Welcome

We Will Be Starting Shortly
Welcome and Introductory Remarks

Bruce Hoffman
Federal Trade Commission
Bureau of Competition
Algorithmic Collusion

Session moderated by:

Ellen Connelly
Federal Trade Commission
Office of Policy Planning

James Rhilinger
Federal Trade Commission
Bureau of Competition
Algorithmic Collusion

Maurice E. Stucke
University of Tennessee College of Law
Algorithmic Collusion

Ai Deng
Bates White
Algorithmic Collusion

Kai-Uwe Kühn
University of East Anglia
Algorithmic Collusion

Rosa M. Abrantes-Metz
Global Economics Group
Algorithmic Collusion

Sonia Kuester Pfaffenroth
Arnold & Porter
Algorithmic Collusion

Joseph E. Harrington, Jr.
University of Pennsylvania
Algorithmic Collusion

Panel Discussion:

Maurice E. Stucke, Ai Deng, Kai-Uwe Kühn,
Rosa M. Abrantes-Metz,
Sonia Kuester Pfaffenroth,
Joseph E. Harrington, Jr.,

Moderators: Ellen Connelly & James Rhilinger
Break

10:45-11:00 am
Framing Presentation
(prerecorded)

Michael I. Jordan
University of California, Berkeley
Emerging Competition, Innovation, and Market Structure Questions Around Algorithms, Artificial Intelligence, and Predictive Analytics

Session moderated by:

Brian O’Dea
Federal Trade Commission
Bureau of Competition

Nathan Wilson
Federal Trade Commission
Bureau of Economics
Emerging Competition, Innovation, and Market Structure Questions Around Algorithms, Artificial Intelligence, and Predictive Analytics

Panel Discussion:

Robin Feldman, Joshua Gans, Preston McAfee, Nicolas Petit

Moderators: Brian O’Dea & Nathan Wilson
Facial Analysis Technology
Warning Signs

Joy Buolamwini
Algorithmic Justice League | MIT Media Lab
PhD, MIT Pending
Automated Facial Analysis Tasks

1. **DETECT**
   - Is there a face?
   - **YES**
   - **IDENTIFY or VERIFY**
     - Have I seen this face?
     - Am I looking for this face?
   - **NO**
   - **CLASSIFY ATTRIBUTE**
     - What type of face?
     - (age, gender, emotion, etc.)

2. **IDENTIFY or VERIFY**
   - Identify: one-to-many
     - ex. police search
   - Verify: one-to-one
     - Is this the same face?
     - ex. unlock phone, pay with face
The Coded Gaze

Algorithmic bias creating exclusionary experiences discriminatory practices
Coded Gaze Score: 4/13

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age*</th>
<th>Detected</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBM</td>
<td>M</td>
<td>✔️</td>
</tr>
<tr>
<td>Microsoft</td>
<td>X</td>
<td>☒️</td>
</tr>
<tr>
<td>Face++</td>
<td>X</td>
<td>☒️</td>
</tr>
<tr>
<td>Kairos</td>
<td>M</td>
<td>✔️</td>
</tr>
<tr>
<td>Google</td>
<td>NA</td>
<td></td>
</tr>
</tbody>
</table>

Score: 0/4, 1/4, 3/5
Silent Sweep: Over 117 Million US Adults in Face Surveillance Databases

One in two American adults is in a law enforcement face recognition network used in unregulated searches employing algorithms with unaudited accuracy.

The Perpetual Line Up
(Garvie, Bedoya, Frankle 2016)
Real-World Impact

“In two cases [Scotland Yard Report], innocent women were matched with men.”

- Ian Drury, The Daily Mail – May 15 2018
Expanding Use of Technology

GET THE BEST TALENT, FASTER

HIREVUE
HIRING INTELLIGENCE

SEE HOW
### Potential Harms Index

<table>
<thead>
<tr>
<th>INDIRECT HARMED</th>
<th>COLLECTIVE SOCIAL HARM</th>
<th>DIRECT HARMED</th>
<th>SOCIAL HARM</th>
</tr>
</thead>
<tbody>
<tr>
<td>ILLEGAL DISCRIMINATION</td>
<td></td>
<td>UNFAIR PRACTICES</td>
<td></td>
</tr>
<tr>
<td>HIRING</td>
<td></td>
<td>EMPLOYMENT</td>
<td>LOSS OF OPPORTUNITY</td>
</tr>
<tr>
<td>INSURANCE &amp; SOCIAL BENEFITS</td>
<td></td>
<td>HOUSING</td>
<td>ECONOMIC LOSS</td>
</tr>
<tr>
<td>EDUCATION</td>
<td></td>
<td>CREDIT</td>
<td>SOCIAL STIGMATIZATION</td>
</tr>
<tr>
<td>DIFFERENTIAL PRICES OF GOODS</td>
<td></td>
<td>LOSS OF LIBERTY</td>
<td></td>
</tr>
<tr>
<td>INCREASED SURVEILLANCE</td>
<td></td>
<td>INCREASED SURVEILLANCE</td>
<td></td>
</tr>
<tr>
<td>STEREOTYPE REINFORCEMENT</td>
<td></td>
<td>STEREOTYPE REINFORCEMENT</td>
<td></td>
</tr>
<tr>
<td>DIGNATORY HARMS</td>
<td></td>
<td>DIGNATORY HARMES</td>
<td></td>
</tr>
</tbody>
</table>
Gender Shades
Intersectional Accuracy Disparities in Commercial Gender Classification

230+ articles in 37+ countries on MIT Thesis Research findings

Gold Standard Measures of Success Mislead

Data is Destiny
Does your data reflect the world?

BENCHMARK SKEWS
80% PALE 75% MALE
2014
DEEPPFACE
97.35%
ACCURACY ON
GOLD STANDARD
LFW BENCHMARK
(Taigman et al., 2014)

GOLD STANDARD SKEWS
Labeled Faces in The Wild
Released in 2007

~77.5% Male
~83.5% White
(Han and Jain, 2014)
National Benchmarks Not Immune

NIST 2015 IJB-A BENCHMARK

INTERSECTIONAL BREAKDOWN

4.4% Darker Female
20.2% Lighter Female

59.4% Lighter Male
16% Darker Male

SINGLE AXIS

24.6% Female
75.4% Male
Towards Better Evaluation

**PILOR PARLIAMENTS BENCHMARK**

**FIRST GENDER AND SKIN TYPE LABELED GENDER CLASSIFICATION BENCHMARK**

54.4% Male
53.6% Lighter
Testing Commercial AI Systems

How accurate are systems from IBM, Microsoft, and Face++ at determining the gender of faces in inclusive benchmark?
Overall Accuracy

Aggregate performance metrics can mask racial and gender bias

93.7%  90%  87.9%

www.gendershades.org

May 2017 PPB Results
Gender Bias

All companies perform better on men than women

<table>
<thead>
<tr>
<th></th>
<th>Female Faces</th>
<th>Male Faces</th>
</tr>
</thead>
<tbody>
<tr>
<td>FACE++</td>
<td>78.7%</td>
<td>99.3%</td>
</tr>
<tr>
<td>IBM</td>
<td>79.7%</td>
<td>94.4%</td>
</tr>
<tr>
<td><strong>8-21% Error Gap</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

www.gendershades.org

May 2017 PPB Results
Skin Type ~ Racial Bias

All companies perform better on whites than people of color

<table>
<thead>
<tr>
<th>Company</th>
<th>Darker Faces</th>
<th>Lighter Faces</th>
</tr>
</thead>
<tbody>
<tr>
<td>FACE++</td>
<td>83.5%</td>
<td>95.3%</td>
</tr>
<tr>
<td>IBM</td>
<td>77.6%</td>
<td>96.8%</td>
</tr>
</tbody>
</table>

12-19% Error Gap

Darker

Lighter

www.gendershades.org

May 2017 PPB Results
Intersectional Performance

94%  79.2%  100%  98.3%

DARKER MALES  DARKER FEMALES  LIGHTER MALES  LIGHTER FEMALES

May 2017 PPB Results
Intersectional Performance

99.3% 65.5% 99.2% 94.0%

DARKER MALES  DARKER FEMALES  LIGHTER MALES  LIGHTER FEMALES

May 2017 PPB Results
Intersectional Performance

88%  65.3%  99.7%  92.9%

May 2017 PPB Results
Further Disaggregation Uncovers Even Higher Error Rates

May 2017 PPB Results

<table>
<thead>
<tr>
<th></th>
<th>TYPE I</th>
<th>TYPE II</th>
<th>TYPE III</th>
<th>TYPE IV</th>
<th>TYPE V</th>
<th>TYPE VI</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBM</td>
<td>5.1%</td>
<td>7.4%</td>
<td>8.2%</td>
<td>8.3%</td>
<td>33.3%</td>
<td>46.8%</td>
</tr>
<tr>
<td>FACE++</td>
<td>11.9%</td>
<td>9.7%</td>
<td>8.2%</td>
<td>13.9%</td>
<td>32.4%</td>
<td>46.5%</td>
</tr>
<tr>
<td></td>
<td>1.7%</td>
<td>1.1%</td>
<td>3.3%</td>
<td>0%</td>
<td>23.2%</td>
<td>25.0%</td>
</tr>
</tbody>
</table>

**Commercial Error Rates Per Skin Type on Female Labeled Faces in PPB**
Company Responses to Gender and Racial Bias in Commercial AI Systems

IBM and Microsoft engaged researchers

All companies released new products within 7 months of receiving audit results
Self-Reported Improvement

February 2018 Internal IBM Results

98%  96.5%  99.8%  100%

DARKER MALES  DARKER FEMALES  LIGHTER MALES  LIGHTER FEMALES

Self-Reported Results With .99 Treshold
External Follow-Up Evaluation

August 2018 PPB Results

99.4%  83.0%  99.7%  97.6%

Accuracy Determined Using Gender Label Returned By API
Accuracy Doesn’t Mitigate Abuse

IBM USED NYPD SURVEILLANCE FOOTAGE TO DEVELOP TECHNOLOGY THAT LETS POLICE SEARCH BY SKIN COLOR
Regulators Mitigate Abuse

Gender Shades Research Supported Recommendations

• Require Vendors of Facial Analysis Technology To:
  • Implement internal bias evaluation, mitigation, and reporting procedures
  • Regularly report performance on national benchmarks
  • Support independent evaluation from research community

• Require National Institute of Standards & Technology To:
  • Make public demographic and phenotypic composition of benchmarks
  • Report accessible intersectional performance metrics
Regulators Mitigate Abuse

Broader Considerations

• **Consent and Control:** Ensure consumers have meaningful opportunity to consent or refuse capture of face and ability to control use of face data – (Require companies like Facebook Provide Face Purge Option)

• **Transparency:** Require disclosure when facial analysis technology is in use and information about storage and use of face data

• **Due Process:** Provide mechanisms for redress and contestation of decisions made with or informed by facial analysis technology

• **Heightened Privacy:** Recognize that face images are identifying information, and enable processors to determine consumers’ precise geolocation information
For More Information Contact

comms@ajlunited.org
Lunch
1:00-2:15 pm
Fairness and Intelligibility in Machine Learning Systems

Jenn Wortman Vaughan
Microsoft Research
The Age of AI

NIPS Registrations

![Image showing NIPS Registrations graph from 2002 to 2016]
New Challenges

Online Ads for High-Paying Jobs Are Targeting Men More Than Women

New study uncovers gender bias

When Algorithms Discriminate

The online world is shaped by forces beyond our control, determining the stories we read on Facebook, the people we meet on OkCupid and the search results we see on Google. Big data is used to make decisions about health care, employment, housing, education and policing.

Amazon just showed us that 'unbiased' algorithms can be inadvertently racist

Google apologises for Photos app racist blunder

Do Google's 'unprofessional hair' results show it is racist?

Leigh Alexander

When it Comes to Policing, Data Is Not Benign

Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Justin Aragon, Jeff Lennie, Surya Mattu and Lauren Kirchner, ProPublica

May 29, 2016

ON A SPRING AFTERNOON IN 2014, Brisha Borden was running late to pick up her god-sister from school when she spotted an unlocked kid's blue Huffy bicycle and a silver Razor scooter. Borden
Microsoft’s AI Principles

- Fairness
- Reliability & Safety
- Privacy & Security
- Inclusiveness
- Transparency
- Accountability
FATE: Fairness, Accountability, Transparency, and Ethics in AI
What are machine learning and AI?
AI

Computers doing things that we would normally think of as intelligent
MACHINE LEARNING
Systems that learn from DATA and EXPERIENCE instead of being explicitly programmed

AI
Computers doing things that we would normally think of as *intelligent*
MACHINE LEARNING
Systems that learn from DATA and EXPERIENCE instead of being explicitly programmed
Types of Machine Learning

- **Supervised learning**: Use labeled data to learn a general rule mapping inputs to outputs
- **Unsupervised learning**: Identify hidden structure and patterns in data; cluster data points
- **Reinforcement learning**: Perform a task, such as driving a vehicle or playing a game, in a dynamic environment, learning through trial and error
Why might a machine learning system be unfair?
The Machine Learning Pipeline

- Task Definition
- Dataset Construction
- Model Definition
- Training Process
- Testing Process
- Deployment
- Feedback
Task Definition

- Task Definition
- Dataset Construction
- Model Definition
- Testing Process
- Training Process
- Deployment
- Feedback
Task Definition

(a) Three samples in criminal ID photo set $S_c$.

(b) Three samples in non-criminal ID photo set $S_n$

Figure 1. Sample ID photos in our data set.

(Wu and Zhang, 2016)
Dataset Construction

- Task Definition
- Dataset Construction
- Model Definition
- Testing Process
- Deployment
- Feedback
- Training Process
Data: Societal Bias

Amazon scraps secret AI recruiting tool that showed bias against women

SAN FRANCISCO (Reuters) - Amazon.com Inc’s (AMZN.O) machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.
Data: Societal Bias

(Caliskan et al., 2017)
Data: Societal Bias

He is a nurse. She is a doctor.

O bir hemşire. O bir doktor.

O bir hemşire. O bir doktor.

She's a nurse. He's a doctor.
Data: Skewed Sample

(Buolamwini and Gebru, 2018)
Data: Labeler Bias

More States Opting To 'Robo-Grade' Student Essays By Computer

June 30, 2018 · 8:13 AM ET
Heard on Weekend Edition Saturday

TOVIA SMITH
Model Definition

Feedback → Task Definition → Dataset Construction → Model Definition

Model Definition → Training Process

Training Process → Testing Process

Testing Process → Deployment

Deployment → Feedback
price of house  =  \( w_1 \) * number of bedrooms
+ \( w_2 \) * number of bathrooms
+ \( w_3 \) * square feet
+ a little bit of noise
Model: Assumptions

Artificial Intelligence Is Now Used to Predict Crime. But Is It Biased?
The software is supposed to make policing more fair and accountable. But critics say it still has a way to go.
Training Process

1. Task Definition
2. Dataset Construction
3. Model Definition
4. Testing Process
5. Deployment
6. Feedback
7. Training Process
Training Process

price of house = $w_1 \times$ number of bedrooms
+ $w_2 \times$ number of bathrooms
+ $w_3 \times$ square feet
+ a little bit of noise
Trainig Process

```python
if schedule['Lambda_SKCC'] <= self.total_iter:
    start = time.time()

    shp_SKCC[::] = np.outer(W_d_C, W_d_C)
    shp_SKCC[::, :, bool_diag_CC] = W_a_C * W_d_C
    shp_SKCC *= Y_K[None, :, None, None]
    shp_SKCC *= W_S[None, None, None, :]
    post_shp_SKCC = shp_SKCC + Y_SKCC

    if mask.ndim == 2:
        mask_NN = mask
        zeta_TNC = np.dot(mask_NN, Theta_NC)
        zeta_TCC = np.sum(Theta_NC, axis=0)[None, None, None]
    else:
        mask_TNN = mask
        zeta_TNC = np.einsum('tij,jd->tid', mask_TNN, Theta_NC)
        zeta_TCC = np.einsum('tid,ic->tcd', zeta_TNC, Theta_NC)
        zeta_SCC = np.einsum('tcd,ts->scd', zeta_TCC, Psi_TS)
        post_rte_SKCC = d + zeta_SCC[:, None, None, :]
```
Testing Process

1. **Task Definition**
2. **Dataset Construction**
3. **Model Definition**
4. **Training Process**
5. **Testing Process**
6. **Deployment**
7. **Feedback**

The process is cyclic, starting from **Task Definition** and moving through **Dataset Construction**, **Model Definition**, **Training Process**, **Testing Process**, **Deployment**, and ending with **Feedback** to initiate the cycle again.
Testing: Metrics

Translation tutorial: 21 fairness definitions and their politics

Arvind Narayanan
(Computer scientist, Princeton University)

Computer scientists and statisticians have devised numerous mathematical criteria to define what it means for a classifier or a model to be fair. The proliferation of these definitions represents an attempt to make technical sense of the complex, shifting social understanding of fairness. Thus, these definitions are laden with values and politics, and seemingly technical discussions about mathematical definitions in fact implicate weighty normative questions. A core component of these technical discussions has been the discovery of trade-offs between different (mathematical) notions of fairness; these trade-offs deserve attention beyond the technical community.
# Testing: Metrics

<table>
<thead>
<tr>
<th></th>
<th>Unqualified</th>
<th>Qualified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject</td>
<td>TN</td>
<td>FN</td>
</tr>
<tr>
<td>Hire</td>
<td>FP</td>
<td>TP</td>
</tr>
</tbody>
</table>
What is the probability that a woman is qualified given that you choose to hire her? What about a man?

Predictive parity requires (almost) equal values of

\[
\frac{TP}{TP + FP}
\]
Testing: Metrics

<table>
<thead>
<tr>
<th></th>
<th>Unqualified</th>
<th>Qualified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject</td>
<td>TN</td>
<td>FN</td>
</tr>
<tr>
<td>Hire</td>
<td>FP</td>
<td>TP</td>
</tr>
</tbody>
</table>

What is the probability of hiring a woman if she is unqualified? What about a man?

False positive rate balance requires (almost) equal values of

\[
\frac{FP}{FP + TN}
\]
Testing: Metrics

<table>
<thead>
<tr>
<th></th>
<th>Unqualified</th>
<th>Qualified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject</td>
<td>TN</td>
<td>FN</td>
</tr>
<tr>
<td></td>
<td>FP</td>
<td>TP</td>
</tr>
</tbody>
</table>

What is the probability of rejecting a woman if she is qualified? What about a man?

False negative rate balance requires (almost) equal values of

\[
\frac{FN}{FN + TP}
\]
Testing: Metrics
Testing: Metrics

RESPONSE TO PROPUBLICA: DEMONSTRATING ACCURACY EQUITY AND PREDICTIVE PARITY

The website ProPublica recently published a story that focused on the scientific validity of COMPAS, raising questions about racial bias. As a result of the article and the subsequent national attention that it garnered, Northpointe launched an in-depth analysis of the data samples used by ProPublica. Drawing from the results of our analysis of ProPublica’s data, Northpointe unequivocally rejects the ProPublica conclusion of racial bias in the COMPAS risk scales.

Predictive modeling is a specialized field within statistics and the appropriate use and interpretation of valid predictive models require a solid understanding of the techniques and methodological nuances common to this type of work. Our detailed review of how ProPublica conducted their analysis revealed several statistical and technical errors such as misspecified regression models, mis-defined classification terms and measures of discrimination, the incorrect interpretation and use of model errors, and more. These errors led to a false conclusion of racial bias; we do not
Testing: Metrics

A computer program used for bail and sentencing decisions was labeled biased against blacks. It’s actually not that clear.

By Sam Corbett-Davies, Emma Pierson, Ari Feller and Sharad Goel
October 17, 2016

(Kleinberg et al., 2016; Chouldechova, 2017)
Deployment: Context

(Phillips et al., 2011)
Feedback

Task Definition

Dataset Construction

Model Definition

Testing Process

Deployment

Training Process
Feedback Loops

Use history of drug-crime reports and arrests to predict future crime locations…

More historic arrests in Black and Hispanic areas

More policing in these areas

More arrests in these areas
So what can we do?
Strategies to Mitigate Harms

- Prioritize fairness at every stage of the ML pipeline
- Think critically about implicit assumptions made at each stage
- Pay attention to potential biases in the data source and data preparation process
- Check if test data matches the deployment context
- Involve diverse stakeholders and gather multiple perspectives
- Acknowledge our mistakes and learn from them
Transparency vs. Intelligibility
What is Transparency?

- In policy circles, transparency represents two distinct ideas
  - People should be able to understand and monitor how AI systems work
  - Those who deploy AI systems should be honest and forthcoming about how and when they are being used

- In machine learning circles, the former is called “intelligibility” or “interpretability,” and literal transparency can work against it!
Transparency ≠ Intelligibility

- Exposing ML source code doesn’t tell us much
- Exposing model internals can stop people from noticing when a model makes a mistake because of information overload

(Poursabzi-Sangdeh et al., 2018)
Why intelligibility?

— Accountability: An applicant wants to know why she was denied a loan.

— Trust: A model deployed in a school predicts that a student is likely to drop out. Knowing the factors relevant for the prediction could help his teacher decide whether to believe it and how to intervene.

— Bias assessment: A model matches candidates to jobs. By understanding characteristics of the training data, an employer may see that female candidates are underrepresented, leading to potential bias.

— Robustness: A data scientist sees unexpected predictions from a model she has trained. Knowing why these predictions were made could help her debug the model.
Intelligibility via “Simple Models”

Point Systems
(Jung et al., 2017; Ustun & Rudin, 2015)

<table>
<thead>
<tr>
<th>AGE</th>
<th>RANGE</th>
<th>SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-20</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>21-25</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>26-30</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>31-50</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>51 and older</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PAST COURT DATES MISSED</th>
<th>RANGE</th>
<th>SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>4+</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

\[ y = f_1(x_1) + \ldots + f_d(x_d) \]

Generalized Additive Models
(Lou, Caruana, et al., 2012&2013)

Classic methods: decision trees, rule lists (if-then-else), rule sets, sparse linear models, …
Intelligibility via Post Hoc Explanations

Simple Explanations of a Single Prediction
(e.g., Ribeiro et al., 2016; Lundberg and Lee, 2017)

Simple Approximations of a Full Model
(e.g., Lakkaraju et al., 2017)
Data Intelligibility: Datasheets for Datasets

Gebru et al., (2018)

Data Collection Process

How was the data collected? (e.g., in-person surveys, email surveys, software program, web interface, API)

The new images for this dataset were obtained from the Faces in the Wild database collected by Tenenbaum & Bartlett 9.

The images in this dataset were gathered from news articles on the web using software to crawl news articles.

Who was involved in the data collection process? (e.g., students, instructors, others) (if applicable, were they compensated? e.g., with money, with course credit)

Over what time frame was the data collection? Does the collection time frame match the creation time frame of the instances?


Data Intelligibility: Datasheets for Datasets

- Questions cover dataset motivation, composition, collection process, pre-processing, distribution, maintenance, legal concerns, and ethical concerns

- Sample use cases:
  - Post with public datasets to inform potential users about the make-up and origin of the data
  - Include with a company’s internal-use datasets to provide relevant information to future users from across the company
# No One-Size-Fits-All Solution

<table>
<thead>
<tr>
<th></th>
<th>Audit a single prediction</th>
<th>Understand model globally</th>
<th>Make better decisions</th>
<th>Debug models</th>
<th>Assess bias</th>
<th>Inspire trust</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEOs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Approach A</td>
</tr>
<tr>
<td>Data scientists</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Approach C</td>
</tr>
<tr>
<td>Lay people</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulators</td>
<td>Approach B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Approach A**: Suitable for CEOs
- **Approach B**: Suitable for Regulators
- **Approach C**: Suitable for Data scientists
No One-Size-Fits-All Solution

- Why is the explanation needed? What is your goal?
- What is being explained? Prediction or whole system?
- To whom should the system be intelligible?
- Does the explainer have access to system internals?
- Does the explainer have access to the training data?
- What is the dimensionality or scale of the system?
- What type of data is used? Feature vectors? Text?
- Could giving away too much open up the system to manipulation?
- Could giving away too much reveal proprietary information?
Takeaways

- There is no one-size-fits-all solution to fairness, transparency, or intelligibility
- These principles cannot be treated as afterthoughts; they must be considered at every stage of the machine learning pipeline
- Technology can be part of the solution, if used with care
- It is important to involve diverse stakeholders and gather multiple perspectives
- We should admit our mistakes and learn from them
Thanks!

http://jennwv.com
jenn@microsoft.com
Wrapping Up and Looking Ahead: Roundtable Discussion of Key Legal and Regulatory Questions in the Field

Session moderated by:

Ellen Connelly
Federal Trade Commission
Office of Policy Planning

Benjamin Rossen
Federal Trade Commission
Division of Privacy and Identity Protection
Wrapping Up and Looking Ahead: Roundtable Discussion of Key Legal and Regulatory Questions in the Field

Panel Discussion:

Justin Brookman, Pam Dixon, Salil Mehra, Joshua New, Nicol Turner-Lee

Moderators: Ellen Connelly & Benjamin Rossen
Closing Remarks

Danielle Holley-Walker
Howard University School of Law
Thank You

Join Us In December