WEBINAR SERIES
A Conversation With ...

Michelle Seng Ah Lee
PhD Researcher, Compliant & Accountable Systems Research Group
University of Cambridge

Challenges of Algorithmic Fairness in Financial Services
The FDPi was created by CAIA to:

✓ Provide financial professionals with the knowledge necessary to succeed in an industry disrupted by the advent of big data and machine learning.

✓ Advocate for the highest levels of professional ethics and standards.

✓ Establish the FDP Charter as a global professional designation in the area of financial data science.

EARN YOUR FDP DESIGNATION
A globally-recognized charter is awarded to FDP charter holders

TWO ONLINE CLASSES*
Choose either Python or R
Can be completed before or after the exam

ONE COMPREHENSIVE EXAM
Offered twice per year
March & November

VALUE ADD
Employers increasingly seek to find professionals to have the skills to apply data science tools to solve their most challenging problems
FDP EXAM

1. Introduction to Data Science & Big Data
5. Machine Learning: Classification & Clustering
7. Data Mining & Machine Learning: Naïve Bayes & Text Mining

A self-study program developed by financial professionals, for financial professionals.

Exam candidates now have one of two test options for the FDP exam

Option 1: At a Prometric test center near you
Exam window: October 12 – November 8, 2020

Option 2: Through Remote proctoring from your home/office
Exam Day: December 1, 2020

Learn more at www.fdpinstitute.org
Financial Data Professional Institute

Challenges of Algorithmic Fairness in Financial Services

Why is it important, and what’s next?

Michelle Seng Ah Lee
AI Ethics Lead, Deloitte UK
PhD Researcher, University of Cambridge
Compliant & Accountable Systems Research Group

Keith Black
Managing Director, Content Strategy, CAIA

Mirjam Dekker
PM, FDP Institute

www.fdpinstitute.org
July 29, 2020
Challenges of Algorithmic Fairness in Financial Services

Why is it important, and what’s next?

Introduction

The Problem Statement

Pre-AI: racial discrimination in lending (ex. US Mortgage)

Problem is not the algorithmic design

Why can’t we just make algorithmics fair?

Fairness complexities

Fairness tests

Proposal: benchmarking

Trade-off between financial inclusion and minority denial rates

Why are algorithms biased against back applicants?

If algorithms can triangulate race, what data can we use?

Example: spelling error and default risk

Findings

Why do people make spelling errors?

Phd future research
Challenges of algorithmic fairness in financial services

Why is it important, and what’s next?

Michelle Seng Ah Lee, PhD Candidate
About me

BA Political Science, Symbolic Systems (Decision-making and rationality)

S&O Consulting, Deloitte US

AI team, Risk Analytics, Deloitte UK

MSc Oxford, Digital Ethics Lab

Started PhD at Cambridge

UK Chapter Lead, DataKind

DataKind UK
The problem statement

MO COMPARE Motorists fork out £1,000 more to insure their cars if their name is Mohammed

By Ben Leo 22nd January 2018, 12:17 am | Updated: 22nd January 2018, 11:51 pm

Top firms such as Admiral and Marks & Spencers have been dragged into an insurance race row after giving far lower quotes for drivers with traditionally English names like John.

Facial-Recognition Software Might Have a Racial Bias Problem

Depending on how algorithms are trained, they could be significantly more accurate when identifying white faces than African American ones.

Apple Card Investigated After Gender Discrimination Complaints

A prominent software developer said on Twitter that the credit card was “sexist” against women applying for credit.

Your online, mobile, and social behaviour are now data-points used by fintech startups and governments in scoring credit-worthiness

Imagine a world where authoritarian governments and fintech companies monitor everything you do, amass huge amounts of data on almost every interaction, and award you a single score that measures how "trustworthy" you are.
Pre-AI: racial discrimination in lending (ex. US mortgage)

### Data collection
- **Selection bias**
  - Limited marketing / outreach to minority groups

### Pre-application
- **Human bias**
  - Black applicants quoted lower loan amounts, received less assistance and information (paired audit study)

### Decision-making
- **Disparate impact**
  - 8% difference in denial rates even after controlling for all other “legitimate” features

### Learning
- **Missing counterfactual**
  - No information on whether or not those denied a loan would have defaulted
Problem is not the algorithm design

...and therefore cannot be solved algorithmically

**Data collection**
- **Selection bias**
  Requires a change in strategy to increase outreach and advertisement in high-minority areas

**Pre-application**
- **Human bias**
  Requires a change in process to provide similar service to similar individuals (vs. at individual discretion)

**Decision-making**
- **Disparate impact**
  Requires an understanding of **why** the existing credit risk process results in disadvantageous outcomes for minorities (e.g. income volatility) and to measure this difference

**Learning**
- **Missing counterfactual**
  Requires an active analysis of whether there are denials of credit-worthy applicants
Why can’t we just make algorithms fair?

<table>
<thead>
<tr>
<th>Definition</th>
<th>Philosophy</th>
<th>Contextual Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choose the model that makes everyone better off in aggregate: The most accurate model gives people what they “deserve” by minimising errors</td>
<td><em>Expected consequentialism</em></td>
<td>• Difference in distribution could be a reflection of inequality in other markets</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Historical discrimination can inaccurately reflect outcomes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• An algorithm is limited by the historical data available (selection bias)</td>
</tr>
<tr>
<td>Demographic parity, or group fairness:</td>
<td><em>Strict egalitarianism</em></td>
<td>• In most use cases, difficult or impossible to theorise a strong causal link and formalise its graphical representation</td>
</tr>
<tr>
<td>$P(\tilde{Y}</td>
<td>A = 0) = P(\tilde{Y}</td>
<td>A = 1)$.</td>
</tr>
<tr>
<td>Counterfactual fairness:</td>
<td><em>David Lewis, cause and effect</em></td>
<td>• Can lead to reverse discrimination and inaccurate predictions</td>
</tr>
<tr>
<td>$P(\tilde{Y}_{A=x,a}(U) = Y</td>
<td>X = x, A = a) = P(\tilde{Y}_{A=x,a}(U) = y</td>
<td>X = x, A = a)$.</td>
</tr>
<tr>
<td>Individual fairness:</td>
<td><em>Mixed</em></td>
<td>• Fails to address discrimination that may already be embedded in the data</td>
</tr>
<tr>
<td>$\tilde{Y}(X^{(i)}, A^{(j)}) = Y((X^{(i)}, A^{(j)})$.</td>
<td></td>
<td>• Market does not exist in a vacuum; impossible to isolate what is outside of people’s control</td>
</tr>
<tr>
<td>Equalized odds:</td>
<td><em>Dworkin’s theory of Resource Egalitarianism</em></td>
<td>• Fails to address discrimination that may already be embedded in the data</td>
</tr>
<tr>
<td>$P(\tilde{Y} = 1</td>
<td>A = 0, Y = y) = P(\tilde{Y} = 1</td>
<td>A = 1, Y = y), y \in (0, 1)$.</td>
</tr>
<tr>
<td>Equalized opportunity:</td>
<td><em>Dworkin’s theory of Resource Egalitarianism</em></td>
<td></td>
</tr>
<tr>
<td>$P(\tilde{Y} = 1</td>
<td>A = 0, Y = 1) = P(\tilde{Y} = 1</td>
<td>A = 1, Y = 1)$.</td>
</tr>
</tbody>
</table>
Each of these is presented as a one-size-fits-all precondition of fairness, but doesn’t it depend on the context?
**Fairness tests**

Approach in literature on fairness not useful in providing actionable insights

<table>
<thead>
<tr>
<th>Model</th>
<th>Equality of Opportunity (EOP)</th>
<th>False Positive Error Rate Balance (FPERB)</th>
<th>Equal odds (EO)*</th>
<th>Positive Predictive Parity (PPP)</th>
<th>Positive Class Balance (PCB)</th>
<th>Negative Class Balance (NCB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td>4%</td>
<td>4%</td>
<td>4%</td>
<td>33%</td>
<td>2%</td>
<td>3%</td>
</tr>
<tr>
<td>KNN</td>
<td>6%</td>
<td>6%</td>
<td>6%</td>
<td>18%</td>
<td>8%</td>
<td>8%</td>
</tr>
<tr>
<td>CART</td>
<td>6%</td>
<td>6%</td>
<td>6%</td>
<td>21%</td>
<td>16%</td>
<td>18%</td>
</tr>
<tr>
<td>NB</td>
<td>7%</td>
<td>7%</td>
<td>7%</td>
<td>44%</td>
<td>41%</td>
<td>41%</td>
</tr>
<tr>
<td>RF</td>
<td>9%</td>
<td>9%</td>
<td>9%</td>
<td>3%</td>
<td>18%</td>
<td>18%</td>
</tr>
</tbody>
</table>

* EO condition is met if EOP and FPERB are both met.
My proposal: benchmarking

Which model’s trade-off are you most comfortable with?

Aggregate group benefit vs. impact on protected groups

- Logistic regression
- KNN
- CART
- Naive Bayes
- Random forest

Baseline

Aggregate benefit

Negative impact on protected groups
Trade-off between financial inclusion and minority denial rates
Why are the algorithms biased against black applicants?

**Human bias**
- FHA-insured loan type associated with both race and loan outcome

**Embedded proxies**
- It is possible to predict race using the input features
If algorithms can triangulate race, what data can we use?

Example: spelling error and default risk

- Open source peer-to-peer (P2P) lending data set
- Data from 2007-2011 retain the column containing free-form text description of the loan from the borrower (42,309 loans, 15% default rate)

Below is a partial excerpt from one of the descriptions, with non-standard language in *italics* and accidental typos in **bold**:

Borrower added on 07/21/11 paying off house and 4,000.00 back to my saving *acct* bought a 12,000.00 boat and **payed** cash for it. Borrower added on 07/24/11 reason for paying house off i never have more then one loan at a time. Borrower added on 08/02/11 Thanks to the investors you’re making a sure thing with me. my credit rating has **allways** been number one with me. thanks again
Higher number of typos is associated with a greater probability of default.

\[
P(\text{default} \mid 0 \text{ typo}) = 17.35\% \\
P(\text{default} \mid 1 \text{ typo}) = 19.84\% \\
P(\text{default} \mid 2 \text{ typos}) = 22.59\%.
\]

However, whether the typo was orthographic or phonological and the distance of the typo from the correction do not affect the likelihood of default.
Why do people make spelling errors?

Dyslexia  Non-native speaker  Carelessness

How do we isolate the impact of carelessness?
<table>
<thead>
<tr>
<th>Papers</th>
<th>Target journal / conference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. From fairness metrics to Key Ethics Indicators (KEIs): a context-conscious approach to algorithmic ethics in an unequal society</td>
<td>Science and Engineering Ethics Journal</td>
</tr>
<tr>
<td>B. Identifying 6 types of biases in software development lifecycle</td>
<td>Fairness, Accountability, Transparency (FAccT) Conference</td>
</tr>
<tr>
<td>C. Gaps in the open source fairness and explanation toolkits</td>
<td>ACM CHI Conference</td>
</tr>
<tr>
<td>D. Regulating beyond AI and ADM: risk factor framework</td>
<td>European Data Protection Handbook (proposal accepted as a chapter for publication in 2021)</td>
</tr>
</tbody>
</table>
List of my publications: https://michellesengahlee.wordpress.com/publications/

My blog posts on Deloitte Financial Services: https://ukfinancialservicesinsights.deloitte.com/u/102fti4/michelle-lee

Recommended further reading:

- 50 years of (test) unfairness: https://arxiv.org/abs/1811.10104
- Delayed impact of fair machine learning: https://arxiv.org/abs/1803.04383
- The measure and mismeasure of fairness: https://arxiv.org/abs/1808.00023
- 21 definitions of fairness and their politics: https://www.youtube.com/watch?v=jIXluYdnyyk
Q&A
Add your questions to the chat room

WEBINAR SERIES
A Conversation With...
Nigel Noyes
Principal Data Scientist
Quicken Loans

Data Science in Production
Aug 6, 2020 @ 1:00 PM EDT

Webinar:
Making Sense of Machine Learning
August 19, 2020
1pm EDT

Yinidu (Andrew) Li,
Quantitative Research Analyst,
State Street Associates

David Turkington, Sr. Mg. Dir.,
Head of Portfolio & Risk Mgmt
State Street Associates
In Closing

Learn more about the FDP Institute at www.fdpinstitute.org