The Use and Effect of Big Data:
What Fair Housing Advocates Need to Know

Stephen M. Dane
Relman, Dane & Colfax PLLC
1225 19th Street, NW, Suite 600
Washington, D.C. 20036
(202) 728-1888
sdane@relmanlaw.com
Big Data and Civil Rights – New Books

1. Weapons of Math Destruction
   - Cathy O'Neil

2. Algorithms of Oppression
   - Safiya Umoja Noble

3. Automating Inequality
   - Virginia Eubanks

RELMAN, DANE & COLFAIX
WASHINGTON, DC (202) 728-1888
Big Data/Civil Rights
Studies and Reports

  - Focuses on FCRA, ECOA

  - Addresses Civil Rights more generally

- Leadership Conference on Civil Rights, Robinson + Yu: Civil Rights, Big Data, and Our Algorithmic Future, September 2014
  - Civil Rights generally

- National Consumer Law Center (NCLC), Big Data: A Big Disappointment for Scoring Consumer Credit Risk (March 2014)
  - Consumer credit
Barocas & Selbst, Big Data’s Disparate Impact, 104 Cal. L. Rev. 671 (2016)


Big Data – Good or Evil?

- Can be used for Good

- Can be used for Evil
Bad Big Data Algorithms

- Opacity:
  - Invisible models are the rule, clear ones the exception

- Scale
  - Affect large populations

- Damage
  - Harm or work against some category of people
Bad Big Data Algorithms

“Many of these models encoded human prejudice, misunderstanding, and bias into software systems. . . . They tended to punish the poor and the oppressed in our society, while making the rich richer.”

Weapons of Math Destruction:
How Big Data Increases Inequality and Threatens Democracy
By Cathy O’Neil (Crown, 2016)
“Racism is the most slovenly of predictive models. It is powered by haphazard data gathering and spurious correlations . . . and polluted by confirmation bias. In this way, racism operates like many of the [algorithms] described in this book.”

“Racists don’t spend a lot of time hunting down reliable data to train their twisted models. And once their model morphs into a belief, it becomes hardwired. It generates poisonous assumptions, yet rarely tests them, settling instead for data that seems to confirm and fortify them.” (23)
Compare

- Baseball statistics
  - Transparent
  - Assumptions and conclusions available for all to see
  - Everyone has access to data
  - Actual data used, proxy data not used
  - Continuously updated and refined with fresh data
  - Can be replicated
Example – Law Enforcement

- Neighborhood policing and predictive crime models
  - Examples: Zero tolerance, “broken window” policing, stop-and-frisk
  - Assumption in models: “antisocial behavior” causes disorder
  - Use low-level arrest and crime data (e.g., vagrancy, panhandling, small drug charges, nuisance crimes)
  - Vicious feedback loop: Data populated for AA neighborhoods, but not WH neighborhoods
  - Policing spawns new data, which justifies more policing, which spawns more data in AA neighborhoods
Example – Law Enforcement

- What if police spent more time and gathered more data in WH neighborhoods?
  - E.g., traffic violations, jaywalking, loitering, parking violations
- What if models gathered different data? E.g.,
  - White collar crimes, numbers of victims, value of harm, etc.
  - As a group, white collar criminals and financial industry are not targeted for data collection and modeling

Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy
By Cathy O’Neil (Crown, 2016)
Example – Credit

- **FICO Scoring**
  - Non-WMD attributes for lending
    - No proxies: individual’s actual credit history is used, not proxies
    - Have a feedback loop: future credit behavior is captured and fed back into model
    - Somewhat transparent: instructions provided on how to improve score

- **BUT – many other uses of “e-scores” have WMD characteristics**
  - On-line data used as proxies for behavior to be scored
    - E.g., zip code; recent purchases; websites visited; internet surfing patterns
  - E-scores are opaque – don’t know what’s in them
  - Data obtained is frequently in error
  - Applied to thousands/millions of consumers
  - Negative consequences: Denial of service, increased pricing, inferior products
Example – Non-Credit Uses of Credit Data or Scores

- FICO-type credit scoring has moved beyond credit
  - “Creditworthiness” has expanded far beyond its original purpose
  - “Framing debt as a moral issue equivalent to trustworthiness and reliability”
  - Now used as a proxy for:
    - Home insurance
    - Car insurance
    - Employment
    - Dating sites

- Self-reinforcing: People without savings or credit available to them (i.e., the poor) always relegated to “lower” levels of products and services
  - Statistically proven that “credit checks” disproportionately affect low income applicants and applicants of color
  - Yet hardly ever prohibited by law

- Largely unregulated
Big Data Sources

- Often Contain Errors or Mismatches
  - Credit reports contain 5-10 percent errors
  - Big money in data sales, even if rate of “success hit” is small
  - How accurate is data from:
    - Retailers, Advertisers, Smart phone app makers, Sweepstakes, Social Networks, Voting records, Arrest records, Housing and real estate, etc.
    - Consumers often don’t get to see or correct inaccurate data; few complaints
    - Feedback loops with negative consequences
    - Qualities of fairness, justice, equity – Not present in data or feedback results
  - Internet becoming flooded with behavioral data – How will it be used?
Fairness and Other Public Values

- Are NOT factored into WMDs
- Definitions are squishy and hard to quantify
- Which of these values should be sacrificed for efficiency?
- “As e-scores and [unfair algorithms] pollute the sphere of finance [and employment], opportunities dim for the have-nots.”
Some Spurious Correlations Based on Big Data

Per capita cheese consumption correlates with

Number of people who died by becoming tangled in their bedsheets

---

[Graph showing correlation between cheese consumption and deaths from bedsheet tanglings from 2000 to 2009]
Some Spurious Correlations Based on Big Data

Money spent on pets (US) correlates with Lawyers in California

<table>
<thead>
<tr>
<th>Year</th>
<th>Money spent on pets (US) (Billions of dollars)</th>
<th>Lawyers in California (Projections, Bureau of Labor Statistics)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>39.7</td>
<td>128,553</td>
</tr>
<tr>
<td>2001</td>
<td>41.9</td>
<td>131,139</td>
</tr>
<tr>
<td>2002</td>
<td>44.6</td>
<td>132,452</td>
</tr>
<tr>
<td>2003</td>
<td>46.8</td>
<td>134,468</td>
</tr>
<tr>
<td>2004</td>
<td>49.8</td>
<td>136,571</td>
</tr>
<tr>
<td>2005</td>
<td>53.1</td>
<td>139,371</td>
</tr>
<tr>
<td>2006</td>
<td>56.9</td>
<td>141,030</td>
</tr>
<tr>
<td>2007</td>
<td>61.8</td>
<td>145,355</td>
</tr>
<tr>
<td>2008</td>
<td>65.7</td>
<td>148,399</td>
</tr>
<tr>
<td>2009</td>
<td>67.1</td>
<td>149,982</td>
</tr>
</tbody>
</table>

Correlation: 0.998386
The Use and Effect of Big Data:

What Fair Housing Advocates Need to Know

Stephen M. Dane
Relman, Dane & Colfax PLLC
1225 19th Street, NW, Suite 600
Washington, D.C. 20036
(202) 728-1888
sdane@elmanlaw.com