



**Session 1:**

**Big Data and Personal Lines Pricing – the Revolution You Might Have Missed**

**Session 2:**

**Big Data and Claims – She Didn't Really Post That on Facebook, Did She?**

**Session 3**

**Big Data and Life Insurance – Promoting Innovation through Consumer Protection**

**IRES Career Development Seminar**

August 19 and 20, 2019

Birny Birnbaum

Center for Economic Justice

# The Center for Economic Justice

CEJ is a non-profit consumer advocacy organization dedicated to representing the interests of low-income and minority consumers as a class on economic justice issues. Most of our work is before administrative agencies on insurance, financial services and utility issues.

On the Web: [www.cej-online.org](http://www.cej-online.org)

# About Birny Birnbaum

Birny Birnbaum is the Director of the Center for Economic Justice, a non-profit organization whose mission is to advocate on behalf of low-income consumers on issues of availability, affordability, accessibility of basic goods and services, such as utilities, credit and insurance.

Birny, an economist and former insurance regulator, has authored reports and testimony for numerous public agencies and consumer organizations, covering a wide variety of topics, including analysis of insurance markets, insurers' use of big data, market regulation, force-placed insurance, homeowners and flood insurance, consumer credit insurance, title insurance and insurance credit scoring. He has served for many years as a designated Consumer Representative at the National Association of Insurance Commissioners. He is a member of the Federal Advisory Committee on Insurance, chairing the Subcommittee on Affordability and Availability of Insurance.

Birny served as Associate Commissioner for Policy and Research and the Chief Economist at the Texas Department of Insurance. In that role, Birny was responsible for review and approval of rate filings, the development of data collection programs for market surveillance and the analysis of competition in numerous insurance markets.

Prior to his work at the TDI, Birny served as Chief Economist at the Texas Office of Public Insurance Counsel where he provided expert testimony in rate and rule hearings on behalf of insurance consumers before the TDI. While at OPIC, Birny performed the first auto insurance redlining study in Texas.

Birny was educated at Bowdoin College and the Massachusetts Institute of Technology. He holds the AMCM certification.

## Why CEJ Works on Insurance Issues

### ***Insurance Products Are Financial Security Tools Essential for Individual and Community Economic Development.***

CEJ works to ensure ***fair access*** and ***fair treatment*** for insurance consumers, particularly for low- and moderate-income consumers.

### ***Insurance is the Primary Institution to Promote Loss Prevention and Mitigation, Resiliency and Sustainability:***

CEJ works to ensure insurance institutions maximize their role in efforts to reduce loss of life and property from catastrophic events and to ***promote resiliency and sustainability*** of individuals, businesses and communities.

# Personal Consumer Information in Big Data

- Telematics – Auto, Home, Wearable Devices
- Social Media
- Shopping Habits/Purchase History
- Hobbies and Interests
- Demographics/Household Data/Census Data
- Government Records/Property Records
- Web/Mobile Phone Tracking/GPS/Data Harvesting
- Vehicle Registration and Service Records
- Facial Analytics
- Mainstream Credit Files: Loans, Credit Cards
- Alternative Credit Data: Telecom, Utility, Rent Payment

Sources of Data include consumers (via telematics or wearable devices), government, social media platforms, web sites, mobile devices, e-mail/text, data brokers, online data aggregators, aircraft/satellite photos and many others.

## Big Data vs. Artificial Intelligence vs. Machine Learning

Artificial Intelligence: Machines carrying out tasks based on algorithms that mimic cognitive functions of humans. General AI – a human brain – versus Narrow AI – a specific function. AI started in 1956.

Machine Learning: A subset of AI -- ability of machines to receive a set of data and learn for themselves, changing algorithms as they learn more about the information they are processing.

Widespread Machine Learning is new, powered by greater computational power and more data and more speed. Greater computational power means ability to analyze data in a timely fashion. More data provides raw material for machines to learn – e.g., facial recognition. More speed – internet – means the ability to collect data and deploy algorithmic results in real time.

Human –designed algorithms vs. machine-developed algorithms.

# Examples of Insurer Big Data Algorithms in Insurance

## Pricing/Underwriting:

- Price Optimization/Demand Models
- Customer Value Scores
- Telematics,
- Social Media Scores
- Credit Scores
- Criminal History Scores,
- Vehicle Scores,
- FireLine Rating
- Facial Analytics

## Claims:

- Claim Optimization/Demand Models
- Fraud Scores
- Facial Analytics
- Severity Scores
- Telematics

## Big Data Algorithms as Insurance Market Gatekeepers

- Marketing: web searches and web advertising that pre-score and channel consumers to particular products, providers and price-levels.
- Pricing: pre-fill applications and pricing without the consumer providing information, pricing based not just on risk but on price optimization / consumer demand models, real-time competitive options and/or socio-economic characteristics
- Claims: automated, instant claim settlement proposals based on data generated by a vehicle, home telematics or wearable device and utilizing price optimization/consumer demand models to determine amount of claim settlement offer a particular consumer is likely to accept based on his or her personal data.
- Common characteristics – opaque algorithms, little or no disclosure or transparency to consumer, great potential to penalize most vulnerable consumers, limiting loss mitigation role of insurance



## What's So Big about Big Data?

1. Insurers' use of Big Data has huge potential to benefit consumers and insurers by transforming the insurer-consumer relationship and by discovering new insights into and creating new tools for loss mitigation.
2. Insurers' use of Big Data has huge implications for fairness, access and affordability of insurance and for regulators' ability to keep up with the changes and protect consumers from unfair practices
3. The current insurance regulatory framework generally does not provide regulators with the tools to effectively respond to insurers' use of Big Data. Big Data has massively increased the market power of insurers versus consumers and versus regulators.
4. Market forces alone – “free-market competition” – cannot and will not protect consumers from unfair insurer practices. So-called “innovation” without some consumer protection and public policy guardrails will lead to unfair outcomes.

# ***“3 Ways Insurance Industry Makes Better Use of Data Analytics”***

## **In\$urance CIO Outlook**

### **Fraud Detection**

Insurance fraud can bring substantial financial loss to insurance companies, and data science platforms and software can detect fraudulent activity, suspicious links and subtle behavior patterns using multiple techniques. To make this possible, the algorithms should be fed with a constant flow of data. These models rely on the historical cases of fraudulent activity and apply sampling method to analyze them. Besides, predictive modeling techniques are used for the analysis and filtering of fraud instances.

### **Price Optimization**

Price optimization is a complex notion and uses numerous combinations of various methods and algorithms. This process combines the data not related to the expected costs and risk features, loss and expenses, and further analysis. It considers the changes in comparison to the previous year and policy. Hence price optimization is related to the customers' price sensitivity. Price optimization helps to increase the customer's loyalty and comes with a maximization of profit and income.

## **Personalization**

Customers always like personalized and relevant insurance experiences that would match their requirements and expectations. Insurers face the challenge of assuring digital communication with their customers to meet these demands. Highly personalized and relevant insurance experiences are guaranteed with the help of artificial intelligence and advanced analytics, which extracts the insights from a vast amount of data

-

## Big Data Example 1: Predicting Opioid Abuse

*Wall Street Journal*, December 13, 2018: “**Cigna's Algorithms Aim to Predict Opioid Abuse:** Machine learning and predictive analytics is helping the health insurance company identify customers likely to overdose”

Cigna Corp. is using artificial intelligence to predict whether patients might abuse and or overdose on prescription opioids as part of the company's commitment to reducing the substance's use among its consumers, said Mark Boxer, executive vice president and global chief information officer.

Cigna's proprietary algorithms are aided by the use of machine learning, a subfield of artificial intelligence that refers to the science of getting computers to act intelligently without being explicitly programmed.

A combination of 16 datasets are used to inform the algorithms, including data about patients' behavioral health claims, chronic disease history and interactions with pharmacies. The algorithms were built with the help of in-house staff. Over the past few years, Cigna has hired more than 1,000 data scientists, software engineers and analytics experts, Dr. Boxer said.

## **Big Data Example 2: Willis Towers Watson RADAR**

“Willis Towers Watson Radar is a platform of software products for Property & Casualty (P&C) and Life insurer and intermediary pricing teams, including actuaries and underwriters. These products provide powerful management information to support portfolio monitoring and rate setting. Further products are available that perform price optimization and optionally integrate with rating systems to deliver real-time processing capability”

## “What Radar can do for you”<sup>1</sup>

“Provide flexible management information

Enable informed pricing strategy development underpinned by robust analysis of alternatives

Model the impact of potential pricing decisions on volume, profitability and other performance metrics

Highlight weaknesses and cross-subsidies in current pricing structures

Reduce the risk of pricing anti-selection

Allow sophisticated competitor monitoring and analysis

Enable sophisticated demand-based price modeling founded on real world, practical business constraints

Support real-time decision making

Align the pricing and marketing functions through common value metrics and behavioral models

Deliver data and graphic-rich management information that is fully customizable”

---

<sup>1</sup> <https://www.willistowerswatson.com/en-US/Solutions/products/radar>

## “Radar Base

Radar Base provides powerful management information that you can use to develop business plans and summarize results of risk modeling and competitor analysis. With Radar Base, you can test the effectiveness of new prices through scenario modeling, extracting maximum value from your predictive models to help make successful business decisions. And it can provide detailed rate-setting models for individual insurance products.”

## “Radar Optimiser

Radar Optimiser extends Radar Base functionality with sophisticated price optimization capabilities.”

## **Radar Pricing Software 4.5 Tests Fairness**

The new release emphasizes pricing transparency, giving insurers tools to measure fairness in their pricing models and fulfill their commitment to giving customers clear and fair information.

“Increasing regulatory pressure to demonstrate transparency in pricing means insurers must work out the best way to define, monitor and exhibit fairness.

“This latest release of Radar gives insurers the tools they need to measure fairness in their pricing models and deliver on their commitment to give customers clear and fair information so they can make the right decision.”

Radar 4.5 includes an evaluation library component to help insurers assess their pricing choices against several measures of fairness, such as fairness through unawareness, the quota system and conditional group parity, and determine whether or not their prices adhere to or violate any of those particular metrics within their portfolio, Willis Towers Watson reports. Other updated features of Radar include further enhancements to the elastic net machine learning method that the solution uses.



The full Willis Towers Watson Radar product suite comprises the Radar Base modeling and reporting environment, Radar Dashboard for sharing pricing management information across the organization, and Radar Optimizer for price optimization and Radar Live for rate delivery.

All the solution set's components are fully integrated with Emblem, which fits predictive models rapidly to very large and complex data sets to reveal the underlying pattern; and Classifier, which provides detailed categorization and assessment of risk by geography, according to the vendor.

## Big Data Example 3: Facial Analytics

From “The Why and What of Accelerated Underwriting”<sup>2</sup>

Accelerated underwriting with new data sources . . . can cause movement of between risk classes of existing insured/applicant pool.

Multiple new data sources to address the full UW space: Wearable technology, Credit profiles, Criminal histories, Smarter App & Candor Analytics

Acceleration without Automation may leave companies falling short of the ultimate potential to change the paradigm.

Most of the new, emerging commonly suggested alternative data sources can be used to predict/stratify mortality:

Criminal History, Credit Mortality Risk Score, Facial Analytics

---

<sup>2</sup> <https://www.acli.com/-/media/ACLI/Files/Events/MED2018/Mon021918TheWhyandWhatofAcceleratedUnderwritingMaryBahnaNolanonlineversion.ashx?la=en>

**Facial Analytics** is one emerging technology that may be used to verify smoker status, BMI, other diseases and reduce the sentinel effect

Technological advances allow the combining of facial analytics with constantly evolving bio-demographic data to provide insurers with more insight, speed and accuracy than ever before.

While insurance companies have traditionally used chronological age for estimating lifespan, this technology provides a new, scientifically proven method of forecasting mortality based on estimates of the rate at which someone is aging. As no two people age at the same rate, by taking each user's individual traits into account, facial recognition provides more realistic and reliable results.

## **AINow, “Discriminating Systems: Gender, Race and Power in AI”<sup>3</sup>**

***The use of AI systems for the classification, detection, and prediction of race and gender is in urgent need of re-evaluation.***

The histories of ‘race science’ are a grim reminder that race and gender classification based on appearance is scientifically flawed and easily abused. Systems that use physical appearance as a proxy for character or interior states are deeply suspect, including AI tools that claim to detect sexuality from headshots, predict ‘criminality’ based on facial features, or assess worker competence via ‘micro-expressions.’ Such systems are replicating patterns of racial and gender bias in ways that can deepen and justify historical inequality. The commercial deployment of these tools is cause for deep concern.

---

<sup>3</sup> <https://ainowinstitute.org/discriminatingystems.pdf>

## **“Facing Up to Bias in Facial Recognition,” *American Banker*<sup>4</sup>**

Last week the American Civil Liberties Union demanded that Amazon stop selling its Rekognition program to government agencies and police departments. The ACLU said the technology is flawed and that it is worried law enforcement agencies will use the system to track protesters and immigrants.

Recent studies have shown facial recognition systems tend to have higher error rates for women and minorities than white men.

Antony Haynes, associate dean for strategic initiatives and information systems at Albany Law School, pointed out that all artificial intelligence systems have the potential for bias.

“One assumption we make as human beings is that putting something in software makes it somehow objective or neutral or unbiased,” he said. “That couldn’t be further from the truth because a human being has to write the software, provide the training data, and tell the system when it succeeds or fails.”

---

<sup>4</sup> <https://www.americanbanker.com/news/facing-up-to-bias-in-facial-recognition>, May 29, 2018

## Big Data Example 4: Genetic Information

### SwissRe SONAR 2019

The major challenge for life insurers is to obtain adequate and risk-relevant information during the underwriting process, since existing regulation was mostly enacted before the widespread distribution of direct-to-consumer (DTC) genetic tests. Generally, regulation disallows the use of genetic information in underwriting life insurance.<sup>5</sup> This raises the prospect of more customers at higher risk of disease or mortality applying for life insurance, leading to adverse selection. Customers in the know may also fear being denied life cover due to some genetic conditions, leading the insured to withhold such information from the insurer.

Regulation that stimulates genetic information asymmetry will significantly impact insurers' ability to offer attractively priced coverage, and may challenge the way in which insurance risk is considered and managed. Insurers must be able to evaluate relevant consumer information when underwriting, and that includes risk-relevant data from genetic tests.

---

<sup>5</sup> Not in the U.S.

Currently, there seem to be three broad regulatory approaches to access and use of genetic data for risk assessment: none/self-regulation, limitations by law, and outright legal ban. This lack of uniform approach shows the need for ***industry groups and regulators to work together to agree on reasonable self-regulation***, one that balances the interests of consumers while maintaining the ability of insurers to underwrite sustainable products.

As with any new innovation, there will be a challenging transition period in which insurers will need to develop the know-how of capturing and managing the data, design systems to incorporate the data and implement new underwriting approaches. As the results yielded by genetic tests become more accurate and their use becomes more widespread, the way insurers traditionally pool risk to differentiate individual risks may no longer be suitable.

American Council of Life Insurers, *Understanding Risk Classification*

Put simply, it makes sense for the applicant and insurer to be "on the same page," sharing vital information. If someone knows a risk factor about themselves and are applying for coverage, the insurers should know as well.

**From the *Position Paper on Genetic Information*  
U.S. Council for Responsible Genetics from 1996.**

Unlike infectious diseases, genetic conditions exist at a fairly stable incidence in our society. There is no epidemic of genetic conditions. Thus, they are already reflected in the actuarial tables used by insurers to establish rates. It is misleading for insurers to suggest that their financial solvency **will be** jeopardized if they are obligated to insure people at risk for genetic conditions. In fact, insurers have always insured people at risk for genetic conditions.

The insurance industry has offered no compelling reason to specifically exclude this group from the insured pool now. Early identification of risk status may actually lead to insurer cost savings as a result of preventative care and longer life spans during which premiums can be collected.



Recent developments in human genetic science and the technology of testing are not identifying new costly diseases which were not previously accounted for by the insurance industry's' actuarial data. Rather, these developments are only facilitating the identification of those individuals who carry disease-associated genes at earlier times; many of these people will never have a related illness, or will experience a lifetime of the asymptomatic, presymptomatic or minimally symptomatic phases of the condition. It is not, therefore, the cost of financing the care of genetic conditions which is driving the call for access and inclusion of genetic information in insurance practices. There is no reason for insurers to begin to use this new predictive information now, merely because it is available.

## GENETIC DISCRIMINATION SETS A DANGEROUS PRECEDENT

Genetic testing is not only a medical procedure. It is also a way of creating social categories. As a basic principle, we believe that people should be evaluated based on their individual merits and abilities, and not based on stereotypes and predictions about their future performance or health status. In most cases, genetic testing can only reveal information about probabilities, not absolute certainties. We believe that individuals should not be judged based on stereotypes and assumptions about what people in their class or status are like. Insurance or employment practices which employ these stereotypes in underwriting inadvertently reinforce them in other arenas as well.

There is a strong public policy precedent for avoiding the negative social consequences of such a practice. For example, statistics demonstrate that African Americans do not live as long as Americans of Northern European descent, even when one controls for socio-economic factors. And yet no life insurance company in the country rates applicants differentially on the basis of race. To do so would violate deeply held community values about equality and equal access. Skin color, like other genetic traits, is mediated by genes. These lie entirely outside the individual's control. Whereas individuals can exercise choices about whether to smoke, how much exercise they get, and how much fat is in their diets, they cannot change the contents of their genes. To make employment or insurance decisions on the basis of genetic characteristics determined at the moment of conception is to discard cherished beliefs in justice and equality.

## **Big Data Example 5: Aerial Photography at Granular Detail**

Geospatial Intelligence Center<sup>6</sup>

“Our Metro-Maps collection is based on versatile, high-resolution aerial systems and offer deep insight into top US metropolitan areas through 7.5cm GSD vertical and 360° oblique views. Mobile mapping sensors provide street-level intelligence through 360° imagery along with elevation data produced through high-density point clouds. Collections are repeated on an annual basis and more frequently for high growth locations.”

### **Keener Insight Into Property Conditions**

The GIC offers the most comprehensive, diverse and detailed imagery collection along with tools and analytics for keener insight into property condition. Combined with historical data of a surrounding area, predictive analysis can be made for deeper understanding of underwriting risk. The result is streamlined underwriting, allowing you to better serve your customers while reducing operating costs.

---

<sup>6</sup> <http://geointel.org/annual-imagery-program/>

## **Rapid & Effective Disaster Response**

GIC high-resolution aerial imagery is rapidly accessible to first responders, humanitarian organizations, and federal and state agencies, providing them with actionable insight into the situation on the ground.

Funded by the GIC consortium of insurers, this support is provided to help responders assist those in need and save lives during and after disaster

The GIC is a National Insurance Crime Bureau (NICB) initiative in partnership with Vexcel Imaging that is focused on building a national database of high resolution imagery to be used by its member companies that write almost 80 percent of all property/casualty insurance and over 94 percent of all auto insurance in the country, as well as public sector and non-governmental organizations. The Geospatial Intelligence Center has previously mapped the areas hardest hit by hurricanes and disasters and those views are also available through the web map portal.

## Big Data Example 6: Criminal History

**“TransUnion recently evaluated the predictive power of court record violation data (including criminal and traffic violations)**

“Also, as court records are created when the initial citation is issued, they provide insight into violations beyond those that ultimately end up on the MVR—such as violation dismissals, violation downgrades, and pre-adjudicated or open tickets.”

***What is the likelihood that TU Criminal History Scores have a disparate impact against African-Americans? Consider policing records in Ferguson, Missouri.***

## US DOJ Investigation of the Ferguson Police Department

Ferguson's approach to law enforcement both reflects and reinforces racial bias, including stereotyping. ***The harms of Ferguson's police and court practices are borne disproportionately by African Americans, and there is evidence that this is due in part to intentional discrimination on the basis of race.***

Ferguson's law enforcement practices overwhelmingly impact African Americans. Data collected by the Ferguson Police Department from 2012 to 2014 shows that African Americans account for 85% of vehicle stops, 90% of citations, and 93% of arrests made by FPD officers, despite comprising only 67% of Ferguson's population.

## US DOJ Investigation of the Ferguson Police Department (2)

FPD appears to bring certain offenses almost exclusively against African Americans. For example, from 2011 to 2013, African Americans accounted for 95% of Manner of Walking in Roadway charges, and 94% of all Failure to Comply charges.

***Our investigation indicates that this disproportionate burden on African Americans cannot be explained by any difference in the rate at which people of different races violate the law. Rather, our investigation has revealed that these disparities occur, at least in part, because of unlawful bias against and stereotypes about African Americans.***



## Big Data Example 7: Carpe Data

Using proprietary algorithms and proven AI, Carpe Data harnesses the power of emerging and alternative data for insurance carriers around the world.

Claims Activity performs the search and analysis of highly impactful social and web data. It monitors and finds the “needle in the haystack” to provide reliable actionable claimant activity.

Leveraging social data sources is a must for the modern insurer:

- It influences claims investigation and handling
- Its efficient use impacts the ultimate claim cost

## ***Carpe Data Claims Activity – How It Works***

Pictured in Valentine's Day swing dance at local restaurant site

Noted on neighborhood blog as winning the Easter Egg Roll contest

Posted about enjoying a big family picnic

Finishing time in the Turkey Trot 5K

## **Next Generation Indexes**

A suite of indexes targeting dimensions of risk that can be tuned by segment and location.

Cleans and normalizes data that can be sparse, noisy, biased, unstructured, or redundant.

## **Big Data Example 8: ISO ClaimSearch and Related Tools**

### **Fast-track meritorious claims while improving fraud detection**

ISO ClaimSearch<sup>®</sup> is more than the world's largest database of property/casualty claims—it's also a robust claims intelligence platform. Its claims-matching technology is an essential first step in fraud detection, and the results can help facilitate fast-tracking. The platform also provides seamless access to integrated claims fraud-detection and investigative analysis tools.

## Claims Director

### Improve claims triage with our fraud analytics scoring system

ClaimDirector uses industry-based rules to analyze claims and identify fraud indicators. It then calculates a score to help adjusters decide whether to process a claim or triage it for investigation. SIU teams can use the solution to uncover potential fraud and explore suspect factors. The solution uses advanced algorithms to determine the likelihood of claims fraud:

- Compares claims to 1.3 billion records in [ISO ClaimSearch](#)<sup>®</sup> plus NICB data
- Evaluates claims by type, line of business, loss date, and loss type
- Provides scores for both claim and entity
- Revises scores in real time as claims are updated
- Customizes claims triage based on company preferences

## Liability Navigator

Achieve consistent settlements, accurate liability assessments, and better subrogation recovery

Insurers are charged with resolving claims as quickly and fairly as possible—while minimizing leakage and paying only what is owed. Applying accurate liability and value determinations is critical to achieving the right balance. But currently, claims handlers often are unable to identify cases in which a claimant is partially at fault. So insurers miss opportunities to improve the bottom line.

With Liability Navigator<sup>®</sup>, customers report improvements—**from 8 percent to more than 20 percent**—in the application of comparative liability.

## ***Big Data Algorithms Can Reflect and Perpetuate Historical Inequities***

Barocas and Selbst: *Big Data's Disparate Impact*<sup>7</sup>

Advocates of algorithmic techniques like data mining argue that they eliminate human biases from the decision-making process. But an algorithm is only as good as the data it works with. Data mining can inherit the prejudices of prior decision-makers or reflect the widespread biases that persist in society at large. Often, the “patterns” it discovers are simply preexisting societal patterns of inequality and exclusion. Unthinking reliance on data mining can deny members of vulnerable groups full participation in society.

---

<sup>7</sup> [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2477899](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899)

## **Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor***

America's poor and working-class people have long been subject to invasive surveillance, midnight raids, and punitive public policy that increase the stigma and hardship of poverty. During the nineteenth century, they were quarantined in county poorhouses. During the twentieth century, they were investigated by caseworkers, treated like criminals on trial. Today, we have forged what I call a digital poorhouse from databases, algorithms, and risk models. It promises to eclipse the reach and repercussions of everything that came before.

# Algorithmic Bias

Steve Bellovin, “Yes, ‘algorithms’ can be biased. Here’s why. A computer scientist weighs in on the downsides of AI.”<sup>8</sup>

This is what's important: machine-learning systems—“algorithms”—produce outputs that reflect the training data over time. If the inputs are biased (in the *mathematical* sense of the word), the outputs will be, too. Often, this will reflect what I will call “sociological biases” around things like race, gender, and class.

One thing is to exercise far more care in the selection of training data. Failure to do that was the likely root cause of Google Images labeling two African-Americans as gorillas. Sometimes, fixing the training data can help.

Of course, this assumes that developers are even aware of the bias problem. Thus, another thing to do is to test for biased outputs—and some sensitive areas, such as the criminal justice system, simply do not use these kinds of tools.

---

<sup>8</sup> <https://arstechnica.com/tech-policy/2019/01/yes-algorithms-can-be-biased-heres-why/>



Given that ML systems (including facial recognition systems) can produce biased output, how should society treat them? Remember that, often, the choice is not between algorithmic output and perfection but between algorithmic decisions and human ones—and humans are demonstrably biased, too. That said, there are several reasons to be wary of the "algorithmic" approach.

One reason is that people put too much trust in computer output. Every beginning programmer is taught the acronym "GIGO:" garbage in, garbage out. To end users, though, it's often "garbage in, gospel out"—if the computer said it, it must be so. (This tendency is exacerbated by bad user interfaces that make overriding the computer's recommendation difficult or impossible.) We should thus demand less bias from computerized systems precisely to compensate for their perceived greater veracity.

The second reason for caution is that computers are capable of doing things—even bad things—at scale. There is at least the perceived risk that, say, computerized facial recognition will be used for mass surveillance. Imagine the consequences if a biased but automated system differentially misidentified African-Americans as wanted criminals. Humans are biased, too, but they can't make nearly as many errors per second.

Our test, then, should be one called disparate impact. "Algorithmic" systems should be evaluated for bias, and their deployment should be guided appropriately. Furthermore, the more serious the consequences, the higher the standard should be before use.

## **Amazon Created a Hiring Tool Using A.I. It Immediately Started Discriminating Against Women.<sup>9</sup>**

All of this is a remarkably clear-cut illustration of why many tech experts are worried that, rather than remove human biases from important decisions, artificial intelligence will simply automate them. An investigation by ProPublica, for instance, found that algorithms judges use in criminal sentencing may dole out harsher penalties to black defendants than white ones. Google Translate famously introduced gender biases into its translations. The issue is that these programs learn to spot patterns and make decisions by analyzing massive data sets, which themselves are often a reflection of social discrimination. Programmers can try to tweak the AI to avoid those undesirable results, but they may not think to, or be successful even if they try.

---

<sup>9</sup> Jordan Wasserman at, <https://slate.com/business/2018/10/amazon-artificial-intelligence-hiring-discrimination-women.html>

## **“Insurance Companies Are Paying Cops To Investigate Their Own Customers,” BuzzFeed News, August 15, 2019**

When police showed up at Harry Schmidt's home on the outskirts of Pittsburgh, he thought they were there to help. He was still mourning the disappearance of the beloved forest green Ford F-150 pickup that he'd customized with a gun storage cabinet, and he hoped the cops had solved the crime.

Instead, the officers accused him of faking the theft. The Vietnam veteran was now facing up to seven years in prison.

Schmidt was stunned, but he was even more upset when he found out who had turned him in.

Erie Insurance, one of the nation's largest auto insurers, had not only provided the cops with evidence against its own loyal customer — it had actively worked with them to try to convict him of insurance fraud.

Schmidt, a grandfather living on disability benefits from his war-related injuries, had no history of theft or fraud. But he found himself the target of an extraordinary alliance between private insurers and public law enforcement agencies — one that transforms routine claims into criminal evidence, premium-paying customers into suspects, and the justice system into a hired gun for a multibillion-dollar industry. It's an arrangement essentially unheard of in other businesses, and one rife with potential conflicts of interest, as well as grave consequences for law-abiding customers.

BuzzFeed News examined 27 cases around the country in which people were falsely charged with felonies based in whole or in part on evidence insurers provided to law enforcement. In Indiana, State Farm helped detectives craft an arrest warrant for a contractor who was charged with 14 felonies. All charges were ultimately dropped when the evidence turned out to be deeply misleading — but not before the insurance giant’s allegations had destroyed his business. In Georgia, a local prosecutor relied on lab tests provided by an insurer to charge a woman with arson, resulting in a three-year ordeal in which she ended up homeless, only to drop the charges when the test results proved unreliable. And in Wisconsin, a man spent nearly three years in prison based on now-discredited science used by an insurance investigator until his conviction was overturned.

It's impossible to say how often innocent customers are the victims of false charges, because much of the potential evidence is hidden from public view by the companies themselves. In one case, investigators at State Farm withheld several crucial reports contradicting their fraud allegations from the bundle of evidence they handed over the law enforcement. In another, a Farmers manager admitted under oath that there was an "unwritten policy" within the company to withhold evidence from customers that could help prove their innocence.

But by reviewing hundreds of court records from around the country, complaints filed with state regulators, and internal company records, and interviewing dozens of former insurance company employees, BuzzFeed News uncovered a system that has ensnared countless innocent people.

In Florida alone, state law enforcement received roughly 14,500 suspected fraud referrals about homeowners and vehicle claims in the past five years — the vast majority of them from insurance companies. Yet authorities determined that in more than 75% of the cases, there was not enough evidence to move forward with a criminal investigation. And regardless of whether fraud accusations turn out to be false, many sit in databases shared among insurers, where they can haunt people if they ever have to file another insurance claim.



## “The Real Reason Tech Struggles with Algorithmic Bias”<sup>10</sup>

These are mistakes made while trying to do the right thing. But they demonstrate why tasking untrained engineers and data scientists with correcting bias is, at the broader level, naïve, and at a leadership level insincere.

No matter how trained or skilled you may be, it is 100 percent human to rely on cognitive bias to make decisions. Daniel Kahneman’s work challenging the assumptions of human rationality, among other theories of behavioral economics and heuristics, drives home the point that human beings cannot overcome all forms of bias. But slowing down and learning what those traps are—as well as how to recognize and challenge them—is critical. As humans continue to train models on everything from stopping hate speech online to labeling political advertising to more fair and equitable hiring and promotion practices, such work is crucial.

---

<sup>10</sup> Yael Eisenstat at <https://www.wired.com/story/the-real-reason-tech-struggles-with-algorithmic-bias/>

Becoming overly reliant on data—which in itself is a product of availability bias—is a huge part of the problem. In my time at Facebook, I was frustrated by the immediate jump to “data” as the solution to all questions. That impulse often overshadowed necessary critical thinking to ensure that the information provided wasn't tainted by issues of confirmation, pattern, or other cognitive biases.

There is not always a strict data-driven answer to human nature. The belief that simply running a data set will solve for every challenge and every bias is problematic and myopic. To counter algorithmic, machine, and AI bias, human intelligence must be incorporated into solutions, as opposed to an over-reliance on so-called “pure” data.

## **NAIC Activities Related to Big Data and AI**

**2016:** The Market Regulation and Consumer Affairs (D) Committee will:

Explore insurers' use of big data for claims, marketing, underwriting and pricing. Explore potential opportunities for regulatory use of big data to improve efficiency and effectiveness of market regulation. If appropriate, make recommendations no later than the 2016 Fall National Meeting for 2017 charges for the Committee to address any recommendations identified by the 2016 exploration.

**2017:** The Big Data (D) Task Force will:

A. Review current regulatory frameworks used to oversee insurers' use of consumer and non-insurance data. If appropriate, recommend modifications to model laws/regulations regarding marketing, rating, underwriting and claims, regulation of data vendors and brokers, regulatory reporting requirements, and consumer disclosure requirements.

B. Propose a mechanism to provide resources and allow states to share resources to facilitate states' ability to conduct technical analysis of and data collection related to states' review of complex models used by insurers for underwriting, rating, and claims. Such mechanism shall respect and in no way limit states' regulatory authority.

C. Assess data needs and required tools for regulators to appropriately monitor the marketplace and evaluate underwriting, rating, claims, and marketing practices. This assessment shall include gaining a better understanding of currently available data and tools and recommendations for additional data and tools as appropriate. Based upon this assessment, propose a means to collect, house, and analyze needed data.

**2017:** The Innovation and Technology (EX) Task Force will:

A. Provide a forum for the discussion of innovation and technology developments in the insurance sector, including the collection and use of data by insurers and state insurance regulators—as well as new products, services and distribution platforms—in order to educate state insurance regulators on how these developments impact consumer protection, insurer and producer oversight, marketplace dynamics, and the state-based insurance regulatory framework.

B. Develop regulatory guidance, white papers, model laws and/or regulations, or make other recommendations to the Executive (EX) Committee, as appropriate.

C. Discuss regulatory issues that arise with the development of autonomous vehicles; study and, if necessary, develop recommendations for changes needed to the state-based insurance regulatory framework; consider development of a white paper or model legislation, if necessary.

D. Discuss emerging issues related to on-demand insurance applications—in addition to potential implications on the state-based insurance regulatory structure—including, but not limited to, cancellations, nonrenewals, coverage issues, notice provisions and policy-delivery requirements.

E. Coordinate with other NAIC committees and task forces, as appropriate, on technology and innovation issues.

**2019:** The Market Regulation and Consumer Affairs (D) Committee will review state insurance privacy protections regarding the collection, use, and disclosure of information gathered in connection with insurance transactions and make recommended changes, as needed, to certain NAIC models, such as the Insurance Information and Privacy Protection Model Act (#670) and the Privacy of Consumer Financial and Health Information Model Regulation (#672) by the 2020 Summer National Meeting.

**2019:** The Artificial Intelligence (EX) Working Group of the Innovation and Technology (EX) Task Force will study the development of artificial intelligence, its use in the insurance sector, and its impact on consumer protection and privacy, marketplace dynamics, and the state-based insurance regulatory framework. The Working Group will develop regulatory guidance, beginning with guiding principles, and make other recommendations to the Innovation and Technology (EX) Task Force as appropriate by the 2020 Summer National Meeting.”

# **Insurers' / Producers' Use of Personal Consumer Data – the Role of Supervisors**

## Step 1: Establish Values and Principles

“Before we choose our tools and techniques, we must first choose our dreams and values, for some tools serve them while others make them unobtainable.” Tom Bender



## CEJ's Principles/Values for Ethical AI

- **Cost-Based Pricing:** Protect insurer financial condition, provide proper investment risk / mitigation benefit price signals, fair treatment of consumers entering into contracts of adhesion
- **Loss Prevention / Mitigation / Sustainability / Resilience:** Enhance, not undermine the loss prevention potential of insurance
- **Risk Pooling:** Protect risk diversification, availability and affordability of insurance
- **Availability / Affordability –** Address the protection gap for low- and moderate-income consumers and small businesses. Most important tool for individual, business and community recovery and resilience.
- **Fair Competition –** Antitrust enforcement for emerging types of collective pricing and claim settlement practices facilitated by big data algorithms

- Fair Competition – Empower consumers by more symmetric sharing of information between insurers and consumers
- Digital Rights – consumer ownership and consent to identified uses, protection of consumer data, contestability, disclosure of and remediation following data breaches
- Transparency, Explainability and Accountability – ethical and accountable algorithms
- Regulatory and Legal Compliance – compliance with the letter of and the intent of the law

# Values to Guide Ethical Use of AI

## Artificial Intelligence: Australia's Ethics Framework<sup>11</sup>

### Core Principles for AI

1. Generate Net-Benefits
2. Do No Harm
3. Regulatory and Legal Compliance
4. Privacy Protection
5. Fairness
6. Transparency and Explainability
7. Contestability
8. Accountability

---

<sup>11</sup> <https://consult.industry.gov.au/strategic-policy/artificial-intelligence-ethics-framework/>

# Values to Guide Ethical Use of AI

## Transatlantic Consumer Dialogue

Resolution on Fintech: [http://tacd.org/wp-content/uploads/2018/11/TACD\\_resolution\\_fintech\\_final.pdf](http://tacd.org/wp-content/uploads/2018/11/TACD_resolution_fintech_final.pdf)

Resolution on Competition, Privacy and Consumer Welfare: [http://tacd.org/wp-content/uploads/2018/12/TACD-Resolution\\_Comp\\_Dec2018\\_final.pdf](http://tacd.org/wp-content/uploads/2018/12/TACD-Resolution_Comp_Dec2018_final.pdf)

European Commission Consultation: FinTech –A more competitive and innovative European Financial Sector

<http://tacd.org/wp-content/uploads/2017/06/TACD-response-to-the-European-Commission-consultation-on-Fintech.pdf>

## **New York Department of Financial Services Guidance to Insurers:**

“Use of External Consumer Data and Information Sources in Underwriting for Life Insurance”<sup>12</sup>

“Following reports of the emergence of unconventional sources or types of external data available to insurers, including within algorithms and predictive models, the New York State Department of Financial Services (“Department”) commenced an investigation of insurers’ underwriting guidelines and practices in New York related to the use of external data in underwriting for life insurance.

“For purposes of this Circular Letter, external data includes any data or information sources not directly related to the medical condition of the applicant that is used – in whole or in part – to supplement traditional medical underwriting, as a proxy for traditional medical underwriting, or to establish “lifestyle indicators” that may contribute to an underwriting assessment of an applicant for life insurance coverage.

---

<sup>12</sup> [https://www.dfs.ny.gov/industry\\_guidance/circular\\_letters/cl2019\\_01](https://www.dfs.ny.gov/industry_guidance/circular_letters/cl2019_01)

“The Department fully supports innovation and the use of technology to improve access to financial services. Indeed, insurers’ use of external data sources has the potential to benefit insurers and consumers alike by simplifying and expediting life insurance sales and underwriting processes. External data sources also have the potential to result in more accurate underwriting and pricing of life insurance. At the same time, however, the accuracy and reliability of external data sources can vary greatly, and many external data sources are companies that are not subject to regulatory oversight and consumer protections, which raises significant concerns about the potential negative impact on consumers, insurers and the life insurance marketplace in New York.

This circular letter addresses two particular areas of immediate concern with the use of external data sources, algorithms or predictive models that were identified during the Department's investigation. First, the use of external data sources, algorithms, and predictive models has a significant potential negative impact on the availability and affordability of life insurance for protected classes of consumers. An insurer should not use an external data source, algorithm or predictive model for underwriting or rating purposes unless the insurer can establish that the data source does not use and is not based in any way on race, color, creed, national origin, status as a victim of domestic violence, past lawful travel, or sexual orientation in any manner, or any other protected class. Moreover, an insurer should also not use an external data source for underwriting or rating purposes unless the use of the external data source is not unfairly discriminatory and complies with all other requirements in the Insurance Law and Insurance Regulations.

Second, the use of external data sources is often accompanied by a lack of transparency for consumers. Where an insurer is using external data sources or predictive models, the reason or reasons for any declination, limitation, rate differential or other adverse underwriting decision provided to the insured or potential insured should include details about all information upon which the insurer based such decision, including the specific source of the information upon which the insurer based its adverse underwriting decision.



## Unfair Discrimination

Based on its investigation, the Department has determined that insurers' use of external data sources in underwriting has the strong potential to mask the forms of discrimination prohibited by these laws. Many of these external data sources use geographical data (including community-level mortality, addiction or smoking data), homeownership data, credit information, educational attainment, licensures, civil judgments and court records, which all have the potential to reflect disguised and illegal race-based underwriting that violates Articles 26 and 42.

Other models and algorithms purport to make predictions about a consumer's health status based on the consumer's retail purchase history; social media, internet or mobile activity; geographic location tracking; the condition or type of an applicant's electronic devices (and any systems or applications operating thereon); or based on how the consumer appears in a photograph. At the very least, the use of these models may either lack a sufficient rationale or actuarial basis and may also have a strong potential to have a disparate impact on the protected classes identified in New York and federal law.

In light of the Department's investigation and findings, the Department is providing the following principles that insurers should use as guidance in using external data sources in underwriting.

First, an insurer should not use an external data source, algorithm or predictive model in underwriting or rating unless the insurer has determined that the external tools or data sources do not collect or utilize prohibited criteria. An insurer may not simply rely on a vendor's claim of non-discrimination or the proprietary nature of a third-party process as a justification for a failure to independently determine compliance with anti-discrimination laws. The burden remains with the insurer at all times.

Second, an insurer should not use an external data source, algorithm or predictive model in underwriting or rating unless the insurer can establish that the underwriting or rating guidelines are not unfairly discriminatory in violation of Articles 26 and 42. In evaluating whether an underwriting or rating guideline derived from external data sources or information is unfairly discriminatory, an insurer should consider the following questions:

(1) Is the underwriting or rating guideline that is derived, in whole or in part, from external data sources or information supported by generally accepted actuarial principles or actual or reasonably anticipated experience that justifies different results for similarly situated applicants?

(2) Is there a valid explanation or rationale for the differential treatment of similarly situated applicants reflected by the underwriting or rating guideline that is derived, in whole or in part, from external data sources or information?

Importantly, even if statistical data is interpreted to support an underwriting or rating guideline, there must still be a valid rationale or explanation supporting the differential treatment of otherwise like risks. The second part of this inquiry is particularly important where there is no demonstrable causal link between the classification and increased mortality and also where an underwriting or rating guideline has a disparate impact on protected classes.

Data, algorithms, and models that purport to predict health status based on a single or limited number of unconventional criteria also raise significant concerns about the validity of such models.

An insurer may establish guidelines and practices to assess an applicant's health status and identify individuals at higher mortality risk if based on sound actuarial principles or if related to actual or reasonably anticipated experience. However, the data, algorithms, and predictive modeling used by the insurer must comport with the principles set forth above and all other relevant requirements in federal and New York law. An insurer may not rely on external data or external predictive algorithms or models unless the insurer has determined that the external data or predictive model is otherwise permitted by law or regulation and is based on both sound actuarial principles or experience and a valid explanation or rationale.

## Consumer Disclosure/Transparency

Transparency is an important consideration in the use of external data sources to underwrite life insurance. Pursuant to Insurance Law § 4224(a)(2), insurers must notify the insured or potential insured of the right to receive the specific reason or reasons for a declination, limitation, rate differential or other adverse underwriting decision. An adverse underwriting decision would include the inability of an applicant to utilize an expedited, accelerated or algorithmic underwriting process in lieu of a traditional medical underwriting. Where an insurer is using external data sources or predictive models, the reason or reasons provided to the insured or potential insured must include details about all information upon which the insurer based any declination, limitation, rate differential or other adverse underwriting decision, including the specific source of the information upon which the insurer based its adverse underwriting decision.

An insurer may not rely on the proprietary nature of a third-party vendor's algorithmic processes to justify the lack of specificity related to an adverse underwriting action. Insurers must also provide notice to and obtain consent from consumers to access external data, where required by law or regulation. The failure to adequately disclose the material elements of an accelerated or algorithmic underwriting process, and the external data sources upon which it relies, to a consumer may constitute an unfair trade practice under Insurance Law Article 24.

# **New York DFS Rule Regarding Use of Education and Occupation Insurance Regulation 150**

During this investigation, the superintendent learned that some, but not all, insurers in the state use an individual's education level attained and/or occupational status in establishing initial tier placement. The insurers' consideration of these factors has resulted in cases where classes of insureds have been placed in less favorably rated tiers, which may lead to higher premiums, without adequate substantiation that an individual's level of education attained and/or occupational status relates to his or her driving ability or habits such that the insurer would suffer a greater risk of loss. The insurers failed to provide sufficient support for the existence of the necessary relationship for the use of occupational status or any convincing evidence to support the necessary relationship for the use of an insured's level of education attained, whether alone or in combination with occupational status. As a result, the insurers failed to establish that their use of education and/or occupation in establishing initial tier placement was not unfairly discriminatory.

(b) Use of Education Level Attained in Multi-Tier Programs. As a result of the superintendent's determination following investigation that certain insurers' use of education level attained in initial tier placement results in unfairly discriminatory rates, an insurer shall not use an individual's education level attained as a factor in either initial tier placement, tier movement, or the establishment of the rates at all, unless the insurer demonstrates to the satisfaction of the superintendent that its use of education level attained in initial tier placement or tier movement shall not result in a rate that violates Insurance Law Article 23.

(c) Use of Occupational Status in Multi-Tier Programs.

(1) Except as provided in paragraph (2) of this subdivision, an insurer shall not use occupational status as a factor in either initial tier placement, tier movement, or the establishment of the rates unless the insurer demonstrates to the satisfaction of the superintendent that its use of occupational status in initial tier placement or tier movement shall not result in a rate that violates Insurance Law Article 23.



(2) An insurer may use occupational status as a factor in initial tier placement and tier movement provided that the insurer demonstrates to the satisfaction of the superintendent that its use of the factor shall not result in a rate that violates Insurance Law Article 23 and that each of the following conditions is satisfied:

(i) the insurer demonstrates to the satisfaction of the superintendent that each particular occupation grouping has a reasonable relationship to an insured's driving ability or habits such that an insurer would predictably suffer a greater or lesser risk of loss;

(ii) an insured's income shall not be a risk characteristic, whether directly or indirectly;

(iii) an unemployed person who was previously employed, including a retired person, shall remain in the group designated for his or her previous occupation regardless of the amount of time since such person was last employed;

- (iv) the insurer shall establish a group separate from its other occupational groups that contains homemakers and those insureds who have never been employed, which group shall be a neutral factor for impacting rates;
- (v) where the insured's initial tier placement is influenced by occupational status, the insurer shall provide tier movement rules that address a change in the insured's occupational status at the time of renewal providing the insured with the opportunity to reduce his or her premium;
- (vi) any rate differential permitted under this paragraph shall be commensurate with the related reduction of loss costs and associated premiums; and
- (vii) where insufficient data exists with respect to a particular occupation and the insurer wishes to align that occupation in a group with a similar one, the insurer shall have a sufficiently reasonable explanation for placing the two occupations in the same group.

(e) Every insurer that uses education level attained and/or occupational status in accordance with this section as a factor in its multi-tier program shall provide a written notice acceptable to the superintendent at least annually to the first-named insured that conspicuously explains how an insured may notify the insurer or its agent of any update in the insured's education level attained and/or occupational status that would result in a reduction of premium.

# Supervisory Oversight of Insurers'/Producers' Use of Personal Consumer Data

## Educate

- Establish and publish consumer digital rights
- Declare Values / Establish Principles for Insurers' / Producers' Fair Treatment of Consumers

## Investigate

- Survey insurers/producers on types, sources and uses of personal consumer data – what data from what sources is used for marketing, pricing, sales, claims settlement/anti-fraud
- Collect data on consumer market outcomes

## Evaluate

- Are insurers/producers using certain data for impermissible purposes?
- Are consumer outcomes showing disparate impact against protected classes or otherwise contrary to public policy?

## Prevent Consumer Harm

- Establish guidelines for ethical collection, use, protection of personal consumer data
- Establish guidelines for ethical algorithm development and use
- Promote / celebrate best practices / outcomes consistent with core values

## Redress Consumer Harm

- Identify consumer harm resulting from inappropriate use of consumer personal information / big data algorithms / machine learning / AI
- Stop harmful practices / remediate consumer harm

## Operationalize

- Identify and secure necessary human and digital resources

## **Ethical Algorithms: Minimizing Disparate Impact in Insurance Models**

### One Tool: Consider Prohibited Risk Classes in Model Development

***Step 1: Include race, religion and national origin – or proxies for these characteristics if actual individual characteristic unknown – as independent variables – control variables – in the model.***

By using the characteristics as independent variables in the development of the model, the remaining independent variables' contribution (to explaining the dependent variable) is shorn of that part of their contribution that is a function of correlation with the prohibited characteristics. For the independent variables other than race, religion and national origin, what remains is a more accurate picture of the remaining independent variables' contribution to the target outcome.

***Step 2: Omit race, religion and national origin when the model is deployed.***

New York Times, August 10, 2015: Algorithms and Bias: Q. and A. With Cynthia Dwork

Q: Some people have argued that algorithms eliminate discrimination because they make decisions based on data, free of human bias. Others say algorithms reflect and perpetuate human biases. What do you think?

A: Algorithms do not automatically eliminate bias. . . .Historical biases in the . . .data will be learned by the algorithm, and past discrimination will lead to future discrimination.

Fairness means that similar people are treated similarly. ***A true understanding of who should be considered similar for a particular classification task requires knowledge of sensitive attributes, and removing those attributes from consideration can introduce unfairness and harm utility.***

Q: Should computer science education include lessons on how to be aware of these issues and the various approaches to addressing them?

A: Absolutely! First, students should learn that design choices in algorithms embody value judgments and therefore bias the way systems operate. They should also learn that these things are subtle: For example, designing an algorithm for targeted advertising that is gender neutral is more complicated than simply ensuring that gender is ignored. They need to understand that classification rules obtained by machine learning are not immune from bias, especially when historical data incorporates bias.



## Illustration of One Technique to Minimize Disparate Impact

Let's create a simple model to predict the likelihood of an auto claim:

$$b_0 + b_1X_1 + b_2X_2 + b_3X_3 + e = y$$

Say that  $X_1$ ,  $X_2 + X_3$  are miles driven, driving record and credit score and we are trying to predict  $y$  – the frequency of an auto claim.

Let's assume that all three  $X$ s are statistically significant predictors of the likelihood of a claim and the  $b$  values are how much each  $X$  contributes to the explanation of claim.

$b_0$  is the “intercept” – a base amount and  $e$  is the error term – the portion of the explanation of the claim not provided by the independent variables.

## What Happens When We Explicitly Consider A Variable For Race?

$$b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \mathbf{b_4R_1} + e = y$$

$R_1$  is a control variable – by including race in the model development, the correlation of the  $X$ s to race is statistically removed and the new  $b$  values are now the contribution of the  $X$ s, independent of their correlation to race, to explaining the likelihood of a claim

When the model is deployed, the variable for race is removed – the  $X$ s remain, but the  $b$  values now minimize disparate impact.

## **Why is a Statistical Test for Disparate Impact Consistent with Actuarial Justification Used by Insurers?**

Actuarial justification is a statistical test – that a particular characteristic of the consumer, vehicle, property or environment is correlated with a particular outcome, like pure premium (average claim cost). The same statistical test can be used to evaluate and minimize disparate impact. Stated differently – if a particular correlation and statistical significance is used to justify, say, insurance credit scoring, those same standards of correlation and statistical significance are reasonable evidence of disparate impact and unfair discrimination on the basis of prohibited factors.

## **Ethical Algorithms: Reasonable and Necessary for Insurance Pricing and Claims Settlement Models**

1. Minimizes Disparate Impact – Stop the Cycle of Perpetuating Historical Discrimination.
2. Promotes Availability and Affordability for Underserved Groups
3. Improves Cost-Based Insurance Pricing Models
4. Improve Price Signals to Insureds for Loss Mitigation Investments
5. Help Identify Biases in Data and Modelers / Improve Data Insights
6. Improve Consumer Confidence of Fair Treatment by Insurers

## **Flawed Premise: Greater Risk Segmentation Equals More Accurate Pricing, Better Matching of Price to Risk.**

Micro risk segmentation is a function of greater reliance on more complex predictive algorithms. Greater segmentation – more variables, more data sources – introduces new risks – modeling risk, data bias risk, unfair discrimination risk. Greater segmentation does not create greater “accuracy,” where “accuracy” purports to be better matching price to risk.

The concept of ever more “accurate” and granular risk segmentation must mean greater and greater disparity between the most and least favored consumers with great implications for availability and affordability of insurance and likely burdens on consumers in already-underserved communities under the guide of “matching price to risk” or “fighting fraud.”

# Antitrust and Competition Concerns With Data Brokers and Vendors of Algorithms

Increased antitrust scrutiny and reinvigorated competition analysis is needed to address the market power and potential collusion mechanisms of data and algorithm vendors.

Dozens or hundreds of companies are engaging in practices that have historically required supervisory oversight to exempt the practices from antitrust laws, including the collection and sharing of insurers' experience and the provision of collective pricing guidance. Vendors that collect exposure and claims data from insurers, combine these data with other, non-insurance data to provide pricing or claim settlement tools present mechanisms for collective pricing and claim settlement valuations, also known as collusion.

## Ethical Algorithms -- Sources

Pauline T. Kim, "Auditing Algorithms for Discrimination"

<https://www.pennlawreview.com/online/166-U-Pa-L-Rev-Online-189.pdf>

Claire Whitaker, "Ethical Algorithms"

<https://www.kdnuggets.com/2019/03/designing-ethical-algorithms.html>

Erin Russel, "The Ethical Algorithm"

<https://www.cognitivetimes.com/2019/01/the-ethical-algorithm/>

Barocas and Selbst

[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2477899](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899)

Kroll, et al, "Accountable Algorithms:

[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2765268](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2765268)

Virginia Eubanks, *Automating Inequality: How High Tech Tools Profile, Police and Punish the Poor*

Selbst and Barocas, "The Intuitive Appeal of Explainable Machines

[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3126971](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3126971)

Levy and Barocas, "Designing Against Discrimination in Online Markets

[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3084502](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3084502)

New York Times, "Algorithms and Bias, Q and A with Cynthia Dwork," 10 August 2015

<https://www.nytimes.com/2015/08/11/upshot/algorithms-and-bias-q-and-a-with-cynthia-dwork.html>

Martin, Kirsten E. M., What Is an Ethical Algorithm (And Who Is Responsible for It?) (October 21, 2017). Available at SSRN:

<https://ssrn.com/abstract=3056692> or <http://dx.doi.org/10.2139/ssrn.3056692>

Kirsten Martin, "Ethical Implications and Accountability of Algorithms"

<http://kirstenmartin.net/wp-content/uploads/2013/11/Martin-JBE-Ethics-and-Accountability-of-Algorithms.pdf>

Kirsten Martin, DATA AGGREGATORS, BIG DATA, & RESPONSIBILITY ONLINE

<http://kirstenmartin.net/wp-content/uploads/2013/11/AAPOR-Martin-Info-Value-Chain-v2.pdf>

AIandBigData:Ablueprintforhumanrights,socialandethicalimpactassessmentAlessandroMantelero

<https://reader.elsevier.com/reader/sd/pii/S0267364918302012?token=3836947F0CAD3C145A1F273E3CBE6C38F67E777DD7E4D590548F481916130DAACA8D57BED4667BD1FE1F4D8FC80E7C56>