| | 8/8/2023 | | |
|---|---|---|---|
| 9:10-10:15 | **Opening Keynote** | | |
| | | Annette Zimmermann | The Generative AI Deployment Rush: How to Democratize the Politics of Pace |
| 10:35-12:30 | **Proceedings Session 1** | | |
| | Talk | Darren Cook, Miri Zilka, Heidi Desandre, Susan Giles and Simon Maskell | Protecting Children from Online Exploitation: Can a trained model detect harmful communication strategies? |
| | Talk | Tunazzina Islam, Ruqi Zhang and Dan Goldwasser | Analysis of Climate Campaigns on Social Media using Bayesian Model Averaging |
| | Talk | Johanna Walker, Gefion Thuermer, Elena Simperl and Julian Vincens | AI Art and Misinformation∗ Approaches and strategies for media literacy and fact checking |
| | Talk | Michael Feffer, Michael Skirpan, Hoda Heidari and Zachary Lipton | From Preference Elicitation to Participatory ML: A Critical Survey & Guidelines for Future Research |
| | Talk | Saumik Narayanan, Guanghui Yu, Chien-Ju Ho and Ming Yin | How does Value Similarity affect Human Reliance in AI-Assisted Ethical Decision Making? |
| | Student Program Lightning Talk | Saumik Narayanan | Exploring the Effect of AI Assistance on Human Ethical Decisions |
| | Student Program Lightning Talk | Jess Westbrook | Queering Futures with Data-Driven Speculation the design of an expanded-mixed methods framework integrating qualitative, quantitative, and practice-based modes |
| | Student Program Lightning Talk | Maria Lungu | Evaluating regulatory structures for the expansion of artificial intelligence: a panel assessment of U.S state legislation |
| | Student Program Lightning Talk | Valerie Chen | Are Model Explanations Useful in Practice? Rethinking How to Support Human-ML Interactions |
| | Student Program Lightning Talk | Kathleen Cachel | Ranked Candidate Fairness in Preference Aggregation |
| | Student Program Lightning Talk | Daniel Affsprung | The ELIZA Defect: Constructing the Right Users for Generative AI |
| | Student Program Lightning Talk | Emmie Hine | Governing Silicon Valley and Shenzhen: Assessing a New Era of Artificial Intelligence Governance in the US and China |
| | Student Program Lightning Talk | Jinhwa Kim | Safety in Conversational Systems |
| | Lightning Talk | Charles Evans and Atoosa Kasirzadeh | User Tampering in Reinforcement Learning Recommender Systems |
| | Lightning Talk | Shalaleh Rismani, Renee Shelby, Andrew Smart, Renelito Delos Santos, Ajung Moon and Negar Rostamzadeh | Beyond the ML Model: Applying Safety Engineering Frameworks to Text-to-Image Development |
| | Lightning Talk | Thomas Gilbert, Nathan Lambert, Sarah Dean, Tom Zick and Aaron Snoswell | Reward Reports for Reinforcement Learning |
| | Lightning Talk | William Seymour, Xiao Zhan, Mark Coté and Jose Such | A Systematic Review of Ethical Concerns with Voice Assistants |
| | Lightning Talk | Julia Barnett | The Ethical Implications of Generative Audio Models: A Systematic Literature Review |
| | Lightning Talk | Tyler Cook | Robust Artificial Moral Agents and Metanormativity |
| | Lightning Talk | Ryosuke Ueda, Koh Takeuchi and Hisashi Kashima | Mitigating Voter Attribute Bias for Fair Opinion Aggregation |
| | Lightning Talk | Nathanael Jo, Sina Aghaei, Jack Benson, Andres Gomez and Phebe Vayanos | Learning Optimal Fair Decision Trees: Trade-offs Between Interpretability, Fairness, and Accuracy |
| | Lightning Talk | Jindi Zhang, Luning Wang, Dan Su, Yongxiang Huang, Caleb Chen Cao and Lei Chen | Model Debiasing via Gradient-based Explanation on Representation |
| | Lightning Talk | Sruthi Gorantla, Anay Mehrotra, Amit Deshpande and Anand Louis | Sampling Individually-Fair Rankings that are Always Group Fair |
| | Lightning Talk | Alexandre Nanchen, Lakmal Meegahapola, William Droz and Daniel Gatica-Perez | Keep Sensors in Check: Disentangling Country-Level Generalization Issues in Mobile Sensor-Based Models with Diversity Scores |
| | Lightning Talk | Valeriia Cherepanova, Steven Reich, Samuel Dooley, Hossein Souri, John Dickerson, Micah Goldblum and Tom Goldstein | A Deep Dive into Dataset Imbalance and Bias in Face Identification |
| 2:00-3:00 | **Proceedings Session 2** | | |
| | Talk | Yilun Zhou | Iterative Partial Fulfillment of Counterfactual Explanations: Benefits and Risks |
| | Talk | Ira Globus-Harris, Varun Gupta, Christopher Jung, Michael Kearns, Jamie Morgenstern | Multicalibrated Regression for Downstream Fairness |
| | Talk | Peter Henderson, Eric Mitchell, Christopher Manning, Dan Jurafsky and Chelsea Finn | Self-Destructing Models: Increasing the Costs of Harmful Dual Uses of Foundation Models |
| | Talk | Mackenzie Jorgensen, Hannah Richert, Elizabeth Black, Natalia Criado and Jose Such | Not So Fair: The Impact of Presumably Fair Machine Learning Models |
| 3:00-4:10 | **Keynote Panel** | | |
| | | Roxana Daneshjou, Atoosa Kasirzadeh, Kate Larson, and Gary Marchant. Moderated by Alex John London | Large Language Models: Hype, Hope, and Harm |
| 4:30-6:00 | **Proceedings Session 3** | | |
| | Talk | Eran Tal | Target specification bias, counterfactual prediction, and algorithmic fairness in healthcare |
| | Talk | S J Bennett, Caroline Claisse, Ewa Luger and Abigail C. Durrant | Unpicking Epistemic Injustices in Digital Health: On the Implications of Designing Data-Driven Technologies for the Management of Long-Term Conditions |

| | | Talk | Ming Ying Yang, Gloria Hyunjung Kwak, Tom Pollard, Leo Anthony Celi and Marzyeh Gh | Evaluating the Impact of Social Determinants on Health Prediction in the Intensive Care Unit |
|---|---|---|---|---|
| | | Talk | Hubert D. Zając, Natalia R. Avlona, Finn Kensing, Tariq O. Andersen and Irina Shklovski | Ground Truth Or Dare: Factors Affecting The Creation Of Medical Datasets For Training AI |
| | | Student Program Lightning Talk | Bhargavi Ganesh | The Role of Governance in Bridging AI Responsibility Gaps |
| | | Student Program Lightning Talk | Vishwali Mhasawade | Advancing Health Equity Using Machine Learning |
| | | Student Program Lightning Talk | Lingwei Cheng | Algorithm-Assissted Decision Making and Racial Disparities in Housing: A Study of the Allegheny Homelessness Assessment Tool |
| | | Student Program Lightning Talk | Kevin Bryson | Designing Interfaces to Elicit Data Issues for Data Workers |
| | | Student Program Lightning Talk | Chinasa Okolo | Navigating the Limits of AI Explainability: Designing for Novice Technology Users in Low-Resource Settings |
| | | Student Program Lightning Talk | Chahat Raj | True and Fair: Robust and Unbiased Fake News Detection via Interpretable Machine Learning |
| | | Student Program Lightning Talk | Aidan Kierans | Benchmarked Ethics: A Roadmap to AI Alignment, Moral Knowledge, and Control |
| | | Lightning Talk | Harry Jiang, Lauren Brown, Jessica Cheng, Mehtab Khan, Abhishek Gupta, Deja Workman, Alex Hanna, Johnathan Flowers and Timnit Gebru | AI Art and its Impact on Artists |
| | | Lightning Talk | Nathan Dennler, Anaelia Ovalle, Ashwin Singh, Luca Soldaini, Arjun Subramonian, Huy Tu, William Agnew, Avijit Ghosh, Kyra Yee, Irene Font Peradejordi, Zeerak Talat, Mayra Russo and Jess De Jesus De Pinho Pinhal | Bound by the Bounty: Collaboratively Shaping Evaluation Processes for Queer AI Harms |
| | | Lightning Talk | Pamela Robinson | Action Guidance and AI Alignment |
| | | Lightning Talk | Charlotte Bird, Eddie Ungless and Atoosa Kasirzadeh | Ethical and Social Risks of Generative Text-to-Image Models |
| | | Lightning Talk | Emanuele Albini, Shubham Sharma, Saumitra Mishra, Danial Dervovic and Daniele Magazzeni | On the Connection between Game-Theoretic Feature Attributions and Counterfactual Explanations |
| | | Lightning Talk | Ian Hardy, Jayanth Yetukuri and Yang Liu | Adaptive Adversarial Training Does Not Increase Recourse Costs |
| | | Lightning Talk | Shubham Sharma, Sanghamitra Dutta, Emanuele Albini, Freddy Lecue, Daniele Magazz | REFRESH: Responsible and Efficient Feature Reselection guided by SHAP values |
| | | Lightning Talk | Carlos Mougan, Jose M. Alvarez, Salvatore Ruggieri and Steffen Staab | Fairness Implications of Encoding Protected Categorical Attributes |
| | | | | |
| | 8/9/2023 | | | |
| 9:00-10:00 | **Panel Discussion** | | | |
| | | | Patrick J Fowler, Hoda Heidari, and Christo Wilson. Moderated by Sanmay Das | AI *for* society: Developing, Deploying, and Auditing Public-Facing AI |
| 10:35-12:10 | **Proceedings Session 4** | | | |
| | | Talk | Glen Berman | Machine Learning practices and infrastructures |
| | | Talk | Agathe Balayn, Mireia Yurrita, Jie Yang and Ujwal Gadiraju | Fairness Toolkits, A Checkbox Culture? On the Factors that Fragment Developer Practices in Handling Algorithmic Harms |
| | | Talk | Anaelia Ovalle, Arjun Subramonian, Vagrant Gautam, Gilbert Gee and Kai-Wei Chang | Factoring the Matrix of Domination: A Critical Review and Reimagination of Intersectionality in AI Fairness |
| | | Talk | Vincent Straub, Deborah Morgan, Youmna Hashem, John Francis, Saba Esnaashari and Jonathan Bright | A multidomain relational framework to guide institutional AI research and adoption |
| | | Student Program Lightning Talk | Pranav Venkit | Towards a Holistic Approach: Understanding Sociodemographic Biases in NLP Models using an Interdisciplinary Lens |
| | | Student Program Lightning Talk | Olalekan Akintande | Algorithmic Bias: When stigmatization becomes a perception |
| | | Student Program Lightning Talk | Jessica Woodgate | Ethical Principles for Reasoning about Value Preferences |
| | | Student Program Lightning Talk | Luca Nannini | Explainability in Process Mining: A Framework for Improved Decision-Making |
| | | Student Program Lightning Talk | Mubarak Hussain | Can AlphaGo be apt subjects for Praise/Blame for "Move 37"? |
| | | Student Program Lightning Talk | Deborah Morgan | Anticipatory regulatory instruments for AI systems - A comparative study of regulatory sandbox schemes |
| | | Student Program Lightning Talk | Terkura Thomas Mchia | Examining the Ethics of Brain-Computer Interfaces: Ensuring Safety, the Rights and Dignity of Personhood. |
| | | Lightning Talk | Keziah Naggita, Julienne LaChance and Alice Xiang | Flickr Africa: Examining Geo-Diversity in Large-Scale, Human-Centric Visual Data |
| | | Lightning Talk | Rachel Hong, Tadayoshi Kohno and Jamie Morgenstern | Evaluation of targeted dataset collection on racial equity in face recognition |
| | | Lightning Talk | Shiva Omrani Sabbaghi, Robert Wolfe and Aylin Caliskan | Evaluating Biased Attitude Associations of Language Models in an Intersectional Context |
| | | Lightning Talk | Pranav Narayanan Venkit, Sanjana Gautam, Ruchi Panchanadikar, Ting-Hao Huang and | Unmasking Nationality Bias: A Study of Human Perception of Nationalities in AI-Generated Articles |
| | | Lightning Talk | Tom Williams and Kerstin Sophie Haring | No Justice, No Robots: From the Dispositions of Policing to an Abolitionist Robotics |

| | | | |
|---|---|---|---|
| | Lightning Talk | Violet Turri and Rachel Dzombak | Why We Need to Know More: Exploring the State of AI Incident Documentation Practices |
| | Lightning Talk | Shalaleh Rismani and Ajung Moon | What does it mean to be a responsible AI practitioner: An ontology of roles and skills |
| | | | |
| 2:00-3:00 | **Proceedings Session 5** | | |
| | Talk | Shin-Shin Hua and Haydn Belfield | Effective Enforceability of EU Competition Law Under AI Development Scenarios: a Framework for Anticipatory Governance |
| | Talk | Christie Lawrence, Isaac Cui and Daniel Ho | The Bureaucratic Challenge to AI Governance: An Empirical Assessment of Implementation at U.S. Federal Agencies |
| | Talk | Matteo Fabbri | Self-determination through explanation: an ethical perspective on the implementation of the transparency requirements for recommender systems set by the Digital Services Act of the European Union |
| | Student Program Lightning Talk | Arna Woemmel | Public attitudes toward ethical AI in courts: A Vignette Survey and Deliberation Experiment |
| | Student Program Lightning Talk | Behrad Taghibeyglou | How to promote equitable sleep care among people experiencing homelessness: An AI-enabled person-centred computer vision-based solution |
| | Student Program Lightning Talk | Nora Freya Lindemann | Sealed Knowledges: A Critical Approach to the Usage of LLMs as Search Engines |
| | Student Program Lightning Talk | Charvi Rastogi | Investigating the Relative Strengths of Humans and Machine Learning in Decision-Making |
| | Student Program Lightning Talk | Abiola Azeez | A Responsible Artificial Intelligence Approach to Epistemic Credibility in Artificial Intelligence Systems |
| | Lightning Talk | Avi Schwarzschild, Max Cembalest, Karthik Rao, Keegan Hines and John Dickerson | Reckoning with the Disagreement Problem: Explanation Consensus as a Training Objective |
| | Lightning Talk | Avijit Ghosh, Pablo Kvitca and Christo Wilson | When Fair Classification Meets Noisy Protected Attributes |
| | | | |
| 3:00-4:10 | **Keynote 2** | | |
| | | Jamie Morgenstern | Changing distributions and preferences in learning systems |
| 4:30-6:00 | **Proceedings Session 6** | | |
| | Talk | Elizabeth Edenberg and Alexandra Wood | Disambiguating Algorithmic Bias: From Neutrality to Justice |
| | Talk | Dafna Burema, Nicole Debowski-Weimann, Alexander von Janowski, Jil Grabowski, Miha | A sector-based approach to AI ethics: Understanding ethical issues of AI-related incidents within their sectoral context |
| | Talk | Elizabeth Seger, Aviv Ovadya, Divya Siddarth, Ben Garfinkel and Allan Dafoe | Democratising AI: Multiple Meanings, Goals, and Methods |
| | Talk | Renee Shelby, Shalaleh Rismani, Kathryn Henne, Ajung Moon, Negar Rostamzadeh, Pa | Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction |
| | Student Program Lightning Talk | Joshua Brand | Exploring the Moral Value of Explainable Artificial Intelligence Through Public Service Postal Banks |
| | Student Program Lightning Talk | Hubert Zając | It Takes A Village To Raise An AI-based System - The Interdisciplinary Work Of Realising Clinical AI in Denmark and Kenya |
| | Student Program Lightning Talk | Anastasia Siapka | AI-driven Automation as a Pre-condition for Eudaimonia |
| | Student Program Lightning Talk | Hetvi Jethwani | Multi Value Alignment: four steps for aligning ML/AI development choices with multiple values |
| | Student Program Lightning Talk | Anna Marie Rezk | "Way too good and way beyond comfort": The trade-off between user perception of benefits and comfort in media personalisation |
| | Student Program Lightning Talk | Anna Schmitz | Towards formalizing and assessing AI fairness |
| | Student Program Lightning Talk | Soroush Ebadian | Is Sortition Both Representative and Fair? |
| | Student Program Lightning Talk | Matteo Fabbri | How and to which extent will the provisions of the Digital Services Act of the European Union impact on the relationship between users and platforms as information providers? |
| | Lightning Talk | Jayanth Yetukuri, Ian Hardy and Yang Liu | Towards User Guided Actionable Recourse |
| | Lightning Talk | Przemyslaw Grabowicz, Nicholas Perello and Kenta Takatsu | Learning from Discriminatory Training Data |
| | Lightning Talk | Karan Bhanot, Ioana Baldini, Dennis Wei, Jiaming Zeng and Kristin Bennett | Stress-testing Bias Mitigation Algorithms to Understand Fairness Vulnerabilities |
| | Lightning Talk | Guusje Juijn, Niya Stoimenova, Joao Reis and Dong Nguyen | Perceived Algorithmic Fairness using Organizational Justice Theory: An Empirical Case Study on Algorithmic Hiring |
| | Lightning Talk | Ranjita Naik and Besmira Nushi | Social Biases through the Text-to-Image Generation Lens |
| | Lightning Talk | Junaid Ali, Matthäus Kleindessner, Florian Wenzel, Kailash Budhathoki, Volkan Cevher a | Evaluating the Fairness of Discriminative Foundation Models in Computer Vision |
| | Lightning Talk | Philippe Lammerts, Philip Lippmann, Yen-Chia Hsu, Fabio Casati and Jie Yang | How do you feel? Measuring User-Perceived Value for Rejecting Machine Decisions in Hate Speech Detection |
| | | | |

| 8/10/2023 | | | |
|---|---|---|---|
| 9:00-9:40 | **Proceedings Session 7** | | |
| | Talk | Spencer Rarrick, Ranjita Naik, Varun Mathur, Sundar Poudel and Vishal Chowdhary | GATE: A Challenge Set for Gender-Ambiguous Translation Examples |
| | Talk | Alan Chan, Herbie Bradley and Nitarshan Rajkumar | Reclaiming the Digital Commons: A Public Data Trust for Training Data |
| | Lightning Talk | Katherine Collins, Matthew Barker, Mateo Espinosa Zarlenga, Naveen Raman, Umang E | Human Uncertainty in Concept-Based AI Systems |
| | Lightning Talk | Donal Khosrowi, Finola Finn and Elinor Clark | Diffusing the Creator: Attributing Credit for Generative AI Outputs |
| | Lightning Talk | Sourojit Ghosh and Aylin Caliskan | ChatGPT Perpetuates Gender Bias in Machine Translation and Ignores Non-Gendered Pronouns: Findings across Bengali and Five other Low-Resource Languages |
| | Lightning Talk | Charvi Rastogi, Marco Tulio Ribeiro, Nicholas King, Harsha Nori and Saleema Amershi | Designing for Human-AI Collaboration in Auditing LLMs with LLMs |
| | Lightning Talk | Harvineet Singh and Rumi Chunara | Measures of Disparity and their Efficient Estimation |
| 11:40-12:50 | **Closing Keynote** | | |
| | | Paola Ricaurte Quijano | AI for/by the majority world: From technologies of dispossession to technologies of radical care |