the **MSU Ethics Intiative** *presents*

The Ethics of Big Data and Artificial Intelligence Conference

Flash Talk Speakers

2022

MICHIGAN STATE UNIVERSITY

Flash Talk Speaker Abstracts

Johannes Bauer..... 3

Bobby Brathwaite..... 3

Hyesun Choung..... 4

Sarah Gretter..... 4

Geoffrey Henebry..... 5

Marianne Huebner..... 5

Ike Iyioke..... 6

Stephanie Jordan..... 6

Zach Kaiser..... 7

Dennis Kennedy..... 7

Elizabeth Mack..... 8

Maria Molina..... 8

Ellen Moll..... 9

Annette O' Connor..... 9

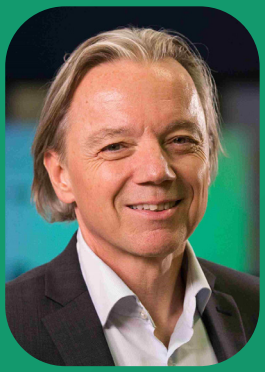
Emilee Rader..... 10

Isaac Record..... 10

Dean Rehberger..... 11

Anjana Susarla..... 12

Jiliang Tang..... 11



Johannes Bauer

The Role of Ethics in Establishing Guardrails for AI Innovation

Innovation creates novelty by combining and recombining existing knowledge. AI opens a vast space of combinatorial innovation opportunities for exploration by humans and machines. At the beginning of an innovation journey, the social and economic consequences and repercussions of new technologies are unknown. This short presentation explores the potential and limitations of ethics to guide AI innovation in the presence of deep uncertainty. It argues that ethics can provide guardrails to channel AI innovation in beneficial directions. Implementing appropriate governance mechanisms will be challenging but will be a critical to harness the full benefits of AI.

Bobby Brathwaite

Ethical Considerations When Utilizing Foreign Language Text for Natural Language Processing

This presentation utilizes event-data generated using Natural Language Processing (NLP) from a study on religious violence that examines English and foreign-language media sources. The talk will provide a quick overview of some of the study's findings that show variation in the measurement of religious violence when using different language sources and highlight ethical implications associated with media bias and political culture when utilizing machine learning approaches for textual analysis.

Hyesun Choung

Public Perceptions of the Social and Ethical Implications of AI

In collaboration with Drs. Prabu David and Arun Ross, a team of researchers has been working together on the topic of ethics and trust in AI. Our goal is to understand the social and ethical implications of various AI applications, including voice assistant and facial recognition technologies, and develop metrics and tools to build trustworthy AI systems. In this flash talk, we will discuss the relevance of virtue ethics along with a multi-faceted approach to the governance of AI, specifically in light of the ongoing controversy surrounding the use of facial recognition by various agencies.

Sarah Gretter

The Pedagogy of Ethics in Tech Education

As the talent for tech workforce increases and tech education programs grow in higher education, the pedagogy of tech education has emerged as key to the larger conversation around Diversity, Equity, and Inclusion in the tech industry. How might tech programs embed ethics as they prepare future developers and tech leaders? In this flash talk, Dr. Gretter will discuss the current landscape of tech education and associated pedagogies, and will share helpful tools for instructors to consider the pedagogy of ethics in tech education and beyond.

Geoffrey Henebry

Engaging RCR Discussions for a Flattened World of Scholarship

How do we prepare early-career scientists for RCR topics given cross-cultural norms and expectations in an era of big data and international collaborations? In leading RCR discussions with Geography graduate students, there can be wide divergence in their understandings of what constitutes unacceptable scientific practices in US context. Given my various editorial roles, it is clear plagiarism has become more widely spread as STEM publishing has flattened due to increased access to the internet, online publications, open access journals, the push for open science, and more. I share an approach effective for identifying and engaging class discussions on troublesome topics.

Marianne Huebner

Ethical Data Science Practice

Data scientists make data-driven decisions that require the collection of data and model building approaches that can have serious implications for health, security, politics, society. The computational perspective is not enough. AI and Machine learning raise questions of accountability. Data quality standards may not have been considered. The outcome may be determined by the way the classifier is trained. Who takes responsibility when things go wrong? Ethical data science takes a holistic approach that includes people from a range of perspectives (computational, statistical, scientific, ethical and legal, human) who collectively and collaboratively solve problems.

Ike Tyioke

Big Data and AI Vs Small Data and Native Intelligence

The talk compares Big Data and AI which are characteristic of advanced economies of the Global North, with 'Small Data and Native Intelligence' which refers to the knowledge systems that characterize Indigenous peoples, e.g., in Africa. It is a cautionary tale of how the digital gap between the Global North and the South could widen even more if we do not apply ways of addressing the challenges and barriers to equitable international research and development by promoting collaborative work that is inclusive, elevates underrepresented voices and groups, and demonstrates fairness of opportunity.

Stephanie Jordan

Environmental, Social and Design Justice: New Perspectives on Big Data, Robotics and AI in the Climate and Ocean Sciences

Uniting justice-oriented traditions in computing fields and diversity efforts within the sciences illuminates clearly that the future of work can only work if we take seriously big data's physical side: how specific geographies are instrumented, maintained and repaired to support a healthy ecosystem and good community relations. Drawing from over a decade of community-engaged participatory design, ethnographic observations and interviews with climate and ocean scientists at sea, I highlight new best practices for sharing geography and building relationships across research, industry and public spheres that make big data efforts adaptable and impenetrable to the ongoing crises that surround science today.

Zach Kaiser

The Computable Subjectivity and its Ethical Implications

Questions of ethics in the collection and use of data abound in scholarly literature on technology. These questions, however, often explore bias and privacy while perhaps obscuring a much bigger question that looms over the seemingly ever-expanding ability of various technologies to capture data about us and make inferences based on that data: what does it mean for oneself to be understood as both computable and computing—as made of nothing but data, and functioning by computational processes? What are the ethical implications of this computable subjectivity?

Dennis Kennedy

When Technology Outpaces Law: Applying Backward-looking Principles and Approaches to Fast-moving Big Data and AI Issues

Common law builds on past cases. In the US, new technologies are often covered by decades-old statutes. Technology raises issue that are no longer local, but global. And the pace of technological change continues to accelerate. How will law and regulation keep up with the pace or can they? What are the consequences and what approaches might work? Are Big Data and AI both part of the problem and part of the solution?

Elizabeth Mack

Ethics Considerations for an Age of Autonomous Systems

Artificial intelligence (AI) is systems or machines that are able to perform human-like tasks and improve performance iteratively based on information that is collected. In this sense, autonomous vehicles (AVs) or self-driving cars, which are vehicles that use systems to respond to objects and events, are a form of AI. Although these systems are expected to yield many benefits for society as a whole, the forthcoming era of autonomous vehicles and transport systems does pose ethical challenges. This flash talk will provide an overview of the ethical considerations related to AVs that need to be addressed.

Maria Molina

Calibrating User Trust for a More Responsible Design of AI Systems

Research suggests users rely on cognitive heuristics or mental short cuts when assessing AI systems. This leads users to either over trust AI because of the belief that AI is more objective and accurate compared to humans or under trust it because of the belief that AI lacks human sensitivity. As machine learning experts are working on technologies to make AI better, we also need to design such technologies from a user perspective so that user trust matches the system's actual capacity. This flash talk will discuss different mechanisms to calibrate user trust in AI in the context of content moderation.

Ellen Moll

Two Directions for Student Data

This talk will outline some key ethical and equity issues related to predictive analytics and other methods of using student data, and apply feminist technoscience theoretical frameworks that could be useful in imagining the university as a model for how institutions ethically gather and use data.

Annette O' Connor

User-Center Research Synthesis for Climate-Smart Agriculture

Our long-term goal is to enable rapid and accurate knowledge transfer of research to society. We aim to transform research synthesis using artificial intelligence tools to search, filter, and extract information for research-based smart-climate solutions for sustainable agriculture. The research literature around smart climate agriculture has all the characteristics of big data: it is voluminous, has great value, is varied in form, is being produced at velocity, and with considerable concerns about the veracity of extracted data. Therefore we need to create a research synthesis system that reduces the need for humans to do menial tasks like searching and filtering information and extracting information from publications and enables focus on interpretation and adoption.

Emilee Rader

An Argument for Limiting Collection of Data

"Notice and choice," the dominant model for governing digital data collection and use, assumes that if proper transparency is provided people will only use platforms that have data practices they agree with. But, in reality, widespread data collection and use of machine learning enables inferences to be generated that are hard for people to anticipate. This talk summarizes characteristics of people's awareness of and assumptions about what data are collected about them and how those data are used, and makes an argument about how specific limits on data collection and inferences could help people make privacy choices that are more aligned with their preferences.

Isaac Record

Ubiquitous and Unique: Synthetic Media and the Biasing of Imagination

The promise of synthetic media such as deepfakes and DALL-E is that they enable ordinary computer users the practicable possibility to create and share their imaginations in rich multimedia form. While scholars rightly worry that synthetic media may undermine our trust in video evidence, destabilize politics, and contribute to a culture of objectifying women, I am persuaded that a further and more insidious threat is that biased training sets allow some to express their imagination more easily than others, and moreover will encourage some creators to accept biased versions of their imagined worlds. This is a form of epistemic injustice.

Dean Rehberger

Considering Ethics through Knowledge Graph Infrastructure

Ethics are not simply an addition to Knowledge Graphs but must help to organize and develop the infrastructure. Our Enslaved.org project brings together recent developments in data management, applications, and publication; advances in digital humanities scholarship on the lives of the enslaved; and the anti-racist imperative of the moment. The initiative leverages Linked Open Data (LOD) using Semantic Web standards, techniques, and tooling that include the use of Wikibase, Web Ontology Language (OWL), Resource Description Framework (RDF), and a graph database, to create an innovative, centralized hub to engage historical data about the transatlantic slave trade extracted from a wide variety of primary sources.

Jiliang Tang

Trustworthy AI: to be Fair or to be Robust

Robustness and fairness of machine learning models are the two crucial dimensions to achieve trustworthy artificial intelligence. In recent years, the robustness and fairness of the model have attracted extensive research attention. It is often assumed that these two dimensions are independent, therefore the majority of existing research only focuses on one of them. In this talk, I will discuss that robustness and fairness are not independent: enhancing robustness introduces fairness issues; while ensuring fairness makes models less robust.

Anjana Susarla

Identifying and Evaluating Credible Healthcare Information Sources Online Using Responsible Artificial Intelligence

Studies suggest that one in three US adults use the Internet to diagnose or learn about a health concern. Health literacy is well recognized as a challenge for public health, with many adults lacking the requisite skills to engage successfully in the management of their healthcare. For patients engaging in searches for health information on digital media platforms, health literacy divides can be exacerbated both by their own lack of knowledge and by algorithmic recommendations, with results that disproportionately impact disadvantaged populations, minorities, and low health literacy users. I will present AI-augmented methods to identify credible and representative sources of healthcare information applied to a large corpus of videos and their metadata from a social media platform, YouTube.

Assessing Bias and Fairness in YouTube Videos on Healthcare: Analytics using Face and Speech Recognition Methods

Even though social media offers an excellent conduit for encouraging patients to increase their knowledge on health literacy, there is a significant challenge in producing fair and transparent recommendations that address the needs of every user. We provide methods to assess representativeness, bias, and inclusion in videos for chronic health conditions on YouTube. We use methods from fair machine learning to examine whether videos about certain health conditions are skewed towards a certain demographic group, and present methods for bias-aware recommendations on social media. The results will help health organizations and providers with best practice guidelines about the retrieval and use of healthcare information on social media.