FAT*2020

Barcelona, Spain

January 27th to 30th, 2020

Floor Plan

Level E

- MR8
- MR7
- MR6
- MR5
- North Shuttle
- Plenary Room
- Orbital Showroom
- Registration Desk
- Orbital Shuttle
- MR12
- MR13
- Nitrogen Bar
- MR14
- Hydrogen Restaurant

Level 0

- MR3
- Nursery
- MR2
- Prayer Room
- MR1
- Orbital Bar
- Orbital Showroom
- WC
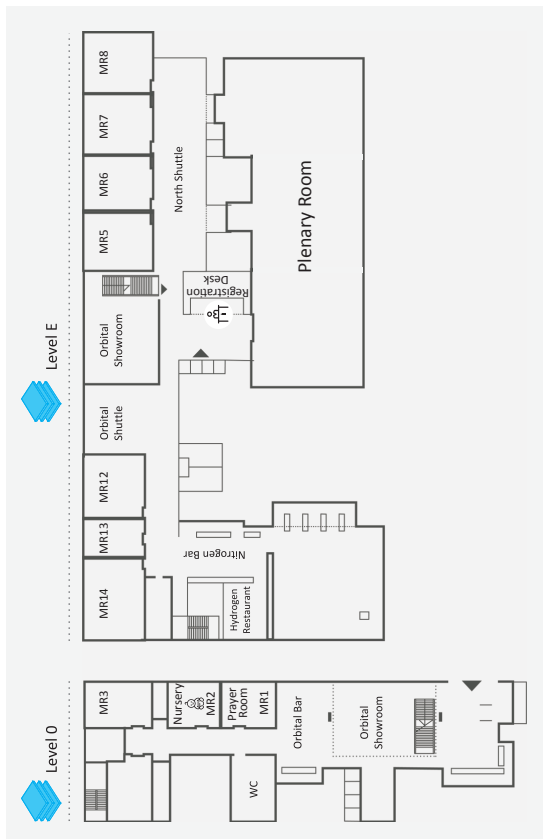
Meet**Barceló**
HOTEL GROUP

Welcome

Welcome to **ACM FAT*2020** in Barcelona!

While FAT* is only in its third year, it has become the focus for excellent research into fairness, accountability, and transparency of computer systems. For 2020, our goal has been to sustain and further improve the high quality of computer science research in this domain, and to continue developing a diverse, inter-disciplinary research community. To this end, we have appointed dedicated selection committees securing excellence in computer science (CS), social sciences and humanities (SSH), and legal scholarship (LAW), thus creating a rich program centered on computer science with strong and diverse cross-disciplinary components.

We are particularly delighted to host this conference in Europe for the first time, and have done our best to reflect and emphasize issues and research in Europe on these topics.

Thank you for attending ACM FAT*2020. Enjoy the conference, the hotel, the food, and the (hopefully relatively) warm weather in the fantastic city of Barcelona!

Mireille Hildebrandt and Carlos Castillo
ACM FAT*2020 General Co-Chairs

https://fatconference.org/2020

Diversity and Inclusion —

We want to provide an open, diverse, safe and inclusive atmosphere where everyone feels comfortable to freely express who they are, and to voice their own opinions and points of view. Please:

- **communicate clearly and respectfully** while acknowledging others' feelings;

- **do not imply that there are things that everybody should know** as we all have different backgrounds;

- **refrain from offensive or disparaging comments**, particularly those related to gender identity and expression, sexual orientation, race, ethnicity, language, neuro-type, size, ability, class, religion, culture, subculture, political opinion, age, skill level, occupation, or background;

- **connect with the diversity and inclusion chair** if you become aware of disrespectful behavior in the above sense;

- **alert organizers** if you notice a dangerous situation or someone in distress.

The Policy Against Harassment at ACM Activities applies during this conference; see https://fatconference.org/2020/inclusion.html for details and contact information.

Program Schedule

Monday 27th

Time			
9:00 - 10:30	**Doctoral Consortium**		
	Room: MR5+MR6		
10:30 - 10:45	Coffee break ☕		
10:45 - 12:15	**Doctoral Consortium**		
	Room: MR5+MR6		
12:15 - 13:00	Lunch for Tutorial and Doctoral Consortium Attendees 🍴		
	Tutorials 1		
13:00 - 14:30	Implications Tutorial	Hands-On Tutorial	Hands-On Tutorial
	Explainable AI in Industry: Practical Challenges and Lessons Learned	Experimentation with Fairness-Aware Recommendation Using LibRec-Auto	AI Explainability 360
	Room: Plenary	Room: MR8	Room: MR7
	Translation Tutorial	Translation Tutorial	
	Policy 101: An Introduction to Participating in the Policymaking Process	Positionality-Aware Machine Learning	
	Room: MR5+MR6	Room: MR12	
14:30 - 15:00	Coffee break ☕		

🕐 Tutorials 2

Invited Tutorial	Hands-On Tutorial (cont.)	Hands-On Tutorial (cont.)
What Does "Fairness" Mean in (Data Protection) Law?	Experimentation with Fairness-Aware Recommendation Using LibRec-Auto	AI Explainability 360
Room: Plenary	Room: MR8	Room: MR7
Translation Tutorial	Implications Tutorial	Translation Tutorial
Leap of FATE: Human Rights as a Complementary Framework for AI Policy and Practice	Assessing the Intersection of Organizational Structure and FAT* Efforts within Industry	From the Total Survey Error Framework to an Error Framework for Digital Traces of Humans
Room: MR5+MR6	Room: MR12	Room: MR3

15:00 - 16:30

16:30 - 16:45 Short Break / Refreshments ⏸

Tutorials 3

Translation Tutorial	Invited Tutorial	Translation Tutorial
The Meaning and Measurement of Bias: Lessons from NLP	Gender: What the GDPR does not Tell Us (But Maybe You Can?)	Two Computer Scientists and a Cultural Scientist Get Hit by a Driver-Less Car …
Room: Plenary	Room: MR8	Room: MR7
Hands-On Tutorial	Translation Tutorial	
Probing ML models for Fairness with the What-If Tool and SHAP	Can an Algorithmic System be a "Friend" to a Police Officer's Discretion?	
Room: MR5+MR6	Room: MR12	

16:45 - 18:15

8:45 - 9:00	Opening Remarks and Tuesday Overview
9:00 - 10:50	**Research Track** Plenary room **Session 1: Accountability** **Session 2: Explainability 1**
10:50 - 11:20	Coffee break ☕
11:20 - 12:15	**Research Track** Plenary room **Session 3: Auditing/Assessment 1**
12:15 - 13:15	Keynote by Ayanna Howard
13:15 - 14:15	Lunch 🍴
14:15 - 16:05	**Research Track** Plenary room **Session 4: Fairness 1** **Session 5: Ethics and Policy**
16:05 - 16:35	Coffee Break ☕
16:35 - 18:05	**Research Track** Plenary room **Session 6: Values** **Session 7: Data Collection**
18:05 - 18:30	Pause ⏸
18:30 - 19:30	Reception 🍸
19:30 - 21:00	Catalan Culture Event

8:45 - 9:00	**Wednesday Overview**
9:00 - 10:50	**Research Track** Plenary room ☰ **Session 8: Fairness 2** **Session 9: Cognition and Education**
10:50 - 11:20	Coffee break ☕
11:20 - 12:15	**Research Track** Plenary room ☰ **Session 10: Auditing/Assessment 2**
12:15 - 13:15	Keynote by Yochai Benkler
13:15 - 14:15	Lunch 🍴
14:15 - 14:50	**Research Track** Plenary room ☰ **Session 11: Sensitive Attributes**
14:50 - 15:00	Short Break to Switch Rooms ⏸

CRAFT 1			
Fairness, Accountability, Transparency in AI at Scale	Bridging the Gap from AI Ethics Research to Practice	Lost in translation: An Interactive Workshop	Ethics on the Ground: From Principles to Practice
Plenary Room	Room: MR8	Room: MR6	OFF-SITE: Telefonica I+D
From Theory to Practice	Burn, Dream and Reboot!	Algorithmically Encoded Identities	Creating Community-Based Tech Policy
Room: MR7	Room: MR12	Room: MR3	Room: MR5

15:00 - 16:30 (applies to CRAFT 1 rows above)

16:30 - 17:00	Coffee break ☕

CRAFT 2			
When Not to Design, Build, or Deploy	Manifesting the Sociotechnical	Lost in translation (cont.)	Ethics on the Ground (cont.)
Plenary Room	Room: MR8	Room: MR6	OFF-SITE: Telefonica I+D
From Theory to Practice (cont.)	Burn, Dream and Reboot (cont.)	Algorithmically Encoded Identities (cont.)	Creating Community-Based Tech Policy (cont.)
Room: MR7	Room: MR12	Room: MR3	Room: MR5

17:00 - 18:30 (applies to CRAFT 2 rows above)

8:45 - 9:00	Thursday Overview
9:00 - 10:50	**Research Track** Plenary room ☰ **Session 12: Fairness 3** **Session 13: Auditing/Assessment 3**
10:50 - 11:20	Coffee break ☕
11:20 - 12:15	**Research Track** Plenary room ☰ **Session 14: Explainability 2**
12:15 - 13:15	Keynote by Nani Jansen Reventlow
13:15 - 14:15	Lunch 🍴
14:15 - 14:50	**Research Track** Plenary room ☰ **Session 15: Fairness 4**
14:50 - 15:00	Short Break to Switch Rooms ⏸

🕐	**CRAFT 3**			
	Rump Session	Centering Disability Perspectives	Ethics on the Ground (cont.)	Deconstructing FAT (cont.)
15:00 - 16:30	Plenary Room	Room: MR13+MR14	Room: MR7	Room: MR5
	Infrastructures	Hardwiring Discriminatory Police Practices	CtrlZ.AI Zine Fair (Starts 13:30)	
	Room: MR8	Room: MR12	OFF-SITE: Flowers by Bornay	
16:30 - 17:00	Coffee break ☕			
17:00 - 18:15	**Town Hall Meeting** Town Hall Part 1 (Form): Organizational Aspects — 15 minutes Town Hall Part 2 (Content): Trends and Blindspots in FAT* Research — 60 minutes			
18:15 - 18:25	**Awards and Closing** 🏅			

Accepted Papers

Papers on this list are sorted by sessions.

Tuesday 28th

Session 1: Accountability

What to account for when accounting for algorithms: A systematic literature review on algorithmic accountability

M. Wieringa

Algorithmic Realism: Expanding the Boundaries of Algorithmic Thought

B. Green; S. Viljoen

Algorithmic Accountability in Public Administration: The GDPR Paradox

S. Kang

Closing the AI Accountability Gap: Defining a SMArTeR Framework for Internal Algorithmic Auditing

D. Raji; A. SMART; R. White; M. Mitchell; T. Gebru; B. Hutchinson; J. Smith-Loud; D. Theron; P. Barnes

Toward Situated Interventions for Fairness, Accountability, and Transparency: Lessons from the field

M. Katell; M. Young; B. Herman; D. Dailey; V. Guetler; A. Tam; C. Binz; D. Raz; P. Krafft

Session 2: Explainability 1

Explainability Fact Sheets: A Framework for Systematic Assessment of Explainable Approaches

K. Sokol; P. Flach

Multi-layered Explanation from Algorithmic Impact Assessments in the GDPR

G. Malgieri; M. Kaminski

The Hidden Assumptions Behind Counterfactual Explanations and Principal Reasons

S. Barocas; A. Selbst; M. Raghavan

Why Does My Model Fail? Contrastive Local Explanations for Retail Forecasting

A. Lucic; H. Haned; M. de Rijke

The Human Body is a Black Box: Supporting Clinical Decision-Making with Deep Learning

M. Sendak; M. Elish; M. Gao; W. Ratliff; M. Nichols; J. Futoma; A. Bedoya; S. Balu; C. O'Brien

📋 Session 3: Auditing/Assessment 1

Assessing Algorithmic Fairness with Unobserved Protected Class Using Data Combination

N. Kallus; X. Mao; A. Zhou

FlipTest: Fairness Testing via Optimal Transport

E. Black; S. Yeom; M. Fredrikson

Implications of AI (Un-)Fairness in Higher Education Admissions: The Effects of Perceived AI (Un-)Fairness on Exit, Voice and Organizational Reputation

F. Marcinkowski; K. Kieslich; C. Starke; M. Lünich

Auditing Radicalization Pathways on YouTube

M. Ribeiro; R. Ottoni; R. West; V. Almeida; W. Meira Jr.

Case Study: Predictive Fairness to Reduce Misdemeanor Recidivism Through Social Service Interventions

K. Rodolfa; E. Salomon; L. Haynes; I. Mendieta; J. Larson; R. Ghani

📋 Session 4: Fairness 1

What "Fairness" means? A linguistic and contextual interpretation from the GDPR

G. Malgieri

Studying Up: Reorienting the study of algorithmic fairness around issues of power

C. Barabas; C. Doyle; J. Rubinovitz; K. Dinakar

POTs: Protective Optimization Technologies
R. Overdorf; B. Kulynych; E. Balsa; C. Troncoso; S. Gürses

Fair Decision Making using Privacy-Protected Data
S. Kuppam; R. McKenna; D. Pujol; M. Hay; A. Machanavajjhala; G. Miklau

Fairness Warnings and Fair-MAML: Learning Fairly with Minimal Data
D. Slack; S. Friedler; E. Givental

📋 Session 5: Ethics and Policy

From Ethics Washing to Ethics Bashing: A View on Tech Ethics from Within Moral Philosophy
E. Bietti

Onward for the freedom of others: Marching beyond the AI Ethics
P. Terzis

Whose Side are Ethics Codes On? Power, Responsibility and the Social Good
A. Washington; R. Kuo

Algorithmic Targeting of Social Policies: Fairness, Accuracy, and Distributed Governance
A. Noriega-Campero; B. Bulle-Bueno; L. Cantu; M. Bakker; L. Tejerina; A. Pentland

Roles for Computing in Social Change
R. Abebe; S. Barocas; J. Kleinberg; K. Levy; M. Raghavan; D. Robinson

📋 Session 6: Values

Regulating Transparency? Facebook, Twitter and the German Network Enforcement Act
B. Wagner; K. Rozgonyi; M. Sekwenz; J. Singh; J. Cobbe

The relationship between trust in AI and trustworthy machine learning technologies
E. Toreini; M. Aitken; A. van Moorsel; K. Elliott; K. Coopamootoo

The philosophical basis of algorithmic recourse
S. Venkatasubramanian; M. Alfano

Value-laden Progress in Machine Learning

R. Dotan; S. Milli

Effect of Confidence and Explanation on Accuracy and Trust Calibration in AI-Assisted Decision Making

Y. Zhang; Q. Liao; R. Bellamy

Session 7: Data Collection

Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning

E. Jo; T. Gebru

Data in New Delhi's predictive policing system

V. Marda; S. Narayan

Garbage In, Garbage Out: Do Machine Learning Application Papers in Social Computing Report Where Human-Labeled Training Data Comes From?

R. Geiger; K. Yu; Y. Yang; M. Dai; J. Qiu; R. Tang; J. Huang

Wesnesday 29th

Session 8: Fairness 2

Bidding Strategies with Gender Nondiscrimination Constraints for Online Ad Auctions

M. Nasr; M. Tschantz

Multi-category Fairness in Sponsored Search Auctions

C. Ilvento; M. Jagadeesan; S. Chawla

Reducing Sentiment Polarity for Demographic Attributes in Word Embeddings using Adversarial Learning

C. Sweeney; M. Najafian

Interventions for Ranking in the Presence of Implicit Bias

A. Mehrotra; L. Celis; N. Vishnoi

The Disparate Equilibria of Algorithmic Decision Making when Individuals Invest Rationally

L. Liu; A. Wilson; N. Haghtalab; A. Kalai; C. Borgs; J. Chayes

Session 9: Cognition and Education

An Empirical Study on the Perceived Fairness of Realistic, Imperfect Machine Learning Models

G. Harrison; J. Hanson; C. Jacinto; J. Ramirez; B. Ur

Artificial mental phenomena: Psychophysics as a framework to detect perception biases in AI models

L. Liang; D. Acuna

The Social Lives of Generative Adversarial Networks

M. Castelle

Towards a more representative politics in the ethics of computer science

J. Moore

Integrating FATE/Critical Data Studies into Data Science curricula: where are we going and how do we get there?

J. Bates; D. Cameron; A. Checco; P. Clough; F. Hopfgartner; S. Mazumdar; L. Sbaffi; P. Stordy; A. de León

Session 10: Auditing/Assessment 2

Recommendations and User Agency: The Reachability of Collaboratively-Filtered Information

S. Dean; S. Rich; B. Recht

Bias in word embeddings

O. Papakyriakopoulos; S. Hegelich; J. Serrano; F. Marco

What does it mean to 'solve' the problem of discrimination in hiring? Social, technical and legal perspectives from the UK on automated hiring systems

J. Sánchez-Monedero; L. Dencik; L. Edwards

Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices

M. Raghavan; S. Barocas; J. Kleinberg; K. Levy

Quantifying the impact of overbooking on a pre-trial risk assessment tool

K. Lum; C. Boudin; M. Price

📖 Session 11: Sensitive Atributes 2

Awareness in Practice: Tensions in Access to Sensitive Attribute Data for Antidiscrimination

A. Rieke; M. Bogen; S. Ahmed

Towards a Critical Race Methodology in Algorithmic Fairness

E. Denton; A. Hanna; J. Smith-Loud; A. Smart

What's Sex Got to Do With Fair Machine Learning?

L. Hu; I. Kohler-Hausmann

Thursday 30th

📖 Session 12: Fairness 3

On the Apparent Conflict Between Individual and Group Fairness

R. Binns

Fairness Is Not Static: Deeper Understanding of Long Term Fairness via Agents and Environments

A. D'Amour; Y. Halpern; H. Srinivasan; P. Baljekar; J. Atwood; D. Sculley

Fair Classification and Social Welfare

L. Hu; Y. Chen

Preference-Informed Fairness

M. Kim; A. Korolova; G. Rothblum; G. Yona

Towards Fairer Datasets: Filtering and Balancing the Distribution of the People Subtree in the ImageNet Hierarchy

K. Yang; K. Qinami; L. Fei-Fei; J. Deng; O. Russakovsky

🔵 Session 13: Auditing/Assessment 3

The Case for Voter-Centered Audits of Search Engines during Political Elections

E. Mustafaraj; E. Lurie; C. Devine

Whose Tweets are Surveilled for the Police: An Audit of a Social-Media Monitoring Tool via Log Files

G. Borradaile; B. Burkhardt; A. LeClerc

Dirichlet uncertainty wrappers for actionable algorithm accuracy accountability and auditability

J. Mena Roldán; O. Pujol Vila; J. Vitrià Marca

Counterfactual Risk Assessments, Evaluation, and Fairness

A. Coston; A. Chouldechova; E. Kennedy; A. Mishler

The False Promise of Risk Assessments: Epistemic Reform and the Limits of Fairness

B. Green

🔵 Session 14: Explainability 2

Explaining Machine Learning Classifiers through Diverse Counterfactual Explanations

R. Mothilal; A. Sharma; C. Tan

Model Agnostic Interpretability of Text Rankers via Intent Modelling

J. Singh; A. Anand

Doctor XAI: An ontology-based approach to black-box sequential data classification explanations

C. Panigutti; A. Perotti; D. Pedreschi

Objectivity in ML Explanations: Why Robustness Matters

L. Hancox-Li

Explainable Machine Learning in Deployment

U. Bhatt; A. Xiang; S. Sharma; A. Weller; A. Taly; Y. Jia; J. Ghosh; R. Puri; J. Moura; P. Eckersley

Session 15: Fairness 4

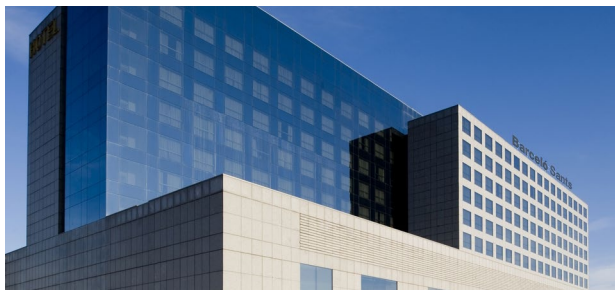Fairness and Utilization in Allocating Resources with Uncertain Demand

K. Donahue; J. Kleinberg

The Effect of Competition and Regulation on Error Inequality in Data-Driven Markets

H. Elzayn; B. Fish

Measuring Injustice in Machine Learning: Difference, Discrimination, and Disability

A. Lundgard

📍 **Barceló Sants**
Plaça dels Països Catalans, s/n
08014 Barcelona – Spain

• Airport:
El Prat (16 km away, 20 minutes from the hotel)

• Public Transportation and Taxis:
AVE Train Station – Barcelona Train Station – Taxi stop in front of the hotel

• City Centre:
5 minutes by train, 15 minutes by car

FAT*2020

Barcelona, Spain