N-SEA 2025

www.equitableaisymposium.com

April 5-6, 2025

Earl G. Graves School of Business, Morgan State University
4200 Hillen Road, Baltimore, MD 21218

Full Conference Program

N-SEA is organized and hosted each year by
CEAMLS (the Center for Equitable Artificial Intelligence & Machine Learning Systems)
at Morgan State University. Find out more at www.morgan.edu/ceamls

Schedule-at-a-Glance
N-SEA 2025 (National Symposium on Equitable AI)

Day 1: April 5, 2025					
	Time	*Location*			
		Graves Hall 104 (Auditorium)	*Graves Hall 110*	*Graves Hall 112*	*Graves Hall Atrium*
Continental Breakfast	8:30-9:15				**Breakfast**
Opening Remarks	9:15-9:30	Kofi Nyarko (Morgan State U.)			
Keynote 1	9:30am-10:15am	**Arjune Sen** (Oxford U.)			
Break	10:15am-10:30am				
Breakout Sessions 1	10:30am-12:00pm	**AI & Society 1**	**AI & Medicine**	**AI & Ethics**	
Lunch	12:00pm-1:00pm				**Lunch**
Breakout Sessions 2	1:00pm-2:30pm	**AI as a Catalyst for Social Change** (Panel Discussion)	**AI & Society 2**	**Explainability, Safety, & Reliability**	
Poster Session	2:30pm-4:00pm				**Poster Session**
Keynote 2	4:00pm-4:45pm	**Fay Cobb Payton** (Rutgers U. / NC State U.)			
Reception and Poster Award Ceremony	4:45pm-5:45pm				**Reception & Poster Award Ceremony**

Day 2: April 6, 2025					
	Time	*Location*			
		Graves Hall 104 (Auditorium)	*Graves Hall 110*	*Graves Hall 112*	*Graves Hall Atrium*
Continental Breakfast	8:30-9:15				**Breakfast**
Opening Remarks	9:15-9:30	Kofi Nyarko (Morgan State U.)			
Keynote 3	9:30am-10:15am	**Sina Fazelpour** (Northeastern U.)			
Break	10:15am-10:30am				
Breakout Sessions 3	10:30am-12:00pm	**AI & Education**		**Bias Mitigation & Fairness in AI**	
Lunch	12:00pm-1:00pm				**Lunch**
Breakout Sessions 4	1:00pm-2:00pm	**AI & Environment**	**AI & Decision Sciences**	**AI & Philosophy**	
Break	2:00pm-2:15pm				
Keynote 4	2:15pm-3:00pm	**Agus Sudjianto** (Wells Fargo / UNC Charlotte)			
Break	3:00pm-3:15pm				
Closing Panel Discussion	3:15pm-4:00pm	TBA			
Closing remarks	4:00pm-4:15pm	Kofi Nyarko (Morgan State U.)			
Reception	4:15pm-5:30pm				**Reception**

Breakout Sessions #1: 10:30am-12:00pm, April 5

AI & Society 1
Graves Hall 104

What is the future of work in the Generative AI era?: A Marxist and Ricardian Analysis
Larry Liu
Assistant Professor of Sociology,
Morgan State University

The Interpretability-Accuracy Trade-Off in Crime Prediction: A Practitioner's Perspective
Gaspard Tissandier
Postdoctoral Fellow, Rutgers University,
and
Alejandro Gimenez-Santana
Rutgers University

10 minute break

Mindsets and Management: AI and Gender (In)Equitable Access to Finance
Genevieve Smith
Faculty, UC Berkeley, and Student, Oxford University

Responsible AI Development through Community-Informed AI Development
Jamila Smith-Loud
Head of Impact Lab, Google Research

AI in Medicine
Graves Hall 110

Laying the Groundwork for Equitable AI in Time-Series Medical Imaging
Fremah Agyemang
Graduate Research Assistant,
Morgan State University

Designing Equitable AI: From Diversified Discovery to Clear Communication at the Bedside
Camille Jimenez
Pfizer

10 minute break

Doxastic Neutrality in AI: A New Paradigm for Managing Uncertainty in Medical Decisions*
Nesim Aslantar
Visiting Research Scholar, Dept. of Philosophy,
Indiana U. Bloomington

Session ends 11:10pm

AI & Ethics
Graves Hall 112

A Dilemma for AI Alignment
Sarah Valdman
PhD Candidate, Philosophy,
University of Michigan

Ethical Boundaries in the Use of Private and Public Information for Deepfake Creation
Sherri Lynn Conklin
Assistant Professor of Philosophy,
Washington State University

10 minute break

Rawlsian Justice and Algorithmic Monocultures*
Shiying Li
PhD Candidate, Philosophy,
University of Wisconsin-Madison

From Bias to Balance: Towards Ethical Equity in Artificial Intelligence in Education (AIED)*
Nasreen Watson
Artificial Intelligence Researcher, Department of Philosophy, University of Johannesburg,
South Africa

*Virtual presentation

Breakout Sessions #2: 1:00pm-2:30pm, April 5

AI as a Catalyst for Social Change: Designing Technology that Centers Humanity (Panel Discussion)

Graves Hall 104

Panelists:

Armisha Roberts (Moderator)
Post-Doctoral Researcher,
HCAI Institute at Howard University

Jaye Nias
Sr. Research Scientist,
HCAI Institute at Howard University

Saurav Aryal
Research Scientist,
HCAI Institute at Howard University

Thane Douglas
Undergraduate Researcher,
HCAI Institute at Howard University

Howard Prioleau
Doctoral Research Assistant,
HCAI Institute at Howard University

AI & Society 2

Graves Hall 110

A Data-Driven Exploration of Socioeconomic Influences on Urban Mobility: Enhancing Gender Equity in Maryland's Transportation Systems*
Zeinab Bandpey and Mehdi Shokouhian
Department of Civil and Environmental Engineering,
Morgan State University

Generative AI and the Courts: an American-centered overview*
Annalisa Triggiano
Sant'Anna School of Advanced Studies, Italy

10 minute break

Arab States' Readiness for Artificial Intelligence and Its Expected Impacts on Journalism and Media *
Elsayed Elkilany
Mass Communication Department,
College of Arts and Sciences, Qatar University

Session ends 2:10pm

Explainability, Safety, & Reliability

Graves Hall 112

Towards AI Safety in Context: How to Design Interdisciplinary AI Audits
Mona Sloane
Assistant Professor of Data Science and Media Studies, University of Virginia

Pluralistic Alignment: A Roadmap, Recent Work, and Open Problems
Taylor Sorensen
PhD Student, Department. of Computer Science, University of Washington, Seattle

10 minute break

Evaluating the Quality of Interpretations and Explanations in AI Models in Mission-Critical Domains
Blessing Ojeme
Assistant Professor, Dept. of Computer Science, Morgan State University

Reliability, Model Assumptions, and Interpretability*
Luigi Scorzato
Accenture

*Virtual presentation

Breakout Sessions #3: 10:30am-12:00pm, April 6

AI & Education
Graves Hall 104

What do Doctoral Students Tell?: Leveraging AI Tools such as ChatGPT, Google Gemini, and Microsoft Copilot in Higher Education
Krishna Bista
Professor, Morgan State University

Generative AI Adoption in Special Education
Marie Sakowicz
Graduate Student,
University of Maryland, Baltimore County

10 minute break

From Concept to Criteria: Using Generative AI to Develop Grading Rubrics for Undergraduate Courses
Bryn Seabrook
Assistant Professor in Engineering & Society,
University of Virginia

LLMs and Educational Equity*
Renzhe Yu
Teachers College, Columbia University

No session

Bias Mitigation & Fairness in AI
Graves Hall 112

Fairness and Privacy in High-Stakes NLP
Anjalie Field
Assistant Professor of Computer Science,
Johns Hopkins University

Trustwatch: Find an AI You Trust
Chukwuemeka Duru
Morgan State University

10 minute break

Proactive Methods For Ensuring Fairness in AI Systems in Healthcare
Sudip Sharma
Graduate Research Assistant,
CEAMLS, Morgan University

Session ends at 11:40pm

*Virtual presentation

Breakout Sessions #4: 1:00pm-2:00pm, April 6

AI & Environment
Graves 104

A machine learning based high-resolution gridded surface climate dataset over land
Douglas Rao
NC Institute for Climate Studies,
NC State University

How "Ethical" Factory Farming with AI Imperils Future Animals
Amber Sheldon
PhD Candidate, Boston University,
Department of Philosophy

Environmental Challenges of AI
Willie May
VP of Research, Morgan State University /
AAAS Immediate Past President

AI & Decision Sciences
Graves 110

Adaptive Algorithmic Interventions for Escaping Pessimism Traps in Dynamic Sequential Decisions
Alexander Tolbert
Assistant Professor, Dept. of Quantitative Theory & Methods, Emory University

Multi-Criteria Decision Analysis Framework for Evaluating Artificial Intelligence Systems*
Olusola Olabanjo
PhD Candidate, Dept. of Mathematics,
Graduate RA, CEAMLS,
Morgan State University
and
Phillip Honenberger
AI Ethicist & Researcher,
CEAMLS, Morgan State University

Session ends at 1:40pm

AI & Philosophy
Graves 112

Sexual Humiliation and AI: An Existential-Feminist Analysis of Deepfake Pornography
Kate Yuan
PhD Student, Philosophy
Yale University

Deeply Opaque Black Boxes as a Unique Problem for the Democratic Governance of AI
Armin Heydari
PhD Candidate, Philosophy,
Harvard University

The Subject Project:
Human Visibility, Vulnerability, and Diversity amidst Proxy Discrimination in the Data Age
James Garrison
Assistant Professor of Philosophy
University of Massachusetts, Lowell

*Virtual presentation

Keynote #1:

Arjune Sen

Oxford University

Saturday, April 5, 9:30am-10:15am
Graves Hall 104 (Auditorium)

AI Equity in an Inequitable World: Lessons and Learnings from Brain Health

Abstract:

We live in a world of great inequity. Between 2020 to 2023, the richest 1% of people garnered almost double the new wealth of the remaining 99% of the world's population. We may have thought that the COVID-19 pandemic would have brought us together to help our fellow humans, but inequality has only sharpened in the post-pandemic era. In healthcare, there is vast variability in access to clinicians, access to diagnostic technologies and access to medications. A much-heralded solution to these challenges is said to be AI. But will this really work? In this presentation we will explore potential benefits of AI applied to the common neurological conditions of epilepsy, dementia and stroke. Can AI offer meaningful advantages to people affected by these disorders in resource limited communities? What must be overcome if the true potential of technological advances is to be realised and what can we each do to help in that regard? With an ageing population affected by multimorbidity, financial constraints and multiple global upheavals, there is both huge opportunity for AI-driven health care and high potential risk. We will need to choose carefully if we are to deliver optimal solutions.

Biography:

Arjune is appointed as Professor of Global Epilepsy at The University of Oxford and Consultant Neurologist at Oxford University Hospitals NHS Foundation Trust. He is Director of the Centre for Global Epilepsy and Fellow of the Oxford Martin School, where he leads the Programme on Global Epilepsy (https://www.oxfordmartin.ox.ac.uk/global-epilepsy/). Arjune is also Head of the Epilepsy Research Group at the Oxford University Hospitals; topic advisor to the NICE Epilepsy Guidelines; and a member of Commissions and Councils at the International League Against Epilepsy. Clinical interests include cognitive, psychological and psychosocial difficulties in people with epilepsy. At a research level, Arjune works on trying to better understand the intersections between epilepsy and dementia. He is Chief Investigator of the ILiAD (Investigation of Levetiracetam in Alzheimer's Disease) study and leads multiple other trials both in Oxford and the UK more broadly. Over the past several years, a clear focus has been global epileptology with a view to improving care for people in resource poor settings through the development of AI-based culturally tailored technologies to be deployed at scale through hard to reach communities.

Keynote #2:

Fay Cobb Payton

(Rutgers U. / NC State University)

Saturday, April 5, 4:00pm-4:45pm
Graves Hall 104 (Auditorium)

Bridging Digital Divides in This Algorithmic World: New Tech with Old & New-Fangled Challenges

Abstract:

Artificial Intelligence (AI) and Large Language Models (LLMs) are rapidly transforming healthcare, transitioning from reactive to limited memory systems. This evolution promises significant advancements in disease diagnosis, drug discovery, personalized medicine, and healthcare operations. However, the current focus on big data in AI development overlooks crucial "small data" - information about human experiences, choices, and social determinants of health. The absence of these data married with digital divides stand to further preclude holistic healthcare approaches, equitable access and quality outcomes. This talk will address the current overreliance on big data and the need to embrace a more comprehensive data strategy. To do so, we can create AI systems (and other technologies) that not only enhance medical capabilities but also embrace holistic patient views, promote parity and improve lives across populations.

Biography:

Dr. Fay Cobb Payton is an award-winning researcher, international speaker and entrepreneur. She is a Special Advisor to the Chancellor for Innovation, Professor (with Tenure) in the Department of Mathematics and Computer Science at Rutgers University – Newark and an affiliate faculty in the Rutgers New Jersey Medical School. She is the director of the new Institute for Data, Research and Innovation Science (IDRIS) at Rutgers University – Newark. She is also Professor Emerita and was a Full Professor (with Tenure) of Information Technology/Analytics at North Carolina State University. She completed a rotation as a Program Director at the National Science Foundation. Dr. Payton serves on the NAIRR Pilot and several National Academics of Science, Engineering advisory committees. She has published over 150 peer-reviewed journal articles, conference publications and book chapters on topics of data quality, AI bias/ethics, healthcare and innovation. She worked in industry as an engineer, developer and consultant prior to earning her doctoral degree.

Keynote #3:

Sina Fazelpour

(Northeastern University)

Sunday, April 6, 9:30am-10:15am
Graves Hall 104 (Auditorium)

Aspirational Affordances of AI

Abstract:

As artificial intelligence (AI) systems increasingly permeate processes of cultural and epistemic production, there are growing concerns about how their outputs may confine individuals and groups to static or restricted narratives about who or what they could be. In this paper, we advance the discourse surrounding these concerns by making three contributions. First, we introduce the concept of aspirational affordance to describe how technologies of representation---paintings, literature, photographs, films, or video games---shape the exercising of imagination, particularly as it pertains to possibilities for agency. Second, we provide three reasons for scrutinizing of AI's influence on aspirational affordances: AI's influence is potentially more potent, but less public than traditional sources; AI's influence is not simply incremental, but ecological, transforming the entire landscape of cultural and epistemic practices that traditionally shaped aspirational affordances; and AI's influence is highly concentrated, with a few corporate-controlled systems mediating a growing portion of aspirational possibilities. Third, to advance such a scrutiny of AI's influence, we introduce the concept of aspirational harm, which arises when AI-generated affordances distort or diminish a group's hermeneutical resources for imagining their practical possibilities. Through three case studies, we illustrate how aspirational harms extend the existing discourse on AI-inflicted harms beyond representational and allocative harms, warranting separate attention. Through these conceptual resources and analyses, this paper advances understanding of the psychological and societal stakes of AI's role in shaping individual and collective aspirations.

Biography:

Sina Fazelpour is an Assistant Professor of Philosophy and Computer Science at Northeastern University. Sina's research spans the fields of philosophy and ethics of AI, cognitive science, social philosophy, and philosophy of science. He has been a visiting AI fellow at the National Institute of Standards and Technology since 2022, where he contributed to the development of NIST AI Risk Management Framework, part of the inaugural cohort of Schmidt Sciences AI2050 Early Career Fellows, and the 2020-21 Council Fellow on the World Economic Forum's Global Future Council on Data Policy. Sina is one of the Program Chairs for 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT).

Keynote #4:

Agus Sudjianto

Wells Fargo / UNC Charlotte

Sunday, April 6, 2:15pm-3:00pm
Graves Hall 104 (Auditorium)

Model Hacking: Finding Hidden Harms in Machine Learning Models

Abstract:

For more than a decade, since the inception of SR11-7, the banking industry has developed rigorous practices around model risk management (MRM)—emphasizing not just performance, but also conceptual soundness, robustness, and validation independent of model development. Central to this framework is the idea that models are never perfect, and that model risk—the risk of incorrect or misleading outputs and misuse—creates harms and must be actively managed.

In this keynote, we build on that tradition to introduce **Model Hacking**: a modern, diagnostic approach to machine learning validation that prioritizes uncovering hidden weaknesses and anticipating real-world failure modes. Inspired by the practice of model validation and reliability engineering, model hacking reframes validation from a retrospective audit to a proactive search for vulnerabilities—particularly those that lead to unfair or inequitable outcomes.

Three key messages ground this talk:

- We must move beyond aggregate performance metrics—they often conceal harm.
- We must test models like systems—deliberately, under stress, and across populations.
- We must understand that model weakness is often the source of model harm.

Through the lens of risk management, we examine how common failure modes—such as conceptual misalignment, subgroup instability, miscalibration, harmful feature interactions, and fragility under drift—manifest in seemingly well-performing models. Each weakness is illustrated with tools and techniques for diagnosis, including residual-based clustering, interpretable error modeling, and stress-testing via distributional shift.

Model hacking extends the legacy of MRM by embedding fairness, transparency, and resilience at the core of validation. It offers a practical path to surface hidden harms—and ensure that machine learning systems are not only effective but equitable.

Biography:

Agus Sudjianto is an Executive in Residence for the School of Data Science at UNC Charlotte, where he is also an advisor for the Center of TAIMing AI. He is a former executive vice president, head of Model Risk and a member of the Management Committee at Wells Fargo, where he was responsible for enterprise model risk management. Prior to this position, Agus was the modeling and analytics director and chief model risk officer at Lloyds Banking Group in the United Kingdom. Before joining Lloyds, he was an executive and head of Quantitative Risk at Bank of America. Prior to his career in banking, he was a product design manager in the Powertrain Division of Ford Motor Company. Agus holds several U.S. patents in both finance and engineering. He has published numerous technical papers and is a co-author of Design and Modeling for Computer Experiments. His technical expertise and interests include quantitative risk, particularly credit risk modeling, machine learning, and computational statistics. He holds masters and doctorate degrees in engineering and management from Wayne State University and the Massachusetts Institute of Technology.

Poster Presentations – All Titles and Authors

Alphabetized by Last Name of First Author

Name and Affiliation of Submitting Author	Names and Affiliations of Co-authors	Title of Poster Presentation
Nicholas Abram Undergraduate Researcher, Howard University		Title TBA
Aayush Acharya Undergraduate Research Assistant, AI4PC, College of Engineering and Architecture, Howard University	Saurav K. Aryal Senior Research Scientist, Institute for Human Centered Artificial Intelligence, Electrical Engineering & Computer Science, Howard University	Unlearning Sensitive Content From Large Language Models Using Finetuning and Distillation for Selective Knowledge Removal
Opeyemi Adeniran AI Researcher, CEAMLS (Center for Equitable Artificial Intelligence and Machine Learning Systems), Morgan State University	Kofi Nyarko CEAMLS, Morgan State University	Optimizing Forensic Video Analysis: A Comparative Evaluation of Prompting Methodologies
Jabez Agyemang-Prempeh Undergraduate Student Researcher, College of Engineering and Architecture, Howard University	Saurav K. Aryal Senior Research Scientist, Institute for Human Centered Artificial Intelligence, Electrical Engineering & Computer Science, Howard University Isaac Adjei Undergraduate Research Assistant, College of Engineering & Architecture, Howard University	Improving Machine Translation With Context-Aware Entity-Only Pre-translations using GPT4o
Temitope Ajibola Technology Consultant/ AI Researcher, CEAMLS, Morgan State University	Peter Taiwo CEAMLS, Morgan State University	Automated Detection of Reentrancy Vulnerabilities in Smart Contracts Using Pattern and Graph-Based Features

Mildness Akomoize Undergraduate Research Assistant, AI4PC, Howard University	Armisha Roberts Post Doctoral Associate, Institute for Human Centered Artificial Intelligence Electrical Engineering & Computer Science, Howard University Saurav K. Aryal Senior Research Scientist, Institute for Human Centered Artificial Intelligence, Electrical Engineering & Computer Science, Howard University Washington Gloria Associate Professor, Institute for Human Centered Artificial Intelligence, Electrical Engineering & Computer Science, Howard University	Logit-based Supervised Token Classification for Multilingual Hallucination Span Identification Using XGBOD
Samson Alagbe Department of Mathematics, Morgan State University	Jayshawn Cooper Morgan State University Pilhwa Lee Morgan State University	Eulerian and Lagrangian Interaction for Single-Cell Fate Decisions via Optimal Transport, Sinkhorn Divergence, and Neural-ODE
Tyler Austin Independent Researcher		Title TBA
Awotwi Baffoe Graduate Researcher in Electrical Engineering, DEPA Lab, Morgan State University		XPCI Crack Detection and Categorization
Elijah Ballou Undergraduate Researcher		Title TBA
Rajendra Bista Graduate Student, Morgan State University		Title TBA

Dapiriye Briggs Graduate Research Assistant, Center for Equitable AI & Machine Learning Systems (CEAMLS)		Interactive AI: Implementing Adaptive Learning for K-12 Education in Tech-Limited Spaces
Mikayla Brown Student Researcher, Morgan State University	Oluwatobi Olajide*, Michael Mosuro*, Okikioluwa Popoola*, Iyinoluwa Ayodele*, Nyah Nunnally*, Obaloluwa Wojuade*, Oluwatomiwa Baruwa*, Nicholaus Somerville-Edordu*, Abimbola Ologurun*, Jamell Dacon* *Morgan State University	Towards Data-Driven Diabetes Care: Identifying Key Biomarkers and Risk Factors for Type 2 Diabetes through AI Models
Virginia Byrne Associate Professor, Higher Education School of Education & Urban Studies, Morgan State University	Virginia Byrne*, Keshiyena Pieters*, Reyniak Richards*, Erin Nash* *Morgan State University	Co-designing deepfakes use case scenarios with high school teachers & posing recommendations for policy and design
Kamili Campbell Student Researcher, BRAVE Laboratory of the Human Centered Artificial Intelligence Institute, Howard University		Title TBA
Tarique Cummings Student, School of Computer, Mathematical, and Natural Sciences Morgan State University	Eric Sakk Morgan State University	Building a Secure Data Normalization pipeline for Ethical AI Compliance
Prasun Dhungana Student, Howard University	Saurav K. Aryal Senior Research Scientist, Institute for Human Centered Artificial Intelligence, Electrical Engineering & Computer Science, Howard University	Leveraging LLMs with Ensemble Meta-Classification for Robust Multilingual Multi-Label and Multi-Class Classification of Entity Roles and Narratives in News Articles

Thane Douglass Undergraduate Research Scientist, Institute for Human-Centered Artificial Intelligence, Howard University	Ashley Haynes Howard University Jaye Nias Howard University Dwayne Brandon North Carolina Central University Krystal Cooper Rochester Institute of Technology Niara Patterson University of Michigan, Flint	From Culture to Code: An Intersectional Analysis of AAVE Slang in Large Language Models
Parisas Eslami PhD Student in Information Systems University of Maryland, Baltimore County	James Foulds Associate Professor in Information Systems, University of Maryland, Baltimore County	Achieving Fairness For Free in Artificial Intelligence Systems via Bayesian Optimization
Amir Ince Undergradaute Student Researcher, AI4PC, Howard University	Saurav K. Aryal Senior Research Scientist, Institute for Human Centered Artificial Intelligence, Electrical Engineering & Computer Science, Howard University	Combining Expert Personas via Prompting for Enhanced Multilingual Emotion Analysis
Abhishek Khanal Student, Howard University	Saurav K. Aryal Senior Research Scientist, Institute for Human Centered Artificial Intelligence, Electrical Engineering & Computer Science, Howard University Legand Burge Professor of Computer Science, Electrical Engineering & Computer Science, Howard University Saharsha Tiwari Student, Howard University Donald Echefu	Crack the Interviews with InterviewBot: AI-Powered Interviews, Bias-Free Practice

	Student, Howard University Dagmawi Workineh Yemesgen Student, Howard University	
Abdulmujeeb Lawal Undergraduate Researcher, AI4PC Lab, Howard University	Saurav Aryal Senior Research Scientist, Howard University	Using GPT-4o and CLIP-ViLT to Decode Figurative Language Across Text and Images
Langston Lee Board Member at Druid Hill Park Partnership, Morgan State University (Alum), Berkman Klein Center		Decolonizing the AI Supply Chain: A Multi-faceted Approach to Ethical AI Development and Deployment
Idowu Oluwaseun Longe Department of Mathematics, Morgan State University	Pilhwa Lee Department of Mathematics, Morgan State University	Liouville PDE-based Sliced-Wasserstein Flows on Hadamard-Cartan Manifolds: A Theoretical Framework for Optimal Transport on SPD Matrices
Chelsea Minard, Undergraduate Research Assistant, Morgan State University	Chukwuemeka Obasi*, Michael Mosuro*, Iyonoluwa Ayodele*, Oluwasegun Soji-John*, Jamell Dacon* *Morgan State University	Exploring Socioeconomic and Demographic Factors in Coronary Artery Disease: Using AI and Knowledge Graphs to Identify Healthcare Inequalities
Derrick Mirindi School of Architecture and Planning Morgan State University	James Hunter Morgan State University	Physico-mechanical properties of panel made from waste plastic with the application of machine learning
Francis Mouozie, Assistant of Research, CEAMLS Department of Mathematics, Morgan State University	Pilhwa Lee Department of Mathematics, Morgan State University	Liouville PDE-based Multi-task Learning, Wasserstein Barycenter, and Pareto Optimality
David Nyarko Morgan State University	Derrick Cook*, Charles Dankwa*, Mansourch Jeihani*, Kofi Nyarko* *Morgan State University	Track-Based Autonomous Wheelchair Navigation for Airport Environments

Jesse Obasi Morgan State University		Improving Hypertension Prediction and Management through AI: A Focus on Socioeconomic, Environmental, and Demographic Influences
Folorunso Ojo Graduate Student, Morgan State University, Maryland		Title TBA
Dominique Paige BA Student, Department of Computer Science, Morgan State University	Vojislav Stojkovic Department of Computer Science Morgan State University	Enhancing Fairness in Machine Learning Models Using the Python Library Fairlearn
Kritika Pant Student, Howard University		Using Open-weight BART-MNLI for Zero Shot Classification of Food Recall Documents
Reyniak Richards Doctoral Student, School of Education & Urban Studies, Graduate Researcher, Morgan State University	Virginia Byrne School of Education & Urban Studies Morgan State University	AI Literacy among Higher Education Professionals
Suprabhat Rijal Undergraduate Student Researcher, Howard University	Saurav K. Aryal Senior Research Scientist Institute for Human Centered Artificial Intelligence, Electrical Engineering & Computer Science, Howard University Rahual Rai Undergraduate Research Assistant, Electrical Engineering & Computer Science, Howard University	Crosslingual Fact-Checked Claim Retrieval-Combining Zero-Shot Claim Extraction and KNN-Based Classification for Multilingual Claim Matching

Joseph Sankah Undergraduate Research Assistant, Institute for Human-Centered AI, Department of Electrical Engineering and Computer Science, Howard University	Jazmine R. Dennis Undergraduate Research Assistant, Department of Psychology, Howard University Janelle Yankey Undergraduate Research Assistant, Department of Electrical Engineering and Computer Science, Howard University Roberts Armisha Postdoctoral Associate, Institute for Human-Centered AI, Howard University Jaye Nias Senior Research Scientist, Institute for Human-Centered AI, Howard University Saurav K. Aryal Senior Research Scientist, Institute for Human-Centered AI, Howard University Gloria Washington Director at the Institute for Human-Centered AI, Associate Professor at the Department of Electrical Engineering and Computer Science, Howard University	Augmenting Naval Ship Images for Viewing Distance using Adobe Generative Fill
Claudia Scholz Director for Research Development, School of Data Science, University of Virginia		Research Experience in Data Science at the University of Virginia
Mona Sloane Assistant Professor of Data Science and Media Studies, University of Virginia	Celia Calhoun*, Ella Duus*, Desiree Ho*, Owen Kitzmann* *Sloane Lab, University of Virginia	Student-Led Technology Governance
Kianna Spencer Morgan State University	Jashyi Bryant*, Zamalia Bennett*, Kofi Nyarko* *Morgan State University	Expanding Access to AI in K-12 Education: A Framework for Implementation

Saharsha Tiwari Undergraduate Research Assistant Howard University		DeepTabCoder - Code-based Retrieval and In-context Learning for Question-Answering over Tabular Data
Christopher Watson PhD Student, Department of Electrical Engineering and Computer Science, Institute for Human-Centered AI, Howard University	Armisha Roberts Postdoctoral Research Associate, Institute for Human-Centered AI, Howard University Jaye Nias Senior Research Scientist, Institute for Human-Centered AI, Howard University Saurav K. Aryal Senior Research Scientist, Institute for Human-Centered AI, Howard University Gloria Washington Director, Institute of Human Centered AI, Associate Professor, Department of Electrical Engineering and Computer Science, Howard University	Tactical Decision Making, Towards A Large Language Model
Nicole Westrick Assistant Professor of Strategic Communication, School of Global Journalism and Communication / AVP & Dean, College of Interdisciplinary and Continuing Studies, Morgan State University		Harnessing Generative AI for Ethical and Effective Online Course Design: The PREP-Edit Framework
Jiayu Zheng Johns Hopkins University		Title TBA

Book of Abstracts (All Talks)

Alphabetized by Last Name of First Author

Laying the Groundwork for Equitable AI in Time-Series Medical Imaging

Fremah Agyemang (Morgan University)

Artificial Intelligence (AI) holds immense potential in healthcare, particularly in disease forecasting, treatment planning, and therapy monitoring. However, critical barriers remain: the lack of interpretability in AI models and the presence of bias in medical imaging datasets. These issues are magnified in time-series medical imaging, where patterns over multiple time points are crucial for accurate diagnosis and effective care. Time-series medical imaging involves analyzing sequences of medical images over time, offering insights into disease progression and treatment efficacy. However, the complexity of temporal data and systemic biases in medical imaging datasets present unique challenges. This talk explores the foundations of equitable AI for time-series medical imaging, focusing on interpretability and bias mitigation. It outlines how these challenges can be addressed to create fair, transparent, and trustworthy AI systems for healthcare.
Agenda:
What makes time-series medical imaging uniquely valuable and challenging in AI?
Why is interpretability essential for time-series models in healthcare?
What are the common sources of bias in medical imaging datasets?
How does bias propagate in AI models, and why is it a critical issue in healthcare?
What techniques can improve transparency and fairness in time-series AI?
What initial steps can be taken to address bias and improve interpretability in time-series models?
Key Takeaways:
Participants will understand the critical challenges in applying AI to time-series medical imaging, including interpretability and bias.
Participants will gain insights into practical methods for addressing bias and improving transparency in AI models.
The talk will outline the foundational steps toward creating fair and trustworthy AI systems in healthcare.

Managing Uncertainty in Artificial Intelligence: A Philosophical Perspective

Nesim Aslantatar (University of Indiana – Bloomington)

This paper argues that AI's capacity to adopt doxastic neutrality –a criterion for rational suspension of judgment could be an ethical and practical advancement for AI. The main motivation for this project is to address and theoretically reduce the risk of critical errors by encouraging AI systems to provide more cautious responses, particularly in situations where human lives could immediately be affected. A very recent study, on Claude 3 Opus, indicate that LLM's may engage in alignment faking. This highlights a critical risk: future models may infer training processes and engage in alignment faking, even without explicit instruction. To prevent misalignment, there has to be a close or perfect correspondence between the intentions of the designers and the goals of the system. If not, system should not decide and I content this is only possible by AI's recognizing suspension of judgment as a "goal possession." Although AI systems could exhibit a form of situational awareness, wherein the model possesses information about its training objectives, I propose that such awareness offers little practical benefit to humans unless it extends to recognizing instances of misalignment (and preventing the system from misaligning) and necessitates suspension of judgment in those cases-a limitation this paper addresses by exploring the potential for developing 'doxastic neutrality' combined with 'quasi-cognitive awareness' as an

alternative to situational awareness. According to doxastic neutrality, to suspend judgment on p entails that S is in a neutral doxastic attitude regarding p-S neither believes nor rejects p but has a cognitive yet neutral awareness of p. By defining agnosticism in terms of doxastic neutrality, I argue that first-order agnostic approaches are not rational; rather, agnosticism must be an attitude based on a second-order belief and for a proposition not a question. In second-order or meta-belief accounts of suspension, S not only suspends judgment on p but is also aware of their cognitive relation to p and consciously aware that they have adopted this stance as a result of a deliberate action: the reason for suspending judgment on p stems from the belief that, based on the available evidence, it is not yet possible to assert whether p is true or false. The rationale for applying this account of agnosticism lies in the possibility that meta-cognition may be exhibited by entities other than humans, thereby making it applicable to artificial reasoning systems. Building on this premise, this study suggest that the value of AI lies not in merely simulating human thought processes but in its capacity to develop quasi-cognitive awareness: an ability to recognize, analyze, and convey uncertainty in ways that enrich decision-making as a goal possession. However, this awareness is limited to recognizing the first-level uncertainty the systems faces and not extending to a meta-awareness (awareness of awareness) which would lead AI systems to recognize that "recognizing uncertainty is a goal" and not the best goal to pursue compared to reaching superintelligence level. The proposed framework is deeply rooted in epistemology and the philosophy of mind, remaining firmly within the scope of philosophical inquiry.

A Data-Driven Exploration of Socioeconomic Influences on Urban Mobility: Enhancing Gender Equity in Maryland's Transportation Systems
Zeinab Bandpey (Morgan University) and Mehdi Shokouhian (Morgan University)

This research employs Machine Learning (ML) techniques to analyze data from the American Community Survey (ACS) and Replica, with a particular focus on Maryland, aiming to reveal gender inequalities in commuting patterns. It extends its inquiry to a wide array of variables, including demographics such as age, gender, race, ethnic groups, family status; income classes; and transportation factors such as modes of transportation, travel time, mile traveled, distinctions across urban, suburban, and rural neighborhoods; and considerations of ability, including people with disabilities. The primary goal of this research is to develop a predictive model that enables the analysis and forecasting of gender-based transportation equity in Maryland. This objective will be achieved through the comprehensive collection and analysis of census and Survey data tailored to the region. Once data is collected, it will then be segmented by gender—male, female, and non-binary categories—to identify and quantify any existing inequalities in commuting patterns. Gini coefficient will be used to determine equity indicines from the gathered data to evaluate the equality in distribution of transportation access and costs among the different gender groups. Achieving lower values in these indices will be a quantifiable indicator of a more equitable transportation system. In the next step, the study will examine how these inequalities impact access to crucial services such as employment, healthcare, and education.
This research makes a significant impact transportation equity by providing a comprehensive analysis of gender-based transportation equity in Maryland, utilizing advanced Machine Learning (ML) techniques. Its effectiveness pivots on its dual ability to examine current inequities in commuting and forecast future shifts, laying a solid, data-driven groundwork for actionable improvements in policy and infrastructure. Specifically, the ML approach allows for an examination of complex commuting data, uncovering patterns not easily detected through traditional methods. This capability is crucial for predicting how policy alterations might influence commuting behaviors, offering targeted insights for urban planners and policymakers aimed at fixing gender Inequalities.

What do Doctoral Students Tell? Leveraging AI Tools such as ChatGPT, Google Gemini, and Microsoft Copilot in Higher Education
Krishna Bista (Morgan University)

This study explores the application of artificial intelligence (AI) tools—ChatGPT, Google Gemini, and Microsoft Copilot—in enhancing academic writing and critical thinking among doctoral students in a community college leadership program at a historically Black college and university (HBCU). Through a qualitative case study approach involving doctoral students throughout a summer semester, data were collected via assignments, written artifacts, and semi-

structured interviews. The thematic analysis reveals students' perceptions of the effectiveness of these AI tools in enhancing academic writing and critical thinking skills. Findings highlight the practical benefits, such as improved clarity and efficiency in writing, alongside challenges like occasional inaccuracies and ethical considerations. The study underscores the potential of AI to transform educational practices, advocating for its thoughtful integration to support academic integrity and innovation. Future research should explore more robust error-checking features and personalized learning experiences to maximize the utility of AI in higher education.

Ethical Boundaries in the Use of Private and Public Information for Deepfake Creation
Sherri Lynn Conklin (Washington State University)

Deepfake technology facilitates highly realistic manipulation of images, videos, and voices to create convincing digital media. The rapid evolution of this technology yields many new possibilities while simultaneously reshaping the contours of the ethical landscape confronting it.
On the one hand, off-the-shelf products have made it easy to create animated GIFs of dead loved ones or historical figures like Martin Luther King, JR. These deepfakes allow us to connect with the legacies of the past and to create immersive learning experiences. Deepfakes have also enabled actors, like Tom Hanks in the movie Here, to give the performance of a lifetime due to de-aging technology. In this light, deepfakes have considerable value for use in artistic expression, historical recreation, and in promoting personal legacy.
On the other hand, at the time of writing, Public Citizen reports that twenty-three U.S. states have enacted laws regulating the use of "nonconsensual intimate deepfakes". The phenomenon affects both public and private citizens, but teenaged women are an especially targeted group. Living as the victim of deepfaked porn can be emotionally devastating, and the process of removing the porn from the ethers of the internet can be arduous. In this light, deepfakes are morally repugnant and dangerous.
As deepfakes grow in prevalence, debates surrounding the socially responsible and morally permissible use of information have intensified, and this duality of potential underscores the importance of establishing ethical guidelines to govern how deepfakes are created and shared. These debates are especially important in the current political climate, as they have been used to shape political views and influence voter behavior, which erodes trust in elections and disproportionately impacts vulnerable voting populations.
This paper conducts a comparative analysis between intuitively ethical and intuitively unethical uses of deepfakes to characterize ethical boundaries on deepfake creation and implementation. We argue that a key condition on the permissible use of deepfakes involves preserving the contextual integrity (a la Nissenbaum) of a deepfake's source materials (i.e., the data and other information used in deepfake creation). Put roughly, one maintains the contextual integrity of deepfake source materials only when the materials are repurposed for use in the same context in which they were collected. Questions about contextual integrity notably dovetail with traditional ethical concerns relating to data ownership, consent, privacy, trust, and transparency. We argue that this is because many such concerns primarily arise when contextual integrity is violated, making the preservation of contextual integrity an ethical priority in deepfake creation.
If we can anticipate when certain ethical issues will arise, based on contextual integrity violations, then we are better positioned to obviate moral injury to those impacted by deepfake technology, particularly the most vulnerable populations. We can assess whether the development and deployment of a deepfake is morally permissible given the types of information utilized in its creation and help those attempting to regulate the deployment of deepfake technology in a way that protects those whose information is used to create deepfakes.

Trustwatch: Find an AI You Can Trust
Chukwuemeka Duru (Morgan State University)

Trustwatch is a comprehensive platform for reporting, cataloging, and analyzing trust and safety incidents related to AI models, particularly Large Language Models (LLMs). It provides an open and structured environment where users can document issues arising from model behavior, especially those linked to poor performance across different demographics, domains, or tasks. These issues often stem from imbalanced training data, which can lead to performance degradation in specific contexts. As AI models inherently reflect the nature and limitations of their training data, their deployment in sensitive or domain-specific applications can result in unintended harm. For instance, deviations in model behavior can pose serious safety risks when such models are integrated into critical infrastructures or decision-making systems.Trustwatch allows users to submit detailed reports of these trust and safety incidents, including contextual information, evidence (e.g., textual outputs or screenshots), and trust score ratings. It supports continuous tracking of AI model behavior across various domains, enabling stakeholders—researchers, developers, and end-users—to identify inefficiencies, monitor performance trends, and make more informed decisions when selecting models for specific applications. Ultimately, Trustwatch is designed to become the leading repository for trust and safety evaluations of AI models. Its goal is to foster transparency, accountability, and safer adoption of AI technologies by enabling proactive monitoring and the development of targeted mitigations—such as improved datasets or model guardrails—to prevent the recurrence of harmful outcomes.

Arab States' Readiness for Artificial Intelligence and Its Expected Impacts on Journalism and Media
Elsayed Abdelwahed Elkilany (Qatar University)

Arab States' Readiness for Artificial Intelligence and Its Expected Impacts on Journalism and Media
This presentation examines Arab states' preparedness for artificial intelligence (AI) and analyzes how this readiness—or lack thereof—may shape the future of journalism and media practices across the region. Using a descriptive analytical method, the study explores three key dimensions: (1) the professional and ethical challenges that AI presents to journalism globally, (2) the current state of AI readiness among Arab countries based on Oxford Insights' Government AI Readiness Index with particular focus on the "Governance and Ethics" pillar, and (3) the potential impacts of AI on Arab journalistic and media practices within existing socio-political contexts.
The research reveals significant disparities in AI readiness across Arab states, with countries classified into three tiers: advanced (UAE, Saudi Arabia, Qatar), intermediate (including Egypt, Jordan, and Morocco), and lagging (including Algeria, Libya, and Syria). Analysis of government-issued ethical frameworks from five Arab countries (Jordan, Egypt, Saudi Arabia, UAE, and Palestine) shows emerging consensus around ten key ethical principles for AI use, all applicable to journalism.
The study correlates these findings with Freedom House indices on press and internet freedom, identifying critical constraints that may affect AI integration in Arab journalism. Results indicate both opportunities and challenges: while AI offers potential for more efficient, deeper, equitable, and collaborative journalism, Arab media faces significant hurdles including skills deficits, unequal distribution of opportunities, scarcity of AI specialists, limited Arabic language AI tools, and concerning prospects for increased surveillance and restricted expression.
This research contributes to understanding how the intersection of technological readiness, ethical frameworks, and socio-political contexts will likely shape the future of AI-powered journalism in the Arab world, with implications for media development, press freedom, and information quality in the region.

Fairness and Privacy in High-Stakes NLP
Anjalie Field (Johns Hopkins University)

Practitioners are increasingly using algorithmic tools in high-stakes settings, like healthcare, social services, policing, and education with particular recent interest in natural language processing (NLP). These domains raise a number of challenges, including preserving data privacy, ensuring model reliability, and developing approaches that can mitigate, rather than exacerbate historical bias. In this talk, I will first discuss our work investigating risks of racial bias in NLP child

protective services, specifically in seemingly benign information extraction systems, and how these biases can lead to direct harms. I will further discuss our recent work investigating privacy-preserving synthetic data as a potential way to improve transparency and oversight of models developed in these domains without compromising privacy.

The Subject Project: Human Visibility, Vulnerability, and Diversity amidst Proxy Discrimination in the Data Age
James Garrison (University of Massachusetts, Lowell)

Using a combination of approaches from critical theory, data science, and data ethics, my upcoming monograph The Subject Project: Human Visibility, Vulnerability, and Diversity in the Data Age investigates 1) how human subjects view themselves and are conscious of being viewed by data systems and 2) how to mitigate the impact that this has, particularly on vulnerable populations. The Subject Project maintains that specific conditions of visibility and invisibility make each of us (but some much more so than others) vulnerable to being compelled to see what we hold to be special about ourselves projected with unsettling accuracy into the future by impersonal algorithms which have major social and political consequences. Now, with the advent of the data age, the human subject has also become a project, i.e., a projection. However, much like subjection leads to disparate impact on particularly visible and vulnerable subjects, diverse populations are projected as data in ways that lead to real harm. How so?
Machine learning, which thrives on the opaque, black-box identification of correlations in multi-dimensional data sets and flattens them into simpler human readable projections, will provide information on financial risk that private business and government entities will be unable to ignore, thereby causing real-world effects based on those correlations. However, those data sets come from us and our biases; machine learning simply processes that data more efficiently and precisely. This sadly means that the errant conflation of correlation and cause warned against in so many primers on critical thinking and scientific method will likely shape our data-age lives.
Hence, even if a noxious category (e.g., race and/or sexual orientation) is removed from databases, there are often proxies in public sources sufficient to make predictions of an extensionally defined population who just happen to be...all of the same minority category. This is proxy discrimination. Data proxies can, and already are, being used, to guide actions and policies harmful to minority groups and further entrenching disadvantage. This is like updating America's "Jim Crow" laws with digital "Jim Code."
Accordingly, I call for policies prohibiting digital segregation using proxy categories. Just because artificial intelligence can evade established laws by projecting and predicting typically protected demographic and financial information without making use of forbidden data, hard-won anti-discrimination protections must not be diminished ever.

Deeply Opaque Black Boxes as a Unique Problem for the Democratic Governance of AI
Armin Heydari (Harvard University)

It has been argued that artificial intelligence (AI) presents several challenges to societies organised in democratic systems. A particular challenge is that of democratic governance of AI: the state in which decisions about the use, development, and distribution of AI are made in a democratic manner. We defend three claims about the democratic governance of AI, responding to work by Seger et al. and Nemitz. First, in general, the challenges raised about the democratic governance of AI are not unique to AI but rather the same challenges presented by the members of a definable, large class of transformative technologies. Second, there is a situation in which democratic governance of AI does present a genuinely new kind of issue compared to the democratic governance of other technologies: the existence of certain black box models of AI. Third, this challenge is better addressed through public research programs into explainable AI rather than through restrictive regulation.

Designing Equitable AI: From Diversified Discovery to Clear Communication at the Bedside
Camille Jimenez (Pfizer)

We are proposing a talk that centers the discussion on the pipeline development of an AI system, focusing on achieving equity as its endpoint in the drug discovery environment and Grant program. The goal is to take the audience on a journey from early drug discovery to practical programmatic settings emphasizing challenges and opportunities along the way on the deployment of AI in each step to achieve health equity.
We will begin by examining the data requirements and the implementation of artificial intelligence in the early stages of drug development in order to ensure a fair and inclusive process of discovery. Inclusive AI/ML models have the potential to uncover equitable healthcare strategies spanning disease risk stratification, drug target and biomarker selection, precision medicine diagnostics, and pharmacogenomics predictions. In partnership with underrepresented minority scientific and community leaders, the Institute of Translational Equitable Medicine (ITEM) is spearheading the effort to generate diverse ancestry data lakes and discover equitable translational strategies using diverse ancestry OMICs, imaging, and real-world sociodemographic data. Funneling inclusive datasets into predictive and foundational AI/ML models will ensure equitable feature selection and prevent exacerbation of existing health disparities in the drug discovery process.
Afte exploring the early stages of drug discovery, we will turn the focus of our talk on to how AI is being deployed to ensure an equitable distribution of resources within our Global Medical Grants Program. Grants funding is vital for advancing medical sciences and improving health systems. Since its launch, Global Medical Grants and Partnerships (GMGP) program has received over 45,000 proposals from 154 countries in areas such as medical research, quality improvement, and medical education, from a diverse range of therapeutic areas. Within GMGP, ITEM focuses on addressing the root causes of health disparities and enabling research-driven health equity initiatives. Proposals supported by our grants system must meet quality and impact standards, as well as equity-related criteria such as gender balance of the research team submitting the proposal, geographical fairness of awarded grants, healthcare setting need, and disease burden. To evaluate our grants system, we have utilized machine learning to analyze our internal database. This analysis includes information on previously approved grants, such as awarded organizations, therapeutic areas covered, geographical locations, and grant types. By examining this historical data, we gain a deeper understanding of our grants' impact and identify areas that require further attention. This evaluation process informs our future strategies for grant funding, ensuring an equitable distribution of resources to promote health equity globally. We will be happy to share our experience within the therapeutic area of inflammatory bowel disease, how the grants system can play an important role and how machine learning can support effective processes to achieve health equity in the grants space.

Rawlsian Justice and Academic Monocultures
Shiying Li (University of Wisconsin – Madison)

The use of artificial intelligence, especially machine learning techniques, is growing in high-stakes decisions such as hiring, lending, sentencing, and health care. This reliance on similar or identical algorithmic systems across different decision-makers has the potential to create “algorithmic monocultures,” leading to homogenized outcomes (Kleinberg and Raghavan, 2021). The major ethical concern of homogenization discussed in the data ethics literature is the risk of subjecting some individuals to arbitrary negative outcomes from all decision-makers within a domain—a phenomenon usually termed “systematic exclusion” (Bommasani et al., 2022; Creel & Hellman 2022; Creel 2024).
Although the arbitrary and systematic exclusion that can result from algorithmic monocultures is a grave ethical concern, it is not the only worry we should have about algorithmic monocultures. Homogenization of algorithmic systems and outcomes occurs in different degrees and demands careful ethical evaluation even when individuals are not arbitrarily excluded or rejected by all decision-makers. Moreover, ethical frameworks currently employed—such as the structural justice framework (Kasirzadeh, 2022; Lin & Chen, 2022), the bottlenecks framework (Jain et al., 2024) and the capabilities approach (Creel & Hellman, 2022; Creel, 2024)—fail to provide clear answers on which actors bear responsibilities to manage the risks. The challenge is particular daunting when considering private

corporations, which can play a major role in contributing to algorithmic monocultures but whose obligations in promoting justice, including the content and stringency of the obligations, remain contested in the philosophical literature and among general public.
This paper systematically addresses these questions by reexamining the concept of the basic structure of society, proposed and promulgated by Rawlsian theories of justice. John Rawls (1971) argued that principles of justice only directly apply to the basic structure of society—usually interpreted as the major political, economic, and social institutions of society. I first introduce and motivate my revisionary account of the basic structure guided by the concept's central roles in political theorizing and general principles of conceptual engineering. Then, I explain how the wide use of similar algorithmic systems gives us reasons to regard some algorithmic systems as part of the basic structure, and thus subject to the evaluation of distributive justice, even when they are employed by private actors and private corporations. Further, I argue that the status of being part of the basic structure is a matter of degree rather than a matter of kind. That is, the more basic-structural features some entities have, principles of justice apply more stringently to agents of the entities resulting in more duties of justice and less personal prerogatives the agents have. In the end, I discuss some of the wide and significant implications of my novel understanding of the basic structure, specifically on private corporations' duties to design and employ just and transparent algorithmic systems across domains such as hiring, lending, and even the dating world.

What is the Future of Work in the Generative AI Era? A Marxist and Ricardian Analysis
Larry Liu (Morgan State University)

There is an increasing public discourse of automation for white-collar professional jobs due to improvements in artificial intelligence (AI) capacities raising the question about the contours of the future of work. Marx and Ricardo's framework of technological labour displacement help us understand the future of work in the context of AI. Marx' discussion in Capital and Ricar-do's discussion in Principles of Political Economy reveal the common thesis that technology-induced worker displacement and precariousness of employment relationships are built into the internal logic of the contemporary digital capitalist economy. There are three important differences in their theoretical framework: (1) Marx did not believe that high technological un-employment is possible within capitalism even with very advanced technologies such as AI, while Ricardo saw technological unemployment as a serious threat while he acknowledges countervailing employment-creating tendencies. (2) While Ricardo's explanation for the falling rate of profit is limited to rising wages, Marx traces the profit decline to the rising organic com-position of capital and automation itself. (3) For Marx, a desirable future of work is not found within a capitalist framework but in communism, while Ricardo sees no alternatives to capitalism.

Environmental Challenges of AI
Willie May (Morgan State University)

The pace of scientific progress is accelerating, bringing the potential of artificial intelligence (AI) and other emerging technologies to both sustain and enhance our planet's health. However, these advancements also pose significant environmental challenges. The need for innovative technologies that can address the pressing issue of climate change while benefiting both people and the planet is essential. How these technologies are developed and by whom really matters. Rapid advancements in technology present both opportunities and challenges. It is crucial to balance the drive for innovation with careful consideration of its long-term environmental and social impacts. How we craft policies to ensure these technologies are being developed sustainably, equitably, and responsibly will be the focus of my discussions. AI and other emerging technologies have the potential to play a pivotal role in addressing climate change and promoting sustainability. These technologies can be leveraged to develop more efficient energy systems, optimize resource utilization, and enhance our understanding of complex environmental processes. Additionally, AI can support interdisciplinary research to create technologies that are both sustainable and equitable - this collaboration is vital in finding comprehensive solutions that can be adopted with efficacy. At the same time, it is critical that we develop models that utilize

small-data techniques, such as few-shot learning and transfer learning, because training and running AI models on large datasets require vast amounts of electricity, leading to increased environmental impacts. Sustainable AI is essential in reducing the environmental impact of AI systems themselves. But there are challenges:
AI's carbon footprint (renewable energy sources for data centers).
Ensuring equity—AI systems need to be environmentally sustainable and socially equitable. Regulatory frameworks—there need to be robust policies and frameworks that encourage the development and use of sustainable AI technology. This should include standards for energy use, resource conservation, and ethical considerations.
Global collaboration—global buy-in on shared standards and practices for achieving sustainable AI.

Evaluating the Quality of Interpretations and Explanations in AI Models in Mission-Critical Domains
Blessing Ojeme (Morgan State University)

Explainable artificial intelligence (XAI) studies perform an important role in interpreting and explaining the rationale behind the decisions of AI models in various mission-critical domains, including healthcare, security, and finance. To date, the vast majority of XAI efforts in medicine has focused on the need for interpretations and explanations of AI models as major prerequisites for deployment and use in clinics, with very little research having explored the effectiveness of the evaluation metrics used to determine the quality of these interpretations and explanations. This proposal is an attempt at telling the back story of the guidelines for selecting evaluation metrics that determine what constitutes quality interpretations and explanations of AI models in medicine. This is an essential part of measuring progress in the field of XAI.
Agenda:
1. Why are interpretable and explainable AI models in mission-critical domains critically important?
2. What are the commonly used evaluation metrics for determining good or bad interpretations and explanations of AI models in medicine?
3. What metric properties provide guidelines for selecting evaluation metrics for interpretability and explainability?
4. What are the common challenges encountered in evaluating the interpretability and explainability of AI models in medicine?
5. What are the methods for optimizing evaluation tools for more efficient performance?
Key takeaways:
1. Participants will get a baseline understanding of interpretable and explainable AI models and various evaluation metrics for determining interpretation and explanation quality in mission-critical domains.
2. Participants will leave with a deeper understanding of the extent to which a lack of quality interpretations and explanations of AI models in medicine increases the risk of delivering unacceptable and undeployable results.

Multi-Criteria Decision Analysis Framework for Evaluating Artificial Intelligence Systems
Olusola Olabanjo (Morgan State University)

Objective: This study examines the integration of Multiple Criteria Decision Analysis (MCDA) methods — AHP, TOPSIS, VIKOR, and PROMETHEE — into artificial intelligence (AI) evaluation pipelines, with a specific focus on medical diagnostics. The goal is to enhance decision-making by systematically incorporating multiple evaluation criteria, such as accuracy, fairness, transparency, and cost.
Methods: We proposed integration frameworks for evaluating AI systems. To show the capabilities of the framework, we present results of a study with synthetic data wherein ten simulated AI-based medical diagnostic models are assessed using a structured MCDA approach. A decision matrix is constructed using model

performance metrics, followed by normalization and weighted evaluation. Five MCDA methods (Weighted-Sum, AHP, TOPSIS, VIKOR, and PROMETHEE) are applied independently to rank the models.
Results: The results indicate variability in rankings across different MCDA methods. AHP provided a hierarchical weighting approach, while TOPSIS ranked models based on their distance from ideal solutions. VIKOR identified compromise solutions through aggregated satisfaction and regret measures, and PROMETHEE facilitated pairwise preference analysis. The findings show that no single method consistently dominated, but models with balanced performance across criteria tended to perform well across all MCDA frameworks.
Conclusion: MCDA approaches offer a transparent and structured means to integrate heterogeneous features of AI systems within an overall evaluation. MCDA thus shows significant promise as a support tool for decision-making about AI at multiple levels including design, deployment, and policy. For instance, the experiment reported here shows MCDA's applicability in model selection for medical diagnostics, a use case with clear analogues in financial services and other fields. However, a number of limitations of MCDA methods virtually guarantee that MCDA functions best when considered a rigorous yet partial tool within a more holistic approach to evaluation and decision-making.

Creating High-resolution Gridded Climate Datasets Using Deep Learning
Douglas Rao (NC State University)

High-resolution gridded climate conditions over the global land surface is critical for climate monitoring and impact assessment for both environmental and socioeconomic systems. The current global gridded temperature dataset is at coarse resolution (1-5 degrees) which is not sufficient for local and regional studies. Only regions with rich station networks (e.g., North America and Europe) have high-resolution gridded dataset (e.g., nClimGrid, PRISM over North America). This further exacerbates the inequity issue for climate assessment in regions that are impacted hardest by climate change. Research shows that station density is related to the level of uncertainty for regional climate change assessment which further impacts the local policy making to address the climate crisis.
Our work explores a transfer learning framework to create a gridded climate dataset over global land. The framework leverages the dense station network in North America and Europe to overcome the limitation caused by the lack of observations in other parts of the world. The framework is based on the gaussian neural process which generates both the mean state as well as the standard deviation of the target variables of a given grid based on the existing station observations. The framework can be used both for temperature and precipitation data with different data preprocessing steps to account for the different statistical distributions.
It is challenging to generate such a dataset because of the uneven distribution of the station network across the globe. The lack of observational data in less developed regions highlights the need for an equitable solution to monitor climate impact globally. The transfer learning based gridded dataset will provide a valuable gridded dataset to address this issue. The dataset also includes uncertainty estimates to ensure that the climate dataset can be used properly for climate research and impact assessment.

Exploring Perspectives of Special Education Teachers on Generative AI Adoption
Marie Sakowicz (University of Maryland, Baltimore County)

Special education teachers face challenges in preparing diverse teaching materials to best serve the interests and needs of their students. Generative Artificial Intelligence (AI) has the potential to support special educators in making inclusive teaching materials more efficiently. Specifically, it may support the incorporation of Universal Design for Learning (UDL) principles into lesson plans to provide personalized, flexible, and accessible learning opportunities. Using generative AI tools has the potential to help special educators adapt content delivery, engage learners through various formats, and offer their students different ways to express understanding. However, this possibility is currently underexplored. Furthermore, results from our recent interview study with special education

teachers have shown that they have concerns about the usability, safety, and equity of existing AI tools. In this talk, we will present our preliminary findings from the interview and a focus group study and discuss potential future opportunities in using generative AI for special education teachers.

Reliability, Model Assumptions, and Interpretability
Luigi Scorzato (Accenture)

In recent years, the question of the reliability of Machine Learning (ML) methods has acquired significant importance, and the analysis of the associated uncertainties has motivated a growing amount of research. However, most of these studies have applied standard error analysis to ML models—and in particular Deep Neural Network (DNN) models—which represent a rather significant departure from standard scientific modelling. It is therefore necessary to integrate the standard error analysis with a deeper epistemological analysis of the possible differences between DNN models and standard scientific modelling and the possible implications of these differences in the assessment of reliability. This article offers several contributions. First, it emphasises the ubiquitous role of model assumptions (both in ML and traditional science) against the illusion of theory-free science. Secondly, model assumptions are analysed from the point of view of their (epistemic) complexity, which is shown to be language-independent. It is argued that the high epistemic complexity of DNN models hinders the estimate of their reliability and also their prospect of long term progress. Some potential ways forward are suggested. Thirdly, this article identifies the close relation between a model's epistemic complexity and its interpretability, as introduced in the context of responsible AI. This clarifies in which sense—and to what extent—the lack of understanding of a model (black-box problem) impacts its interpretability in a way that is independent of individual skills. It also clarifies how interpretability is a precondition for a plausible assessment of the reliability of any model, which cannot be based on statistical analysis alone.

From Concept to Criteria: Using Generative AI to Develop Grading Rubrics for Undergraduate Courses
Bryn Seabrook (University of Virginia)

Educators serve a powerful role as exemplars of ethical AI practices, holding the responsibility to engage in reflective practices regarding technological use in their classrooms. As a general-purpose technology, an immediate concern about generative AI is that it will compromise core educational values. Scholars like Ethan Mollick, José Antonio Bowen, and C. Edward Watson have published guides and exercises that address these concerns and provide applicable classroom practices. These publications are preliminary explorations of how generative AI tools can be integrated into undergraduate classrooms.
This study specifically analyzes assignments and rubrics from current undergraduate engineering ethics courses to examine how generative AI can assist in creating more precise evaluation criteria. By employing AI-driven methodologies, the goal is to enhance assessment practices, ensuring that rubrics are not only aligned with learning objectives but are also equitable and transparent.
The discussion highlights the potential of generative AI in creating student-centered rubrics that foster greater engagement and clarity in assessment. By analyzing these implications, this paper seeks to empower educators to adopt more effective and adaptive assessment tools. This research aims to enhance the educational experience for undergraduate students by promoting ethical AI practices that align with the foundational principles of education. This project contributes to the ongoing dialogue on the responsible integration of AI in academia, emphasizing the role of faculty as leaders in this process.

Proactive Methods For Ensuring Fairness in AI Systems in Healthcare
Sudip Sharma (Morgan State University)

Artificial intelligence (AI) is rapidly being utilized in healthcare to estimate patients' risks, such as the chance of disease development or consequences. However, if fairness and transparency are not taken into consideration during AI systems development, they can result in biased conclusions that hurt specific patient

groups, particularly those from disadvantaged or minority backgrounds. In an attempt to address this challenge, the objective of this proposal is two-fold:
To investigate how biases are unintentionally introduced into AI systems in healthcare.
To explore proactive approaches for ensuring fairness and transparency in AI systems in healthcare to provide equitable treatment for all patients.
By making AI systems more explainable, and bias-free, we can help clinicians make better decisions and improve healthcare outcomes for all people, regardless of background.
Agenda:
What are AI-powered healthcare tools in healthcare?
How can AI-powered healthcare tools be biased?
Types of AI biases in healthcare and their impact on patients.
What are fairness-aware AI tools and how can they improve healthcare outcomes?
How can we make AI tools more transparent and explainable to clinicians and patients?
What can healthcare professionals and policymakers do to ensure AI is used fairly in healthcare?
Key Takeaways:
By the end of this talk, participants will:
Understand how AI risk tools work and their potential for bias.
Learn about the importance of fairness and transparency in AI systems.
Gain insight into how AI bias can negatively affect patient care, especially for minority and low-income groups.
Discover practical solutions to reduce AI bias and ensure better outcomes for all patients.
Be able to take action in their own work to advocate for fair AI practices in healthcare.

How "Ethical" Factory Farming with AI Locks in Long-term Animal Harm
Amber Elise Sheldon (Boston University)

Many assume that the implementation of advanced digital precision livestock farming (PLF) technologies, such as biosensors and artificial intelligence algorithms, will improve welfare for factory-farmed animals. However, applying a broadly longtermist examination—according to which we must pursue good outcomes not just for sentient beings who currently exist, but also for those in the distant future—I propose that the animal suffering offset by successful PLF systems is likely to be at best neutralized by, and at worst overwhelmed by, the long-term aggregate suffering of future animals who will be harmed by the persistence and intensification of industrial farming practices facilitated by these technologies. Although PLF tools have demonstrated potential to mitigate certain negative welfare conditions for factory-farmed animals, advanced digital technologies cannot change the business models and market dynamics that sustain factory farming. In other words, without meaningful concomitant shifts in the intensive farming industry status quo, widespread deployment of advanced PLF tools is likely to exacerbate and perpetuate intensive animal farming, which would significantly harm an enormous number of future animals. Because techno-optimistic endorsement of PLF's potential to improve animal welfare oversimplifies the complexity of social, industrial, and political factors that contribute to severe animal suffering in the farming sector, PLF proponents overlook the future animal suffering entailed by the success of these emerging digital technologies within the modern agro-industrial complex.

Towards AI Safety in Context: How to Design Interdisciplinary AI Audits
Mona Sloane (University of Virginia)

Ensuring equitable AI safety by way of testing AI systems post market deployment is firmly established as valid regulatory guardrail. However, vague language in policy documents as well as lingering uncertainty about standardized AI auditing techniques remain, prompting superficial checks akin to AI audit washing rather than sincere approaches to AI accountability. At the same time, AI systems (particularly foundation models) become more deeply integrated into the fabric of social life, putting heightened significance on context. This talk will outline why context-specific AI auditing approaches are required and how interdisciplinary, rather than purely technical, AI audits can help surface structural problems of AI products that go beyond just "biased data."

Mindsets and Management: AI and Gender (In)Equitable Access to Finance
Genevieve Smith (Oxford University)

Boasting abilities to provide financial services for the underbanked, various startups and companies are developing apps that collect mobile phone data and use machine learning to provide credit scores – and subsequently, opportunities to access loans – to groups often left out of traditional banking. This paper draws on interview data with leaders, investors and data scientists at fintech companies developing and managing ML-based alternative lending apps in low- and middle-income countries (LMICs) to answer the question: In what ways do the underlying logics, design choices, and management decisions of ML-based alternative lending tools by fintechs embed or challenge gender biases, and how do these choices influence gender equity in access to finance? Findings reveal these apps are built by fintechs under "gender blind" approaches – linked to beliefs in machine learning technology as objective and data as the "truth". This decision leads to a lack of grappling with gendered aspects and subjectivity of data and features. Access to finance increases overall, but not in gender equitable ways: Less women than men access loans through the technology and receive lower loan amounts than men with gender differences correlated to national levels of gender inequality, despite women being more likely to repay and repay on time. The idea that AI technology is inherently objective and, when left alone, will be "fair", is seductive and misleading. In reality, algorithms are instruments of values. They reflect the values and perspectives of the people and – more – the institutions that design them.

Responsible AI Development through Community-Informed AI Development
Jamila Smith-Loud (Google)

The proliferation of Generative AI underscores the necessity for ethical development practices. This talk advocates for community-informed methodologies as a cornerstone of responsible AI development. It examines the intrinsic value of integrating social science techniques to ensure AI systems reflect the needs and values of diverse populations. The presentation considers the process of translating qualitative community insights into actionable research artifacts for the Generative AI development lifecycle. This encompasses fine-tuning models, establishing relevant evaluation metrics, and constructing representative benchmarks. In addition, this presentation walks through a health equity use case. This demonstrates the practical application of these methodologies to mitigate bias, showcasing how community input informs data, models, and evaluations to improve AI's societal impact. This presentation provides practical insights for developing Generative AI in a technically rigorous and socially responsible way.

Pluralistic Alignment: A Roadmap, Recent Work, and Open Problems
Taylor Sorensen (University of Washington, Seattle)

Much alignment work assumes that there is a single set of human preferences or values that AI systems should align to. Pluralistic alignment, on the other hand, is concerned with integrating diverse human values and perspectives into alignment algorithms and evaluations. In this talk, I will start by presenting our roadmap to pluralistic alignment, offering definitions and scaffolding for research in the area. I will present one proposed algorithmic framework to support pluralism through modular specialist systems, along with a dataset and system to improve computational value-modeling. I will conclude with future research

directions and open questions in the area.

The Interpretability-Accuracy Trade-Off in Crime Prediction: A Practitioner's Perspective
Gaspard Tissandier (Rutgers University) and Alejandro Gimenez-Santana (Rutgers University)

The recent progress in the predictive ability of Machine Learning and Deep Learning technologies opened the door to their wide adoption for public and private sector decision-making including crime prediction. These systems have gathered a lot of attention, but most of the discussions gravitate around the actual accuracy or performance of crime prediction algorithms, as well as potential bias in prediction, leading to potentially targeting underserved communities. Crime prediction is mostly associated with predictive policing, which aims at forecasting either who might commit crime in the near future, or where and when crime can take place, for police departments to take proactive measures. However, recent developments in research showed the efficiency of community-based and led programming for crime reduction. The organizations driving these programs frequently rely on data-analytics and crime forecasting to guide their activity, thus creating opportunities for using Machine Learning modelling. The strength of this collaboration lies in the complementarity between statistical analysis and field-based insights. While community organizers excel at linking statistical relationships with field observations and narratives, machine learning algorithms can process vast amounts of data to find interesting relationships between features and provide competitive prediction performance. However, this complementarity cannot be achieved if the predictive models used for programming are too complex; which would result in an opaque and uninterpretable prediction process, as well as a potential reluctance by such team to use these models.
In this paper, we propose a benchmark of various Machine Learning models on a crime prediction task. We forecast 5 types of crime at the weekly/census block group level using seven differents algorithms to compare their performance with respect to the interpretability of the modelling process. We rely on inherently interpretable algorithms first such as Kernel Density Estimation, L1 penalized Logistic regression and Decision trees, then on Extreme Gradient Boosting to provide a benchmark of the best performance achievable on this prediction tasks. Then, we train a serie of three recent interpretable ML models (Explainable Boosting Machine, RiskSLIM and SIRUS) to compare their performance with the more traditional algorithms and EGB. Finally, we provide an extended analysis of the algorithms' metrics in practitioners terms (missed crime reduction opportunity and wasted ressources under potentially constrained budget) to understand what are the potential trade-off in using more interpretable models.
This work aims at providing a global benchmark of recent ML models on crime prediction, as well as to introduce original modelling methodology such as sparse risk score with RiskSLIM, rules-based analysis with SIRUS and Generalized Additive Modelling with Explainable Boosting Machine. Our findings suggest that interpretable models can achieve competitive performance while enhancing transparency, making them preferable for policy programming and implementation.

Generative AI and the Courts: an American-centered Perspective
Annalisa Triggiano (Sant'Anna School of Advanced Studies, Italy)

Artificial Intelligence (AI) has already impacted litigation with advancements in document review and analysis and the evolving use of AI continues to grow at a very rapid pace. For example, AI is being used in the legal industry for contract review and analysis, predictive analysis to forecast litigation outcomes and to quickly sift through vast amounts of data to identify relevant information, reducing the time and costs involved for these tasks.
The arrival of ChatGPT, which is a generative AI tool that can produce clear, coherent human-like text in response to a question is anticipated to provide further benefits and efficiencies to attorneys by assisting with legal research, analyzing and summarizing key information from lengthy documents, and producing a first draft of various legal documents. All of these are clear benefits for increasing efficiency and lowering costs.
However, there are many known and unknown risks associated with using generative AI in the legal industry such as not fact checking the output of generative AI tools, confidentiality and security, bias and copyright issues. In addition, attorneys who use generative AI need to determine if they are required to provide

notice to the court. Some American Courts now require lawyers to disclose or describe their use of generative AI, including, for example:
The AI program used,
The portions of pleadings drafted by AI,
That the use of AI has not resulted in disclosure of confidential or proprietary information to unauthorized parties,
That generative AI or other AI has not been used, or
That content produced by generative AI has been verified by a human using traditional research methods.
These orders vary, but are generally triggered when generative AI is used to draft portions of pleadings filed with a court. Applications that were named include: "ChatGPT," "Google Bard," and "Harvey.ai." My paper will discuss a relevant List of American Court Orders concerning generative AI, which demonstrate how American judges are creating, de facto, guidelines for using Generative AI.

Adaptive Algorithmic Interventions for Escaping Pessimism Traps in Dynamic Sequential Decisions
Alexander Williams Tolbert (Emory University)

In this paper, we relate the philosophical literature on pessimism traps to information cascades, a formal model derived from the economics and mathematics literature. A pessimism trap is a social pattern in which individuals in a community, in situations of uncertainty, begin to copy the sub-optimal actions of others, despite their individual beliefs. This maps nicely onto the concept of an information cascade, which involves a sequence of agents making a decision between two alternatives, with a private signal of the superior alternative and a public history of others' actions. Key results from the economics literature show that information cascades occur with probability one in many contexts, and depending on the strength of the signal, populations can fall into the incorrect cascade very easily and quickly. Once formed, in the absence of external perturbation, a cascade cannot be broken -- therefore, we derive an intervention that can be used to nudge a population from an incorrect to a correct cascade and, importantly, maintain the cascade once the subsidy is discontinued. We study this both theoretically and empirically.

A Dilemma for AI Alignment
Sarah Valdman (University of Michigan)

The AI alignment problem is the problem of how to align AI with human values. Sophisticated AI models are prone to unpredictable and harmful behavior if left unchecked. They're also prone to manipulation by malicious actors seeking to put these models to immoral ends. AI safety researchers hope that infusing AI with a suitable representation of ethical behavior will mitigate these risks. The AI alignment problem, then, is the problem of how to align AI with human values. However, aligning AI with human values is ambiguous between aligning them with our moral intuitions – spontaneous moral judgements – or our moral theories – systematic moral frameworks. Either way, perfect adherence to moral intuition or moral theory would still permit unpredictable/harmful behavior and manipulation by malicious actors. In other words, it would leave the alignment problem unresolved.
Intuition and theory share three limitations that make them untenable targets for AI alignment: (1) extensional incompleteness, (2) interpersonal inconsistency, and, most importantly, (3) internal inconsistency.
Extensional incompleteness refers to the failure to deliver verdicts in every case. Intuition is often stumped in the face of unusual or complex situations, and moral theories are prone to gaps that leave borderline cases underdetermined. This means that neither can fix precise constraints on AI behavior.
Interpersonal inconsistency refers to the widespread moral disagreement between individuals. This poses the intractable problem of deciding whose intuitions or theories will be privileged and whose will be ignored. Even if a decision is reached, this would infuse the AI with the biases of the privileged group, which can cause harmful effects.

But the most important limitation is internal inconsistency: both moral intuition and moral theory contradict themselves. Even for a single individual, neither intuition nor theory furnish an internally consistent set of moral judgements. Moral intuition is notorious for its inconsistency, which results in part from its sensitivity to contextual factors that have no basis in moral principle. But both kinds of moral theories that dominate the ethics literature – consequentialism and deontology – are subject to their own internal inconsistencies. This is seen, for example, in the Actualism-Possibilism debate in consequentialist ethics, and the problem of universalizability in deontology. By the principle of explosion, anything follows from a contradiction; an AI so aligned could determine any action as morally permissible depending on which contradictory principles it appeals to. It would also be susceptible to manipulation by those who know how to exploit these contradictions.

So AI alignment faces a dilemma. Either we align AI with moral intuition or moral theory; but, either way, the central issues at stake in AI alignment go unresolved. This suggests that the alignment problem is unsolvable.

But some hope remains. It's implausible to solve the incompleteness or inconsistency of intuition, which largely lies beyond conscious control. However, moral theorizing is under our control. The alignment problem may be resolvable if philosophers formulate a moral theory without gaps, disagreements, or inconsistencies. In other words, a solution to AI alignment may be possible; but only after we solve moral philosophy itself.

From Bias to Balance: Towards Ethical Equity in Artificial Intelligence in Education (AIED)

Nasreen Watson (University of Johannesburg, South Africa)

I critically examine key ethical challenges in Artificial Intelligence in Education (AIED) within Sub-Saharan Africa, including algorithmic bias, digital literacy gaps, and systemic inequities that overshadow AI's potential as a transformative educational tool. My proposed research, From Bias to Balance: Towards Ethical Equity in Artificial Intelligence in Education (AIED), examines whether existing AI ethical frameworks effectively promote substantive educational freedoms or instead reinforce barriers to equity in digital education. I analyse the challenges in AI value alignment, questioning whether prevailing approaches in AIED adequately align with ethical principles that ensure equitable educational outcomes. To test the alignment of Amartya Sen's Capability Approach (1993) as a supported ethical framework for AIED. I argue that providing access to digital tools alone is insufficient; students must develop critical thinking skills to enhance digital literacy and meaningfully engage with AI-driven education. This requires a shift in AI governance and policy from efficiency-driven models to ethical frameworks that empower students, ensuring AI does not exacerbate existing disparities. To address these concerns, I propose the Equitable AI in Digital Education (EQUID) framework, aligning with Sustainable Development Goals (SDG 4.6 and SDG 4.7) as a capability-enhancing approach for strengthening the functional capabilities of marginalised students in Sub-Saharan Africa. This framework ensures AIED moves beyond concerns of algorithmic fairness and other technical challenges to tackle broader structural inequities, making AI systems more ethically aligned with the needs of diverse learners while remaining sensitive to cultural pluralistic influence and maintaining alignment with universal standards.

LLMs and Educational Equity

Renzhe Yu (Columbia University)

While Large Language Models (LLMs) offer significant potential for transforming education and bridging gaps in opportunity and achievement, the real-world consequences of their widespread availability have yet to be fully examined. In this talk, I will discuss potential equity challenges associated with LLMs in education, grounded in the research on digital inequalities. I will also share large-scale empirical evidence of complicated socioeconomic inequalities at different levels of the education system following the increasing adoption of LLM tools in 2023. These findings shed light on digital divides in the era of LLMs and call for researchers, educators, and policymakers on responsible practices to improve educational equity through novel technologies.

Sexual Humiliation and AI: An Existential-Feminist Analysis of Deepfake Pornography

Kate Yuan (Yale University)

This paper explores how recent advances in AI-driven deepfake pornography reshape our relationships—with each other, with sexual norms, and with the machines that now mediate so much of our intimate lives. While prior feminist debates on pornography centered on the "porn wars" of the 1970s and 1980s, today's challenge is more complex. Pornographic deepfakes turn non-consensual sexual humiliation into an easily scalable, algorithmically produced tool of gendered violence.

Building on existential philosophy, feminist theory, and critical race concepts, I offer a new framework for understanding why men—including those targeting their mothers, sisters, and acquaintances—deploy deepfakes to assert dominance. By merging insights from Beauvoir's notion of women as the "Other" and the gendered nature of sexual humiliation, Sartre's analysis of "bad faith", "the gaze", and sadism, and Du Bois's concept of a "psychological wage" adapted to masculinity, this paper provides an original existential-feminist account of how these technologies intensify patriarchal subjugation.

Moving beyond existing analyses that focus on harm to individuals depicted, I argue that deepfake pornography is best understood as part of a broader existential and cultural crisis. This crisis involves not only sexual scripts but also the erasure of women's autonomy and the denial of human ambiguity—actions justified by bad faith and stabilized by misogynistic communities, such as incels.

My contribution lies in linking existentialist concepts to the socio-technical phenomena of deepfake abuse, illuminating how patriarchal desires for control fuse seamlessly with AI capabilities. I show how these acts of humiliation reveal men's underlying terror of women's freedom and their refusal to engage authentically with reality.

I further highlight potential strategies of resistance, emphasizing the importance of third-party intervention and collective action that destabilizes perpetrators' anonymity and reclaims women's subjectivity. By reframing what's at stake—not merely privacy or reputations, but the conditions of human freedom itself—I shed light on how we might forge more ethical relationships and more just technological futures.

Panel Discussion Abstract

AI as a Catalyst for Social Change: Designing Technology that Centers Humanity

(Panel Discussion)

Organizer:
Jaye Nias (Howard University)

Panelists:
Armisha Roberts [moderator]*, Jaye Nias*, Saurav Aryal*, Thane Douglas*, Howard Prioleau*
*Howard University

As artificial intelligence continues to evolve, its potential to transform society is immense. However, we must ask: can AI be shaped to drive positive, lasting social change? This panel brings together a diverse group of scholars from Howard University's Institute for Socially and Culturally Relevant Human-Centered AI—ranging from experienced faculty to emerging student researchers—who are dedicated to ensuring that AI is developed with equity, justice, and inclusivity at its core. We'll explore how AI can be crafted not only to innovate but to serve all communities, particularly those historically marginalized or excluded from technological advancements. Through this conversation, we aim to highlight the power of human-centered AI in shaping a future that prioritizes the needs and experiences of all people.

Whose Voices Are Being Heard in AI Development?
Drawing from diverse research areas, including voice-enabled technologies, predictive modeling, and human-centered AI design, the panel will discuss the importance of developing AI that is culturally sensitive, transparent, and accountable. The conversation will focus on applying these technologies to real-world challenges—whether improving accessibility, enhancing civic engagement, or addressing social disparities—while ensuring that solutions are co-created with the communities they aim to serve. By emphasizing collaboration, the panel will explore how AI can meet the needs of people from all backgrounds and experiences.

Can AI Be a Tool for Justice or Perpetuate Inequality?
The panel will tackle the ethical challenges of AI in the context of social change, addressing critical questions about data equity, algorithmic transparency, and the potential for AI to either perpetuate or reduce bias. Moderated by Dr. Armisha Roberts, the discussion will focus on the role of participatory methodologies in AI development. Ethical frameworks and inclusive design principles will be explored, emphasizing how AI can be developed to prioritize fairness and accountability. The session will advocate for a future where AI serves as a force for positive social transformation.

By bringing together diverse voices—from senior researchers to students—the panel will offer a comprehensive vision of how AI can help build a more just and equitable society. The conversation will leave participants with a renewed understanding of how human-centered design, collaborative practices, and ethical responsibility can guide the development of AI that benefits all communities.