanarayan. *Explainable Computer Vision (eXCV) Workshop at the European Conference on Computer Vision (ECCV) 2024*.
- [2] DiffusionWorldViewer: Exposing and Broadening the Worldview Reflected by Generative Text-to-Image Models. Zoe De Simone, Angie Boggust, Arvind Satyanarayan, and Ashia Wilson. *arXiv 2024*.
- [3] VisText: A Benchmark for Semantically Rich Chart Captioning. Benny J. Tang*, Angie Boggust*, and Arvind Satyanarayan. *The Annual Meeting of the Association for Computational Linguistics (ACL) 2023*. 🏆 **Outstanding Paper Award**
- [4] Active learning increases student performance in science, engineering, and mathematics. Scott Freeman et al. *Proceedings of the National Academy of Sciences (PNAS) 2014*.
- [5] Using Self Determination Theory Principles to Promote Engineering Students' Intrinsic Motivation to Learn. Kyle F. Trenshaw et al. *International Journal of Engineering Education 2016*.
- [6] Culturally Diverse Undergraduate Researchers' Academic Outcomes and Perceptions of Their Research Mentoring Relationships. Angela M. Byars-Winston et al. *International Journal of Science Education 2015*.