CIS399: The Science of Data Ethics	Lecture 2: Jan 22, 2019
Instructor: Michael Kearns & Ani Nenkova	Scribes: Jamison (JJ) Vulopas & Joseph Goodman

Book presentations begin! (All slides are posted, so notes should be viewed as supplements to—not replacements for—the slides.)

1 *Weapons of Math Destruction* by Cathy O' Neil

- How does big data increase inequality and threaten democracy?
- **Weapons of Math Destruction** (WMDs) are often created with good intentions, but often end up hurting individuals/subgroups, especially those labelled as "exceptions." And since these models have no mechanism for incorporating feedback about whether their predictions are on the right track, they won't learn from their mistakes.
  - e.g. mortgage security pricing in 2008
  - e.g. **college rating questionnaires.** ML procedures that incorporate different metrics to rank universities lead to a bad feedback loop that might cause safety schools (rated lower) to reject top applicants, who are less likely to go to that school. One school even paid for admitted students to retake the SAT with the goal of boosting average SAT ratings to help their ranking. Another school even admitted to lying about key reporting metrics, which had caused their ranking to rise by 20 places. (The Obama administration suggested creating a new rankings model.)
- Instances of WMD's in online advertising
  - In a modern marketplace, there is certainly a mutually beneficial (on firm-side and on consumer-side) purpose to targeted, algorithmically-informed online advertising. However, much current advertising targeting and harming vulnerable groups.
  - Discusses *lead generation*: identifying vulnerable, especially susceptible people and "offloading" them to third parties (e.g. DeVry U, U of Pheonix). Some of these for-profit colleges don't add value when it comes to employment. Vulnerable people become even *more* targetable for future nefarious programs.
  - Key question: *Where do we draw the line between predatory behavior and strategic marketing decisions?*
- Instances of WMD's in predictive policing (PredPol)
  - Historical crime data is processed to identify areas with higher crime rates. US Police Forces adopted this software and send cops to specified neighborhoods. However, many of the crimes considered by the algo were *nuisance* crimes (minor, non-violent).
  - So targeted areas were mainly impoverished, minority communities, but these minor crimes were *also* occurring in wealthier neighborhoods. So again, this WMD creates nefarious feedback loop: As police patrol the poorer neighborhoods, more data points are created, and causing the predictive software to disproportionately affect the more vulnerable neighborhoods.

  - Fairness vs/ efficiency in the legal system: there's the presumption of innocent until proven guilty. It's inefficient but it *does* guarantee fairness. On the other hand, WMD's are efficient but ignore fairness. So we must consider fairness when constructing algos, even at the cost of efficiency.

- Instances of WMD's in job applications
  - Some models used to help screen job-applicants are WMD's. These models are, again, created with good intentions (take a large applicant pool and cut it down to reasonable side), but have negative effects.
  - One firm (Kronos) developed a personality test to assess job performance. But if an applicant's personality score is too low, the applicant gets blacklisted from any the application process of company using this model.
    * Problematic because the questions were vague, overly personal, misinterpreted and not necessarily indicative of performance. Again, overly-simplistically predicting *proxies* for performance (rather than performance itself) can be problematic.
    * It's actually illegal to incorporate tests like these in the hiring process. There's an ongoing lawsuit.
  - The author also discusses automatic resume screening. Often, models only detect formatting issues, disproportionately affecting people who, while qualified, might not have access to resources that can polish resumes.
  - This ties back into college admissions system. Some medical schools wanted to implement something similar to screen applicants, predicting how strong of a doctor applicants would be, but the algos had negative effects, e.g. processing foreign-language names as grammatical errors, causing foreign applicants would disproportionately get rejected. Female applicants also were disproportionately rejected because the algo anticipated taking time off for maternity leave.
  - These algorithms are, again, well-intentioned, but *ultimately encode degenerate biases* by being overly-simplistic in accounting for confounding factors.
  - These WMD's also fail to check whether their assumptions were correct, resulting in an incomplete feedback system that only strengthens biases.

- Workplace-related instances of WMD's
  - Scheduling algorithms that attempt to minute-by-minute optimize employee time result in employees being unable to plan their lives until just days in advance, because schedules aren't released early enough, nor are they consistent.
  - Companies (e.g. Starbucks) vowed to remove the implementation of these systems, but they had already constructed managerial pay structures partially based on store performance, i.e. managers were incentivized to continue using these toxic practices.
  - Another example tried to measure how innovative employees were and how communicative they were of good ideas. Fell victim to the same problem of being overly-simplistic, and predicting an approximation of the truth rather than the truth itself. Another example of a well-intentioned model that fails.
  - Some steps are being made to bring regulations to the use of these assessments in education.

- Instances of WMD's in credit
  - Before the rise of these algos, you would go to a banker for a loan, and the banker would consider your finances *and* incorporate subject-matter knowledge about your *context* to determine whether to grant you a loan. Breakthrough in loan-giving was the *FICO score.* Strictly based on finances.
  - With the introduction of more big data, people started to try to make proxies to predict credit scores, eScores.
  - FICO scores are not WMD's, but eScore *proxies* are, because eScores are capturing an *approximation* your financial history, where FICO scores are capturing your true financial history.
  - eScores fall victim to a negative feedback loop, worsened by errors in the data collection process. Factors that are invisible to priveleged applicants might disproportionately affect less-priveleged applicants.
  - Credit scores are also sometimes used to proxy one's trustworthiness, accountability, and virtue. But this approximation is, very often, wrong and strengthens a bias-ridden negative feedback loop.
  - Potential solutions include regulating data usage (e.g. the "Data Science Hippocratic Oath," Mandatory Certificate of Fairness)
- Instances of WMD's in insurance
  - The insurance industry is widely affected by big data. Insurance companies split people into groups to determine pricing schema[1]
  - Adults with poor credit but clean driving records pay *more* than drivers with bad driving history but good credit. WMD's are fine-tuned to squeeze as much money as possible out of the different subgroups.
- Instances of WMD's in employee data usage
  - Data companies used cell phone data to divide people into groups based on behavior.
  - CVS started requiring employees to report health stats, which is a frightening next step in incorporating proxies into decision processes with the *goal of generating more revenue for the principal.*
  - Companies are certainly overusing our data.
- Role of big data in politics
  - *Facebook.* FB featured an "I voted!" banner during election season. Seeing that banner on a friend's profile meant one was more likely to vote. A researcher at FB noticed that showing more political posts in the news feed increased voter participation by 3%.
  - The author is not claiming that FB is a WMD, but it certainly has the potential to become a WMD if used harmfully.

[1] CFA listed 100,000 segments of pricing for different groups. Some people faced a 90% discount while others an 800% markup!

  - *Political Microtargeting.* Campaign teams can harness confirmation bias to cater their campaigns to appeal to specific voters (saying what the voter *wants* to hear). *This undermines democracy.* Bending the political message to match preferences of certain subsets of voters whose profiles you can access also *neglects* other voters, lessening their incentive to vote in the next election cycle.
  - Campaigns can (and should) use big data, but big data should not drive one's campaign.
- *One WMD's trash (output) is another WMD's treasure (input).*
  - When used separately, WMD's can be harmful, but when used in concert with one another, the negative feedback loop can become even more dangerous.
  - We must regulate WMD's. Regulate big data to incorporate the tension between profit for the principal and *fairness* for the general public. Regulate data scientists. Just like doctors must recite the Hippocratic oaths, data scientists should establish a philosophical grounding and be willing to sacrifice some accuracy for fairness.
  - *We are becoming more data-driven. We must incorporate ethics.*

2 *Technically Wrong: Sexist Apps, Biased Algorithms, and Other Threats of Toxic Tech* by Sara Wachter-Boettcher

- Some consequences in predictive technology include
  - Unethical uses of data: Facebook housing advertisers, Uber God Biew, Facebook Fake news
  - Encouragement of bias: Sexist word2vec embeddings, racist AI photo technology, COMPAS
- **Hiring Criteria**
  - In terms of getting a job, almost 80% of the criteria required is related to their ability to fit in whereas only 5% is related to actual work skills
  - This is exasperated by the pipeline towards getting a job. When applying for a job, applicants follow the same route towards job opportunities. This pipeline reduces racial and gender diversity in the workplace because the last step is usually an on-site interview where applicants who dont fit the same tech culture are rejected. Grace Hopper, a woman only tech career fair, is a method to alleviate this problem by supplying an alternate route towards internship and job opportunities.
  - Companies shift blame from themselves to the pipeline by making statements such as few POC and women graduate from tech related fields
  - Furthermore, because of the hostile tech environment perpetuated by the culture, there is a steady flow of interns and people out of the tech industry. A quote from the book describes the situation The industry wants diversity numbers but doesnt want to disrupt its culture to get or keep diverse people.

  - Tech culture has sexism, ableism, racism. An example brought up was a company where men talked about their accomplishments during the past year while women were asked to dance.

- **Lack of Diversity**

  - Part of the lack of diversity is caused by labeling edge cases as non-average and penalizing them. Doing so excludes people with those non-average experiences. This can be seen in tech startups. Often, the founder of the company will have a strong vision in where he wants to take the company. However, his commitment to that vision may stop him from not only understanding others perspectives surrounding issues but will also bring harm to those who dont also fit the vision.

  - Another cause of lack of diversity is the failure to consider different experiences. Examples given during the presentation were: an Etsy post not accounting for the possibility a user would have a female partner instead of only a male partner, a weight loss app not considering the possibility a user would be trying to gain weight instead of lose weight, Facebook only accounting for two genders on its sign-up form, and tumblr spreading an unfortunate #neo-nazis.

  - Furthermore, tech companies design their products for a Normal user that doesnt exists. Combining factors discussed earlier such as excluding edge cases with factors such as the default effect perpetuate a lack of diversity. The default effect is creating default values to match an average user for the platform which often may be a white middle-aged male. However, if users are not being identified correctly, the products that use the corresponding data will be inaccurate. For instance, in a website creation tool, the default picture for a CIO was an older white male, which will not be accurate for many companies. This can also be problematic because users mostly choose the default case, and that pigeonholes people into one stereotype.

- **Causes of Failure**

  - Misleading Users into Providing Data

  - Zuckering is a term defined by when users give more information and data about themselves than they want to. This is a cause of failure and can be seen where users trust the data and algorithms they are using when there may in face be a hidden bias. Companies use the fact that simplistic interfaces make things looks unbiased to get users to supply data they may not want to give or do not realize they are giving in the first place. An example of this is the Quizzes and Apps on Facebook. Using or playing them supplied the app with data that was not relevant to the usage app while the user did not usually know.

  - And as has been seen through this course, when companies use proxies for real values, the overall system can be less accurate especially if a group of individuals games or takes advantage of the proxy. When this happens or a company gets the proxy wrong, it leads to biases. And these biases can create self-perpetuating loops. For instance,

marginalized groups are more vulnerable to surveillance.

  - Systems that do not correct for historical bias increase or mirror the bias in real life. For instance, as discussed in class, in COMPAS, an algorithm for determining the rate of recidivism, black people are incarcerated more beforehand, and that input causes black people to be incarcerated even more.

  - Meritocracy in the tech industry is considered a failure. There is a belief that the tech industry is just based on merit, but that devalues soft skills that keep systems ethical and increases diversity in the workplace.

  - Similar to this mindset, is The Hacker Way tech mindset. The fact that people move fast and break things. This shows in industry that companies value innovation at the expense of societal implications and potential ethical problems.

  - This is continued by saying tech is elitist. There is a thought that everyone in tech is a genius who is superior to their arts and sciences brethren. This devalues humanities and social science backgrounds and once again reduces potential diversity of thought within the workplace.

- **Regulation**

  - Regulating a whole industry is nearly impossible as there isnt even a current universal definition of fairness causing a lack of regulation and problems arising from that in tech

  - Absurdly, were expecting lawmakers, the media and average consumers to understand these wildly different offerings as part of one single, endlessly complex, industry. Thats an impossible task Perpetuating the myth of a monolithic tech industry overtaxes our ability to manage the changes that technology is making to society (p. 187)

- **Everyone should understand the technology products they use**

  - There are many tech products that are used everyday that require people to give a lot of private information that people dont know about. The author says there should be more documentation on what data the programs use. Potential solutions include model cards which talk about how the model performs on various conditions like age, sex, race, and gender. There should also be policies to help regulate online content.

- **Solutions**

  - Dont design for the average user, products should be useable for a wider audience.
  - Tech companies should be more invested in their training data
  - There should be attempts to correct and debias historical bias using debiasing algorithms
  - Decide what it means for a system to be fair. And be more transparent about what type of fairness each algorithm has
  - For diversity, Disrupt the tech culture. Understand the importance of different perspectives.

- Tech cannot be as understanding of the people around if it cannot be diverse from within. An example for this, is Facebook has a larger percentage of black users than black employees. Matching those two figures could help Facebook better meet its userbase
- Make further efforts to diversify and recruit from other colleges
- An incentive for diversifying is diverse companies are more likely to outperform: 15% more likely with gender-diverse companies and 35% race more likely with ethnically-diverse companies