Data Harm Record

Joanna Redden and Jessica Brand

The aim of this document is to provide a running record of 'data harms', harms that have been caused by uses of big data. The goal is to document and learn from where things have gone wrong. The document compiles the examples of harms that have been detailed in previous research and publications. Each listed example contains a link to the original source.

The Data Harm Record pulls together concrete examples of harm that have been referenced in previous work so that we might gain a better 'big picture' appreciation of how people have already been negatively affected by uses of big data. A survey of harms also suggests where things may go wrong in the future and ideally stimulates more debate and interventions into where we may want to change course. The idea is that we can learn a lot by paying attention to where things have gone wrong and by considering data harms in relation to each other.

Please note: This document records harms that have already happened. There is a great deal of research raising concerns about how harm may be caused in the future. Such work is incredibly important, but not a focus of this record.

Background

People working in business, government, politics and for non-profit organizations are all developing new ways to make use of big data. These bodies have always collected and analysed data, but what's changed is the size, scope and methods to analyse data. The digitization of near everything along with major computing advances mean that it is now possible to combine sizes and types of data previously unimaginable, and to then analyse these staggering datasets in new ways to find patterns and make predictions.

There is an abundance of enthusiasm and optimism about how big data can be used for good. Optimism persists for good reason, there is a lot of good that can be done through new uses of big data.[1] There is also growing consensus that with big data comes risks to individuals and society. Previous work has detailed how data analytics can be used in ways that threaten privacy, security, as well as increase inequality and discrimination. The danger with big data is that harms can be caused unintentionally and intentionally. As argued by Cathy O'Neil , this is important to keep in mind as in many cases the algorithmic systems that are leading to harm were developed with very good intentions. The problem is that new big data tools present new ways to sort, profile, exclude, exploit, and discriminate. The complexity, opacity, and proprietary nature of many big data systems mean that often we don't know things have gone wrong until after large numbers of people have been affected.

Another problem is that few people have the skills needed to interrogate and challenge these new big data systems. What recourse do citizens have if they have been wrongfully targeted, profiled, excluded or exploited? Government agencies, civil society organizations and researchers across disciplines are drawing attention to these risks.

Defining data harms

Dictionary definitions of **harm** link it to physical and material injuries, but also to potential injuries, damages and adverse effects.[2] Solove and Citron argue that harm can be understood as 'the impairment, or set back, of a person, entity, or society's interests. People or entities suffer harm if they are in worse shape than they would be had the activity not occurred'.[3]

Building on these definitions, one way to understand **data harms** is as the adverse effects caused by uses of data that may impair, injure, or set back a person, entity or society's interests. While this definition is a start, clearly it is insufficient and will need to be developed given the increasing ubiquity of big data practices all around us.

Our legal and political systems are struggling to come to terms with data harms. Across nations it is becoming easier for corporate and government bodies to share data internally and externally. New data about us is being generated by us and collected by others through new systems. Consider for example the range of data that can be generated and collected through the Internet of Things and also the range of harms that can be caused if the wrong people hack into industrial systems. Increasingly, our digital selves and the digitization of services affect the kind of lives we lead, the opportunities afforded to us, the services we can access and the ways we are treated. All of these developments present new types of risk and harm. For all of these reasons we need to develop a more complex understanding and appreciation of data harms and a means to assess current and future harms, from the perspective of people who are and may be negatively affected by these harms.

Examples of data harms are detailed below. An attempt has been made at categorization, but in some cases the examples listed could fit in several categories simultaneously.

Examples

Commercial uses of data

Potentials for exploitation

Targeting based on perceived vulnerability

Some have drawn attention to how new tools make it possible to discriminate and socially sort with increasing precision. By combining multiple forms of data sets a lot can be learned.[4] Newman calls this 'algorithmic profiling' and raises concern about how much of this profiling is invisible as citizens are unaware of how data is collected about them across searches, transactions, site visits, movements, etc. This data can be used to profile and sort people into

marketing categories, some highly problematic. For example, data brokers combine data sets to identify specific groups. Much of this sorting goes under the radar. Some of it raises serious concerns. In her testimony to Congress, World Privacy Forum's Pam Dixon reported finding brokers selling lists of rape victims, addresses of domestic violence shelters, sufferers of genetic diseases, sufferers of addiction and more.

Another example, in 2015 the U.S. Federal Trade Commission 'charged a data broker operation with illegally selling payday loan applicants' financial information to a scam operation that took millions from consumers by debiting their bank accounts and charging their credit cards without their consent'.[5]

When your personal information gets used against you

Concerns have been raised about how credit card companies are using personal details like where someone shops or whether or not they have paid for marriage counselling to set rates and limits.[6] This has been called 'personalization', or 'behavioural analysis' or 'behavioural scoring' and refers to companies tailoring things to people based on what is known about them. Croll notes that American Express used purchase history to adjust credit limits based on where customers shopped. Croll as well as Hurley and Adebayo , describe the case of one man who found his credit rating reduced from $10,800 to $3,800 in 2008 because American Express determined that 'other customers who ha[d] used their card at establishments where [he] recently shopped have a poor repayment history with American Express'.[7] This event, in 2008, was an early big data example of 'creditworthiness by association' and is linked to ongoing practices of determining value or trustworthiness by drawing on big data to make predictions about people.[8]

Unintentional or intentional discrimination

Discrimination - skin colour, ethnicity, class or religion

Credit Scoring

As companies responsible for credit scoring, background checks, and hiring make more use of big data, an individual's appearance, background, personal details, social network, or socio-economic status may influence their ability to get housing, insurance, access education, or a job.

There are new start-up companies that make use of a range of 'alternative' data points to make predictions about consumers and provide people with credit scores. In addition, traditional credit scoring agencies are making use of big data to develop profiles. While the argument is that these tools could open up the potential for some not served by traditional credit scoring systems to receive credit, there are a range of concerns about how algorithmic scoring may discriminate. For example, a consumer's purchase history could be used, intentionally or unintentionally, as a proxy for ethnicity or religion. If an algorithmic system ends up penalizing one group more than others it may be hard to figure this out given the access issues, opacity and complexity of algorithmic processes. While there are laws in place

for people to review conventional credit scores, there are not yet measures in place for people to interrogate new big data generated scores.

In relation to all of these examples, researchers have raised concerns about how new data driven processes reproduce illegal redlining practices. Historically, redlining was used to discriminate against certain groups of people by denying some groups access, or more expensive access, to housing or insurance. This was often done by 'redlining' certain communities. The issue is that where someone lives is often associated with ethnicity and class. In this way location facilitates racism and inequality. Critics are concerned about how new big data tools can be used to 'redline' given the amount of detail that can be determined about us through our data. Previous research has demonstrated the potential to accurately determine our age, gender, sexuality, ethnicity, religion and political views through the data that can be collected and combined about us.

Relatedly, groups are raising concerns about how new data driven processes may facilitate 'reverse redlining'. This is when a particular group of people is targeted, as was done with sub-prime mortgages. Newman argues that big data was central to the subprime financial crash in 2007 as it played a key role in the manipulation of markets but also in the subprime mortgage industry. Online advertising and data collected about people online was used to direct and target them for sub-prime loans. In 2012 the American Department of Justice reached a settlement with the Wells Fargo Bank concerning allegations that it had 'engaged in a pattern or practice of discrimination against qualified African-American and Hispanic borrowers in its mortgage lending from 2004 through 2009' by pushing these borrowers into more costly sub-prime loans. In the settlement they agreed to provide $184 million in compensation.

The practice of targeting low-income groups continues in the payday loan industry. A U.S. Senate Investigation reports that data brokers have been found selling lists that focus on citizen financial vulnerability. For example, data brokers have compiled the following lists to sell to those interested in targeting such groups: 'Rural and Barely Making It', 'Ethnic Second-City Strugglers', 'Retiring on Empty: Singles', 'Tough Start: Young Single Parents'. One company was found selling a marketing tool to 'identify and more effectively market to under-banked consumers'.[9]

As argued by Madden et al., the fact that those with low-incomes are less likely to take privacy protection measures when online and to also rely more on their mobile phone for online access places them at greater risk than others for online targeting and exploitation.[10] In fact, 'opting out' of being tracked becomes increasingly difficult as technologies become more sophisticated. New tools that make cross-device tracking possible or that are embedded the Internet of Things, mean that the objects we use everyday make more of our lives 'knowable' and trackable and make 'opting out' even harder.[11] Newman raises concerns about how in this age of big data, information inequality is transferred into economic inequality, as companies have more information about citizens that can be used to target and exploit them to their disadvantage.[12]

Citron and Pasquale note that 'evidence suggests that credit scoring does indeed have a negative disparate impact on traditionally disadvantaged groups'. They provide a number of

examples in their article, just one is the case of All-State which was challenged in court and agreed to a multi-million dollar settlement over their scoring procedure which plaintiffs argued 'resulted in discriminatory action against approximately five million African-American and Hispanic customers'.[13] They also raise concerns about how scoring systems and predictive tools may actually create the situations they claim to indicate and "take a life" of their own, for example by labelling someone a poor candidate or unemployable.[14]

In 2015, Christian Haigh, a Harvard undergraduate, discovered that the prices for *The Princeton Review's* online SAT tutoring packages offered to high school students varied depending on where customer's live. Julia Angwin and Jeff Larson of ProPublica investigated Haigh's findings and found that the highest prices were being offered to ZIP codes with a large Asian population and high median income. The *Princeton Review* said that the price difference was not intentional, but as noted by ProPublica, the pricing algorithm clearly did discriminate. Angwin and Larson note that it is significant that in the United States 'unintentional racial discrimination is illegal in housing and employment under the legal doctrine known as 'disparate impact' which prohibits inadvertent actions that hurt people in a protected class'. However this doctrine does not extend to the online world, making it difficult in that country (and others) to take legal action against 'adverse impact' caused by unintentional algorithmic bias.

In 2012, a Wall Street Journal investigation found that Staples Inc. website displayed 'different prices to people after estimating their locations' and that in what appeared to be an 'unintended side effect' Staples tended to show discounted prices to areas with a higher average income and higher prices to areas with lower average incomes.[15]

A 2017 investigation by ProPublica and Consumer Reports showed that minority neighborhoods pay more for car insurance than white neighborhoods with the same risk levels. The study, which compared premiums and payouts in California, Illinois, Texas and Missouri, showed that minority neighborhoods paid 'as much as 30 percent more than other areas with similar accident costs'.

Facial recognition

There are numerous reports of facial recognition systems that have problems identifying people who are not white. Algorithms that are used to focus smartphone cameras, for border security and advertisements sometimes cannot identify, or misidentify, someone who is not white. It has been reported that the problem is that the facial recognition algorithms used across various systems have been trained using datasets that have mostly white faces, that these algorithms have not been exposed to enough diversity and that this problem is also connected to the fact that many of these systems are being developed and tested largely by white people. As argued by Joy Buolamwini, the issue of bias and inaccuracy becomes increasingly important as facial recognition tools are adopted by police and security systems. Examples of problems include the New Zealand case where one man's passport photograph was rejected when a facial recognition program mistakenly identified him as having closed eyes. People have posted reviews online raising questions about the ability of Microsoft's Kinect facial recognition feature to recognize people with darker skin and of HP's tracking webcams 'to see black people'.

Discrimination – gender

A study of Google ads found that men and women are being shown different job adverts, with men receiving ads for higher paying jobs more often.[16] The study, which used a tool called AdFisher to set up hundreds of simulated user profiles, was designed to investigate the operation of Google's ad settings. Although researchers could determine that men and women are being shown different ads, they could not determine why this is happening. Doing so would require access to more information that would need to be provided by advertisers about who they were targeting and by Google about how their system works.

Discrimination - health

Cathy O'Neil has produced a great deal of work demonstrating how unfair and biased algorithmic processes can be. In one example, she tells the story of Kyle Behm, a high achieving university student who noticed that he was repeatedly not getting the minimum wage jobs he was applying for. All of these job applications required him to take personality tests which included questions about mental health. Although healthy when looking for work, Behm did suffer from bipolar disorder and had taken time out previously to get treatment. Behm's father is a lawyer and he became suspicious of the fairness of these tests for hiring. He decided to investigate and found that a lot of companies were using personality tests, like the Kronos test. These tests are used as part of automated systems to sort through applications and in this process decide which applicants proceed and which are 'red-lighted' or discarded. As O'Neil details, these tests are often highly complex, with 'certain patterns of responses' disqualifying people. This example raises a number of ethical questions about the use of health information in automated systems but also about the uses of automated systems in hiring more generally, particularly as it is unlikely that those who have been 'red-lighted' will ever know they were subject to an automated system. O'Neil argues that the increasing use of automated systems to sort and whittle down job applications creates more unfairness as those who know or can pay for help to ensure their applications get to the top of the pile have an advantage.

Loss of privacy

This can happen unintentionally when attempts to release data anonymously do not work. Big data makes anonymity difficult because it is possible to re-identify data that has been anonymized by combining multiple data points.

AOL Example

As detailed by Paul Ohm, in 2006 America Online (AOL) launched 'AOL Research' to 'embrace the vision of an open research community'. The initiative involved publicly releasing twenty million search queries from 650,000 users of AOL's search engine. The data, which represented three months of activity, was posted to a public website. Although the data was anonymized, once the data was posted some users demonstrated that it was possible to identify people's identities using the data which included name, age and address.

Two New York Times reporters Michael Barbaro and Tom Zeller Jr. cross-linked data to identify Thelma Arnold, a sixty-two year old widow from Lilburn Georgia. Her case demonstrates the problems with 'anonymisation' in an age of big data, but also the danger in reading too much into search queries. As Barbaro and Zeller note, Ms Arnold's search queries 'hand tremors', 'nicotine effects on the body', 'dry mouth' and 'bipolar', could lead someone to think she suffered from a range of health issues. Such a conclusion could have negative effects if the organization making that conclusion was her insurance provider. In fact, when they interviewed Arnold, Barbaro and Zeller found that Arnold often does searches for her friends because she wants to help them.

Netflix Example

In 2006 Netflix publicly released one hundred million records detailing the film ratings of 500,000 of its users between Dec. 1999 and Dec. 2005. As Ohm reports, the objective was to launch a competition and for those competing to use this data to improve Netflix's recommendation algorithm.[17] Netflix anonymized the data by assigning users a unique identifier. Researchers from the University of Texas demonstrated not long after this release how relatively easy it was for people to be re-identified with the data.[18] This led to a court case in which Jane Doe argued that the data could be used to out her sexuality.[19] Jane Doe argued that her homosexuality was being revealed by the data as it revealed her interest in gay and lesbian themed films. She argued the data outed her, a lesbian mother, against her wishes and could damage herself and her family. The court case was covered by *Wired* in 2009.

Identity theft, blackmail, reputational damage, distress

Data breaches

Although data breaches are listed under corporate uses of data, they could also be listed here under government uses of data as breaches have happened in both sectors. Solove and Citron argue that 'harm' in relation to data breaches relates to 'a risk of future injury, such as identity theft, fraud, or damaged reputations' and also to a current injury as people experience anxiety about this future risk. They note that the anxiety and emotional distress created about future risk is a harm that people experience 'in the here and now'. Identity theft is a major problem, particularly for those of low-income who lack the resources to pay for legal representation and challenge mistakes due to identity fraud. Further, the sudden loss of income or errors that result from identity fraud can be disastrous for those living from pay cheque to pay cheque. Sarah Dranoff notes that in addition to financial loss, identity theft can lead to 'wrongful arrests, loss of utility service, erroneous information on health records, improper child support garnishments, and harassment by collection agencies'.[20] A number of data breach examples are detailed by Solove and Citron: 1) The Office of Policy Management breach leaked people's fingerprints, background check information, and analysis of security risks, 2) The Ashley Madison breach released information about people's extramarital affairs, 3) The Target breach resulted in leaking credit card information, bank account numbers and other financial data and 4) the Sony breach involved employee email.

Physical injury

Esther Kaplan's investigation into the effects of workplace data monitoring revealed how the monitoring of employees in order to increase their productivity is leading to physical injury in some cases. She interviewed a UPS worker who noted that the physical demands of his job have increased since the company introduced a telematics system. The system monitors employees in real time through tracking devices that include 'delivery information acquisition devices' and sensors on delivery trucks. The pressure to do more work in less time is leading to injury as drivers do not have the time to lift and carry packages properly.[21]

Political uses of Data

Political Manipulation and social harm

The damage that can be done by fake news, bots and filter bubbles have generated much discussion recently. Uses of big data and algorithmic processes in these cases can lead to social and political harm as the information that informs citizens is manipulated, potentially leading to misinformation and undermining democratic and political processes as well as social well-being. A recent study by researchers at the Oxford Internet Institute details the diverse ways that people are trying to use social media to manipulate public opinion across nine countries. They note that this is a concern given the increasing role that social media plays as a key information source for citizens, particularly young people. Further, that social media are fundamental in many countries to the sharing of political information. Civil society groups are 'trying, but struggling, to protect themselves and respond to active misinformation campaigns'.

Woolley and Howard define computational propaganda as involving 'learning from and mimicking real people so as to manipulate public opinion across a diverse range of platforms and device networks'. Bots, automated programs, are used to spread computational propaganda. While bots can be used for legitimate functions, the Internet Institute study details how bots can be used to spam, harass, silence opponents, 'give the illusion of large-scale consensus', sway votes, defame critics, and spread disinformation campaigns. The authors argue that 'computational propaganda is one of the most powerful new tools against democracy'.

Government uses of Data

Data Errors

Big data blacklisting and watch-lists in the U.S. have wrongfully identified individuals. As detailed by Margaret Hu, being wrongfully identified in this case can negatively affect employment, ability to travel, and in some cases lead to wrongful detention and deportation.[22]

Hu details the problems with the American E-Verify programme, which 'attempts to "verify" the identity or citizenship of a worker based upon complex statistical algorithms and multiple databases'. Employers across states use the programme to determine if a person is legally able to work in the U.S. Hu writes that it appears that employers have wrongfully denied employment for thousands. Hu argues that e-verify is problematic due to the unreliability of the data that informs the database screening protocol. The problems with the e-verify programme have also been detailed by Upturn. A study by the American Civil Liberties Union demonstrates that errors are far more likely to affect foreign-born employees and citizens with foreign names. People with multiple surnames and women who change their names after marriage are also more likely to face errors. Harm is further exacerbated by the difficulty in challenging or correcting e-verify errors. As discussed by Alex Rosenblat and others: '[L]ow-wage, hourly workers, whether they are flagged for a spelling error or for other reasons, often lack the time, resources, or legal literacy required to navigate complex bureaucracies to correct misinformation about them in a national database'.

Hu also raises concerns about The Prioritised Enforcement Programme (PEP), formerly the Secure Communities Programme (S-COMM). This is a data-sharing programme between the Federal Bureau of Investigation (FBI), DHS and local law enforcement agencies that requires local agencies to run fingerprints taken from suspects against federal fingerprint databases (ibid: 1770). The programme has made errors. For example, inaccurate database screening results wrongfully targeted 5,880 US citizens for potential detention and deportation, leading critics to question the reliability of PEP/S-COMM's algorithms and data. Furthermore, by using the biometric data of arrestees contained in the S-COMM databases the Immigration and Customs Enforcement (ICE) reportedly may have wrongly apprehended approximately 3,600 US citizens, due to faulty information feeding database screening protocols. As Hu points out, 'error-prone' databases and screening protocols 'appear to facilitate the unlawful detention and deportation of US citizens'.

Hu argues that the big data systems underlying both E-Verify and S-COMM/PEP are causing harm by mistakenly targeting and assigning inferential guilt to individuals. Legally speaking, this kind of digitally generated suspicion is at odds with constitutional rights and there is a growing consensus, at least in the U.S, on the need for substantive and binding due process when it comes to big data governance.

In Australia, the Ombudsman and Senate launched investigations into the Government's automated debt recovery system after numerous complaints of errors and unfair targeting of vulnerable people. The system uses data matching to determine if people have been overpaid their benefits. Onus was placed on those receiving letters to prove an error had been made.

Numerous accounts of errors were published in the press and calls for investigation were taken up by opposition politicians. One case involved a man who was repeatedly sent letters saying he owed the government repayment of $4,000. This turned out to be an error. The man, who suffers from depression and became suicidal, said he successfully convinced the government this was an error only to receive a similar letter a few months later. He again successfully proved this was an error. One of the ombudsman's conclusions was that better project planning and risk management should have been done from the outset.

Cassandra Goldie, Chief Executive of the Australian Council of Social Service, was quoted in the *Guardian* as saying:

> [R]obo-debt has issued thousands of debt notices in error to parents, people with disabilities, carers and those seeking paid work, resulting in people slapped with Centrelink debts they do not owe or debts higher than what they owe … It has been a devastating abuse of government power that has caused extensive harm, particularly among people who are the most vulnerable in our community.

Other examples of failure include attempts to automate welfare services in the U.S. Virginia Eubanks details the system failures that devastated the lives of many in Indiana, Florida and Texas at great cost to taxpayers. The automated system errors led to people losing access to their Medicaid, food stamps and benefits. The changes made to the system led to crisis, hospitalization and as Eubanks reports, death. These states cancelled their contracts and were then sued.

Data Errors – small data

Big data applications used by governments rely on combining multiple data sets. As noted by Logan and Ferguson, 'small data (i.e. individual level discrete data points) … provides the building blocks for all data-driven systems'. The accuracy of big data applications will be affected by the accuracy of small data. We already know there are issues with government data, just two examples: 1) in the United States, in 2011 the Los Angeles Times reported that nearly 1500 people were unlawfully arrested in the previous five years due to invalid warrants and 2) in New York, a Legal Action Center study of rap sheet records 'found that sixty-two percent contained at least one significant error and that thirty-two percent contained multiple errors'. [23]

Harms due to algorithm / machine bias

Research into predictive policing and predictive sentencing shows the potential to over-monitor and criminalize marginalized communities and the poor.[24]

Journalists working with ProPublica are investigating algorithmic injustice. Their article titled 'Machine Bias' in particular, has received a great deal of attention. Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner's investigation was a response to concerns being raised by various communities about judicial processes of risk assessment. These processes of risk assessment involved computer programs that produce scores predicting the likelihood that people charged with crimes would commit future crimes. These scores are being integrated throughout the US criminal justice system and influencing decisions about bond amounts and sentencing. The ProPublica journalists looked at the risk scores assigned to 7,000 people and checked to see how many were charged with new crimes. They found that the scores were 'remarkably unreliable in forecasting violent crime'. They found that only 61%, just over half, of those predicted to commit future crimes did. But the big issue is **bias**. They found that the system was much more likely to flag black defendants as future criminals, wrongly labelling them as future criminals at twice the rate as white defendants. White people were also wrongly labelled as low risk more often than black defendants. The challenge is that these risk scores and the algorithm that determines them is produced by a for profit company, so

researchers were not able to interrogate the algorithm only the outcomes. ProPublica reports that the software is one of the most widely used tools in the country.

Kristian Lum and William Isaac, of the Human Rights Data Analysis Group, recently published an article detailing bias in predictive policing. They note that because predictive policing tools rely on historical data, predictive policing should be understood as predicting where police are likely to make arrests and not necessarily where crime is happening. As noted by Lum and Isaac, as well as by O'Neil, if nuisance crimes like vagrancy are added to these models this further complicates matters and there is an over policing of poor communities, more arrests, and you have a feedback loop of injustice. Lum and Isaac used a range of data sources to produce an estimate of illicit drug use from non-criminal justice, population based data sources which they then compared to police records. They found that while drug arrests tend to happen in non-white low income communities, drug crimes are more evenly distributed across the community. Using one of the most popular predictive policing tools, they find that the tool targets black people twice as much as whites even though their data on drug use shows that drug use is roughly equivalent across racial classifications. Similarly they find that low income households are targeted by police at much higher rates than higher income households.

O'Neil describes how crime prediction software, as used by the police in Pennsylvania leads to a biased feedback loop. In this case the police include nuisance crimes, such as vagrancy, in their prediction model. The inclusion of nuisance crimes, or so-called antisocial behaviour, in a model that predicts where future crimes will occur distorts the analysis and 'creates a pernicious feedback loop' by drawing more police into the areas where there is likely to be vagrancy. This leads to more punishment and recorded crimes in these areas, poor areas where there is likely to be vagrancy. O'Neil draws attention to specific examples of problems: Pennsylvania police use of PredPol, the NYCPD use of CompStat and the Philadelphia police use of Hunchlab.[25]

How can harms be prevented?

Ultimately the goal of this Data Harm Record is to stimulate more debate and critical interrogation of how big data is being used across sectors and areas of life.

The goal is to maintain the Data Harm Record as a running record. Please let us know of any cases you think we should add by sending a message here.

It is hoped that this work contributes to the work of others in this area, many referenced in this publication, who are trying help us gain a better appreciation of: a) how uses of big data are affecting people, b) the kind of datafied world we are creating and experiencing, c) the fact that in this world big data practices affect people differently, and d) how datafication is political and may lead to practices that intentionally or unintentionally discriminate, be unfair, and increase inequality.

There are a range of individuals and groups coming together to develop ideas about how data harms can be prevented.[26] Researchers, civil society organizations, government bodies

and activists have all, in different ways, identified the need for greater transparency, accountability, systems of oversight and due process, and the means for citizens to interrogate and intervene in the big data processes that affect them. It is hoped that this record demonstrates the urgent need for more public debate and attention to developing systems of transparency, accountability, oversight and citizen intervention. For example, O'Neil argues that auditing should be done across the stages of big data projects and include auditing: the integrity of the data; the terms being used; definitions of success; the accuracy of models; who the models fail; the long-term effects of the algorithms being used; and the feedback loops created through new big data applications. Others, like AI Now, note the need for greater involvement with civil society groups, particularly groups advocating for social justice who have long-standing experience identifying and challenging the biases embedded in social systems. Researchers at AI Now have argued that government uses of automated and artificial intelligence systems in the delivery of core services in criminal justice, healthcare, welfare and education should stop until the risks and harms can be fully assessed.

Accompanying a record of data harms, should be a record of data for good that can be used to provide ideas of how to use data in ways that do no harm. Here there are important examples to follow. The Council for Big Data, Ethics, and Society has been releasing case studies that detail how ethical concerns have been managed. Building on this idea, it would be good to develop a record of where and how groups and individuals set out to prevent data harms from the outset.

---

[1] For example see: a) www.datakind.org, b) Gangadharan, SP (2013) 'How can big data be used for social good', *Guardian*, 30 May, available: https://www.theguardian.com/sustainable-business/how-can-big-data-social-good, c) Raghupathi, W and Raghupathi, V (2014) 'Big data analytics in healthcare: promise and potential' *Health Information Science and Systems* 2(3), available: https://www.ncbi.nlm.nih.gov/pmc/articles/ PMC4341817/ d) Mayer-Schönberger, Viktor and Cukier, Kenneth. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. New York: Houghton Mifflin Harcourt, e) Manyika, James, Chui, Michael, Brown, Brad, Bughin, Jacques, Dobbs, Richard, Roxburgh, Charles and Hung Byers, Angela. 2011. "Big Data: The Next Frontier for Innovation, Competition, and Productivity." McKinsey Global Institute, f) Armah, Nii Ayi. 2013. "Big Data Analysis: The Next Frontier." Bank of Canada Review. Summer.

[2] Cambridge Dictionary 'harm', available: https://dictionary.cambridge.org/dictionary/english/harm, Oxford Living Dictionaries 'harm', available: https://en.oxforddictionaries.com/definition/harm

[3] See Citron, D K and Pasquale, F (2014) The scored society: due process for automated Predictions. *Washington Law Review*, 89: 1-33.

[4] See Lyon, D (2015) *Surveillance as Social Sorting: Privacy, Risk and Automated Discrimination*, New York: Routledge.

[5] Federal Trade Commission (2015) FTC charges data brokers with helping scammer take more than $7 million from Consumers' Accounts, 12 August, available: https://www.ftc.gov/news-events/press-releases/2015/08/ftc-charges-data-brokers-helping-scammer-take-more-7-million

[6] Andrews, Lori. 2013. *I Know Who You Are and I Saw What You Did: Social Networks and the Death of Privacy*, New York: Free Press.

[7] As cited in Hurley, M and Adebayo, J (2016) Credit scoring in the era of big data, *Yale Journal of Law and Technology*, 18(1), p.151.

[8] Ibid, p. 151

[9] Office of Oversight and Investigations Majority Staff (2013) A Review of the Data Broker Industry: Collection, Use, and Sale of Consumer Data for Marketing Purposes, Staff Report for Chairman Rockefeller, Dec. 18, available: https://www.commerce.senate.gov/public/_cache/files/0d2b3642-6221-4888-a631-08f2f255b577/AE5D72CBE7F44F5BFC846BECE22C875B.12.18.13-senate-commerce-committee-report-on-data-broker-industry.pdf

[10] Madden, M, Gilman, M, Levy, K and Marwick, A (2017) 'Privacy, Poverty, and Big Data: A Matrix of Vulnerabilities for Poor Americans', *Washington University Law Review*, 95(1)
[11] Whitener, M (2015) 'Cookies are so yesterday; Cross-Device Tracking is In – Some Tips', *Privacy Advisor*, 27 Jan. available: https://iapp.org/news/a/cookies-are-so-yesterday-cross-device-tracking-is-insome-tips/
[12] Newman, N (2014) 'How big data enables economic harm to consumers, especially to low-income and other vulnerable sectors of the population', Public Comments to FTC, available: https://www.ftc.gov/system/files/documents/public_comments/2014/08/00015-92370.pdf
[13] As cited in Citron, D K and Pasquale, F (2014) The scored society: due process for automated Predictions. *Washington Law Review*, 89, p. 15.
[14] Ibid
[15] Valentino-DeVries, J, Singer-Vine, J., and Soltani, A (2012) 'Watched: Websites vary prices, deals based on users' information', *The Wall Street Journal*, 24 Dec., A1
[16] Datta, A, Tschantz, MC and Datta, A (2015) 'Automated Experiments on Ad Privacy Settings', *Proceedings on Privacy Enhancing Technologies*, available: https://www.degruyter.com/view/j/popets.2015.1.issue-1/popets-2015-0007/popets-2015-0007.xml
[17] Ohm, P. (2010). "Broken Promises of Privacy: responding to the surprising failure of anonymization", UCLA Law Review, vol 57 (2010) pp1701–1777
[18] Arvind Narayanan & Vitaly Shmatikov (2008), How to Break the Anonymity of the Netflix Prize Dataset, available: https://arxiv.org/abs/cs/0610105
[19] Singel, R (2009) Netflix spilled your Brokeback Mountain secret, lawsuit claims, *Wired*, 17 December, available: https://www.wired.com/2009/12/netflix-privacy-lawsuit/
[20] Dranoff, S (2014) 'Identity Theft: A Low-Income Issue', *Dialogue*, Winter, available: https://www.americanbar.org/groups/legal_services/publications/dialogue/volume/17/winter-2014/identity-theft--a-lowincome-issue.html
[21] Kaplan, E (2015) 'The Spy Who Fired me', *Harper's*, March, available: https://harpers.org/archive/2015/03/the-spy-who-fired-me/3/
[22] Hu, M. (2015) 'Big Data Blacklisting', *Florida Law Review*, 67: 1735-1809.
[23] Logan, WA and Ferguson, AG (2016) 'Policing Criminal Justice Data', *Minnesota Law Review* 541, available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2761069
[24] See: Sullivan, E and Greene, R (2015) States predict inmates' future crimes with secretive Surveys. AP, Feb. 24, available at: http://bigstory.ap.org/article/; Barocas, S and Selbst, A D (2016) Big data's disparate impact. *California Law Review* 104: 671-732; Starr, S (2016) The odds of justice: actuarial risk prediction and the criminal justice system. *Chance* 29(1): 49-51.
[25] O'Neil, C (2016) *Weapons of Math Destruction*, London: Allen Lane, p. 84-87.
[26] Throughout the record the hyperlinks provided link to individuals and groups whose work raises concerns and also provides recommendations about how to reduce data harms. In addition to those links, some examples of others doing work in this area include those working as part of the FAT / ML Fairness, Accountability and Transparency in Machine Learning group and the Algorithmic Justice League.