CONTACT: Jim Ormond   Angela Daly   Michael Veale
                212-626-0505   +61 3 9627 4853   +44 (0) 20 3108 9736
                ormond@hq.acm.org   press@fatconference.org   press@fatconference.org

ACM FAT* CONFERENCE ASKS: "ARE ALGORITHMIC SYSTEMS FAIR?"

Premier Conference in Exciting and Fast-Growing Research Area to Feature Fairness, Accountability, and Transparency in Socio-technical Systems

New York, NY, January 21, 2020 – The 2020 ACM Conference on Fairness, Accountability and Transparency (ACM FAT*), to be held in Barcelona, Spain from January 27-30, brings together researchers and practitioners interested in fairness, accountability, and transparency in socio-technical 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.

ACM 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; this year over 600 registrations are expected and the amount of papers submitted was doubled. 2020 will mark the first year that ACM FAT* is being presented in Europe.

"More and more, algorithmic systems are making and informing decisions which directly affect our lives," explains ACM FAT* General Co-chair Carlos Castillo, Universitat Pompeu Fabra. "Through a multidisciplinary perspective that encompasses computer science, law and social science, the ACM FAT* Conference not only explores the potential technological solutions regarding potential bias, but also seeks to address pivotal questions about economic incentive structures, perverse implications, distribution of power, and redistribution of welfare, and to ground research on fairness, accountability, and transparency in a legal framework."

Added General Co-chair Mireille Hildebrandt, Vrije Universiteit Brussel and Radboud University Nijmegen, "This year, we have developed a dedicated LAW track and a dedicated SSH track, next to the CS track, highlighting that cross-disciplinary exchanges are core to this conference. We are also excited to introduce the CRAFT initiative as part of ACM FAT*. CRAFT stands for critiquing and rethinking accountability, fairness and transparency, bringing together a broad spectrum of people working in domains affected by algorithmic decision making, aiming to give voice to those who suffer the consequences, while fostering interaction with computer scientists working on technical solutions."

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 the main program content for those who are not able to attend in person. A Doctoral Consortium will support and promote the next generation of scholars working to make algorithmic systems fair, accountable, and transparent.

ACM FAT* 2020 HIGHLIGHTS

Keynote Addresses
"Hacking the Human Bias in AI"
Ayanna Howard, Georgia Institute of Technology
Howard maintains that people tend to overtrust sophisticated computing devices, including robotic systems. As these systems become more fully interactive with humans during the performance of day-to-day activities, the role of bias in these human-robot interaction scenarios must be more carefully investigated. She argues that bias is a feature of human life that is intertwined, or used interchangeably, with many different names and labels – stereotypes, prejudice, implicit or subconsciously held beliefs. In the digital age, this bias has often been encoded in and can manifest itself through AI algorithms, which humans then take guidance from, resulting in the phenomenon of "excessive trust." In this talk, she will discuss this phenomenon of integrated trust and bias through the lens of intelligent systems that interact with people in scenarios that are realizable in the near-term.

"Productivity and Power: The Role of Technology in Political Economy"
Yochai Benkler, Harvard Law School
Benkler argues that the revival of the concept "political economy" offers a frame for understanding the relationship between productivity and justice in market societies. It reintegrates power and the social and material context—institutions, ideology, and technology—into our analysis of social relations of production, or how we make and distribute what we need and want to have. Organizations and individuals, alone and in networks, struggle over how much of a society's production happens in a market sphere, how much happens in nonmarket relations, and how embedded those aspects that do occur in markets are in social relations of mutual obligation and solidarism. These struggles involve efforts to shape institutions, ideology, and technology in ways that trade off productivity and power, both in the short and long term. The outcome of this struggle shapes the highly divergent paths that diverse market societies take, from oligarchic to egalitarian, and their stability as pluralistic democracies.

"Making Accountability Real: Strategic Litigation"
Nani Jansen Reventlow, Digital Freedom Fund
Reventlow asks "How can we make fairness, accountability and transparency a reality?" She points out that litigation is an effective tool for pushing for these principles in the design and deployment of automated decision-making technologies. She also maintains that the courts can be strong guarantors of

our rights in a variety of different contexts and have shown already that they are willing to do so in the digital rights setting. At the same time, she argues that, as automated decisions are increasingly impacting every aspect of our lives, we need to engage the courts on these complex issues and enable them to protect our human rights in the digital sphere. We are already seeing cases being taken to challenge facial recognition technology, predictive policing systems, and systems that conduct needs assessments in the provision of public services. However, we still have much work to do in this space. Reventlow will also explore what opportunities do the different frameworks in this area, and especially European regulations such as the GDPR offer, and how can we maximize their potential?

Accepted Papers (Partial List)
For a complete list of research papers and posters which will be presented at the FAT* Conference, visit .

"Auditing Radicalization Pathways on YouTube"
Manoel Horta Ribeiro, Raphael Ottoni, Robert West, Virgílio A. F. Almeida, Wagner Meira, École Polytechnique Fédérale de Lausanne (EPFL)
Non-profits, as well as the media, have hypothesized the existence of a radicalization pipeline on YouTube, claiming that users systematically progress towards more extreme content on the platform. Yet, there is to date no substantial quantitative evidence of this alleged pipeline. To close this gap, the authors conducted a large-scale audit of user radicalization on YouTube. They analyzed 330,925 videos posted on 349 channels, which they broadly classified into four types: Media, the Alt-lite, the Intellectual Dark Web (IDW), and the Alt-right. According to this radicalization hypothesis, channels in the IDW and the Alt-lite serve as gateways to fringe far-right ideology, here represented by Alt-right channels. Processing 72M+ comments, the authors show that the three channel types indeed increasingly share the same user base; that users consistently migrate from milder to more extreme content; and that a large percentage of users who consume Alt-right content now consumed Alt-lite and IDW content in the past. They also probe YouTube's recommendation algorithm, looking at more than 2M video and channel recommendations between May and July 2019. They find that Alt-lite content is easily reachable from IDW channels, while Alt-right videos are reachable only through channel recommendations. Overall, the authors paint a comprehensive picture of user radicalization on YouTube.

"Fair Decision Making Using Privacy-Protected Data"
Satya Kuppam, Ryan Mckenna, David Pujol, Michael Hay, Ashwin Machanavajjhala, Gerome Miklau, UMass Amherst
Data collected about individuals is regularly used to make decisions that impact those same individuals. The authors consider settings where sensitive personal data is used to decide who will receive resources or benefits. While it is well known that there is a tradeoff between protecting privacy and the accuracy of decisions, the authors initiate a first-of-its-kind study into the impact of formally private mechanisms (based on differential privacy) on fair and equitable decision-making. They empirically investigate novel tradeoffs on two real-world decisions made using U.S. Census data (allocation of federal funds and assignment of voting rights benefits) as well as a classic apportionment problem. Their results show that

if decisions are made using an ϵ-differentially private version of the data, under strict privacy constraints (smaller ϵ), the noise added to achieve privacy may disproportionately impact some groups over others. They propose novel measures of fairness in the context of randomized differentially private algorithms and identify a range of causes of outcome disparities.

"What Does It Mean to 'Solve' the Problem of Discrimination in Hiring? Social, Technical and Legal Perspectives from the UK on Automated Hiring Systems"
Javier Sánchez-Monedero, Lina Dencik, Cardiff University; Lilian Edwards, University of Newcastle
The ability to get and keep a job is a key aspect of participating in society and sustaining livelihoods. Yet the way decisions are made on who is eligible for jobs, and why, are rapidly changing with the advent and growth in uptake of automated hiring systems (AHSs) powered by data-driven tools. Key concerns about such AHSs include the lack of transparency and potential limitation of access to jobs for specific profiles. In relation to the latter, however, several of these AHSs claim to detect and mitigate discriminatory practices against protected groups and promote diversity and inclusion at work. Yet whilst these tools have a growing user-base around the world, such claims of bias mitigation are rarely scrutinized and evaluated, and when done so, have almost exclusively been from a US socio-legal perspective. In this paper, the authors introduce a perspective outside the US by critically examining how three prominent automated hiring systems (AHSs) in regular use in the UK, HireVue, Pymetrics and Applied, understand and attempt to mitigate bias and discrimination.

"Fairness Is Not Static: Deeper Understanding of Long-Term Fairness via Agents and Environments"
Alexander D'Amour, Yoni Halpern, Hansa Srinivasan, Pallavi Baljekar, James Atwood, D. Sculley, Google
As machine learning becomes increasingly incorporated within high impact decision ecosystems, there is a growing need to understand the long-term behaviors of deployed ML-based decision systems and their potential consequences. Most approaches to understanding or improving the fairness of these systems have focused on static settings without considering long-term dynamics. This is understandable; long term dynamics are hard to assess, particularly because they do not align with the traditional supervised ML research framework that uses fixed data sets. To address this structural difficulty in the field, we advocate for the use of simulation as a key tool in studying the fairness of algorithms. We explore three toy examples of dynamical systems that have been previously studied in the context of fair decision making for bank loans, college admissions, and allocation of attention. By analyzing how learning agents interact with these systems in simulation, we are able to extend previous work, showing that static or single-step analyses do not give a complete picture of the long-term consequences of an ML-based decision system.

"Reducing Sentiment Polarity for Demographic Attributes in Word Embeddings Using Adversarial Learning"
Christopher Sweeney, Maryam Najafian, Massachusetts Institute of Technology (MIT)
The use of word embedding models in sentiment analysis has gained a lot of traction in the Natural Language Processing (NLP) community. However, many inherently neutral word vectors describing demographic identity have unintended implicit correlations with negative or positive sentiment,

resulting in unfair downstream machine learning algorithms. We leverage adversarial learning to decorrelate demographic identity term word vectors with positive or negative sentiment, and re-embed them into the word embeddings. We show that our method effectively minimizes unfair positive/negative sentiment polarity while retaining the semantic accuracy of the word embeddings. Furthermore, we show that our method effectively reduces unfairness in downstream sentiment regression and can be extended to reduce unfairness in toxicity classification tasks.

"Integrating FATE/Critical Data Studies into Data Science Curricula: Where Are We Going and How Do We Get There?"
Jo Bates, David Cameron, Alessandro Checco, Paul Clough, Frank Hopfgartner, Suvodeep Mazumdar, Laura Sbaffi, Peter Stordy, Antonio de le Vega de León, University of Sheffield
There have been multiple calls for integrating FATE/CDS content into Data Science curricula, but little exploration of how this might work in practice. This paper presents the findings of a collaborative auto-ethnography (CAE) undertaken by a MSc Data Science team based at [anonymised] Information School where FATE/CDS topics have been a core part of the curriculum since 2015/16. In this paper, we adopt the CAE approach to reflect on our experiences of working at the intersection of disciplines, and our progress and future plans for integrating FATE/CDS into the curriculum. We identify a series of challenges for deeper FATE/CDS integration related to our own competencies and the wider socio-material context. We conclude with recommendations for ourselves and the wider FATE/CDS orientated Data Science community.

"Roles for Computing in Social Change"
Rediet Abebe, Solon Barocas, Jon Kleinberg, Karen Levy, Manish Raghavan, David G. Robinson, Cornell University
A recent normative turn in computer science has brought concerns about fairness, bias, and accountability to the core of the field. Yet recent scholarship has warned that much of this technical work treats problematic features of the status quo as fixed, and fails to address deeper patterns of injustice and inequality. While acknowledging these critiques, we posit that computational research has valuable roles to play in addressing social problems --- roles whose value can be recognized even from a perspective that aspires toward fundamental social change. In this paper, we articulate four such roles, through an analysis that considers the opportunities as well as the significant risks inherent in such work.

CRAFT Sessions
For full descriptions, visit the ACM FAT* [Craft Sessions program page](#).
- When Not to Design, Build, or Deploy
- Fairness, Accountability, Transparency in AI at Scale: Lessons from National Programs
- Creating Community-Based Tech Policy: Case Studies, Lessons Learned, and What Technologists and Communities Can Do Together
- Lost in Translation: An Interactive Workshop Mapping Interdisciplinary Translations for Epistemic Justice

- From Theory to Practice: Where do Algorithmic Accountability and Explainability Frameworks Take Us in the Real World
- Burn, Dream and Reboot! Speculating Backwards for the Missing Archive on Non-Coercive Computing
- Algorithmically Encoded Identities: Reframing Human Classification
- Ethics on the Ground: From Principles to Practice
- Deconstructing FAT: Using Memories to Collectively Explore Implicit Assumptions, Values and Context in Practices of Debiasing and Discrimination-Awareness
- Bridging the Gap from AI Ethics Research to Practice
- Manifesting the Sociotechnical: Experimenting with Methods for Social Context and Social Justice
- Centering Disability Perspectives in Algorithmic Fairness, Accountability & Transparency
- Infrastructures: Mathematical Choices and Truth in Data
- Hardwiring Discriminatory Police Practices: The Implications of Data-Driven Technological Policing on Minority (Ethnic and Religious) People and Communities
- CtrlZ.AI Zine Fair: Critical Perspectives (Off-site)

Tutorials

- Probing ML Models for Fairness with the What-If Tool and SHAP
- AI Explainability
- Experimentation with fairness-aware recommendation using librec-auto
- Leap of FATE: Human Rights as a Complementary Framework for AI Policy and Practice
- Can an algorithmic system be a 'friend' to a police officer's discretion?
- Two computer scientists and a cultural scientist get hit by a driver-less car: Situating knowledge in the cross-disciplinary study of F-A-T in machine learning
- Positionality-Aware Machine Learning
- Policy 101: An Introduction to Participating in the Policymaking Process
- From the Total Survey Error Framework to an Error Framework for Digital Traces of Humans
- The Meaning and Measurement of Bias: Lessons from NLP
- Assessing the intersection of Organizational Structure and FAT* efforts within industry
- Explainable AI in Industry: Practical Challenges and Lessons Learned
- Gender: What the GDPR does not tell us (But maybe you can?)
- What does "fairness" mean in (data protection) law?

###