MICHIGAN INSTITUTE FOR DATA SCIENCE
UNIVERSITY OF MICHIGAN

Future Leaders Summit 2024

APRIL 8-10, 2024

Ann Arbor, Michigan

For more information:

midas.umich.edu/future-leaders-summit-2024

About

Data science and AI are having a significant impact on society in uncountable ways, leading to huge benefits in many cases. Yet, increasingly complex analytical pipelines working with poorly understood heterogeneous data sets can give rise to harms in many ways. Furthermore, there could be deleterious systemic effects such as the magnification of disinformation or surveillance capitalism. There has been tremendous recent interest in understanding and managing these concerns. Together with nearly forty young scholars, the Summit will explore in-depth topics in this broad area, including, but not limited to:

- Equity and fairness, particularly in automated decision making
- Explainability of analytical results
- Reproducibility and replication of scientific results
- Systemic issues, particularly those impacting marginalized populations
- Responsible AI in science and engineering. In 2022, MIDAS established a large postdoctoral training program (the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship, a Schmidt Futures Program). With this, we are dedicated to promoting responsible data science and AI for natural sciences and engineering.

2024 Program Schedule

Event Venue: Michigan League
911 N. University Ave.,
Ann Arbor MI 48109

Day #1: Monday, April 8th

8:20 AM	**Program kick-off** *(Hussey Rm., 2nd Floor)* Light morning refreshments served

Annual Ethical AI Symposium

All programming in Ballroom (2nd Floor) unless noted otherwise

8:45 AM	Opening Remarks
9:00 AM	Opportunities and challenges in participatory AI Dr. Min Kyung Lee
9:45 AM	The Hidden Governance of AI Dr. Abigail Jacobs
10:30 AM	Break
10:40 AM	Moonshot, Woodstock, Watergate, Punk Rock – Challenges to Applying AI in the Workplace Dr. Brian Martin
11:30 AM	A Conversation on AI Policy and Regulation with Bill de Blasio Moderator: Dr. Merve Hickok
12:15 PM	Lunch & Poster Session #1 *(Vandenberg Rm., 2nd Floor)*
2:00 PM	A Practical Approach to Ethical AI Dr. Michael Tjalve
3:00 PM	Poster Session #2 *(Vandenberg Rm., 2nd Floor)*
3-3:30 PM	Solar Eclipse Viewing Party *(Ingalls Mall, outside the League)*
5:00 PM	Networking Reception *(2nd Floor)*
5:30PM	Dinner for FLS attendees & mentors *(Hussey Rm., 2nd Floor)*

2024 Program Schedule

Day #2: Tuesday, April 9th

All programming in the Hussey Room (2nd Floor)

8:45 AM	Mentoring Session: Preparing for the Job Market
10:15 AM	Break
10:30 AM	Research Discussions *(Domains)*
12:00 PM	Lunch with U-M Faculty
1:00 PM	Mentoring Session: Mock Interviews
3:00 PM	**Break**
3:15 PM	Campus Walk
4:00 PM	Research Discussions *(Methodologies)*
5:30 PM	Dinner for FLS attendees with MIDAS staff and postdocs

2024 Program Schedule

Day #3: Wednesday, April 10th

All programming in the Hussey Room (2nd Floor)

8:45 AM	Mentoring Session: Networking and Increasing Impact of Research
10:00 AM	Break
10:15 AM	Research Discussions *(Challenges & Future Research Directions)*
11:15 AM	A Conversation with Jamal El-Hindi Counsel at Clifford Chance, formerly the inaugural Chief Data Officer of US Treasury and Deputy Director of US Treasury Financial Crimes Enforcement Network.
12:00 PM	Lunch with U-M Faculty
1:00 PM	Mentoring Session: Starting strong in your faculty career
2:30 PM	Program closing

Mentors

Dr. Brittany Aguilar

Science Associate, Schmidt Sciences

Brittany Aguilar is a Science Associate at Schmidt Sciences. In her role, she works to manage and support scientific programs. She is a leader on the Eric & Wendy Schmidt AI in Science Postdoctoral Fellowship team and helps to support the Science Communication Award. Beyond her pivotal role in these programs, Brittany contributes her expertise to diverse projects encompassing science communication, higher ed training, and a suite of critical skills vital for the intersection of philanthropy and cutting-edge research. Before Schmidt Sciences, Brittany was a Senior Program Manager at the New York Academy of Sciences, where she co-managed three large, international awards programs, and one postdoctoral fellowship program. She has expertise in scientific education and advocacy. She earned a PhD in behavioral neuroscience from Georgetown University and a bachelors in biology from University of California, Irvine.

Dr. Elizabeth Bondi-Kelly

Assistant Professor of Electrical Engineering and Computer Science, College of Engineering, University of Michigan

Elizabeth Bondi-Kelly is an Assistant Professor of Electrical Engineering and Computer Science at the University of Michigan. She has a PhD in Computer Science at Harvard University, where she was advised by Prof. Milind Tambe, and she was formerly a Postdoctoral Fellow at MIT through the CSAIL METEOR Fellowship. Her research interests are focused on artificial intelligence for social impact, particularly spanning the fields of multi-agent systems and data science. Her work, which has been published in venues such as AAAI, AAMAS, AIES, and IJCAI, has applications in conservation and public health, and has been deployed to support conservation efforts. She also founded and currently leads Try AI, a 501(c)(3) nonprofit committed to increasing diversity, equity, inclusion, and belonging in the field of AI through community-building educational programs, largely focused on AI for social impact.

Mentors

Dr. Bill Currie

Associate Dean, Research and Engagment, Professor of Environment and Sustainability, School for Environment and Sustainability and Professor of Environment, Program in the Environment, School for Environment and Sustainability and College of Literature, Science, and the Arts

Bill Currie studies how physical, chemical, and ecological processes work together in the functioning of ecosystems such as forests and wetlands. He studies how human impacts and management alter key ecosystem responses including nutrient retention, carbon storage, plant species interactions, and plant productivity. Dr. Currie uses computer models of ecosystems, including models in which he leads the development team, to explore ecosystem function across the spectrum from wildland to heavily human-impacted systems. He often works in collaborative groups where a model is used to provide synthesis. He is committed to the idea that researchers must work together across traditional fields to address the complex environmental and sustainability issues of the 21st century. He collaborates with field ecologists, geographers, remote sensing scientists, hydrologists, and land management professionals.

Dr. Jamal El-Hindi

Clifford Chance; Former U.S. Treasury Financial Crimes Enforcement Network (FinCEN) Deputy Director

Jamal El-Hindi is a former U.S. Treasury Financial Crimes Enforcement Network (FinCEN) Deputy Director, having previously served as the Office of Foreign Assets Control (OFAC) Associate Director for Program Policy and Implementation. Jamal was also Treasury's inaugural Chief Data Officer, with responsibility for enhancing management and use of all Treasury data. At FiCEN, Jamal led the operational, policy, and strategic planning aspects of the bureau, overseeing rulemaking, guidance and interpretation efforts, and counselling on enforcement, regulatory analysis, and industry outreach. Jamal has demonstrated experience with respect to new payment methods, including prepaid access and virtual currency. In his role at OFAC, Jamal led Treasury engagements with industry and other stakeholders regarding compliance with OFAC sanctions, rulemaking and licensing.

Mentors

Dr. Arya Farahi

Assistant Professor, Dept of Statistics and Data Science, University of Texas, Austin

Arya Farahi joined The University of Texas at Austin in 2021 as an assistant professor. Previously he was a Data Science Fellow at the Michigan Institute for Data Science at the University of Michigan. His research contributes to the fields of astroinformatics and urban informatics; and is focused on understanding and mitigating the unexpected and not-well understood consequences of AI models, including algorithmic bias and uncertainty quantification, in real-world settings. He was a Schmidt Science Fellow finalist, recipient of the best student paper award in KDD'18, an awardee of the Michigan Institute for Computational Discovery and Engineering (MICDE) fellowship, and recipient of >$400k grant funding. He is an active member of several international projects and collaborations, including the Dark Energy Survey (DES), the COsmostatistics INitiative (COIN), and XMM-XXL Consortium, among others. He is also a Statistics Without Borders volunteer.

Dr. Kent Foster

Director, Innovation + Society, Microsoft

Kent Foster is a seasoned professional with a rich history in the software industry, currently serving as the Microsoft Director of Innovation & Society at Microsoft. With a focus on responsible AI research and advancing translational technology solutions, Kent has been instrumental in fostering partnerships with key universities to address systemic inequities and benefit society. A graduate of the University of Michigan, Kent holds an MPP from the Ford School of Public Policy and holds an undergraduate degree in Far Eastern Languages and Literature (Mandarin Chinese). Kent has a commitment to innovation, education, and societal impact working through university research collaboration including the upcoming MIDAS call for proposals with a major focus is "responsible AI and AI policy", in collaboration with Microsoft.

Mentors

Dr. R. Stuart Geiger

Assistant Professor, Dept of Communication and the Halıcıoğlu Data Science Institute, Affiliate Faculty, Institute for Practical Ethics, Computer Science & Engineering, and Computational Social Science, University of California, San Diego

Geiger studies the relationships between science, technology, and society — not only how science and technology have substantial impacts on society, but also how they are social institutions in themselves. Much of his work focuses on machine learning, particularly in how user generated platforms like Twitter and Wikipedia are moderated. He has examined how values and biases are embedded in these technologies and how communities make decisions about how to use or not use them. He is a methodological pluralist and specializes in mixed methods, such as combining the rich and thick descriptions of cultural context that come from qualitative methods with large-scale quantitative and computational methods from Natural Language Processing. Geiger also studies the development of data science itself as an academic and professional field. He earned his Ph.D in 2015 at the UC Berkeley School of Information and the Berkeley Center for New Media, then was the staff ethnographer at the UC Berkeley Institute for Data Science. He joined UCSD in 2020, jointly appointed as faculty in the Department of Communication.

Dr. H.V. Jagadish

Director, Michigan Institute for Data Science, University of Michigan

H. V. Jagadish is the Director of the Michigan Institute for Data Science, Edgar F. Codd Distinguished University Professor, and Bernard A. Galler Collegiate Professor of Electrical Engineering and Computer Science at the University of Michigan in Ann Arbor. Before his professorship, he was Head of the Database Research Department at AT&T Labs. Dr. Jagadish's research focuses on two themes: the usability of database systems, query models, and analytics processes to inform decision-makers, especially with big and heterogeneous data that go through many transformations; data equity systems that center around issues of representation, diversity, fairness, transparency, and validity. Dr. Jagadish is an elected ACM Fellow and AAAS Fellow. His many academic scholarship roles include establishing the ACM SIGMOD Digital Review and founding the Proceedings of the Very Large Database Endowment (PVLDB), serving on the boards of the Computing Research Association (CRA) and the Very Large Database Endowment.

Mentors

Dr. Min Kyung Lee

Assistant Professor, School of Information, University of Texas, Austin

Min Kyung Lee is an assistant professor in the School of Information at the University of Texas at Austin. Dr. Lee has conducted some of the first studies that empirically examine the social implications of algorithms' emerging roles in management and governance in society, looking at the impacts of algorithmic management on workers as well as public perceptions of algorithmic fairness. She has proposed a participatory framework that empowers community members to design matching algorithms for their own communities. Her current research on human-centered AI is inspired by and complements her previous work on social robots for long-term interaction, seamless human-robot handovers, and telepresence robots. Dr. Lee is a Siebel Scholar and has received the Allen Newell Award for Research Excellence, research grants from NSF and Uptake, and five best paper awards or honorable mentions in venues such as CHI, CSCW, DIS and HRI. She is an associate editor of the ACM Transactions on Human-Robot Interaction. Her work has been featured in media outlets such as the New York Times, New Scientist, Washington Post, MIT Technology Review and CBS. She received a PhD in Human-Computer Interaction and an MDes in Interaction Design from Carnegie Mellon University.

Dr. Michael Tjalve

Chief AI Architect, Tech for Social Impact at Microsoft Philanthropies; Assistant Professor, Linguistics, University of Washington

Michael Tjalve is Chief AI Architect on the Tech for Social Impact team in Microsoft Philanthropies where he works with nonprofits and humanitarian organizations around the world on building technology solutions that help them amplify their impact and address some of today’s biggest societal challenges. He's Assistant Professor at University of Washington where he teaches AI in the humanitarian sector and ethical innovation and he serves as tech advisor to Spreeha Foundation and World Humanitarian Forum.

Mentors

Dr. Elizabeth Yakel

C Olivia Frost Collegiate Professor of Information, School of Information; Faculty Associate, Inter-University Consortium for Political and Social Research, Institute for Social Research, University of Michigan

Elizabeth Yakel, PhD, is C. Olivia Frost Collegiate Professor of Information at the University of Michigan School of Information. Her research interests include digital archives and curation specifically data reuse; teaching with primary sources; and the development of standardized metrics to enhance repository processes and the user experience. Professor Yakel is currently a co-PI on two research projects. The first, “Measuring and Improving the Efficacy of Curation Activities in Data Archives,” funded by the Institute for Museum and Library Services, investigates how curatorial actions impact the reuse of digital collection. The second, "Developing Evidence-based Data Sharing and Archiving Policies," funded by the National Science Foundation, answers three key questions: How can data repositories best allocate their limited resources for different aspects of data archiving and processing? What is the most effective way of making data usable by the broadest audience? What data sharing policies most effectively achieve stakeholders' transparency and innovation goals?

Maryam Berijanian

Michigan State University

Holding dual B.Sc. degrees in Mechanical and Aerospace Engineering from Sharif University of Technology, Iran, and an M.Sc. in Robotics and Mechatronics from the University of Twente, Netherlands, Maryam has a strong foundation in engineering and technology. Furthering her research in Germany in data analysis at RWTH Aachen University's Institute for Rail Vehicles and Transport Systems and in computer vision and deep learning at the Institute of Imaging & Computer Vision, she is now a Ph.D. student studying in the fields of Generative AI, Computer Vision, and Natural Language Processing. Achieving a perfect cumulative GPA of 4.0, she has been recognized with prestigious awards and memberships, including a research grant in Germany, the Engineering Distinguished Scholar fellowship from Michigan State University, two scholarships from University of Twente, and memberships in Phi Kappa Phi and Tau Beta Pi honor societies.

Ethical AI in Digital Pathology: Privacy-Preserving Image Synthesis with Unsupervised Stain Translation

In the rapidly evolving field of digital pathology, the ethical implications of AI technologies are of high concern. This study introduces an innovative unsupervised many-to-many stain translation framework for histopathology images, leveraging an enhanced GAN model with an edge detector to preserve tissue structure while generating synthetic images. Our method addresses two critical ethical challenges in AI: privacy and the reliance on low-cost labor for image annotation. First, by utilizing artificially generated images, our approach circumvents the privacy issues inherent in using real patient data, thereby safeguarding individual confidentiality—an essential consideration in medical research. Second, the reliance on extensive annotated datasets for deep learning applications often implicates ethical concerns regarding the exploitation of low-cost labor for manual image annotation. Our framework mitigates this issue by generating high-quality, realistic synthetic images, reducing the dependency on manually annotated datasets. Empirical results underscore the effectiveness of our approach; incorporating generated images into the training datasets of breast cancer classifiers resulted in performance improvements, demonstrating the technical feasibility and ethical advantages of our method. This research not only contributes to the advancement of digital pathology through AI but also emphasizes the importance of ethical considerations in the development and application of AI technologies.

Isabela Bertolini Coelho

University of Maryland

Isabela Bertolini Coelho is currently a Ph.D. student in Survey and Data Science at the Joint Program in Survey Methodology at the University of Maryland. She holds a Master of Science in Statistics from the University of Sao Paulo and a bachelor's degree in Statistics from the University of Campinas. Additionally, she attended the Sampling Program for Survey Statisticians at the University of Michigan and the International Program in Survey and Data Science at the University of Mannheim. Her main areas of interest include sampling techniques, complex sample data analysis, small area estimation methods, combining samples, and differential privacy.

A Comparative Analysis between AI and Human Coding in Survey Research

Privacy is central to discussions surrounding data protection and ethical considerations in both survey methodology and AI. Understanding stakeholders' attitudes, perceptions, and participation levels toward privacy is crucial to identifying the barriers to adopting formal privacy models in sample survey data, especially for official statistics. Large language models (LLMs) have emerged as powerful tools in various domains, including survey research. In this study, we present a comparative analysis between LLM-generated codifications and human-coded responses to open-ended questions regarding privacy. Based on results from a qualitative study conducted with experts on data privacy, our investigation delves into the similarities and disparities between codifications generated by LLMs and those crafted by human coders. Additionally, we examine the extent to which LLMs capture the contextual intricacies of privacy discussions, especially regarding the differentiation between what privacy means in the context of their work and as they experienced it in their personal lives. Furthermore, this study sheds light on the efficacy of LLMs in survey research, particularly in codifying complex concepts such as privacy. It contributes to ongoing discussions surrounding the role of AI in survey methodology.

Brooks A. Butler

Purdue University

Brooks A. Butler is a Ph.D. candidate in the Elmore Family School of Electrical and Computer Engineering at Purdue University studying networked dynamic systems and safety-critical control. He received his M.S. in Computer Science and his B.S. in Applied Physics from Brigham Young University in 2020 and 2019, respectively. Some of his primary research interests include automatic control of networked systems, safety-critical control, robotics, and artificial intelligence.

Collaborative Safety for Multi-agent Systems

The safe coordination of multi-agent systems presents a complex and dynamic research frontier, encompassing various objectives such as ensuring group coherence while navigating obstacles and avoiding collisions between agents. Expanding upon our prior work in distributed collaborative control for networked dynamic systems, we introduce an algorithm tailored for the formation control of multi-agent systems, considering individual agent dynamics, induced formation dynamics, and local neighborhood information within a predefined sensing radius for each agent. Our approach prioritizes individual agent safety through iterative communication rounds among neighbors, enforcing safety conditions derived from high-order control barrier functions (CBFs) to mitigate potentially hazardous control actions within the cooperative framework. Emphasizing explainable AI principles, our method provides transparent insights into decision-making processes via model-based methods and intentional design of individual agent safety constraints, enhancing the interpretability and trustworthiness of multi-agent system behavior.

Lucius Bynum

New York University

I am a PhD Candidate at the NYU Center for Data Science advised by Julia Stoyanovich as part of the Center for Responsible AI and working closely with Joshua Loftus at the London School of Economics. In my research, I use causal inference and statistics to better understand bias and inequality in AI systems, machine learning, and algorithmic decision making. This work includes developing tools for inequality-aware decision making and more wholistic algorithmic fairness, leveraging counterfactual reasoning to improve model explainability and reduce pre-existing disparities, and reimagining how we use causal modeling formalisms to reason about social categories like race and gender. I am also passionate about teaching via public outreach and making educational material in these areas. My research is generously supported by the Microsoft Research PhD Fellowship.

A New Paradigm for Counterfactual Reasoning in Fairness and Recourse

Counterfactuals and counterfactual reasoning underpin numerous techniques for auditing and understanding artificial intelligence (AI) systems. The traditional paradigm for counterfactual reasoning in this literature is the interventional counterfactual, where hypothetical interventions are imagined and simulated. For this reason, the starting point for causal reasoning about legal protections and demographic data in AI is an imagined intervention on a legally-protected characteristic, such as ethnicity, race, gender, disability, age, etc. We ask, for example, what would have happened had your race been different? An inherent limitation of this paradigm is that some demographic interventions — like interventions on race — may not translate into the formalisms of interventional counterfactuals. In this work, we explore a new paradigm based instead on the backtracking counterfactual, where rather than imagine hypothetical interventions on legally-protected characteristics, we imagine alternate initial conditions while holding these characteristics fixed. We ask instead, what would explain a counterfactual outcome for you as you actually are or could be? This alternate framework allows us to address many of the same social concerns, but to do so while asking fundamentally different questions that do not rely on demographic interventions.

César Claros

University of Delaware

César Claros' academic and research journey reflects a deep commitment to advancing the field of machine learning with applications in biomedicine. He commenced his academic journey with a B.S. in Electronics Engineering from Universidad Mayor de San Andrés (2014) and an M.S. in Electrical and Computer Engineering from the University of Delaware (2020). He is currently advancing towards a Ph.D. in Electrical Engineering at the same university since February 2022. His research is pivotal in leveraging deep learning for predicting brain age from brain tissue mechanical properties and enhancing interpretability through tensor factorization. He also explores clinically viable assessments for predicting athletes' risk of post-concussion musculoskeletal injuries and employs machine learning to identify patterns in neuroimaging and pathology biomarkers for dementia. César's work stands at the intersection of machine learning and signal processing, aiming to drive forward the understanding and application of these technologies in the biomedical field.

Interpreting Age Predictions from Brain Maps via Deep Neural Activations and Tensor Decomposition

This work presents a novel method for interpreting 3D convolutional neural networks (CNNs) that estimate clinically relevant attributes from 3D brain maps, aiming to address the challenge of interpretability in deep learning within healthcare. Unlike common image classification interpretability methods, such as GradCAM, which rely on per-instance explanations due to spatial variation, this approach leverages the consistent spatial registration of brain maps to compute dataset-level explanations. By organizing the network's internal activations into a tensor and applying constrained tensor decomposition, the method identifies key spatial patterns and brain regions focused on during prediction. The technique uses reconstruction error to determine the tensor decomposition rank and employs linear models to link activation decompositions to target attributes. Applied to networks estimating chronological age from brain volume and stiffness maps obtained via MRI and T1-weighted MRE scans, the decomposition highlights brain areas known to change with age. This approach offers a means to interpret CNNs in brain mapping and insights into age-related brain structural changes, enhancing the understanding and trustworthiness of deep learning models in healthcare.

Anja Conev

Rice University

I am a PhD candidate at the Computer Science Department of Rice University. My research interests include applied responsible application of artificial intelligence and machine learning in the context of structural computational biology, drug design and immunoinformatics. I am passionate about interdisciplinary research, mentoring students and developing open source software for the research community.

HLAEquity: Examining Biases in Pan-Allele pHLA Binding Predictors

Peptide-HLA (pHLA) binding prediction is essential in screening peptide candidates for personalized peptide vaccines. Machine learning (ML) pHLA binding prediction tools are trained on vast amounts of data and are effective in screening peptide candidates. Most ML models report the ability to generalize to HLA alleles unseen during training ("pan-allele" models). However, the use of datasets with imbalanced allele content raises concerns about biased model performance. First, we examine the data bias of two ML-based pan-allele pHLA binding predictors. We find that the pHLA datasets overrepresent alleles from geographic populations of high-income countries. Second, we show that the identified data bias is perpetuated within ML models, leading to algorithmic bias and subpar performance for alleles expressed in low-income geographic populations. We draw attention to the potential therapeutic consequences of this bias, and we challenge the use of the term “pan-allele” to describe models trained with currently available public datasets.

Diamond Joelle Cunningham

Tulane University

Diamond Joelle Cunningham, MPH (she/her), is a passionate racial health equity researcher. Diamond earned her master's in public health with a concentration in Urban Public Health where she delved into the intricate dynamics of urban environments, exploring topics such as population health disparities, healthcare access in metropolitan areas, and strategies for promoting health equity in densely populated communities. Currently pursuing her PhD at the Tulane School of Public Health and Tropical Medicine, her research focuses on promoting health equity for Black birthing individuals, particularly within the realm of reproductive healthcare. In addition to her research endeavors, Diamond serves as the Student Chairperson for the Bill Anderson Fund, a fellowship designed for Black and Brown doctoral students across various academic disciplines who specialize in hazard and disaster studies. Diamond's approach to her work is multifaceted, integrating personal experience, academic rigor, and practical application to further her commitment to advancing health equity.

Racial Discrimination and Medical Appointment Adherence:
The Black Women's Experiences Living with Lupus (BeWELL) Study

Black/African American women disproportionately suffer from systemic lupus erythematosus (SLE), with higher prevalence, severity, and poorer outcomes compared to White counterparts. Appointment non-adherence contributes to racial disparities in health outcomes, with factors such as racial discrimination potentially leading to missed appointments among Black/African Americans. This study sought to examine whether racial discrimination in medical settings is Associated with missed appointments among Black/African American women living with SLE. Data from the BeWELL Study (2015-2017) involved 438 Black women diagnosed with SLE in Atlanta. Appointment adherence was gauged by asking about missed appointments with their lupus doctor. Participants reported experiences of racial discrimination in medical care, with multivariable logistic regression used to analyze missed appointments in relation to discrimination. Controlling for SLE duration, disease severity (organ damage and disease activity), and other demographic, socioeconomic, and health-related characteristics, racial discrimination was significantly associated with missed appointments (Odds Ratio: 1.33, 95% Confidence Interval: 1.03-1.73). Results from this study suggest that racial discrimination in medical care may result in missed medical appointments among Black women living with SLE. Antiracist interventions at multiple points of engagement within medical systems, from scheduling to the clinical encounter, may enhance appointment adherence among Black/African American women living with SLE.

Matthew R. DeVerna

Indiana University Bloomington

Matthew R. DeVerna is an Informatics PhD candidate at Indiana University, specializing in computational social science within the complex networks and systems track. He collaborates closely with Dr. Filippo Menczer's research group and the Observatory on Social Media. DeVerna's research focuses on digital platform dynamics, the spread of information online, and strategies to curb misinformation. His work includes identifying influential spreaders of low-credibility content, assessing misinformation's impact on vaccine hesitancy, and exploring misinformation interventions. DeVerna's contributions have been recognized by major media outlets, and he aims to further investigate the challenges and opportunities presented by generative AI in digital spaces.

Fact-checking information generated by a large language model can decrease news discernment

Fact checking can be an effective strategy against misinformation, but its implementation at scale is impeded by the overwhelming volume of information online. Recent artificial intelligence (AI) language models have shown impressive ability in fact-checking tasks, but how humans interact with fact-checking information provided by these models is unclear. Here, we investigate the impact of fact-checking information generated by a popular large language model (LLM) on belief in, and sharing intent of, political news in a preregistered randomized control experiment. Although the LLM performs reasonably well in debunking false headlines, we find that it does not significantly affect participants' ability to discern headline accuracy or share accurate news. Subsequent analysis reveals that the AI fact-checker is harmful in specific cases: it decreases beliefs in true headlines that it mislabels as false and increases beliefs in false headlines that it is unsure about. On the positive side, the AI fact-checking information increases sharing intents for correctly labeled true headlines. When participants are given the option to view LLM fact checks and choose to do so, they are significantly more likely to share both true and false news but only more likely to believe false news. Our findings highlight an important source of potential harm stemming from AI applications and underscore the critical need for policies to prevent or mitigate such unintended consequences.

Majid Farhadloo

University of Minnesota, Twin Cities

Majid Farhadloo is a PhD candidate in Computer Science at the University of Minnesota – Twin Cities. His research interests lie in developing explainable and efficient methods in knowledge discovery and responsible spatial data science with applications in healthcare and oncology. His thesis explores the development of spatially-lucid artificial intelligence classifiers by leveraging recent advances in geospatial artificial intelligence (GeoAI) and spatial omics data. These classifiers aim to incorporate spatial arrangements among data points in learning samples to enhance decision-making processes. Preliminary results of his work have been presented at prestigious conferences such as SIGKDD and SIAM DM. Majid earned his Master's in Computer Science from the University of Minnesota – Twin Cities and his Bachelor's degree in Computer Science from California State University – Fresno.

Spatially-Lucid Classifiers for Oncology and other GeoAI Applications

High-risk applications of Geo-AI must show that their models are safe, transparent, and spatially lucid (i.e., explainable using spatial concepts) to end users. The goal of spatially lucid artificial intelligence (AI) classification approach is to build a classifier to distinguish two classes (e.g., responder, non-responder) based on the their spatial arrangements (e.g., spatial interactions between different point categories) given multi-category point data from two classes. This problem is societally important for many applications, such as generating clinical hypotheses for designing new immune therapies for cancer treatment. This problem is challenging due to an exponential number of category subsets which may vary in the strength of their spatial interactions. Most prior efforts on using human selected spatial association measures may not be sufficient for capturing the relevant spatial interactions (e.g., surrounded by) which may be of biological significance. In addition, the related deep neural networks are limited to category pairs and do not explore larger subsets of point categories. To overcome these limitations, we propose a Spatial-interaction Aware Multi-Category deep neural Network (SAMCNet) architecture and contribute novel local reference frame characterization and point pair prioritization layers for spatially explainable classification. Experimental results on multiple cancer datasets (e.g., MxIF) show that the proposed architecture provides higher prediction accuracy over baseline methods. A real-world case study demonstrates that the proposed work discovers patterns that are missed by the existing methods and has the potential to inspire new scientific discovery.

Emily Fletcher

Purdue University

Emily Fletcher is a digital archaeologist interested in software development, data management, decolonizing archaeology, and technological innovation. She attended Kalamazoo College for her undergraduate studies, where people are often surprised to learn she double majored in computer science and history. She worked as a software developer for a year before beginning graduate studies at Purdue University in 2019. In Emily's research, she writes software to bring new life to archaeological legacy data (records from previous research). She specifically focuses on the Gulkana Site, an important but understudied Alaska Native heritage site where people created a variety of copper tools roughly a thousand years ago. By using AI and collaboration, she hopes that her software can make data about this site easier for archaeologists and descendants to interact with. Outside of her dissertation, Emily also participates in broader efforts to explore ethical data management practices for cultural heritage data.

Processing Archaeological Field Notebooks

Although archaeological field notebooks are created as a resource for future archaeologists to reference in their research, the labor required to digitize handwritten notes presents a barrier to their incorporation in state-of-the-art computational analyses. In this research, I explore if image preprocessing can improve the accuracy of text extracted from handwritten field notebooks by Handwritten Text Recognition. I apply image preprocessing to scans of handwritten field notebooks from the 1970s excavations of the Gulkana Site, a pre-contact Northern Dene site in Alaska's Copper River Basin. These documents contain important data regarding native copper innovation that occurred at the Gulkana Site, but their current state has prevented analysis of that data.

Neil S. Gaikwad

Massachusetts Institute of Technology

Neil Gaikwad, a PhD candidate and a fellow at MIT's Dalai Lama Center for Ethics and Transformative Values. He will soon join the faculty in Data Science and CS at the University of North Carolina, Chapel Hill. Neil's research in Responsible AI and Algorithmic Alignment brings a computational lens to scientifically informing societal decisions concerning sustainable development. This research, published in prominent AI and HCI conferences, has been showcased in the United Nations, The New York Times, Bloomberg, and WIRED. Neil has been recognized with the Facebook Research Fellowship, Human Rights & Technology Fellowship, MIT Graduate Teaching Award, and the Karl Taylor Compton Prize, MIT's highest student award. Additionally, he was honored as a Rising Star by Stanford University and the University of Chicago. He has mentored over 20 students, some of whom have significantly influenced the scholarship and discourse on AI fairness. Neil is an alum of Carnegie Mellon University's School of Computer Science.

Algorithmic Alignment: Value-Sensitive Designs of Human-AI Collaboration for Global Social Inclusion

The rise of AI has brought about a significant transformation in how algorithms engage with societal values, reshaping computational systems and human societies alike. However, despite its widespread adoption, AI innovation often overlooks individuals confronting poverty and heightened public health risks, especially in the face of climate change. To tackle these sustainability challenges, I introduce Public Interest Computing research, which centers on Responsible AI and Algorithmic Alignment, aiming to redefine Human-AI collaboration rooted in social norms. Illustrating through both theoretical grounding and practical examples, I present methods for integrating ethics and values into human-AI systems for societal decision-making. Firstly, by employing new social and democratic learning mechanisms to facilitate ethical decision-making, machine learning preferences gathered from 1.3 million individuals. Secondly, by developing value-sensitive design mechanisms that enhance the agency of historically marginalized communities in algorithmic decision-making for climate change adaptation policy, including addressing pressing issues like farmer suicides affecting 300,000 individuals. Thirdly, by redesigning socially and ethically responsible AI data market systems with incentive-compatible interactions to address equity concerns in data ecosystems. Public Interest Computing prioritizes ethics in human-AI collaboration from the inception rather than as an afterthought, offering a pathway to design technologies that are not only computationally efficient but also fair, value-sensitive, and accessible for everyone around the world.

Katherine R. Garcia

Rice University

Katherine (Katie) Garcia is a Ph.D. candidate in the Department of Psychological Sciences at Rice University in the Human-Computer Interaction & Human Factors (HCIHF) research interest group. She is currently in her fourth year of graduate studies (second year at Rice) in the Human-Automation Collaboration (HAC) Lab led by Dr. Jing Chen. Katie received her B.A. in Psychology and Cognitive Sciences with a minor in Engineering Design and Neuroscience from Rice in 2020, and her M.S. in Psychology with a certificate in Modeling and Simulation Engineering from Old Dominion University in 2022. Katie holds the positions of President of Rice Human Factors and Ergonomics Society (HFES), Social Chair of Psychological Sciences Graduate Student Association (PsycGSA), and Lab Manager of the HAC Lab. Her research interests include AI capabilities, flood warnings, decision-making, transportation, and cybersecurity, to name a few.

How Do People Understand AI Capabilities in Autonomous Vehicles?

The success of autonomous vehicles (AV) depends on artificial intelligence (AI). AI is responsible for sensing the driving environment, and planning, navigating, and executing a path for the vehicle. However, human involvement is crucial to ensure AV safety, especially when AI fails. This study used a think-aloud methodology to study how drivers perceive AI capabilities in AVs when identifying different road-sign images. Participants were tasked with rating how both themselves and AI classify six unique road-sign images with four manipulation types (original/no manipulation, projected gradient descent cyberattack, physical cyberattack, and scrambled manipulation). In order to understand their reasoning, half of the participants were prompted to speak their thoughts during the study, while the other half were not required. The results showed that participants accurately perceived the AI to correctly to classify the original images and not correctly classify the scrambled ones, as predicted. However, they overestimated the AI's capabilities when handling cyberattacks, even when trying to discern the differences from the originals. Participants may perceive the AI to have similar capabilities to their own. These findings suggest that drivers may not appropriately trust or understand AI in completing critical tasks, displaying the need for more explainable AI in AVs.

Ryan Gifford

The Ohio State University

I am a multi-disciplinary, PhD candidate with a passion for Data-Science. In my research, I develop AI solutions to complex, real-world problems. Being a Machine Learning researcher in a Cognitive Systems lab has given me a unique perspective on the development of AI. On the Machine Learning side, real world data is messy, incomplete and one solution never fits all when it comes to modeling. Embedding domain knowledge into the model increases performance and aligns the model to how clinicians go about decision making in the real world. On the cognitive systems engineering side, it's important to consider not just the training of a machine learning algorithm, but its role in the overarching system it will be a part of. Model explainability and putting predictions in the context of the underlying data become critical as we think about not just machine performance, but the performance of the human-machine team.

CNN Trees for Explainable Time-Series Classification

In this research we propose the CNN Tree algorithm, an intrinsically explainable model for time-series classification. The CNN Tree leverages the explainable structure of a Decision Tree and the power of Deep Learning to extract discriminative features from raw data. Recent techniques for explainable time-series classification rely on post-hoc explanations, which are not faithful to the true decision processes of the model they are trying to explain. As an alternative, the CNN Tree is explainable by design and shows hierarchical decision processes using both important time ranges and variables. We tested the CNN Tree with one private and nine open-source datasets; the CNN Tree has better or equivalent accuracy as state-of-the-art explainable AI models while providing faithful explanations.

Bhanu Teja Gullapalli

University of California, San Diego

I'm a PhD candidate at the University of California, San Diego, working on utilizing multimodal physiological data from wearable devices, along with individuals' demographic characteristics (age, comorbidities, etc.), to develop digital biomarkers in the field of substance use addiction. Primarily, I work with mobile sensor data collected in both hospital settings and the real world to predict various aspects of the addiction cycle for different drugs. Secondly, I aim to bridge the gap between the sensed biomarkers and intervenable actions to effectively help individuals break the cycle of addiction. In terms of social life, I enjoy listening to people and stories of any form. In my free time, I like to map stars, draw, and write short stories.

Harnessing Digital Biomarkers of Substance Abuse and Addiction with Large scale Mobile Sensor Data

Mobile sensor devices equipped to monitor electrophysiological signals provide information about various health metrics. However, there exists a significant gap in their applicability to substance use, despite well-documented medical research on the cycle of addiction and changes in mental and physical states. My research focused on building biomarkers to monitor addiction states in opioids and cocaine. Initially, I demonstrate that monitoring breathing and ECG signals of a cocaine-dependent person during a drug binge session provides information on states of drug craving and euphoria. Subsequently, I illustrate how the intrinsic relationship between these states can be leveraged by models to enhance predictions. Similarly, I utilize wearable signals from medical-grade devices to monitor opioid administration. I demonstrate that incorporating domain knowledge, particularly the pharmacokinetics of the drugs, into purely data-driven models can enhance the reliability of opioid monitoring. I observe that the performance of these models is highly dependent on the population group, based on their dependence and drug usage patterns. Consequently, I develop an opioid screening model to differentiate opioid misusers from prescription users using cognitive and psychophysiological data. The findings from my research represent an initial step towards building digital biomarkers for better understanding and treating substance use disorders.

Yifei Huang

University of Illinois at Chicago

Yifei is a fifth-year Ph.D. candidate in Mathematics at University of Illinois Chicago, specializing in statistics. Her research focuses on optimal design theories in experimental design and how they can be applied to healthcare, engineering, and big data problems. She is also interested in clinical trial design and biostatistics. She is keen on exploring how data science and artificial intelligence (AI) can enhance healthcare outcomes and quality.

ForLion: A New Algorithm for D-optimal Designs with Mixed Factors Experiments

In this project, we address the problem of designing experiments with discrete and continuous (mixed) factors under general parametric statistical models. We propose the ForLion algorithm to search for optimal designs under the D-criterion. Simulation results show that the ForLion algorithm will reduce the number of distinct experimental settings while keeping the highest possible efficiency.

Zach Jacokes

University of Virginia

I am a PhD candidate at the University of Virginia studying data science with a concentration in psychology and neuroscience. My work involves examining the neurological and behavioral features of autistic individuals with particular interest in the differences between autistic males and females. I am motivated by a strong desire to collect and analyze data sustainably, ethically, and without subjecting the results to undue bias.

Collider Bias in Autism Research

Autism Spectrum Disorder (ASD) spans a wide array of phenotypic expressions that make it a difficult condition to study. Other factors complicating ASD research include a sex-wise diagnostic disparity (boys are almost four times more likely to receive an ASD diagnosis than girls), cultural biases around ASD traits, and dataset imbalances these issues can cause. This study examines the extent of the selection bias present in an in-progress ASD data collection effort and the issues with drawing generalizable conclusions from this dataset. In particular, this dataset is subject to collider bias, whereby the population of interest is artificially sampled in a way that can affect both the exposures (independent variables) and the outcomes (dependent variables). When the exposures include such variables as neuroanatomical feature size and neuronal interconnectivity between brain regions and outcomes include performance on behavioral surveys, there exists several key factors along the causal pathway between these that clearly impact their association. This study examines how artificially selecting autistic participants with low needs (measured by autism severity score) can act as a collider between exposures and outcomes.

Lavender Jiang

New York University

Lavender Jiang is a third year Data Science PhD student at New York University, co-advised by Eric Oermann and Kyunghyun Cho. She work on natural language processing for clinical notes and is interested in representation learning. She is honored to receive medical fellowship from NYU Langone Health and AIML PhD fellowship from Apple.

Language Model Can Guess Your Identities from De-identified Clinical Notes

Although open data accelerates research, machine learning for healthcare has a limited open data due to concerns about patient privacy. Health Insurance Portability and Accountability Act of 1996 (HIPAA) was created to improve data portability and it allows disclosing "de-identified health information" via Safe Harbour, which requires removing 18 types of identifiers and ensuring the individuals cannot be re-identified. A conventional approach is to detect any tokens that is deemed to be relevant to HIPAA protected identifiers and remove or replace those tokens appropriately. Since it is time-consuming to do so manually, people often view the detection part as the problem of named entity recognition (NER) and remove the detected entities appropriately. However, annotators could miss implicit contextual identifiers, giving rise to the possibility that a de-identifier achieves perfect precision and recall, yet still produce re-identifiable notes. We formalize the de-identification problem using PGM, and show that it is impossible to achieve perfect de-identification without losing all utility. Empirically, we de-identified clinical notes using NER-based de-identifiers, and finetuned a public BERT model to predict annotated demographic attributes from the de-identified notes. We show that it can recover gender, borough, year, month, income and insurance with above random chance with as few as 1000 labelled examples. These predicted attributes can be further used for re-identifying patients. Using the fully finetuned predictions, the probability of being uniquely identified is around 3 in a thousand. Using the 1000-example-finetuned predictions, the probability of being uniquely identified is around 380 in a million.

Wenxin Jiang

Purdue University

My research interest is focused on SE4AI, and Responsible AI. I currently work on advancing AI safety/security and utility, more specifically on improving reusability, trustworthiness, and security of pre-trained deep learning models (PTMs) from model registries/hubs, such as Hugging Face.

Enhancing Trustworthiness and Reusability in Pre-trained Deep Learning Model Ecosystems

Deep neural networks are being adopted as components in software systems. Creating and specializing deep neural networks from scratch has grown increasingly difficult as state-of-the-art architectures grow more complex. Following the path of traditional software engineering, deep learning engineers have begun to reuse pre-trained models (PTMs) and fine-tune them for downstream tasks and environments. However, unlike in traditional software, where reuse practices and challenges have been extensively studied, the knowledge foundation for PTM ecosystems remains underdeveloped. My research addresses this gap through a series of defect studies, case studies, and interviews, aiming to unearth detailed insights into the challenges and practices in PTM ecosystems. Utilizing mining software repository techniques, I've extracted, analyzed, and interpreted the rich data within PTM packages. My work first adopts the methodologies from traditional software engineering to understand the challenges and practices of deep learning software. I have also published two open-source datasets of PTM packages, aiming to support further research on this problem domain. My work focuses on enhancing the trustworthiness and reusability of PTMs. This involves improving transparency through comprehensive metadata extraction, identifying potential anomalies within the ecosystem, and developing optimized model selection strategies to support reuse.

Đorđe Klisura

University of Texas at San Antonio

Đorđe Klisura is a Graduate Research Assistant and a second-year Ph.D. student in Information Technology, with a focus on cybersecurity, at the University of Texas at San Antonio. He received his B.Sc. degree in Computer Science and M.Sc. in Data Science from the University of Primorska, Slovenia, in 2020 and 2022, respectively. He completed his Master's thesis at the University of Turku, in Finland, during the fall semester of 2022. His areas of research include deep learning and large language model approaches for enhancing the security of computer systems and identification and characterization of cyber attacks.

Unmasking Database Vulnerabilities: Zero-Knowledge Schema Inference Attacks in Text-to-SQL Systems

Relational databases are integral to modern information systems, serving as the foundation for storing, querying, and managing data efficiently and effectively. Advancements in large language modeling have led to the emergence of text-to-SQL technologies, significantly enhancing the querying and extracting of information from these databases, while also raising concerns about privacy and security. Our research explores extracting the database schema elements underlying a text-to-SQL model. It is noteworthy that knowledge of the schema can make attacks such as SQL injection easier. To this end, we have developed a novel zero-knowledge framework designed to probe various database schema elements without access to the schema itself. The text-to-SQL models process specially crafted questions to produce an output that we use to uncover the structure of the database schema. We apply it to specialized text-to-SQL models fine-tuned on text-SQL pairs and general-purpose language models (e.g., GPT3.5). Our current results show an average recall of 0.83 and a precision of 0.79 for fine-tuned models in uncovering the schema. This research embeds ethical and responsible AI use considerations, recognizing the importance of transparency in AI-driven systems. This work precedes future experiments, where we will explore regenerating training data used by fine-tuned text-to-SQL systems.

Olivia Krebs

University of Wisconsin-Madison

Olivia Krebs is a Biomedical Engineering Ph.D. candidate at Case Western Reserve University and an NSF graduate research fellow. She is a member of the Integrated Diagnostics and Analytics (IDiA) Laboratory led by Dr. Pallavi Tiwari, having joined in November 2021. Her research focuses on developing computational and machine learning techniques to extract and model quantitative features from medical imaging for the characterization of cancer, with an emphasis on glioblastoma and computational athology. In developing these tools, she aims to enable improved assessment of risks associated with clinical outcomes and to translate the technologies to support clinicians in accurate disease assessment.

Sex-specific pathological markers related to the spatial organization of inflammatory cells predict overall survival in glioblastoma

Overall survival (OS) in glioblastoma (GB) patients has been observed to depend on patient sex and, in part, immunological differences between males and females. This study investigated the relationship between the tumor immune microenvironment and OS in GB. Sex-specific survival models were developed utilizing spatial organization features of inflammatory cells extracted from digitized images of hematoxylin and eosin-stained resected GB tumor tissue. The inflammatory cell-based measurements were used to construct three survival risk-stratification models for male, female, and combined (male + female) cohorts. Patient-specific risk scores derived from these survival models were assessed using Kaplan-Meier estimates. The risk groups stratified by the sex-specific survival models were analyzed for differential expression of relevant cancer biology and treatment response pathways. Our findings indicate organizational histological features of inflammatory cells when trained separately for male and female GB patients, may be independently prognostic of OS. These findings suggest the potential of sex-specific immune-based approaches for constructing more accurate, patient-centric risk-assessment models.

Jessica Leivesley

University of Toronto

Jessica is a quantitative ecologist and postdoctoral research fellow at the University of Toronto in the Department of Statistical Sciences. She is co-supervised by Vianey Leos-Barajas and Dak de Kerckhove. Jessica completed her PhD at the University of Toronto in the Ecology department in 2023 where she was interested in understanding the impact of climate change on growth, survival, and maturation of freshwater turtles through integrated modelling of long-term individual-based data. After this, she joined her current position as a part of the first cohort of Schmidt AI in Science fellows where she works to integrate machine learning into fisheries stock assessments. This is collaborative work with research scientists at the Ontario Ministry of Natural Resources and Forestry.

From sonar to species: using wideband acoustics and machine learning to classify fish species

Canada's recreational fishery contributed $7.9 billion to the national economy in 2015, and in Ontario alone freshwater recreational and commercial fisheries represent a $2.2 billion industry. To maintain sustainable and resilient fisheries, managers must have accurate information on the current status of stock health, population-size, and fish communities for many water bodies at a given time. Generally, this information is gathered through resource-intensive and lethal sampling methods. Current hydroacoustic methods can assess individual fish sizes but species identities cannot be discerned. The recent development of wideband acoustic transducers which emit a wide range of frequencies in a single ping may allow more information on body form to be extracted and thus may aid in species identification. In this study, we created a labelled dataset of acoustic responses of two fish species by tethering individual fish under a transducer emitting 249 frequencies between 45kHz and 170kHz. We then applied three different bespoke machine learning algorithms (deep, recurrent, and residual neural networks) to acoustic backscatter measures at each frequency and tested their ability to correctly classify the two fish species. We found that on unseen data all three methods had over 85% balanced classification accuracy. Further, extracting SHAP values for the deep neural network showed that there is not a single range of frequencies that are important for distinguishing the species, but rather the most important frequencies are distributed across the range of frequencies used. Eventually, these algorithms can be integrated into current abundance or biomass models and allow users to propagate classification uncertainty into these models. Overall, the use of wideband acoustics in conjunction with machine learning techniques offers the potential to drastically reduce the resources needed and costs associated with monitoring fish stocks.

Yaqi Li

University of Oklahoma

I am an assistant professor in the Department of Pediatrics at the University of Oklahoma Health Science Center. I obtain my Ph.D. degree in quantitative psychology. My research expertise lies in developing and applying algorithms or statistical models to conduct time series data analysis and model selection, as well as the application of machine learning/deep learning to social and behavioral science. I also have research interests in the research of data quality, missing data, and causality

Enhance Data Quality in Child Welfare

For almost all scientific research, the accuracy and reliability of findings, as well as the performance of predictive models depend upon the quality of the data used. As highlighted by Arias et al. (2020) in their study, even minor errors within research datasets can significantly impact the overall accuracy of results. However, within child welfare area, the issue of data quality has been relatively overlooked. The quality of child welfare outcomes is intricately linked to the quality of data, Nonetheless, the quality of child welfare outcomes is intricately tied to the quality of data, especially in the context of automated decision-making in service delivery. For instance, Predictive Risk Modeling (PRM), a predictive model utilized to automate decisions regarding child maltreatment, has faced criticism for generating biased decisions in practice, largely due to data fraud with errors. The forthcoming presentation will elucidate the results of data quality evaluations conducted on a nationwide child welfare database. Additionally, strategies to identify potential factors contributing to suboptimal data quality will be addressed. The presentation aims to address critical gaps in understanding and addressing data quality issues within the child welfare area, ultimately aiming to improve the effectiveness and fairness of decision-making in this area.

Tony Liu

University of Pennsylvania

Tony Liu (he/him) is an incoming assistant professor of Computer Science at Mount Holyoke College. He is completing his PhD at the University of Pennsylvania, co-advised by Lyle Ungar and Konrad Kording. His research is at the intersection of causal inference and machine learning, with applications to healthcare policy evaluation and smartphone mobile sensing for mental wellness. At UPenn, he has been a Center for Teaching and Learning graduate teaching fellow and an instructor for CIS 1920: Python Programming over three semesters. Tony is also a part-time scientist and program manager at Roblox, where he conducts research on online civility and coordinates academic collaboration, technology transfer, and external communication for Roblox Research. He received his BA in Computer Science with highest honors from Williams College.

Automated Detection of Causal Inference Opportunities: Regression Discontinuity Subgroup Discovery

The gold standard for the identification of causal effects are randomized controlled trials (RCT), but RCTs may not always be feasible to conduct. When treatments depend on a threshold however, such as the blood sugar threshold for diabetes diagnosis, we can still sometimes estimate causal effects with regression discontinuities (RDs). In practice however, implementing RD studies can be difficult as identifying treatment thresholds require considerable domain expertise -- furthermore, the thresholds may differ across subgroups (e.g., the blood sugar threshold for diabetes may differ across demographics), and ignoring these differences can lower statistical power. Finding the thresholds and to whom they apply is an important problem currently solved manually by domain experts, and data-driven approaches are needed when domain expertise is not sufficient. Here, we introduce Regression Discontinuity SubGroup Discovery (RDSGD), a machine-learning method that identifies statistically powerful and interpretable subgroups for RD thresholds. Using a medical claims dataset with over 60 million patients, we apply RDSGD to multiple clinical contexts and identify subgroups with increased compliance to treatment assignment thresholds. As treatment thresholds matter for many diseases and policy decisions, RDSGD can be a powerful tool for discovering new avenues for causal estimation.

Stephanie Milani

Carnegie Mellon University

Stephanie Milani is a fifth-year PhD student in the Machine Learning Department at Carnegie Mellon University, where she is advised by Fei Fang. Her research focuses on responsible reinforcement learning, including interpretability, transparency, and human involvement. She co-organized the NeurIPS MineRL Diamond and BASALT competitions on sample-efficient reinforcement learning and learning from human feedback, respectively. She has also served as the Logistics Chair for the ICML WiML Un-Workshop to support diversity in machine learning. Previously, she graduated from UMBC with a B.S. in Computer Science and a B.A. in Psychology. Her personal website is: https://stephmilani.github.io/

MAVIPER: Learning Decision Tree Policies for Interpretable Multi-Agent Reinforcement Learning

Many recent breakthroughs in multi-agent reinforcement learning (MARL) require the use of deep neural networks, which are challenging for human experts to interpret and understand. However, existing work on interpretable reinforcement learning (RL) has shown promise in extracting more interpretable decision tree-based policies from neural networks, but only in the single-agent setting. To fill this gap, we propose the first set of algorithms that extract interpretable decision-tree policies from neural networks trained with MARL. The first algorithm, IVIPER, extends VIPER, a recent method for single-agent interpretable RL, to the multi-agent setting. We demonstrate that IVIPER learns high-quality decision-tree policies for each agent. To better capture coordination between agents, we propose a novel centralized decision-tree training algorithm, MAVIPER. MAVIPER jointly grows the trees of each agent by predicting the behavior of the other agents using their anticipated trees, and uses resampling to focus on states that are critical for its interactions with other agents. We show that both algorithms generally outperform the baselines and that MAVIPER-trained agents achieve better-coordinated performance than IVIPER-trained agents on three different multi-agent particle-world environments

Harsh Parikh

Johns Hopkins University

Harsh Parikh is a postdoctoral researcher at the Johns Hopkins Bloomberg School of Public Health, where he specializes in developing cutting-edge causal inference methodologies that are accurate, trustworthy, and sensitive to specific domain requirements. Harsh is committed to bridging the research-to-practice gap by actively collaborating with leading experts in various fields. Among his notable collaborations are projects with neurologists at Massachusetts General Hospital, aimed at improving seizure management protocols for critically ill patients. He also partners with epidemiologists at Columbia University Irving Medical Center to devise effective treatments for opioid use disorder. He earned his Ph.D. from Duke University's Department of Computer Science, where he worked on machine learning-aided causal inference, earning the prestigious Amazon Graduate Fellowship for his work. His dissertation was recognized with an Outstanding Dissertation award. Harsh also holds an MS in Economics from Duke University and a BTech in Computer Science and Engineering from IIT Delhi.

Characterizing Underrepresented Population in Randomized Controlled Trials

Randomized controlled trials (RCTs) serve as the cornerstone for understanding causal effects, yet extending inferences to target populations presents challenges due to effect heterogeneity and underrepresentation. Our work addresses the critical issue of identifying and characterizing underrepresented subgroups in RCTs, proposing a novel framework for refining target populations to improve generalizability. We introduce an optimization-based approach, Rashomon Set of Optimal Trees (ROOT), to characterize underrepresented groups. ROOT optimizes the target subpopulation distribution by minimizing the variance of the target average treatment effect estimate, ensuring more precise treatment effect estimations. Notably, ROOT generates interpretable characteristics of the underrepresented population, aiding researchers in effective communication. Our approach demonstrates improved precision and interpretability compared to alternatives, as illustrated with synthetic data experiments. We apply our methodology to extend inferences from the Starting Treatment with Agonist Replacement Therapies (START) trial -- investigating the effectiveness of medication for opioid use disorder -- to the real-world population represented by the Treatment Episode Dataset: Admissions (TEDS-A). By refining target populations using ROOT, our framework offers a systematic approach to enhance decision-making accuracy and inform future trials in diverse populations.

Rahul Ramesh

University of Pennsylvania

Rahul Ramesh is a Ph.D. student in the department of computer and information science at the University of Pennsylvania and is advised by Pratik Chaudhari. He received his B.Tech from the Indian Institute of Technology Madras in Computer science and Engineering and received the Alumni association prize for graduating at the top of his class. His research attempts build tools to understand data used to train deep networks by studying the geometric structure in the space of learnable tasks.

A Picture of the Space of Typical Learnable Tasks

We develop information geometric techniques to understand the representations learned by deep networks when they are trained on different tasks using supervised, meta-, semi-supervised and contrastive learning. We shed light on the following phenomena that relate to the structure of the space of tasks: (1) the manifold of probabilistic models trained on different tasks using different representation learning methods is effectively low-dimensional; (2) supervised learning on one task results in a surprising amount of progress even on seemingly dissimilar tasks; progress on other tasks is larger if the training task has diverse classes; (3) the structure of the space of tasks indicated by our analysis is consistent with parts of the Wordnet phylogenetic tree; (4) episodic meta-learning algorithms and supervised learning traverse different trajectories during training but they fit similar models eventually; (5) contrastive and semi-supervised learning methods traverse trajectories similar to those of supervised learning. We use classification tasks constructed from the CIFAR-10 and Imagenet datasets to study these phenomena.

Ransalu Senanayake

Arizona State University

Ransalu Senanayake is an Assistant Professor in Computer Science at the School of Computing and Augmented Intelligence, Arizona State University (ASU). He leads the Laboratory for Learning Evaluation of autoNomous Systems (LENS Lab) to make large-scale deep neural networks, especially those deployed in embodied systems and robots, robust and socially compatible. Previously, he was a postdoctoral scholar in the Machine Learning Group at the Dept. of Computer Science, Stanford University. He also worked at the Dept. of Aeronautics & Astronautics Engineering at Stanford University. He completed his PhD in Computer Science at the School of Computer Science in the University of Sydney. He also obtained an MPhil degree in Industrial Engineering from the Hong Kong University of Science and Technology.

Towards Making Embodied Agents Socially Compatible:
Identifying and Mitigating Biases in Large-Scale Vision and Language Models

The deployment of physical embodied AI systems, such as autonomous vehicles, is rapidly expanding. At the heart of these systems, there are numerous computer vision and large language modules that directly influence the downstream decision-making tasks by considering the presence of nearby humans, such as pedestrians. Despite the high accuracy of these models on held-out datasets, the potential presence of algorithmic bias is challenging to assess. We discuss our ongoing efforts at the Laboratory for Learning Evaluation of autoNomous Systems (LENS Lab) in analyzing disparate impacts for groups with different genders, skin tones, body sizes, professions, etc. in large-scale deep neural networks, especially under physical perturbations.

Subhasree Sengupta

Clemson University

I am a postdoctoral researcher in the School of Computing at Clemson University. I will join Florida State University as an Assistant professor in Fall, 2024. My work blends scholarship from Computer-mediated Communication, Social Informatics, Human-AI interaction and Computational Social Science. My present research focuses on understanding ethics in human-AI teamwork and public opinions, sentiments and reactions to the AI ethics debate. I completed my Ph.D. from the School of Information Studies, Syracuse University, in December 2022. Before my Ph.D., I completed my bachelor's in computer science with minors in Mathematics and Statistics from the University of Minnesota, Twin Cities, and my master's in computer science from the University of Southern California, focusing on Artificial Intelligence, machine learning, and social network analysis.

Public Perceptions, Critical Awareness, and Community Discourse on AI Ethics:
Evidence from an Online Discussion Forum

As Artificial Intelligence (AI) becomes increasingly ingrained into society, ethical and regularity concerns become critical. Given the vast array of philosophical considerations of AI ethics, there is a pressing need to understand and balance public opinion and expectations of how AI ethics should be defined and implemented, such that it centers the voice of experts and non-experts alike. This investigation explores a subreddit r/AIethics through a multi-methodological, multi-level approach. The analysis yielded six conversational themes, sentiment trends, and emergent roles that elicit narratives associated with expanding implementation, policy, critical literacy, communal preparedness, and increased awareness towards combining technical and social aspects of AI ethics. Such insights can help to distill necessary considerations for the practice of AI ethics beyond scholarly traditions and how informal spaces (such as virtual channels) can and should act as avenues of learning, raising critical consciousness, bolstering connectivity, and enhancing narrative agency on AI ethics.

Nasim Sonboli

Brown University

Nasim is a Postdoctoral Research Associate in the Data Science Institute (DSI) at Brown University, working with Suresh Venkarasubramanian. Her research interests are the societal aspects of machine learning algorithms, algorithmic fairness, recommender systems, GDPR (General Data Protection Regulations). Previously she was a Postdoctoral Scholar at Tufts University. She got her Ph.D. in Information Science in 2022, working on fairness-aware recommender systems under the supervision of Robin Burke at the University of Colorado Boulder. She has a master's degree in Data Science from DePaul University in 2016 and a bachelor's degree in Software Engineering.

Investigating the tension between Fairness and Data Minimization in Machine Learning Systems

The General Data Protection Regulations (GDPR) are designed to protect personal data from harm, with mandatory adherence within the European Union and varying levels of alignment elsewhere. Complying with GDPR is complex due to the potential contradictions within the regulations themselves. Additionally, operationalizing these regulations in machine learning systems adds additional complexity. Hence, it's crucial to assess the feasibility of simultaneously achieving GDPR compliance in general and in machine learning systems, and to consider potential trade-offs if full alignment proves unattainable. In this research, we study the current research on data minimization in machine learning. We investigate the relationship between data minimization, fairness, and accuracy. Few works have investigated data minimization in machine learning and even fewer research on the conflict of data minimization with other GDPR principles. Our long-term goal is to provide guidelines how to operationalize data minimization in machine learning systems for the computer scientists, practitioners and researchers in academia and industry. We explore the existing tools to implement these regulations in machine learning systems and the advantages and disadvantages of these tools. Additionally, we investigate the potential tradeoffs among GDPR and we provide a roadmap how to navigate them. By exploring these critical aspects, we offer valuable insights for developing machine learning systems that comply with data protection regulations.

Tiffany Tang

University of Michigan

Tiffany is a postdoctoral researcher with Ji Zhu and Liza Levina in the University of Michigan Statistics Department. Her research interests are primarily problem-driven and lie broadly at the intersection of applied statistics, data science, and medicine. She has worked on developing reproducible and interpretable machine learning pipelines for a range of scientific problems including precision medicine and cardiovascular genetics. She will be joining the University of Notre Dame as an Assistant Professor in Fall 2024. Previously, she received her PhD in Statistics from UC Berkeley, where she was advised by Bin Yu.

Interpretable network-assisted prediction via random forests

Machine learning algorithms often assume that training samples are independent. When data points are connected by a network, it creates dependency between samples, which is a challenge, reducing effective sample size, and an opportunity to improve prediction by leveraging information from network neighbors. Multiple prediction methods taking advantage of this opportunity are now available. Many methods including graph neural networks are not easily interpretable, limiting their usefulness in the biomedical and social sciences, where understanding how a model makes its predictions is often more important than the prediction itself. Some are interpretable, for example, network-assisted linear regression, but generally do not achieve similar prediction accuracies as more flexible models. We bridge this gap by proposing a family of flexible network-assisted models built upon a generalization of random forests (RF+), which both achieves highly-competitive prediction accuracy and can be interpreted through feature importance measures. In particular, we provide a suite of novel interpretation tools that enable practitioners to not only identify important features that drive model predictions, but also quantify the importance of the network contribution to prediction. This suite of general tools broadens the scope and applicability of network-assisted machine learning for high-impact problems where interpretability and transparency are essential.

Shantanu Vyas

Texas A&M University

Shantanu is a PhD candidate at Texas A&M University working at the intersection of Human Computer Interaction, applied AI and Design. His work encompasses various projects, such as developing language-guided 3D design tools for children and creating adaptive learning systems for emergency responders in AR/VR environments. With a passion for interdisciplinary collaboration, Shantanu is dedicated to advancing human creativity and comprehension through his research endeavors. Beyond academia, he finds joy in soccer, culinary exploration, and discovering new hobbies.

Fostering Reflective Thinking in Early-Stage Design through Responsible Integration of Generative AI

In the initial stages of design, ambiguity and uncertainty present significant challenges to designers. Although generative AI shows promise in addressing these challenges, its premature application risks hindering creative exploration and inhibiting reflective thinking, both integral to the design process. Our work proposes strategies to responsibly integrate LLMs into the design process, by fostering reflective thinking over immediate solution generation. By reframing the role of LLMs to prompt contextual questioning and surface latent concepts in design problems, we aim to support designers in generating novel ideas while preserving their creative autonomy. We suggest techniques for incorporating explainability into generative design processes, utilizing multi-modal models trained on design language and 3D design concepts to provide explicit rationales for generated design solutions. Through these techniques, our objective is to instill trust in designers regarding solutions generated by AI models and, more importantly, to stimulate reflective thinking processes. Our work seeks to comprehend the responsible utilization of AI to nurture human creativity and critical thinking in the design process without replacing it.

Guanchu Wang

Rice University

Guanchu Wang's research is primarily focused on the field of explainable AI, where he specializes in developing algorithms that elucidate the inner workings of deep neural networks. His research can be widely applied to most foundation models, including multilayer perception, graph neural networks, convolutional neural networks, vision transformers, and large language models. Within the domain of explainable AI, he has contributed through peer-reviewed papers at the top machine learning conferences and journals, including ICML, ICLR, NeurIPS, AAAI, IJCAI, TKDD, and more. On Google Scholar, he has gained over 500 citations, with an h-index of 11 and an i10-index of 11. He was a recipient of the ICML Spotlight Paper in 2022, CIKM Best Demo Paper Award in 2022, and Ken Kennedy Institute Graduate Fellowship in 2023.

Advancing Faithful Explanation Towards Transparent Machine Learning

My doctoral research centers on responsible AI, a critical area that demands the infusion of trust throughout the AI lifecycle. Within this overarching theme, my research delves into explainable AI, which specializes in developing algorithms to explain the behaviors of deep neural networks faithfully. The overarching goal of this thesis is to make the decision-making process within deep neural networks understandable to humans, thereby facilitating the safe deployment of machine learning to high-stake application scenarios. This abstract highlights two significant milestones from my research in explainable AI: 1) Developing Shaplay Value Explanation for DNNs: my seminal work SHEAR focuses on accurately estimating the Shapley value to explain the DNN decision, under a limited sampling budget. In our healthcare project, SHEAR is capable of precisely assessing the impact of gene-gene interaction on Alzheimer's disease. 2) Explaining Large Language Models: we propose a generative explanation framework xLLM for explaining the outputs of large language models (LLMs). xLLM can faithfully explain most existing LLMs, such as the ChatGPT, LLAMA, and Claude, ensuring trustworthy decision-making in AI-driven healthcare.

Haoyu Wang

Purdue University

I am currently a final year Ph.D. student under the advisory of Prof. Jing Gao in School of Electrical and Computer Engineering, Purdue University. My research interests lie in the intersection of data mining, natural language processing, and machine learning, with a strong focus on democratizing AI for broader accessibility.

Democratizing AI for Broader Accessibility

Haoyu Wang's research addresses the critical challenge of democratizing AI, focusing on making AI more accessible through data and parameter efficiency, and ensuring trustworthiness by emphasizing fairness, robustness, and interpretability. His work introduces innovative model compression techniques that facilitate AI deployment on low-resource devices, enhancing global accessibility. Furthermore, his efforts in cross-lingual and multi-lingual understanding aim to overcome language barriers in AI use. By advocating for ethical AI, his research aligns technical advancements with societal needs, ensuring AI's benefits are equitably distributed. This body of work represents a significant step towards accessible, trustworthy AI for all.

Galen Weld

University of Washington

Galen Weld is a PhD Student at the Paul G. Allen School of Computer Science & Engineering at the University of Washington, where he is advised by Tim Althoff and Amy Zhang. His research focuses on developing methods for quantifying the governance of online communities at web scale.

Quantifying Governance of Online Communities at Web Scale

Online communities are powerful tools to connect people and are used worldwide by billions of people. Nearly all online communities rely upon moderators or admins to govern the community in order to mitigate potential harms such as harassment, polarization, and deleterious effects on mental health. However, online communities are complex systems, and studying the impact of community governance empirically at scale is challenging because of the many aspects of community governance and outcomes that must be quantified. In this work, we develop methods to quantify the governance of online communities at web scale. We survey community members to build a comprehensive understanding of what it means to make communities ‘better,’ then assess existing governance practices and associate them with important outcomes to inform community moderators. We collaborate with communities to deploy our governance interventions to maximize the positive impact of our work, and, at every step of the way, we make our datasets and methods public to support further research on this important topic.

Siyu Wu

Pennsylvania State University

Siyu's works focus on neural symbolic AI and data science. Her current research vision is to use a cognitive architecture to provide a symbolic and external structure for LLM output and reasoning, making the reasoning process more transparent and understandable. She achieves this by integrating a symbolic representation of ACT-R within the LLM framework. In her free time, she enjoys Zumba, Legos, and gardening.

Enhancing LLMs with a Neuro-Symbolic Architecture (ACT-R) for Decision Making

This research uniquely integrates ACT-R's cognitive framework within LLMs to provide structure and clarity to their reasoning, enhancing decision-making transparency and explainability – a step not yet explored in current studies. We first highlight the disparity between Large Language Models (LLMs) and human decision-making, noting LLMs' focus on rapid, intuitive processes and their limitations in complex reasoning and learning continuity. To address these shortcomings, we then propose integrating LLMs with the ACT-R cognitive architecture, a framework that models human cognitive processes. This integration aims to enhance LLMs with human-like decision-making and learning patterns by correlating ACT-R decision-making data with LLM embeddings. The architecture we propose has the potential to enable LLMs to make decisions and learn in ways that more closely mirror human cognition, addressing the critical challenge of aligning machine reasoning with human processes.

Yuchen Zeng

University of Wisconsin-Madison

I am a graduate student pursuing a PhD's degree in the Department of Computer Science at the University of Wisconsin-Madison. I am advised by Prof. Kangwook Lee. My current research interests include large language models and machine learning fairness. I received my Master's degree in Statistics from UW-Madison in 2020, where I was advised by Prof. Miaoyan Wang. Before joining UW-Madison, I completed my Bachelor's degree in Statistics from the School of Mathematical Sciences at Nankai University in 2019.

Investigating Parameter-Efficient Fine-Tuning Methods for Enhancing Fairness in Large Language Models

Recently, there has been a significant increase in the development of large language models (LLMs), which are now extensively used in everyday life. However, the fairness and safety of these models have become significant concerns. Existing studies suggest that parameter-efficient fine-tuning (PEFT) can help alleviate the inherent biases present in LLMs. Our research aims to comprehensively understand PEFT's capabilities through both experimental and theoretical lenses. We demonstrate that Low-Rank Adaptation (LoRA), a popular PEFT method, excels in adapting LLMs for non-language tasks, including processing tabular datasets, a crucial type for fair classification tasks, as evidenced by extensive experiments. Furthermore, we theoretically establish that LoRA can fine-tune a randomly initialized model into any smaller target model, showcasing the potential of PEFT. Through an in-depth exploration of PEFT's practical applications and theoretical underpinnings, our work lay the foundation for future research aimed at enhancing the fairness and safety of LLMs via PEFT.

Thank you to our

Future Leaders Summit 2024 Sponsor