Racial Bias in Facial Recognition Software

Stephanie Kim
About Me

Python, R, data analysis, ETL, NLP, ML & writes deep learning docs.

Founder of Seattle PyLadies & organizer of Seattle Building Intelligent Applications

Developer Advocate at Algorithmia

Stephanie Kim
Why?

- Every application will become an intelligent application.
- Running a business without AI will become unthinkable.

What?

- Productionize machine and deep learning models, using Algorithmia to deploy at scale.
- An open marketplace running over 5,000 algorithms, functions, and models as scalable microservices via a common API.
What We’ll Cover

- Racial bias
- Facial recognition overview
- Real world applications/problems
- Facial recognition 101 with OpenFace
- Celebrity look-a-like use case
- Model failures/What can we do better?
Racial bias in humans can lead to racial bias in AI

Bertrand and Mullainathan (2004)

https://goo.gl/RJmfYh
Two Petty Theft Arrests

VERNON PRATER
LOW RISK 3

BRISHA BORDEN
HIGH RISK 8

Borden was rated high risk for future crime after she and a friend took a kid’s bike and scooter that were sitting outside. She did not reoffend.
## Potential Harms from Algorithmic Decision-Making

<table>
<thead>
<tr>
<th>Individual Harms</th>
<th>Collective Social Harms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illegal Discrimination</td>
<td>Loss of Opportunity</td>
</tr>
<tr>
<td>Unfair Practices</td>
<td></td>
</tr>
<tr>
<td>Hiring</td>
<td></td>
</tr>
<tr>
<td>Employment</td>
<td></td>
</tr>
<tr>
<td>Insurance &amp; Social Benefits</td>
<td></td>
</tr>
<tr>
<td>Housing</td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Credit</th>
<th>Economic Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Differential Prices of Goods</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Loss of Liberty</th>
<th>Social Stigmatization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increased Surveillance</td>
<td></td>
</tr>
<tr>
<td>Stereotype Reinforcement</td>
<td></td>
</tr>
<tr>
<td>Dignatory Harms</td>
<td></td>
</tr>
</tbody>
</table>

[Chart Contents Courtesy of Megan Smith, Former CTO of the United States](https://bit.ly/2Ek9ZwZ)
Real World Applications
Real World Surveillance
Real World Surveillance

- Lack of regulation/accountability
- No transparency for model accuracy
- Facial recognition systems have a higher rate of failure on African Americans.
Facial Recognition 101
Training

M Unique Images

Network ➔ M Embeddings ➔ Triplet Embeddings

- N Anchor Embeddings
- N Positive Embeddings
- N Negative Embeddings

accumulate the gradient for each unique image and then backpropagate.

http://elijah.cs.cmu.edu/DOCS/CMU-CS-16-118.pdf
Isolate Face from Background

- **Haar cascade classifier**
  - OpenCV default
  - Fast, low accuracy

- **Histogram of Oriented Gradients & SVM**
  - Dlib C++ library
  - Faster, higher accuracy, less data

https://goo.gl/CEnamR
Preprocessing

https://goo.gl/D1k7LH
Classification

Neural net does a forward pass to retrieve 128 embeddings for each face used in prediction.
Accuracy

match pairs

Abel Pacheco, 1
Abel Pacheco, 4

mismatch pairs

Abdel Madi Shabneh, 1
Dean Barker, 1

Akhmed Zakayev, 1
Akhmed Zakayev, 3
Abdel Madi Shabneh, 1
Giancarlo Fisichella, 1
True Blood Celebrity Look-a-Like
Model Accuracy
Diversify Training Data
Model Failure (and solutions)

Training Datasets: How is your data extracted?
Model Failure - Training Sets
Model Failure
Labeled Faces in the Wild

Diversity of Testing Datasets Matter
Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification

Joy Buolamwini, Timnit Gebru
Pilot Parliaments Benchmark
<table>
<thead>
<tr>
<th>Gender Classifier</th>
<th>Darker Male</th>
<th>Darker Female</th>
<th>Lighter Male</th>
<th>Lighter Female</th>
<th>Largest Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microsoft</td>
<td>94.0%</td>
<td>79.2%</td>
<td>100%</td>
<td>98.3%</td>
<td>20.8%</td>
</tr>
<tr>
<td>FACE++</td>
<td>99.3%</td>
<td>65.5%</td>
<td>99.2%</td>
<td>94.0%</td>
<td>33.8%</td>
</tr>
<tr>
<td>IBM</td>
<td>88.0%</td>
<td>65.3%</td>
<td>99.7%</td>
<td>92.9%</td>
<td>34.4%</td>
</tr>
</tbody>
</table>
Model Failure - Risk Factors
Build Diverse Teams
I was collecting examples of racial bias in image software and there are SO MANY. Please stop having only white people in your training set.

Bias in AI: offensive to tragic
WRAPUP

Thank you for coming!!
Try out Algorithmia & OpenFace:

OreillyAI NYC
Ask Me Anything!

Twitter: @StephLKim

Email: Stephanie@Algorithmia.com

LinkedIn: https://www.linkedin.com/in/stephanielkim/
Resources

Request the Pilot Parliaments Benchmark dataset: https://bit.ly/2Hxmmqg


Propublica: https://goo.gl/AiHQgV (Note you can donate)

Perpetual Lineup: Georgetown Law: https://goo.gl/G3Hgzy

Perpetual Lineup: Racial Findings: https://goo.gl/Zcm1wy
Resources

Ethnicity Identification from Face Images:
https://bit.ly/2Kx500h

PBS AI Bias: https://goo.gl/32DyxM

OpenFace: Openface paper: https://goo.gl/UsjtDT

Intro to Facial Recognition with OpenFace:
https://goo.gl/FTBWgz

OpenFace on Algorithmia: https://goo.gl/vXGygV