**Discrimination Law: Two Doctrines**

<table>
<thead>
<tr>
<th>Disparate Treatment</th>
<th>Disparate Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formal</td>
<td>Unjustified</td>
</tr>
<tr>
<td>Intentional</td>
<td>Avoidable</td>
</tr>
</tbody>
</table>

“Protected Class”
Dealing with Tainted Examples

• Training data serve as ground truth
  • These would seem like well performing models according to standard evaluation methods

• What the objective assessment *should* have been
  • Accepted and rejected candidates may not differ only in terms of protected characteristics

• How someone *would* have performed under different, non-discriminatory circumstances
  • The difficulty in dealing with counterfactuals and correcting for past injustices
Settling on a Selection of Features

• Does the feature set provide sufficient information to carve-up the population in a way that reveals relevant variations within each apparent sub-group?
  • *Unintentional* redlining

• In other words: How does the error rate vary across the population?
  • Discrimination can be an artifact of statistical reasoning rather than prejudice on the part of decision-makers or bias in the composition of the dataset

• Does the difficulty or cost involved in obtaining the information necessary to bring accuracy rates into closer parity justify subjecting certain populations to worse assessment?
  • Parity = Fair
  • Accurate = Fair
<table>
<thead>
<tr>
<th>Benefit</th>
<th>Harm</th>
</tr>
</thead>
</table>
| • Discovering attractive customers and candidates in populations previously dismissed out of hand → Financial inclusion  
  • Evidence-based and formalized decision-making                      | • Less favorable treatment in the marketplace → Finding specific customers not worth servicing (e.g., firing the customer)  
  • Individualization of risk                                            |
| • Equal treatment in the marketplace → Common level of service and uniform price  
  • Socialization of risk                                                | • Underserving large swaths of the market → Redlining                
  • Informal decision heuristics plagued by prejudice and implicit bias  |

Granularity of the Data

Effects on historically disadvantaged communities
Dealing with “Redundant Encodings”

• In many instances, making accurate determinations will mean considering factors that are somehow correlated with legally proscribed features
  • There is no obvious way to determine how correlated a relevant attribute or set of attributes must be with proscribed features to be worrisome
  • Nor is there a self-evident way to determine when an attribute or set of attributes is sufficiently relevant to justify its consideration, despite the fact that it is highly correlated with these features
Let’s not Forsake Formalization

• These moments of translation are opportunities to debate the very nature of the problem—and to be creative in parsing it
• The process of formalization *can* make explicit the beliefs, values, and goals that motivate a project

Solon Barocas
Microsoft Research
solon@microsoft.com