



Association for
Computing Machinery

Advancing Computing as a Science & Profession

CONTACT: Jim Ormond

212-626-0505

ormond@hq.acm.org

Michael Ekstrand

208-426-5761

press@fatconference.org

Michael Veale

44 (0) 20 3108 9736

press@fatconference.org

ACM FAT* CONFERENCE EXAMINES FAIRNESS, ACCOUNTABILITY AND TRANSPARENCY OF ALGORITHMIC SYSTEMS

Rapidly-Growing Event to Be Livestreamed and Recorded

New York, NY, January 17, 2019 – The 2019 ACM Conference on Fairness, Accountability and Transparency (FAT*), to be held in Atlanta, Georgia from January 29-31, is a unique international and interdisciplinary peer-reviewed gathering that publishes and presents work examining the fairness, accountability and transparency of algorithmic systems. ACM FAT* will host the presentation of research work from a wide variety of disciplines, including computer science, statistics, the social sciences and law. To accommodate the rapidly growing interest in FAT*, portions of the 2019 conference will be recorded and livestreamed. The livestream of the general session will be available on January 30 and 31 from 8:45 AM - 5:15 PM EST at <https://livestream.com/accounts/20617949/events/8521409/player>. Check [Technical Program](#) for program details.

FAT* grew out of the successful Workshop on Fairness, Accountability, and Transparency in Machine Learning (FAT/ML), as well as workshops on recommender systems (FAT/REC), natural language processing (Ethics in NLP), and data and algorithmic transparency (DAT), among others. Last year, more than 450 academic researchers, policymakers and practitioners attended the conference. 2019 will mark the first year that FAT* is an ACM conference.

“Algorithmic and algorithm-assisted decision making can affect our lives in significant ways, and this trend will only continue in the coming years,” explains ACM FAT* Program Co-Chair Alexandra Chouldechova, Carnegie Mellon University. “However, the FAT* conference goes beyond questions of the fairness and transparency of algorithms to encompass the full panoply of interactions between humans and intelligent machines, and how we can ensure that bedrock societal values are inherent in technological progress.”

“In keeping with the mission of FAT, this year’s program reflects a multi-disciplinary and multi-stakeholder effort to address the challenges of AI ethics within a societal context,” added Program Co-Chair Fernando Diaz of Microsoft Research Montreal. “Participants include experts in various disciplines

such as computing, ethics, philosophy, economics, psychology, law and politics. FAT* has really struck a chord, and this year we expect more than 500 participants and we hope many more will partake of the conference through the livestreaming opportunities.”

This year’s program, selected by a scientific committee of 122 top researchers across many disciplines, includes 41 peer-reviewed papers (from 162 submissions) and 13 tutorials from leading experts in various fields (including scientists, lawyers and policymakers). Accepted work is wide in scope, including novel technical approaches to identifying and mitigating social concerns involving computing systems, qualitative and quantitative empirical studies of on-the-ground challenges, critical analysis of current trends from law, ethics, history and philosophy, and interdisciplinary papers translating and building bridges for better understanding and collaboration.

In addition to providing a forum for publishing and discussing research results, the FAT* conference also seeks to develop a diverse and inclusive global community around its topics and make the material and community as broadly accessible as feasible. To that end, the conference has provided over 80 scholarships to students and researchers, subsidizes attendance by students and nonprofit representatives, and will be livestreaming all main program content for those who are not able to attend in person. This year also sees the introduction of a Doctoral Consortium to support and promote the next generation of scholars working to make algorithmic systems fair, accountable, and transparent.

ACM FAT* 2019 HIGHLIGHTS

Keynote Addresses

Deirdre Mulligan

Deirdre K. Mulligan is an Associate Professor in the School of Information at UC Berkeley, a faculty Director of the Berkeley Center for Law and Technology, and a co-organizer of the Algorithmic Fairness and Opacity Working Group, among her other roles. Her research explores legal and technical means of protecting values such as privacy, freedom of expression, and fairness in emerging technical systems. Her book, *Privacy on the Ground: Driving Corporate Behavior in the United States and Europe*, co-authored with Berkeley Law Professor Kenneth Bamberger, is a study of privacy practices in large corporations in five countries. Mulligan and Bamberger received the 2016 International Association of Privacy Professionals Leadership Award for their research contributions to the field of privacy protection.

Jon Kleinberg

Jon Kleinberg is a Professor in the Departments of Computer Science and Information Science at Cornell University. His research focuses on the interaction of algorithms and networks, and the roles they play in large-scale social information systems. His books include *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*, co-authored with David Easley, and *Algorithm Design*, co-authored with Eva Tardos. Kleinberg’s work has been supported by an NSF Career Award, an ONR Young Investigator Award, a MacArthur Foundation Fellowship, a Packard Foundation Fellowship, a Simons

Investigator Award, a Sloan Foundation Fellowship, and numerous public and private grants. Kleinberg received the 2008 ACM Prize in Computing.

Accepted Papers Include

“The Disparate Effects of Strategic Manipulation”

Lily Hu, Harvard University; Nicole Immorlica, Jennifer Wortman Vaughan, Microsoft Research

When consequential decisions are informed by algorithmic input, individuals may feel compelled to alter their behavior in order to gain a system's approval. Models of agent responsiveness, termed "strategic manipulation," analyze the interaction between a learner and agents in a world where all agents are equally able to manipulate their features in an attempt to "trick" a published classifier. In cases of real world classification, however, an agent's ability to adapt to an algorithm is not simply a function of her personal interest in receiving a positive classification, but is bound up in a complex web of social factors that affect her ability to pursue certain action responses. In this paper, the authors adapt models of strategic manipulation to capture dynamics that may arise in a setting of social inequality wherein candidate groups face different costs to manipulation. The authors find that whenever one group's costs are higher than the other's, the learner's equilibrium strategy exhibits an inequality-reinforcing phenomenon wherein the learner erroneously admits some members of the advantaged group, while erroneously excluding some members of the disadvantaged group.

“Who’s the Guinea Pig? Investigating Online A/B/n Tests in-the Wild”

Shan Jiang, John Martin, Christo Wilson, Northeastern University

A/B/n testing has been adopted by many technology companies as a data-driven approach to product design and optimization. These tests are often run on their websites without explicit consent from users. In this paper, we investigate such online A/B/n tests by using Optimizely as a lens. First, the authors provide measurement results of 575 websites that use Optimizely drawn from the Alexa Top-1M, and analyze the distributions of their audiences and experiments. Then, they use three case studies to discuss potential ethical pitfalls of such experiments, including involvement of political content, price discrimination, and advertising campaigns. The authors conclude with a suggestion for greater awareness of ethical concerns inherent in human experimentation and a call for increased transparency among A/B/n test operators.

“On Microtargeting Socially Divisive Ads: A Case Study of Russia-Linked Ad Campaigns on Facebook”

Filipe Ribeiro, Federal University of Ouro Preto; Koustuv Saha, Georgia Institute of Technology; Mahmoudreza Babaei, Max Planck Institute for Software Systems; Lucas Henrique, Zeester; Johnnatan Messias, Max Planck Institute for Software Systems; Oana Goga, Laboratoire d'Informatique de Grenoble, Fabricio Benevenuto, Federal University of Minas Gerais; Krishna P. Gummadi, Max Planck Institute for Software Systems; and Elissa M. Redmiles, University of Maryland

Targeted advertising is meant to improve the efficiency of matching advertisers to their customers. However, targeted advertising can also be abused by malicious advertisers to efficiently reach people susceptible to false stories, stoke grievances, and incite social conflict. The authors examine a specific case of malicious advertising, exploring the extent to which political ads from the Russian Intelligence

Research Agency (IRA) run prior to 2016 US elections exploited Facebook's targeted advertising infrastructure to efficiently target ads on divisive or polarizing topics (e.g., immigration, race-based policing) at vulnerable sub-populations. Among their other findings, the authors show how the enormous amount of personal data Facebook aggregates about users and makes available to advertisers enables such malicious targeting.

“Robot Eyes Wide Shut: Understanding Dishonest Anthropomorphism”

Brenda Leong, Future of Privacy Forum; Evan Selinger, Rochester Institute of Technology

Brenda Leong and Evan Selinger critically examine the trend of designing technologies, especially robots and artificial intelligences, that increasingly look, sound, and behave like human beings, the most well-known of which is Apple's Siri. Part of the essay involves explaining why this trend is occurring—noting what benefits it can bring and, more fundamentally, how it capitalizes on deeply rooted, evolutionary tendencies in human perception and cognition. But the majority of the essay involves clarifying the risks involved—risks that include significant privacy and security concerns, as well dangers related to emotional manipulation, misplaced expectations, and even harms related to perpetuating unfair gender stereotypes. To help technologists, policymakers, and the general public make better decisions when this type of design practice is involved, the authors offer a new taxonomy that pinpoints the central problems that can arise when people treat machines as more human-like than they really are. By implication, this conceptual framework also suggests ways to avoid these complications and pitfalls.

“Clear Sanctions, Vague Rewards: How China's Social Credit System Currently Defines 'Good' and 'Bad' Behavior”

Severin Engelmann, Jens Grossklags, Mo Chen, Technical University of Munich; Felix Fischer, Technical University of Munich and Projects by IF; and Ching-yu Kao, Fraunhofer AISEC

China's Social Credit System (or shehui xinyong tixi) is expected to become the first digitally-implemented nationwide scoring system with the purpose to rate the behavior of citizens, companies, and other entities. Thereby, in the SCS, “good” behavior can result in material rewards and reputational gain while “bad” behavior can lead to exclusion from material resources and reputational loss. Crucially, for the implementation of the SCS, society must be able to distinguish between behaviors that result in reward and those that lead to sanction. In this paper, the authors conduct the first transparency analysis of two central administrative information platforms of the SCS to understand how the SCS currently defines “good” and “bad” behavior.

For a complete list of research papers and posters which will be presented at the FAT* Conference, visit <https://fatconference.org/2019/acceptedpapers.html>.

The proceedings of the conference will be published in the ACM Digital Library.

Tutorial Livestream Links (Partial List)

Room: Georgia Hall 2-3

Tuesday, Jan 29th at 1:00 PM - 6:30PM EST

Livestream Link: <https://livestream.com/accounts/20617949/events/8521403/player>

Time: 1-2:30 PM: [Translation Tutorial: A History of Quantitative Fairness in Testing](#)

Time: 3-3:45 PM: *A New Era of Hate*

Time: 3:45-4:30 PM: [Parole Denied: One Man's Fight Against a COMPAS Risk Assessment](#)

Time: 5-6:30 PM: [Challenges of Incorporating Algorithmic Fairness into Industry Practice](#)

Room: Georgia 4-5

Tuesday, Jan 29th at 1:00 PM - 6:30PM EST

Livestream Link: <https://livestream.com/accounts/20617949/events/8521405/player>

Time: 1-2:30 PM: [Building Community Governance of Risk Assessment](#)

Time: 3-3:45 PM: [Towards a Theory of Race for Fairness in Machine Learning](#)

Time: 3:45-4:30 PM: [Engineering for Fairness: How a Firm Conceptual Distinction between Unfairness and Bias Makes It Easier to Address Un/Fairness](#)

Time: 5-6:30 PM: [Reasoning about \(Subtle\) Biases in Data to Improve the Reliability of Decision Support Tools](#)

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.

###