Discussion of the Book *Weapons of Math Destruction* by Cathy O'Neil

CAS Annual Meeting
November 11, 2020

Panelists

- Moderator – Amy Brener, Director, The CAS Institute
- Panelists:
  - Bob Miccolis – FCAS, Former President, Casualty Actuarial Society
  - Chris Monsour – FCAS, CSPA, Data Robot
  - Louise Francis, FCAS, CSPA, Francis Analytics and Actuarial Data Mining, Inc.
  - Nina Ahmad – PhD, Democratic party candidate for Auditor General, Pennsylvania

Weapons of Math Destruction, by Cathy O'Neil

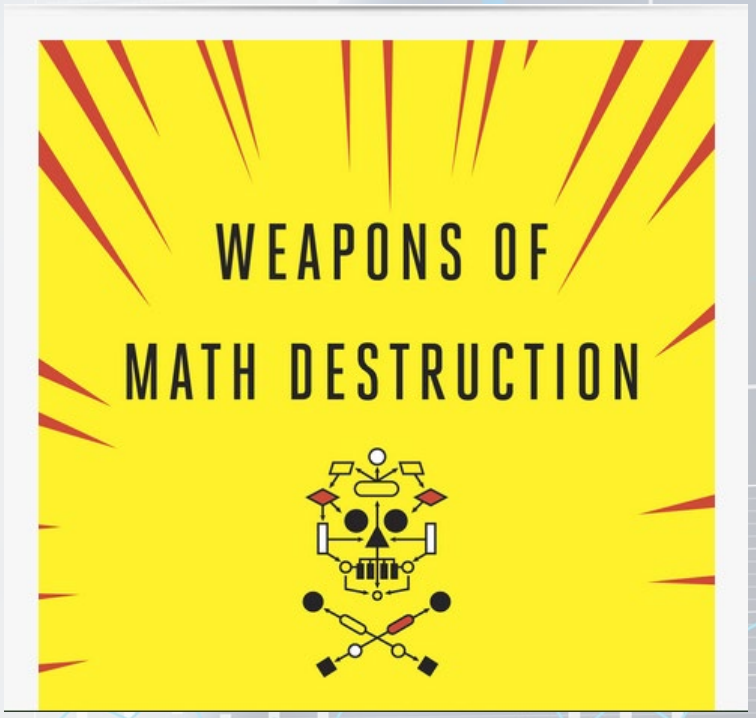


Image from web site:
www.mathbabe.org

The title

- Weapons of Math Destruction a play on the words Weapons of Mass Destruction
- Book is about adverse consequences to society of deployment of predictive models, which have now been widely deployed
- A quote : “thanks to the extraordinary powers that I loved so much, math was able to combine with technology to multiply the chaos and misfortune, adding efficiency and scale to systems that I now recognize as flawed” reflecting author’s experience at a hedge fund and the role of models in the financial crisis

The Book Chapters

- Introduction
  - Brief background on the author and how she came to question models
- Chapter 1 – Bomb Parts: What is in a Model?
  - Have models eliminated bias or camouflaged it? Example LSR-I model used to assess prisoners for parole
- Chapter 2 – Shell Shock: My Journey of Disillusionment
  - Financial weapons of math destruction causing mass financial carnage
- Chapter 3 – Arms Race: Going to College
- Chapter 4 – Propaganda Machine – Online Advertising
  - Discriminatory impact of models used for online advertising
- Chapter 5 – Civilian Casualties – Justice in the Age of Big Data
  - Inclusion of nuisance crimes biases crime prediction models
- Chapter 6 – Ineligible to Serve: Getting a Job

Book Chapter Cont.

- Chapter 7 – Sweating Bullets: On the Job
- Chapter 8 – Collateral Damage: Landing Credit
- Chapter 9 – No Safe Zone: Getting Insurance
  - Use of credit data and e-scores based on data from commercial data vendors, including web usage data in insurance pricing
- Chapter 10 – The Targeted Citizen: Civic Life
  - Ubiquitous use of models by internet technology companies
- Conclusion

Some model issues

- Opacity of models.
  - Models can't be challenged
- Lack of feedback that can be used to correct flawed models
- The models embed assumptions, therefore subjectivity and biases are still present
- Models are scaled (exponential growth in use of the same or similar model). This makes it a weapon of mass damage if it is flawed
- Discriminatory impacts
- Pernicious feedback loop: a low score leads to more low scores

Ethics & Professionalism Standards for WMDs

Best practices for managing the challenges and issues raised by O'Neil's book.

Does Society need to be protected from WMDs

- O'Neil makes a very strong case that **society needs protection** from WMDs.
- **Data Scientists are not regulated**, nor are many of the models and algorithms they develop or use.
- The organizations that use models and algorithms have **little or limited** ethical or professional standards that apply to the **oversight** of their use of data, models, and algorithms.
- **Ethics codes** tend to be rather aspirational and barely effective except when serious unethical behavior is publicly criticized.
- **Professional standards** require clear recognition of who is a qualified professional and effective discipline for those who fail to meet those standards.

Does government regulation makes sense?

- The **essence of government regulation** typically falls on a specific set of rules applied to companies or organizations.
- O'Neil highlights serious examples of **poor or non-existent regulation** of elections, criminal justice, college entrance, employment, insurance, credit, advertising, and social media.
- **Pervasive unethical behavior and inequities** dominate the discussions in her book.
- European government **regulations are related to the storing and use of personal data** but do not directly regulate the professionals.
- Governments can be **severely challenged** to enact effective regulations and have the means to enforce those regulations.

Does government regulation makes sense?

- **What aspects** of data, models, or algorithms should be regulated by a governmental authority?
- How can a governmental authority regulate data, models and algorithms without **both transparency and the advanced training and experience** to review these areas?
- Could data, models, or algorithms be certified by a qualified professional – one who is **bound by professional standards and disciplinary oversight** – and possibly who is **independent?**
- Would technical audits, if required by government regulations, be **effective in reducing abuse and ensuring public trust?**

Are Ethics and Professional Standards sufficient?

- Can the public place trust in, and rely on, professionals who are recognized as qualified professionals, such as through:
  - **Professional License** – issued by government, e.g., professionals in the medical and legal field, accountants, structural engineers, etc.
  - **Certifications** – e.g., real estate, financial planning, various IT specialists, various insurance specialists, etc.
  - **Professional Credentials** – actuaries, financial risk managers, etc.
  - **Academic Degrees** – Economist, Mathematician, Statistician, Data Scientist, etc.
- When is it necessary for government to regulate professionals via a **Professional License** – the most stringent approach to oversight
- Would professionals in Data Science, Predictive Models, Algorithms, etc. agree with **voluntary oversight in order to gain public trust**?

What do Ethics and Professional Standards look like?

- Actuarial organizations provide credentials and membership based on:
  - **Exams** – required to get credentials and for membership.
  - **Continuing Education** – required to maintain the ability to provide certain types of professional services.
  - **Code of Professional Conduct** – A set of precepts which provide a self-regulatory framework for oversight of a professional's work.
  - **Professional Standards** – For detailed guidance applicable to actuarial practices or services provided.
  - **Disciplinary Actions** – a process to remove, revoke, suspend, reprimand or disqualify an actuary.
- New or existing professional organizations could establish some or all of these requirements for qualified professionals.

Does Self-Regulation of Professionals makes sense?

- Ethics and Professional Standards provide a means for self-regulation for qualified professionals:
  - **Public Recognition** – Some level of oversight of professional work could be offered in areas where public trust is needed.
  - **Governmental Control** – Either by direct licensing of professionals or by accepting self-regulation by independent professional bodies.
  - **Marketplace** – Accepted industry recognition of professional bodies.
  - **Professionalism Training and Oversight** – Stressing a professional's responsibilities to the public.
  - **Protecting the Public** – Setting the principles aimed at securing public trust in a professional's work.
- Several actuarial organizations have established these professionalism requirements for their members.

Will Industry support the oversight of potential for O'Neil's WMDs by relying on professionals?

- Will industry accept that **society needs protection** from WMDs?
- Will industry require Data Scientists to **meet professional qualifications** beyond technical skills, in terms of the ethics behind the data, models, and algorithms they develop or use?
- Will industry organizations that use models and algorithms **support ethical and professional standards** that require the oversight of their use of data, models, and algorithms by a qualified professional.
- **Will professionals embrace** the ethical and professional objectives to control the unfair or unjust use of technology.
- Can **professional standards** be developed such that the standards are effective, practical and enforceable.

What ethical or professional standards are in place today which are adaptable to data science, models, algorithms?

- **Actuaries** have **professional standards**, as well as a **Code of Professional Conduct** and a **Discipline Process**.
- Today there are **Actuarial Standards** on **Modeling**, **Data Quality**, **Communications** and **Expert Testimony**.
- The **CAS Institute** also offers **Professional Credentials** for **Predictive Modeling** to its members.
- Membership in the **CAS Institute** is open to anyone with an interest in data science, predictive analytics and other selected fields.
- The **Ethical Principles** of the CAS Institute include **Integrity, Competence & Qualifications, Objectivity & Impartiality, Confidentiality & Compliance, Communication** and **Use of Work Product**.

How can professionals address WMD issues and improve public trust in data science, models and algorithms?

- **Promote professional qualifications** for data science, predictive analytics, modeling, algorithms and other forms of AI.
- **Support** those organizations who have **meaningful discipline authority** to revoke, reprimand, suspend, or counsel certified professionals, including public notices when necessary.
- **Encourage actuaries to be accepted as professionals who can provide certified assessments** through means such as professional reviews or audits, of data, predictive models, algorithms, etc., to evaluate them for biases, unfair treatment, unjustified actions, violations of laws and regulations, etc.
- **Develop professional standards** for certified assessments.

Example - WMD and public trust issues in data science, models and algorithms

- **Auto Insurance Rates – regulated by a state agency** – apply to the insurance company, not to the professional proposing the rates.
- **Actuaries** – are typically involved in the ratemaking process at the company or in the regulatory review by the government.
- Regulations **do not require a qualified actuary** to support the company's rate filing or to review the rates.
- Regulations **do require that rates are not inadequate, redundant nor unfairly discriminatory**.
- **Actuarial guidance and standards** incorporate the regulatory requirements, but a **company is not required to use an actuary**.
- Use of **predictive models to set rates is complex**, rarely transparent and advanced training is needed to evaluate such models.

How to Do AI Without Creating "Weapons of Math Destruction"

CAS Annual Meeting

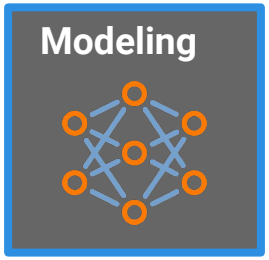
Remembrance Day 2020

DataRobot

THE BIG RISKS ARE NOT NEW

- Systemic Risk
  - 1987 market crash
- Monopolies
  - “Stuck in the system”
- Generalization Risk
  - Audit selection
- Answering the Wrong Question
  - Did we cross-sell product X? vs
  - Does communication strategy A increase the cross-sell of product X?
- Barriers to Change
  - Entrenched interests

AI PROCESS



1. Collect and prepare modeling data

- Structured Data
- Financial Outcomes
- Text & Image data (notes, emails, photographs)
- External Data (Web analytics, 3rd party data)

2. Build and validate multiple model approaches

- GLM's
- Tree-based Models
- Generalized Additive Models (GAM's)
- Deep Learning Models

3. Deploy and integrate models into business decision making processes

4. Monitor models overtime - proactive response to changes

AI PROCESS

Data

Modeling

Insights

1. Collect and prepare modeling data

2. Build and validate multiple model approaches

3. Deploy and integrate models into business decision making processes

API | Batch Scoring | Code Export | Guardrails / Humble AI

4. Monitor models overtime - proactive response to changes

Drift Detection | Accuracy Tracking | Challenger Models | Model Refresh

MAINTAIN CONTROL OF YOUR WORK PRODUCT!

- STAY INVOLVED IN DEPLOYMENT
  - Throwing a model “over the wall” to the IT department to deploy and not thinking about it again is the cause of many problems
- Ensure updates can be put into production readily
- Ensure tracking of drift in input features
  - And immediate attention if this happens
- Use fallback decisioning process for exceptional inputs
- Track accuracy AND other KPIs
  - For example, in the policing case, surveys to determine whether people feel safer
  - Single metrics are always dangerous (no guardrails)
  - Test the system (e.g., no loops in inquiry routing!)

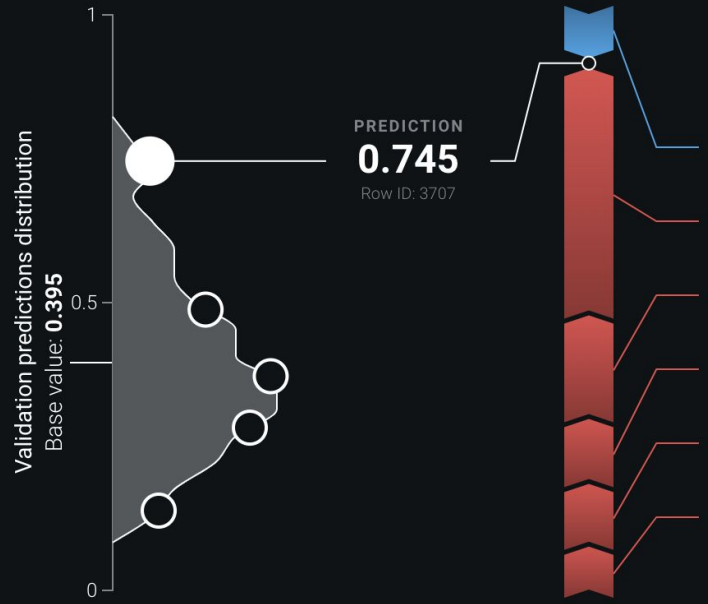
EXPLAINABILITY IS CRUCIAL

- For the impacted individual (e.g., credit applicant)
- For the tactical decision-maker (e.g., loan officer)
- For the strategic decision-maker (e.g., credit portfolio mgmt)
- For YOU
  - Because your business partners or customers should be asking lots of questions

- Good model-agnostic tools are available to INTERROGATE models
  - SHAP Explanations
  - Partial Dependence Plots for subsets of data
  - ICE plots

SHAP Explanations from an XGBoost Classifier

Prediction explanations preview for validation data

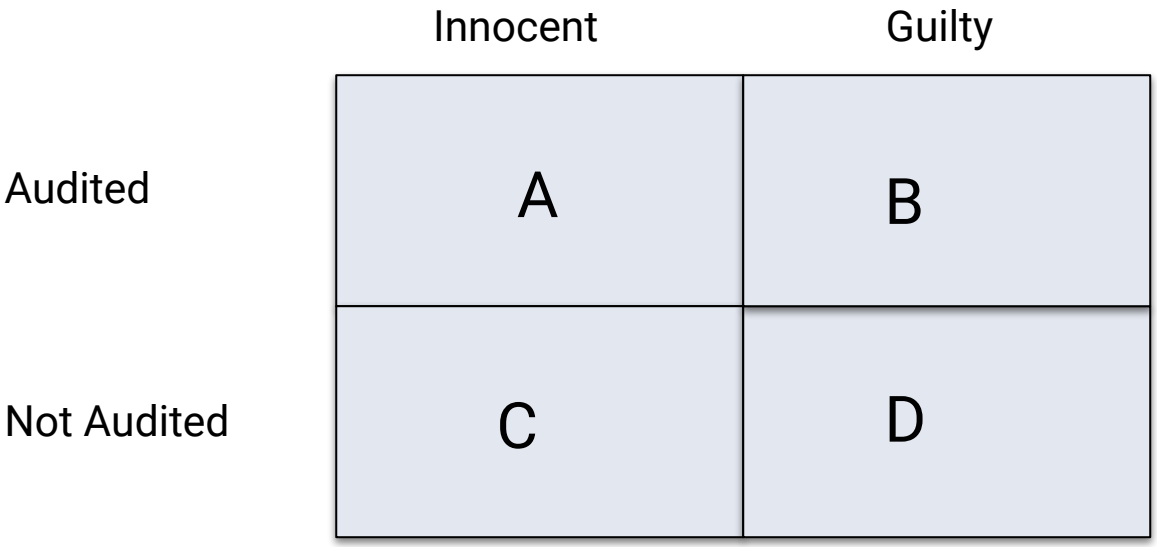


Contribution	Feature	Feature Value
-0.154	diag_2_desc	Paroxysmal supraventricular tachycard…
0.775	number_inpatient	2
0.325	Sum of all other features	
0.204	medical_specialty	Missing
0.197	discharge_disposition_id	Discharged/transferred to home with h…
0.150	diag_1_desc	Congestive heart failure, unspecified

FAIRNESS METRICS

- Lots of mutually incompatible criteria
- Tax audit example
  - Equal Representation (No Disparate Impact)
    - Same proportion audited from each group
  - Equal Efficiency
    - Same proportions of the audited are guilty (within each group)
  - Equal Errors (multiple ways to define)
    - Same probability in each group of an innocent person being audited
    - Same probability of a guilty person not being audited

THE FAIRNESS PARADOX



If we fix A/C, B/D, and A/B, then we have fixed all the proportions, in particular (A+C)/(B+D), which is a fact, not something we control, and which may not in fact be the same for all groups

FAIRNESS METRICS FOR A SCORE

- Again lots of mutually incompatible criteria
- Tax Audit Again
  - Possibilities based on a score (again many ways)
    - Each group is identically distributed across the score bands
    - Within each score band, each group has the same actual proportion of tax cheats
    - Among the innocent, same average score in each group
    - Among the guilty, same average score in each group

SUMMARY

- Deployment and Monitoring Are Crucial
- A Model Need Not and Should Not Be a Black Box
- Fairness Needs Careful Thought

USEFUL REFERENCES

European Commission White Paper on Artificial Intelligence (Feb 2020) https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

Lundberg, Scott M., and Su-In Lee. "A unified approach to interpreting model predictions." Advances in Neural Information Processing Systems. 2017. https://arxiv.org/abs/1705.07874

Kleinberg, Jon, Sendhil Mullainathan, and Manish Raghavan, “Inherent Trade-Offs in the Fair Determination of RIsk Scores”. 2016 https://arxiv.org/abs/1609.05807

Bookstaber, Richard, *A Demon of Our Own Design,* 2007 (if you are interested in better understanding the complex-system nature of the financial markets)

The Role of Models in the Financial Crisis

The Role of Models in the Financial Crisis

- Francis, “The Financial Crisis: An Actuary’s View” in Essays, The Current Financial Crisis, Lessons Learned and Future Implications, 2008, https://www.soa.org/library/essays/rm-essay-2008-toc
- Francis, “The Financial Crisis: Why Won’t We Use the F(raud) Word?” in Essays, Part 2, Systemic Risk, Financial Reform, and Moving Forward From the Financial Crisis, 2011, https://www.soa.org/library/essays/fin-crisis-essay-2011-toc
- Francis, “Data and Disaster: The Role of Data in the Financial Crisis”, submitted to 2010 Data Management, Quality and Technology Call for Papers, *Casualty Actuarial Society Forum*, Spring 2010. With Virginia Prevosto, www.casact.org
- Francis, “Banking on Robbery: The Role of Fraud in The Financial Crisis”, 2010 *Casualty Actuarial Society Forum*, Volume 2, www.casact.org

Francis-Prevosto “Data and Disaster”, 2010 eForum

- Explored role of data in the financial crisis
- Illustrate that data was available
  - Much of analysis was exploratory
  - Some data mining was illustrated
- Could have detected problems
  - Due diligence could have uncovered fraud
  - Provide warning of deterioration on mortgage quality

“Banking on Robbery” Overview

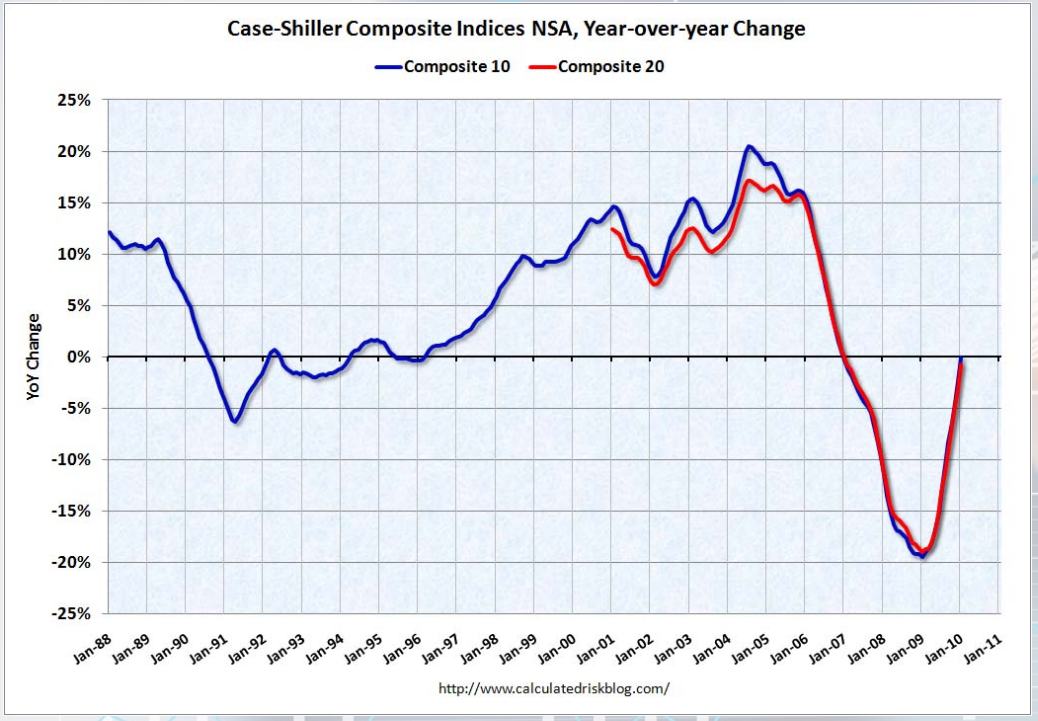
- Objective: Highlight the role of fraud in the Financial Crisis
- Some Fraud History: The S&L Crisis
- The Subprime Crisis
- The Madoff Ponzi Scheme
- The Mathematics of Fraud
- The Fraud Survey
- Conclusions

The Subprime Crisis

	Cumulative Default Rates @12/31/07								
	Development Age								
Year	1.000	2.000	3.000	4.000	5.000	6.000	7.000	8.000	9.000
1999	0.013	0.076	0.131	0.179	0.202	0.223	0.231	0.236	0.239
2000	0.015	0.084	0.144	0.177	0.202	0.214	0.221	0.225	
2001	0.019	0.090	0.148	0.191	0.209	0.221	0.228		
2002	0.011	0.066	0.111	0.135	0.151	0.158			
2003	0.008	0.050	0.081	0.103	0.114				
2004	0.009	0.048	0.064	0.089					
2005	0.010	0.074	0.136						
2006	0.026	0.128							
2007	0.040								

	Age-toAge Factors								
	Development Age								
Year	12-24	24-36	36-48	48-60	60-72	72-84	84-96	96-108	Tail
1999	5.869	1.714	1.371	1.128	1.101	1.035	1.024	1.012	
2000	5.573	1.719	1.233	1.141	1.059	1.033	1.018		
2001	4.876	1.644	1.285	1.099	1.056	1.029			
2002	6.150	1.691	1.213	1.116	1.052				
2003	6.049	1.627	1.276	1.107					
2004	5.570	1.344	1.383						
2005	7.577	1.845							
2006	5.005								
Average	5.834	1.698	1.294	1.118	1.067	1.032	1.021	1.012	
Selected	5.800	1.700	1.300	1.100	1.067	1.032	1.021	1.012	1.0453
Age to Ultimate	16.779	2.893	1.702	1.309	1.19	1.115	1.08	1.058	1.0453

Housing Prices Never Go Down?



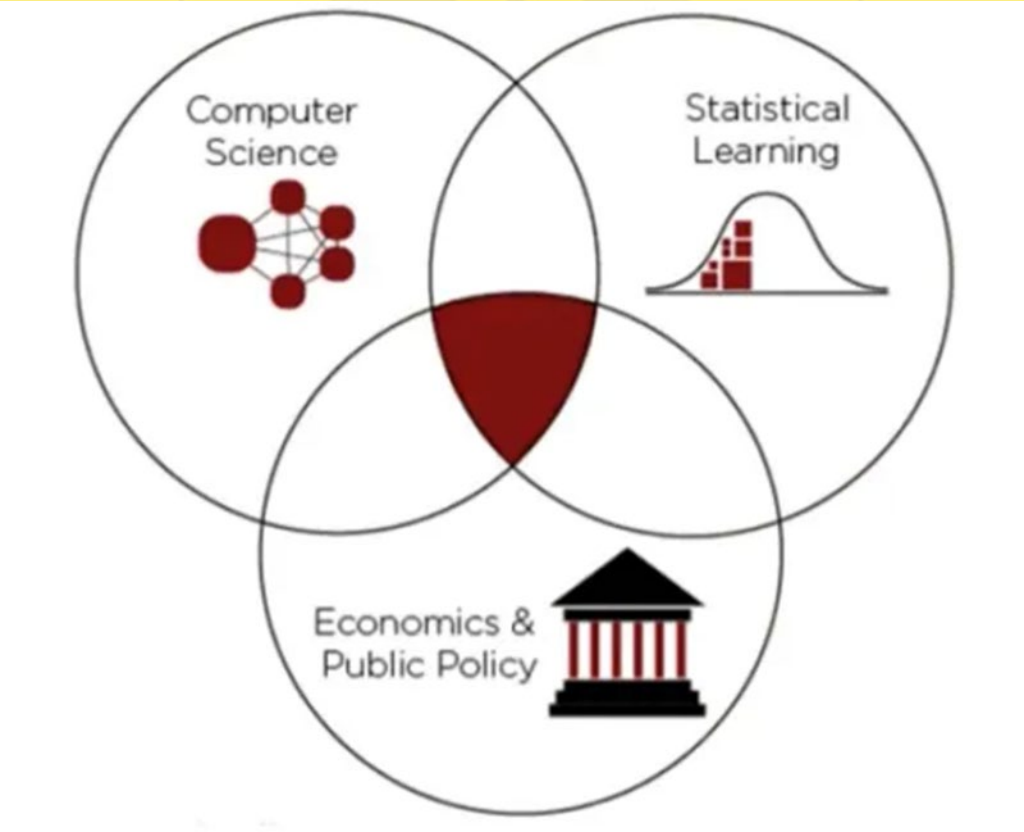
Results: The evidence of bubbles and fraud was there

- Abundant data was available to determine
  - that there was a housing bubble
  - that mortgages were deteriorating
  - that mortgage fraud was occurring and was rapidly increasing
  - that pools of subprime mortgages were granted high quality ratings that they did not deserve
- That Madoff was committing fraud

Fraud and Systemic Risk Regulation

- The evidence presented in this paper suggests that fraud regulation needs to be a key component of Systemic Risk Regulation.
- The SEC needs a “chief criminologist”, i.e., someone experienced in fraud detection and prosecution.
- More FBI resources are needed to investigate and prosecute financial fraud.
- Regulators must search for and prosecute fraud.
- Increasing the emphasis on enforcement and on detecting fraud before it creates a system-wide crisis can be accomplished without any new legislation
- legislative changes in the late 1990s and early 2000s appears to have removed some barriers to fraud.
- if fraud is not addressed, future crises will occur.

Big Data & Public Policy



Source: Data Science and Public Policy, University of Chicago

Big Data & Public Policy

- My mantra has been that while I want to be data driven and evidence based, each datapoint represents a human being, a family, a lived experience.
- When data scientists lose sight of individuals who happen to be exceptions and produces faulty results that negatively impact that individual, putting them in the wrong group, denying them a job or a mortgage-there is no recourse.
- Cathy O'Neil's book, "Weapons of Math Destruction" underscores the clear and present danger of the "big data" industry if allowed to run amok.

Algorithms

- Algorithms are best thought of as digital recipes. They are a set of rules: perform an operation, in a logical order, and have a dependable outcome.
- Algorithms begin and end with human interaction. Individuals are necessary to both start the process and do something useful with the outputs.
- Algorithms are harnessing volumes of macro- and micro-data to influence decisions affecting people in a range of tasks, from making movie recommendations to helping banks determine the creditworthiness of individuals.
- Troubling examples in which the reality of algorithmic decision-making falls short of our expectations.

Limitations of algorithms

- Models by nature are simplifications: no model can include all the real-world complexity or nuance of humans.
- A model's blind spots reflect the judgements and priorities of its creators; models are opinions embedded in mathematics:
  - From data we choose to collect
  - Questions we choose to ask
- Models may classify information based on online proxies for the sensitive attributes, yielding a bias against a group even without making decisions directly based on one's membership in that group. Examples: zip code as proxies for race, or height and weight as proxies for gender.

Mass Incarceration

- America has the largest prison population in the world.
- The US incarcerates more than 25% of the world's prison population (2.3 million inmates).
- The US general population only accounts for 5% of the world.
- No group is more targeted than black men. 1 in 3 Black men in America will serve time in prison and in some states, Black men have been imprisoned for drug charges at rates 20 to 50 times greater than their white counterparts.
- Women are the fastest growing incarcerated population in the US.

Algorithms Impeding Policy to End Mass Incarceration

- Predictive crime models such as PredPol target geography rather than the individual.
- Key inputs:
  - Type of crime
  - Location of crime
  - Time crime committed
- Desired outcome: Police spending more time in high-risk zones foiling burglars and serious crime; benefits community.

Algorithms Impeding Policy to End Mass Incarceration

- Problem:

  - Police do not focus exclusively on Part 1 crimes (violent crime including homicide, arson, assault etc.)
  - Focus is broadened to include Part 2 “nuisance” crimes (vagrancy, panhandling, selling and consuming small quantities of drugs). This is a conscious choice by the police.
  - Nuisance crimes or antisocial behavior (ASB) are endemic to impoverished areas

Pernicious feedback loop

- Nuisance data flows into the predictive model; over policing of those neighborhoods occur resulting in arrests-e.g. kids on the street corner drinking from a brown bag; suburban neighborhoods, teenagers have the luxury to commit such violations undisturbed.
- These low-level crimes populate the model with additional data points making it a high crime area which justifies more policing.
- Outcome: Prisons overpopulated by people guilty of low-level crimes, or unable to pay bail, from under-invested neighborhoods with residents who are primarily people of color.

Outcomes of Pernicious feedback loop

- Geography is a proxy for race since we live in largely segregated cities
- While model is “color blind” result of using the model is not!
- Criminalization of poverty.
- We continue to perpetuate mass incarceration.

Basis of Tracking Nuisance Crimes

- Stems from the “broken window” policing (which was based on tolerant community policing based on local norms)
- Bizarrely led to “zero tolerance” campaigns (famous in New York City)
- Resulted in young men of color being incarcerated for minor offenses
- Also created the orthodoxy of zero tolerance resulting in nuisance data being included to generate policing models
- Input data ⟶ obtain series of responses ⟶ calibrate to achieve objective
- Including nuisance data in the predictive models in addition to the violent crime data, created a fuller more detailed portrait of lawlessness.

Basis of Tracking Nuisance Crimes

- The type of crimes best predicted by the model-Nuisance crimes!
- Because: a chronically inebriated person has a favorite corner, just like a homeless person has a park bench
- But a car thief or a burglar will move strategically to anticipate the movements of police
- The crime maps track poverty.
- High number of arrests feeds the confirmation bias that poor people are commit the most crimes.

Targeting a different type of crime

- Financial crimes of committing fraud and bribing that devastated the global economy in the 2000s
- Millions of people lost their homes, jobs, healthcare
- But due to powerful lobbying, finance is under policed.
- Policing white collar crime requires a different skill set than the beat cop.
- The industry spends heavily on politicians to make themselves invulnerable.
- Law enforcement has made the choice to police the poor and criminalize poverty with the assistance of big data.

Dynamic Model vs. Toxic models

- Baseball is a good example of where mathematical models are used to predict wins and offers a contrast to the toxic models of Big Data.
- These models are fair because of transparency; no opacity as in other models (debacle of mortgage backed securities)
- The public has access to the stats
- Statistical rigor since each successive performance is part of the feedback loop to refine the models
- No proxies used; pertinent data is used such as strikes, hits etc. used to crate feedback loop to refine the predictive model (as opposed to models that substitute stand-in data or proxies such as correlations with zip codes or language patterns)

Big Data and Public Policy

- How to create appropriate feedback loops to improve the results of statistical systems?
- Predictive policing systems are increasingly used to determine how to allocate police across a city in order to best prevent crime. Observed crime data (arrest counts) are used to update the model, and the process is repeated. Such systems have been shown susceptible to runaway feedback loops, where police are repeatedly sent back to the same neighborhoods regardless of the true crime rate.
- (a)Predicting crimes based on arrest data really predicts arrests and not crimes and (b) by sending officers out based on predictions from a model and then using the resulting arrest data to update the model, you're liable to get into a feedback loop where the model results start to diverge from reality. If police don't see crime in a neighborhood because the model told them not to go there, this can cause a feedback loop.
- Appropriate filtering of the inputs fed into the system can counteract runaway feedback
- Additionally issues to address include bias in the observation and reporting of crime.

Mathematical Models should be our tools, not our masters

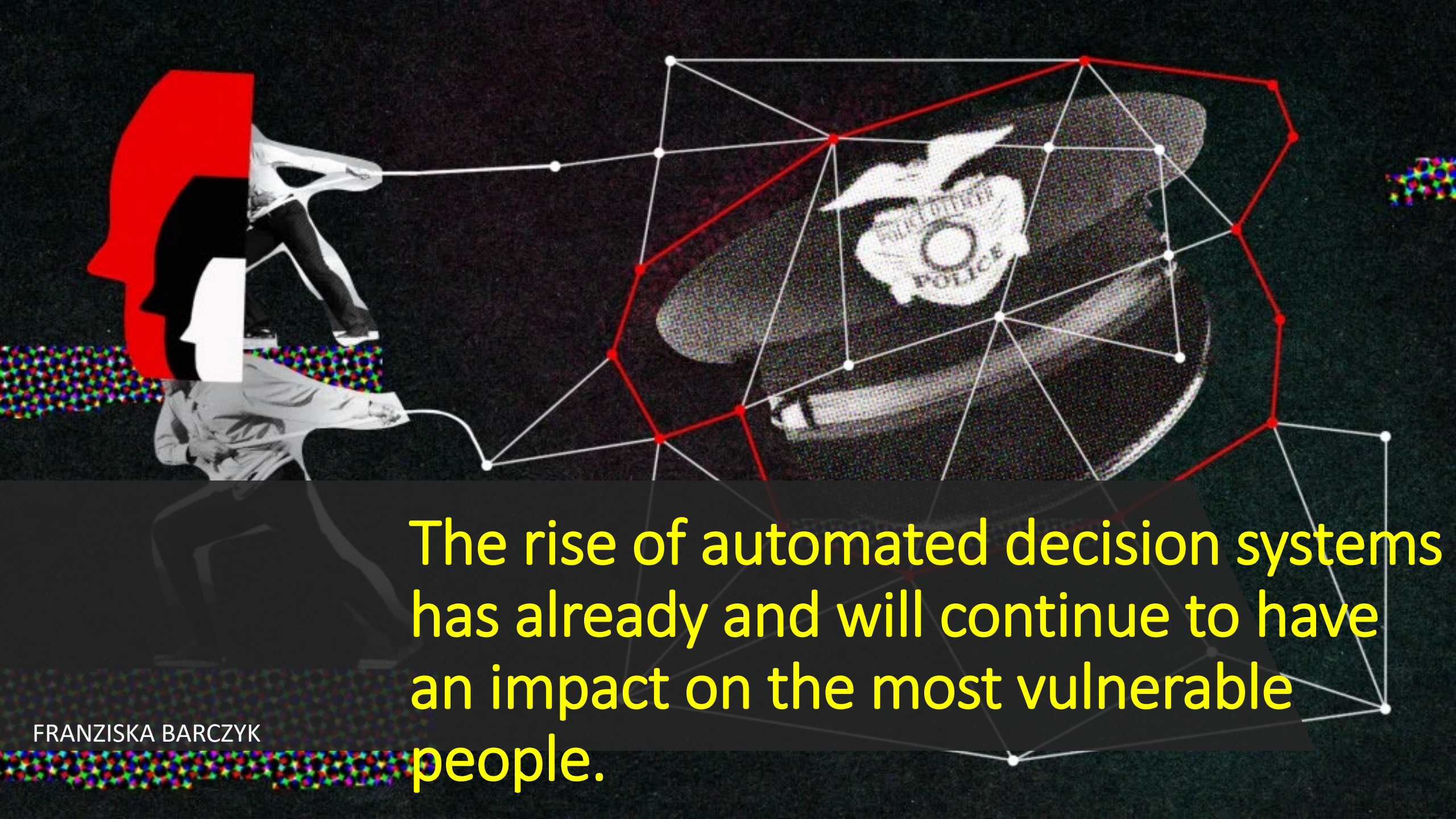
- Algorithms embedded in digital and social technologies can encode societal biases, accelerate the spread of rumors and disinformation, amplify echo chambers of public opinion, hijack our attention, and even impair our mental wellbeing.
- Always be on guard against bias, violations of data privacy, and the potential for harm and misuse.
- A critical part of the solution lies in getting better at diversity in engineering hiring; keeping bias out of algorithms should be part of the “corporate” culture.

Key elements of a public agency algorithmic impact assessment

- Agencies should conduct a self-assessment of existing and proposed automated decision systems, evaluating potential impacts on fairness, justice, bias, or other concerns across affected communities.
- Agencies should develop meaningful external researcher review processes to discover, measure, or track impacts over time
- Agencies should provide notice to the public disclosing their definition of “automated decision system,” existing and proposed systems, and any related self-assessment and researcher review processes before the system has been acquired
- Agencies should solicit public comments to clarify concerns and answer outstanding questions
- Governments should provide enhanced due process mechanisms for affected individuals or communities to challenge inadequate assessments or unfair, biased, or otherwise harmful system uses that agencies have failed to mitigate or correct.
- (source: ACLU)

Algorithmic audits

- Algorithm auditing must be interdisciplinary
- Regulatory role of government: e.g. consumer alert when credit score is being used to judge or vet individual
- Data collection with a User Opt-in
- Suitably transparent to end-users: Is it likely to be used in a socially acceptable way? Might it produce undesirable psychological effects or inadvertently exploit natural human frailties?
- Is the algorithm being used for a deceptive purpose?
- Is there evidence of internal bias or incompetence in its design?
- Is it adequately reporting how it arrives at its recommendations and indicating its level of confidence?



The rise of automated decision systems has already and will continue to have an impact on the most vulnerable people.

FRANZISKA BARCZYK