LEADING EXPERTS IN PHILOSOPHY, ECONOMICS, LAW AND MORE CONVENE AT AIES 2019 TO ASSESS IMPACT OF ARTIFICIAL INTELLIGENCE ON SOCIETY

Multidisciplinary Conference Will Tackle Issues of Value Alignment, Data Handling and Bias, Regulations, and Workforce Displacement

New York, NY, January 24, 2019 – ACM, the Association for Computing Machinery; AAAI, the Association for the Advancement of Artificial Intelligence; and SIGAI, the ACM Special Interest Group on Artificial Intelligence, today announced the second annual ACM/AAAI Conference on Artificial Intelligence, Ethics and Society (AIES). The AIES conference provides a platform for research and discussions from the perspectives of several disciplines to address the challenges of AI ethics within a societal context, featuring participation from experts in computing, ethics, philosophy, economics, psychology, law and politics. The conference is planned for January 27-28, 2019 in Honolulu, Hawaii.

"AI is evolving rapidly, and as a society we're still trying to understand its impact – both in terms of its benefits and its unintended consequences," said conference co-chair Vincent Conitzer, of Duke University. "The AIES conference was designed to include participation from different disciplines and corners of society, in order to offer a unique and informative look at where we stand with the development and the use of artificial intelligence."

The AIES conference was chaired by a multi-disciplinary program committee to ensure a diversity of topics. Conference sessions will address algorithmic fairness, measurement and justice, autonomy and lethality, human-machine interaction and AI for social good, among other focuses. AIES presenters and attendees include representatives from major technology and non-technology companies, academic researchers, ethicists, philosophers, think tanks, and members of the legal profession.

More than 300 individuals are expected to attend. The conference is sponsored by the Berkeley Existential Risk Initiative, DeepMind Ethics and Society, Google, National Science Foundation, IBM Research, Facebook, Amazon, PWC, Future of Life Institute, and the Partnership on AI.

AIES HIGHLIGHTS

Invited Talks

"Guiding and Implementing AI"
Susan Athey, Stanford University
Athey will address theoretical perspectives on how organizations should guide and implement AI in a way that is fair and that achieves the organization's objectives. The role of insights from statistics and causal inference for this problem is examined first, followed by the framework to incorporate considerations about designing rewards for AI and human decision-makers, and designing tasks and authority to optimally combine AI and humans to achieve the most effective incentives.

"How We Talk About AI (and Why It Matters)"
Ryan Calo, University of Washington
Calo will explore how AI should be discussed, what's at stake with our rhetorical choices and the interplay between claims about AI and law's capacity to channel AI in the public interest. Not only do our rhetorical choices influence public expectations of AI, they implicitly make the case for or against specific government interventions.

"The Value of Trustworthy AI"
David Danks, Carnegie Mellon University
In this talk, Danks will unpack the notion of 'trustworthy' from both philosophical and psychological perspectives, as it might apply to an AI system. He will argue that there are different kinds of (relevant, appropriate) trustworthiness, depending on one's goals and modes of interaction with the AI. There is not just one kind of trustworthy AI, even though trustworthiness (of the appropriate type) is arguably the primary feature that we should want in an AI system.

"Specifying AI Objectives as a Human-AI Collaboration Problem"
Anca Dragan, University of California, Berkeley
Estimation, planning, control, and learning are giving us robots that can generate good behavior given a specified objective and set of constraints. In this talk, Dragan will explore how humans enter this behavior generation picture, and discuss two complementary challenges: 1) *how* to optimize behavior when the robot is not acting in isolation, but needs to coordinate or collaborate with people; and 2) *what* to optimize in order to get the behavior we want.

Best Paper Award (sponsored by the Partnership on AI)

"Learning Existing Social Conventions via Observationally Augmented Self-Play"
Alexander Peysakhovich and Adam Lerer, Facebook

In this paper, the authors discuss how a simple procedure that combines small amounts of imitation learning with self-play can lead to agents that can learn social conventions.

Select Papers:

"(When) Can AI Bots Lie?"
Tathagata Chakraborti, IBM Research AI, and Subbarao Kambhampati, Arizona State University
This paper investigates how fabrication, falsification and obfuscation of information can be used by an AI agent to encourage behavior in the hopes of achieving some greater good. Results of a thought experiment indicate that public perception is positive toward lying for the greater good, but poses several unresolved ethical and moral questions with regards to the design of autonomy.

"Active Fairness in Algorithmic Decision Making"
Michiel Bakker, Alejandro Noriega-Campero, Bernardo Garcia-Bulle and Alex Pentland, MIT
The authors propose an alternative active framework for fair classification in order to achieve fairness measures, and suggest two such methods, where information collection is adapted to group and individual needs respectively. Their results suggest that by leveraging their additional degree of freedom, active approaches can substantially outperform currently used randomization-based classifiers.

"Shared Moral Foundations of Embodied Artificial Intelligence"
Joe Cruz, Williams College
The author suggests in his paper that apprehensions regarding AI systems' inability to align with the moral values of human beings are overstated. Embodied Cognition research programs will be implemented to achieve intelligence in its full generality and adaptiveness in AI systems.

"Killer Robots and Human Dignity"
Daniel Lim, Duke Kunshan University
This paper focuses on arguments against the use of Lethal Autonomous Weapon Systems (LAWS) based on human dignity, and suggests that a non-contingent ban on LAWS based on human dignity fails.

"The Right to Confront Your Accuser: Opening the Black Box of Forensic DNA Software"
Jeanna Matthews, Marzieh Babaeianjelodar, Stephen Lorenz, Abigail Matthews, Clarkson University; Mariama Njie, Iona College; Nathan Adams, Forensic Bioinformatic Services; Dan Krane, Wright State University; Jessica Goldthwaite and Clinton Hughes, Legal Aid Society
This paper examines the Forensic Statistical Tool (FST), a forensic DNA system developed in 2010 by New York City's Office of Chief Medical Examiner (OCME), used in more than 1,300 criminal cases, and the first analysis of the impact of an undisclosed function capable of dropping evidence that could be beneficial to the defense.

"Framing Artificial Intelligence in American Newspapers"
Ching-Hua Chuan, Wan-Hsiu Tsai and Su Yeon Cho, University of Miami

This paper examines how artificial intelligence (AI) was covered in five U.S. newspapers over the course of nine years using a content analysis based on framing theory. Results indicate that business and technology were the primary topics in news coverage of AI, and that benefits were discussed more frequently than its risks, but risks of AI were generally discussed with greater specificity.

"Guiding Prosecutorial Decisions with an Interpretable Statistical Model"
Zhiyuan Lin, Alex Chohlas-Wood and Sharad Goel, Stanford University
In one of the first large-scale empirical analyses of pre-arraignment detention, the authors examine police reports and charging decisions for approximately 30,000 felony arrests, finding that 45% of arrested individuals are never charged for any crime but still typically spend one or more nights in jail before being released. A statistical model is proposed to help prosecutors identify cases soon after arrest that are likely to be ultimately dismissed in an effort to reduce such incarceration.

For a complete list of research papers and posters which will be presented at the AIES Conference, visit http://www.aies-conference.com/accepted-papers/. The proceedings of the conference will be published in the AAAI and ACM Digital Libraries.

About ACM
ACM, the Association for Computing Machinery (www.acm.org)**,** is the world's largest educational and scientific computing society, uniting computing educators, researchers and professionals to inspire dialogue, share resources and address the field's challenges. ACM strengthens the computing profession's collective voice through strong leadership, promotion of the highest standards, and recognition of technical excellence. ACM supports the professional growth of its members by providing opportunities for life-long learning, career development, and professional networking.

###