ETHICAL ISSUES IN THE BIG DATA INDUSTRY - IMPlications FOR INSURANCE

Kirsten Martin, PhD
George Washington University
School of Business

U.S. Department of Treasury
Federal Insurance Office’s (FIO)
Federal Advisory Committee on Insurance (FACI)

August 17, 2017
Three Areas of Concern – Big Data

1. Source of Big Data
2. Analysis of Big Data
3. Marketing Use of Big Data
Sourcing Issues
1. Treatment of Individuals
2. Environmental Damage
3. Poor Quality

Manufacturing Issues
1. Treatment of Individuals
2. Pollution
3. Poor Quality

Use Issues
1. Distribution
2. Harm to Consumer
3. Novel 2\textsuperscript{nd} Use
1. SOURCE OF BIG DATA
But …. 

0. Data is public information without any expectation of privacy.

1. Someone else did it.  *I* didn’t actually do anything wrong (e.g., breach confidentiality/laws/privacy expectations), so why should *I* be held accountable? (Nike, Wal-Mart, Apple, Kathy Lee Gifford Argument).

2. Everyone else is doing it (Connor Argument; Lance Armstrong c. 2013 Argument)
65% of respondents have complete or moderate trust in insurance companies (EY 2014)

7% of respondents have confidence that data aggregators protect their data; 50% believe they should not have any data (Pew 2014)
2. ANALYSIS OF BIG DATA
Two problems

Policy as Hidden

Policy as Quickly Replicated
Algorithms as producing an answer

Current Approach to Algorithmic Decision Making

Source Data
Data set including people or information on which to make a decision

Algorithm
Weights + factors to take into consideration

Training Data
Information used to ‘train’ an algorithm if using machine learning.

Outcome
Answer or decision. E.g., Risk assessment
Algorithms as producing an answer

**Source Data**
Best estimates of the factors available about individuals to possibly be used in the decision

**Algorithm**
Rules, policy, principles, ethical norms, laws suggesting the relative importance of factors to a decision

**Training Data**
History of the contextual decision as told by individuals who tracked and recorded the decision.

**Outcome**
Best approximation of intended output. E.g., Risk assessment

Acknowledge Unjust Biases Throughout
**Algorithms as producing an answer**

**Source Data**
*Best estimates of the factors available about individuals to possibly be used in the decision*

**Algorithm**
*Rules, policy, principles, ethical norms, laws suggesting the relative importance of factors to a decision*

- What factors are appropriate & fair for this context?
- What level of ‘accuracy’ is fair for this decision?

**Training Data**
*History of the contextual decision as told by individuals who tracked and recorded the decision.*

- What are the appropriate rules/policies to apply in this context?
- What are the ethical norms?

**Outcome**
*Best approximation of intended output. E.g., Risk assessment*

- How is effective defined for this decision?
- Is the outcome biased unjustly?

**Acknowledge Missing Questions**
- What historical reference points are appropriate & fair for this decision?
- What unjust biases exist in the construction of the historical data?
Two problems

Policy as Hidden

Less Oversight

Policy as Quickly Replicated

Greater Impact
3. MARKETING USE OF BIG DATA

Product, promotion, pricing
Use of Data

Pivotal Decisions – allocation of social goods

- Who deserves the insurance
- Who should have access to financial protections
- Whose claim should be questioned?
- Who should take on greater costs for the same service

“Accuracy” v. “Good Decision”

New Information
- ‘True’ Information
- More Information
- Deeper Information
- Biased Information

Digital marketing manipulation (Calo)

“An insurer might target a chain-smoking motorcycle buff with an action-packed video game designed to help him quit — while appealing to his profile as an adrenaline junkie.”

http://www.statnews.com/2015/12/15/insurance-big-data/

E.g., False Claim Detection:
http://www.tellius.com/machine-learning-transforming-insurance-industry/
E.g., False Claim Detection:
“With a person-centric [versus claim-centric] approach, the beneficiary’s claim history and behavior across multiple sources (such as using a person’s social graph to find similar behavior patterns among individuals that he or she is connected to, and similar claims that were reported by the same person) are analyzed.”

http://www.tellius.com/machine-learning-transforming-insurance-industry/

Pivotal Decisions – allocation of social goods

- Are there biases in the data such that some demographics have more accurate estimates?
- Are some groups afforded better terms in claims adjustments based on this data?

“Accuracy” v. “Good Decision”

- Are contacts an appropriate factor in determining claims?
- Is the claim decision reviewable? Justified?

Digital marketing manipulation (Calo)

- Are individuals particularly vulnerable during this negotiation or promotion?
- Does the data we use give us an unfair advantage?
IN SUM....
Three Areas of Concern – Big Data

1. **Source of Big Data**
   - *Difference in trust* between supplier of data and insurance company
   - *Difference in the importance of trust* in their businesses

2. **Analysis of Big Data**
   - *Policy as hidden*
   - *Policy as quickly replicated*

3. **Marketing Use of Big Data**
   - *Pivotal decisions*
   - *Good versus accurate decision*
   - *Consumer manipulation*