9-2015

Era of Big Data: Danger of Descrimination

Andra Gumbus
Sacred Heart University, gumbusa@sacredheart.edu

Frances Grodzinsky
Sacred Heart University, grodzinskyf@sacredheart.edu

Recommended Citation
Gumbus, A. & Grodzinski, F. (2015). Era of big data: Danger of descrimination. *ACM SIGCAS Computers and Society, 45*(3), 118-125. doi: 10.1145/2874239.2874256

Era of Big Data: Danger of Descrimination

Andra Gumbus
Professor Management
Sacred Heart University
5151 Park Ave
Fairfield, CT 06825
gumbusa@sacredheart.edu

Frances Grodzinsky
Professor Computer Science
Sacred Heart University
5151 Park Ave
Fairfield, CT 06825
grodzinskyf@sacredheart.edu

ABSTRACT

We live in a world of data collection where organizations and marketers know our income, our credit rating and history, our love life, race, ethnicity, religion, interests, travel history and plans, hobbies, health concerns, spending habits and millions of other data points about our private lives. This data, mined for our behaviors, habits, likes and dislikes, is referred to as the "creep factor" of big data [1]. It is estimated that data generated worldwide will be 1.3 zettabytes (ZB) by 2016. The rise of computational power plus cheaper and faster devices to capture, collect, store and process data, translates into the "datafication" of society [4]. This paper will examine a side effect of datafication: discrimination.

Categories and Subject Descriptors
K.4.1. **[Computers and Society]**: Ethics

General Terms
Human Factors

Keywords
Big Data, Discrimination, Human Resources, Privacy

INTRODUCTION

We live in a world of data collection where organizations and marketers know our income, our credit rating and history, our love life, race, ethnicity, religion, interests, travel history and plans, hobbies, health concerns, spending habits and millions of other data points about our private lives. This data, mined for our behaviors, habits, likes and dislikes, is referred to as the "creep factor" of big data [1].

It is estimated that data generated worldwide will be 1.3 zettabytes (ZB= 2^70) by 2016. The rise of computational power plus cheaper and faster devices to capture, collect, store and process data, translates into the "datafication" of society [4].

This paper will examine a side effect of datafication: discrimination. The first part will analyze discriminatory practices based on profiling. Next, it will relate privacy concerns to discriminatory practices, and finally, it will examine the impact of Big Data on Human Resource departments within organizations.

According to the National Institute of Standards and Technology Big Data is data which: "exceed(s) the capacity or capability of current or conventional methods and systems" [2]. A proponent of Big Data, Alex Pentland, Director of the Media Lab Entrepreneurship Program at MIT, describes Big Data as a new asset that we are just beginning to understand. He believes that it is a quantitative measure of human behavior that can be effectively used to solve human problems [2]. Other proponents think that modern economic activity is dependent on Big Data for the functioning of our global economy. Bringing together pools of data to analyze patterns and make informed decisions is the basis for competition and growth as well as enhanced productivity and value creation in business.

Data from industrial goods are being analyzed to provide better service and design of products based on actual use. "The ability to "now cast" using real time data enables prediction and theory testing never before possible in applications in the public sector and in personal location data" [3]. While we acknowledge that developments in the use of Big Data may have the capacity to promote social good, we claim that they also can also perpetuate harm with results that are inequitable or discriminatory when applied to protected classes. Big data analytics can lead to outcomes that go against civil liberties like fair housing, employment, credit and consumer protection.

In their book *Big Data,* Mayer-Schonberger and Cukier describe this new age of big data on page 97.

> Today we are a numerate society because we presume that the world is understandable with numbers and math, and we take for granted that knowledge can be transmitted across time and space. Future generations may have a big data consciousness and the presumption that there will be a quantitative component to everything. … in the new age of data, all data will be regarded as valuable [4].

Sectors such as online advertising, health care utilities, transport, logistics and public administration are using big data to stimulate innovation and productivity growth. Data driven R&D provides enhanced research and development; data-intensive product development uses data as a product or as a component of a product; data–driven processes can optimize production or delivery processes; data-driven marketing improves efforts by targeting ads and personalizing recommendations; and finally data is used to improve management practices and approaches [5].

1. BIG DATA SOCIETY: DISCRIMINATORY PRACTICES

Zwitter identified three categories of Big Data stakeholders. First are the collectors who determine what is collected and how long it is kept; next the utilizers who define and redefine the purpose for use of the data, and finally those of us who generate data. He defines data generators as those who input or record data voluntarily or unknowingly [6]. Data generators are at a disadvantage by not knowing who is collecting data about them and by not knowing how that data is being used. Power inequality exists between the generator and the collector and utilizer, both of whom have greater power than the generator. The Internet of Things (IoT) and global data exacerbates the power imbalance benefitting corporate entities who know how to generate intelligence from data [6].

Under civil rights law, discrimination can occur when there is disparate treatment, with disparate impact. Disparate treatment results from treating a person differently on the basis of race, gender, age, religion or other protected classes. Disparate impact results from a policy or practice that has a disproportionate negative effect on a protected class [7]. Existing anti-discrimination laws in the United States prohibit use of data that will discriminate based on health or disability. For example, employers cannot legally refuse to hire or fire someone who has an illness. However, there is nothing to stop employers with access to data from determining the probability of illness or disease based on health and eating habits. These employees could then be viewed as expensive, a potential insurance risk and therefore non-desirable. How is this done in company practice?

> The analyst involved, whether inside or outside the firm could easily mask the use of health-predictive information. A firm could conclude a worker is likely to be diabetic and a “high cost worker” given the cost of medical care. Given the proprietary nature of the information involved, the most the firm will tell the un-hired or fired worker is the end result: the data predicted that cost to the firm was greater than value (if a rationale is offered)...Secrecy is a discriminator’s best friend: unknown unfairness can never be challenged, let alone corrected (page 1421) [8].

Applications of Big Data are designed to differentiate between different types of people and make distinctions that separate desirable from undesirable individuals when it comes to credit risk, mortgage awards, credit card issuance or customer pricing. The mining of behavioral data carries the risk of the statistical problem of false positives when individuals are placed in a group that grants them undeserved privileges or false negatives: when individuals are placed in a category that inadvertently harms them. When this occurs some are disadvantaged and some have an unfair advantage despite the assertion that data mining algorithms have a 99% accuracy rating. As a result, the resulting misinterpretations may constitute wrong treatment for hundreds of thousands of people who might fall in that 1%.

Much of the problem of discriminatory practice has to do with how the results of Big Data analysis are interpreted and used. The sheer quantity of produced data has given rise to an industry of companies that will help you make sense of analysis results. Those who encourage us to believe that correlations are infallible may be ignoring the fact that their use in particular contexts may be dangerous. For example, some results may give rise to the possibility of profiling based on age, race, sexual orientation or other characteristics and behaviors which when correlated could lead to discriminatory practices. As a result, disparate impact or unequal treatment of an identified class compared to similar groups could result from data analytics. Murphy [9] reveals that job applicants are being profiled using references, prior employment, credit rating, driving record, criminal record credit history, Facebook pages and other sources that can impact hiring decisions that breach employment laws [9]. When correlations lead to policy based on profiled categories the possibility for discrimination exists. Nathan Newman believes,

> Economic inequality is driven by inappropriate use of big data which can coincide with the economic downturn and loss of income for average households. There are other factors contributing to inequality such as de-unionization, globalization and the automaton of unskilled jobs, but when combined with data consolidation the harm to low income and other vulnerable segments of the population increases [10].

His view is supported by FTC Chairwoman Edith Ramirez who stated that big data has the capacity to reinforce disadvantages faced by low income and underserved communities and called for greater transparency and accountability to make sure that low income populations would not receive differential treatment through digital redlining and discrimination by algorithm. Existing disparities can be exacerbated by the segmentation of customers to determine what products are marketed to them, what prices are quoted and what level of service they receive. Conscientious policy makers should ensure that Big Data be used for economic inclusion, not exclusion [11].

Big Data platforms enable racial profiling in subtle and invisible ways by targeting home address and other characteristics as a proxy for race. Online discrimination steered approximately 30,000 Black and Hispanic lenders into costly subprime mortgages during 2004 – 2009 and charged them higher fees than white lenders [10]. These targeted customers were disproportionally Black and Latino and were offered mortgages that had 30% higher interest rates compared to White borrowers. Unethical companies can target vulnerable less educated populations to mislead them with scams of harmful offers. The data industry uses the term “sucker lists” or “suffering seniors” who have been identified as targets for unethical and misleading scams. Algorithmic profiling allows companies to discriminate and categorize consumers into profiled groups in ways that may harm them with price discrimination and other unwelcome exploitive marketing practices [10]. Major corporations such as Staples, Home Depot and other financial services organizations use user location to display different prices to different customers. Instead of benefitting the low income population with lower prices, they did the reverse charging low income people higher prices and giving higher income people better deals. Credit card companies have similar practices offering different deals based on locations and presumptions about income levels. When retailers obscure prices and discriminate, economic models show that prices are higher than if consumers knew all the prices [10]. Price obfuscation strategies foster economic inequality and harm the least well off.

The asymmetry of power between data mining companies and individuals results in a data imbalance between the data have's (government and large corporations) and have not's. Perfect personalization or profiling can result in policies that discriminate for products and services or pricing of products. Buchta [1] explains that this gives companies the potential to create a perfect bubble for each consumer, presenting him/her with only information that algorithms dictate are of value. A cost benefit analysis falls short on the benefits when the harms are factored into this equation. She calls for greater regulation of the data gathering industry, more transparency, notice and choice for consumers [1]. Pasquale and Citron note (page 1419) that

> Of great concern is the collection and analysis of a critical mass of data. Our lives are starting to become an open book for those powerful or rich enough to score our profiles…Will individuals hesitate to join mental health support groups… will they refrain from joining political groups once they realize their affiliations on social media are a detriment to their careers? [8]

Posts on social network sites, locations from smartphones, sensors in our homes and on our bodies create a "nearly ubiquitous data collection capability that can erode our civil liberties and foster discrimination" [12]. Google searches for people with African-American sounding names were more likely to display ads with the word "arrest" which could lead to unfair and inaccurate perceptions of the person. The Chicago police department mined social networks and found 400 people who a model deemed likely to be involved in violent crime. Innocent people run a greater risk of being profiled by computer algorithms [12].

As our data is collected, interpreted and used without our consent, questions about fairness arise. What actually affects our lives in society? To make this point, Helbing asks the following:

- …How can you be sure you are getting your loan for fair conditions, and do not pay a higher interest rate because someone in your neighborhood defaulted?
- Can you afford to live in a multicultural quarter or should you move to a neighborhood to get a reasonable loan?
- Is there a tariff on your health insurance or do you pay more because your neighbors do not jog?
- Should you drink that extra glass of wine, eat red meat or will your mortgage rate go up?
- Would there be a right way of living or would everyone be discriminated against for some behavior or get rewards for other behaviors? [13]

The answers to these questions are elusive. We do not know how much information is collected about us, how long it is kept and how it is used. At present, users have no control over what is collected about them, and this makes it difficult, therefore, to judge whether we have been victims of discriminatory practices. The consumer Watchdog writes that "…consumers deserve clear understandable standards for use of their information" [14].

2. PRIVACY ISSUES

The advent of the IoT means that virtually anything connected to the Internet (TV, phone, tablet, refrigerator, camera, and car) provides data in the IoT movement [15]. Baker identified four major shifts in data collection that erode privacy. They are: invasiveness, variety, integration and scope. Government and businesses collect increasing amounts of data irrespective of privacy boundaries. Data sources are expanding as social media and machine data proliferate. Data is gathered for knowledge's sake not just under the guise of better customer service, marketing, or security. With privacy regulation much of the data is de-identified stripped of name and address or other identifying markers. The problem lies in the re-identification which is very easy to perform using mobile device ID and IP addresses. Data gathered with an IP address can predict a zip code which can be used as a proxy for race and income. The concept of personally identifying information such as social security number and credit card numbers is changing now that we can directly identify individuals based on the volume of data they generate. Computer scientists at Carnegie Mellon predicted full nine digit social security numbers for 8.5 % of people born in the US between 1989 and 2003 [16].

Under Fair Information Practice Principles (FIPP) privacy policy, companies must give consumers notice of data they collect, why they collect it and who they will share it with. They are supposed to use the data only for the purposes for which it was collected and not secondary uses. The Center for Democracy and Technology states that privacy laws should empower people to make informed choices about how their data is collected, stored, used, shared and maintained. The Computer & Communications Industry Association recommends a balance between the benefits and concerns of Big Data. It believes the focus should be on harms that occur from misuse and implications from who is using data, under what terms and for what purpose. Public interest groups also call for special protection for sensitive categories such as financial information, health, race, ethnicity, geo-location, age and data collected in the educational context [17].

How do we protect ourselves from the arbitrariness that can result from informational injustice when data is mined inaccurately? One approach is to legislate or establish a government agency with standards and certification procedures or punishments for violation to guard against false conclusions from data mining. Another is to equip individuals with the ability to correct data or run their own scenarios using various algorithms to run simulations in order to see what predictions result. Helbing describes this as a transparent and participatory approach where results can be verified or falsified, enabling trust in the algorithms and enhanced quality of healthy data results. Citizens control their data and participate in the value generated by their data. They can comment, correct, and determine what kinds of data are used for what purpose enhancing privacy and self-determination [13].

3. WHEN BIG DATA MEETS HUMAN RESOURCES

The human resource function is responsible for guaranteeing that the organization does not discriminate in employment practices and for making sure that state and federal laws as well as company policies are enforced. Overall however, legal and ethical issues are not widely discussed in the research on HR data mining; however, topics of privacy and equality can be found in the literature. Avoiding discrimination and treating people equitably means avoiding unfair treatment based on membership in a group. Discrimination can be blatant or hidden. Stereotyping is a problem in mining data to make HR decisions when unfair and unequal treatment is based on algorithms assigning classification and segmenting individuals into groups based on data. This often occurs without the knowledge of the owner of the data.

What is the role of data in making human resource decisions and what possible discrimination can result from the use of data when recruiting, selecting and making employment decisions? Anything but raw data may not be free from human bias. Human bias could affect what data is collected, what variables are included, what sources are used, what is mined versus what is ignored, and what questions are prioritized. Cold data could also be polluted and corrupted by ingrained company practices or design of algorithms [15]. How do we make sense of all this data? The new companies, whose sole purpose is to help organizations realize business value from their data, do not usually address the ethical issues.

Human resource departments assemble data on factors such as employee attrition and hiring, compensation and benefits, ethnic, gender, cultural, and nationality distributions. By applying advanced analytical techniques on the data, human resource professionals can get business insight, predict changes, and make informed decisions at operational and strategic levels [18]. Online analytical processing and data mining focus on past performance; predictive analytics forecasts on future behavior in order to guide decisions. Data mining tells us what has happened while predictive analytics advise us on appropriate response action. Key activities such as trends, metrics, and performance indices are portrayed in scorecards and dashboards. Advanced analytics can answer human resource questions such as whether capital investments contribute to business performance, how much human resource activities impact employee performance, or what skills the organization will need to meet future opportunities [18].

Big Data has entered the field of human resource management where analysis of the data guides the hiring, promotion and career planning functions in a new field called "work-force science". This is done through the analysis of email, instant messaging, phone calls, written code and mouse clicks, mined to determine how people work and, who they are connected to in their social network. Personality based assessments and other tools and tests used in selection and hiring decisions can be aggregated to determine worker communication patterns, style, and results. The proponents of work-force science predict that it will lead to efficiency and innovation within companies that traditionally rely on gut feel, interviews and reference checking to make hiring decisions. They believe that the revolution in measurement resulting from Big Data will change organizational and personnel economics. They predict that work-force science will "be applied across the spectrum of jobs and professions, building profits, productivity, innovation and worker satisfaction" [19]. However, worker surveillance raises many questions of employee privacy, ownership of data and the use and interpretation of that data. One ethical problem is that usually there is no informed consent about collection and use of this data even though it is being used to make important career decisions that impact worker livelihood.

In order to search for top talent, human resources go to analytics firms that assess talent and provide scores of a candidate in various fields. For example, a candidate's online contributions can be tracked by Remarkable Hire that provides a hiring score or Talent Bin and Guild that provides lists of potential applicants based on online data [20]. HR departments are using computer games and tests to measure emotional intelligence, memory, creativity, knowledge and cognition and employees' willingness to take risks. Companies like Google who previously used SAT's and GPA scores found that these did not correlate to success at Google [19]. They are now using additional metrics. For example, for a programming job, recruiters looked at how well the person codes; is the code reusable and is it respected among other programmers? Companies are now mainly using work-force science in call centers to analyze hourly workers in order to reduce attrition rates which are common at 100%. In these types of settings the improvement opportunity and cost savings is great. With the cost of hiring averaging $1500 per hire, a company found it could hire 800 instead of 1000 people and still had 500 workers on the job 3 months later. It claimed better customer service and less worker-churn [19].

3.1. Dangers of Big Data in Human Resources

In the area of training and development, Big Data can be used to benefit companies in areas such as: the identification of who might leave the organization; retention of top talent; the ability to identify top potentials for succession planning; the ability to assess what drives performance. Based on these metrics, they can adjust their management style. However, a simple misuse or mistake regarding reward or promotion based on an algorithm can have serious negative consequences for the organization as well as the employees if data is mishandled.

Race, gender, ethnicity, age and other discriminatory hiring practices have plagued HR in the past. Proponents of Big Data analytics advise that the crunching of thousands of bits of data may help to eliminate bias by offering 300 variables giving us a more robust portrait of the candidate. Because of the volume of available data, traditional screens like college attended, recommendations from fellow employees or previous employers can be combined with new screens such as "the sites where a person hangs out, the types of language used to describe technology, and self-reported skills on LinkedIn, projects worked on [19]. Some recruiters are using communication styles as a significant metric: What is the person's communication pattern?

How does he/she present on social media sites, and how does he/she communicate ideas?

We are not sure that this is all good news. The recruiters who use social media sites for data can gain disturbing insights from non-work related sites. For example, a student of ours was denied an internship based on old high school photos posted on Facebook that he had neglected to remove. The practice of using non-traditional screens in HR has resulted in law suits from victims who feel they were denied an opportunity and discriminated against in the employment process. There are some protections in U.S. law to protect potential job candidates: the Human rights Act 1998 provides a respect for private and family life; the Data protection Act 1998 states that data holders not have excessive information nor process it unfairly; the Civil Rights Act of 1964 and 1991 protects discrimination by gender, race, national origin, sexual orientation; the Age Discrimination in Employment Act (ADEA) protects against age discrimination and the Americans with Disabilities Act (ADA) against unfair treatment because of disability.

Besides the existing laws, how can we insure ethical use of Big Data in HR practices involving employment decisions? Kettleborough recommends four considerations. First, quality and accuracy must be assured when making life changing decisions about employees and candidates. Second, there must be enough data to make informed decisions and understand probability, sample size and statistical significance. Third, there must be caution about correlation and causation conclusions, i.e., two items that correlate do not necessarily cause each other. Finally, privacy and anonymity must be safe guarded so that personal data is not used against individuals. Also for internally administered surveys on employee satisfaction and culture, we must guard against using demographic information to identify individuals in a way that might turn honest data into dangerous data [21].

Peoplefluent produced a white paper outlining how HR departments can unlock data's value and be more proactive in preparing their organizations for the era of Big Data. Because of the large number of sources, data integration can be a problem. Those companies who have found effective ways to integrate their data have shown more success [22]. Human Resources can jumpstart data mining efforts and be a role model for other functions in an organization. They recommend using a role-based approach to analyzing people data based on functional roles in HR using the following six roles: compensation manager, chief learning officer, line of business manager, and VP HR/ head of talent management [23]. Compensation managers analyze reward schemes and compensation programs in order to ensure accuracy, fairness and visibility to employees. Learning officers look at training needs and data to ensure that employees have the right tools and raining at the right intervals to perform their jobs. The recruitment function looks at identifying optimal internal and external candidates to accelerate the hiring process. The procurement officer projects contingent workforce needs and look at staffing requirements and sourcing resources. Business managers are concerned with managing performance against company goals. The head of HR and talent management is responsible for data from all functional areas to determine if HR is hitting goals and contributing to organizational success. Using predictive analysis to assess historical data and influence future outcomes can enable HR to drive results strategically and be proactive partners in the business as long as they take measures to avoid discriminatory practices.

4. USING BIG DATA WITH HUMILITY AND HUMANITY

In May, 2014 the White House issued a report recommending government limits on how companies make use of information they gather from online customers. The report makes six policy recommendations including a national data breach law that requires disclosure when personal credit card data is exposed and defines customer rights regarding how their data is used. This protection extends to non-citizens of the US and to students regarding educational data [24]. An important aspect of the report is the acknowledgement that data misuse can be discriminatory. Misuse of data has " The potential to eclipse longstanding civil rights to protections on how personal information is used in housing, credit, employment, health, education and the marketplace" [24]. Assessing human values and recognizing the limitations of Big Data are critical for its ethical use.

Mayer-Schonberger and Cukier [4] predict that the effect on individuals may be the most harmful aspect of our future reliance on Big Data. They caution us that individual expertise matters less in a world where probability and correlation are paramount. "The danger to us as individuals shifts from privacy to probability. This leads to an ethical consideration of the role of free will versus the dictatorship of data. We will need new rules to safeguard the sanctity of the individual" (page 17). The authors warn that the demarcation between measurement and manipulation is blurred by the vast amount of data collected, and by our inability to conceptualize just what constitutes Big Data or how it is being used. Technology has reached a point where vast amounts of information can be captured and recorded cheaply. Data can be collected passively, and because the cost of storage has fallen, it is easier to justify keeping this data. Over the past half century the cost of digital storage has roughly been reduced by half every two years while storage density has increased 50 million fold [4].

In order to combat the dominance of Big Data gathering companies, consumers need more control of their data and possible government interventions to protect them. Strategies used in the past to protect consumers such as notice and consent, opting out, and anonymization are no longer effective based on the volume of data available. Users are easily identified and advertisers can fingerprint Web browser according to their skills. Individuals can be re-identified from anonymous data using zip code, birth date and gender to an 87.1% accuracy [9].

These problems can also be addressed by empowering individuals with access to their data and allowing them to analyze their own data and make conclusions from it. This sharing the wealth strategy can address Big Data privacy concerns by empowering consumers and represents a shift in the business model from organizations owning data to individual control. Consumers become free and independent actors in the marketplace, telling vendors what they want; how they want it, when and at what price [25]. This consumer centric model gives individuals control over management and use of their data, selective disclosure of selective data, control over purpose and duration of use, and correlations permitted by the individual not the end user. It also provides for a high level of security, data portability and accountability and enforcement. The question remains whether we can address challenges of this new business model such as technical feasibility, intellectual property rights, and business incentives to switch to a new paradigm [25].

One strategy was tried by Acxiom, the largest data broker, in 2013. Acxiom let people see what information it had about them in a Web site AboutTheData.com. When accessed, the site revealed core data Acxiom had amassed in an effort toward transparency by data brokers. Critics claim that Acxiom revealed selective facts only and not the analysis the company markets to clients such as categories like "potential inheritor," "adult with senior parent," and "diabetic focus" [25].

Another strategy to introduce humility and humanity into the equation is to ensure algorithmic accountability by having closer human scrutiny of the results of algorithms used to make life - changing decisions. Big Data is supposed to bring greater economic opportunity and convenience to all people not just a preferred few. With human oversight adding "machine-to-man" translation of results, data equality will become a reality. It will give context to analytic results. Predictive recommendations can be reviewed and overruled in essence giving human veto power over the result. Critics of data science may object to human intervention, yet this introduces an element of protection for the individual (page A4) [26].

> In a sense a math model is the equivalent of a metaphor, a descriptive simplification. It usefully distills, but it also somewhat distorts. So at times, a human helper can provide that dose of nuanced data that escapes the algorithmic automation. Often the two can be better than the algorithm alone [26].

Gary King, Director of Harvard's Institute for Qualitative Social science recommends that the creators of the algorithms make adjustments in the design of the calculations to favor the individual in order to reduce the risk of getting a wrong result. It will also improve trust in predictive results if the process were more transparent (page A3).

> The key that will make it work and make it acceptable to society is storytelling. Not so much literal storytelling, but an understandable audit trail that explains how an automated decision was made. How does it relate to us? How much of this decision is the machine and how much is human? [26]

In sharp contrast to Big Data is Open Data which is accessible to everyone. Gurin defines Open Data as available to people, companies, and organizations that can be used to make data driven decisions and solve complex problems. The Open data model includes over 500 companies across business sectors that provide platforms to make government data easier to find and access [26]. Open Data is currently being used in legal services including patent data and competitive intelligence; education including data on value of institutions; energy efficiency; precision agriculture; health care transformation; housing and real estate and transportation analysis. The Open Data 500 study includes companies that earn revenue from a variety of business models serving diverse customers. As the amount of federal, state and local data increases the business opportunities will expand for data that is accessible to everyone. The goal of Open Data is to make all government data open unless privacy or security dictates otherwise [27].

CONCLUSION

Organizations that use Big Data analytics should practice it with customer privacy and integrity of data in mind, and guarantee legal and ethical applications through their policies and procedures on the use of data.

In the eSociety where everything has a score, predictive algorithms determine who has value and will receive critical life changing opportunities determined by score. Without fair and accurate scoring systems data can be biased and arbitrarily assign individuals to a stigmatizing group that affects their opportunities. Advances in artificial intelligence are missing the human element,, and we believe that human values are needed as oversight in the design and execution of scoring systems. We need to consider the consequences when we rely solely on scoring machines to make decisions that may not be fair or just.

Citron studied the scored society using credit score as a case study and found three basic problems with credit scores: opacity or lack of transparency, arbitrary results and disparate impact on women and minorities. Consumers do not know why or how their credit scores change. Different credit bureaus have vastly different scores for the same individual and punish cardholders for paying bills. Biases are embedded in the code and defined parameters of data mining. For example certain occupations can get a low score like service jobs which are held by minorities. Although discrimination was not intended, and may be unintentional, it is discrimination none the less. Credit scores have a negative disparate impact on disadvantaged groups – women and minorities as recent settlements by Allstate typify where five million African-American and Hispanic customers were discriminated against in the denial of insurance based on credit score [28].

Citron recommends regulatory oversight of scoring systems to include: gathering of data into scores, calculating gathered data into scores, disseminating scores to decision makers, and employers and others use of scores in making decisions. Ideally calculations would be public and processes transparent, inspected for fairness and accuracy. Individuals deserve to know how they are rated and who is getting the data. Licensing and audit requirements for sensitive areas that impact employment, insurance or heath care are needed to avoid arbitrariness by algorithm [28]. To this end the FTC addressed the following concerns about predictive algorithms: How are companies using scores? Are they accurate? Can consumers benefit from available scores? How is privacy ensured? Patterns and correlations about race, nationality, sexual orientation and gender that are already covered by discrimination law deserve added scrutiny

FTC Chairwoman Ramirez stated that decisions by algorithm require

> transparency, meaningful oversight and procedures to remediate decisions that adversely affect individuals who have been wrongly categorized by correlation. Companies must be sure that they are not using big data algorithms that are accidently classifying people based on categories that society has decided by law or ethics

not to use such as race, ethnic background, gender and sexual orientation [28].

As we utilize insights gained from Big Data analytics we need to recognize that results have a scope limited by context. Much of the data we generate is collected without a question in mind although it is being used to make predictions about us. Although correlations can be very useful, when it comes to interpreting them and making decisions, we are not willing to give over final decisions affecting individuals in society to a machine alone.

We need to recognize the perils of Big Data when decisions are made about disadvantaged and protected classes. We need to guard against data that reinforces gaps between the rich and poor, haves and have not's and that suppress already disadvantaged people and benefit the wealthy and privileged. We cannot succumb to the powerful allure of data only as precise and reliable, when it can also be unjust and unfair, constraining opportunities for the disadvantaged and perpetuating discrimination. The exponential growth of data has the capacity to bring great value to society but can challenge the ethical and legal systems if the rights of individuals are violated in the process of bringing added value to business.

REFERENCES

[1] Buchta, Heather (2014) How Did Data Get to Be So Big? Inside Counsel. Breaking News. November 25, http://www.insidecounsel.com/2014/11/25/how-did-data-get-to-be-so-big Accessed 6/1/15.

[2]MIT Technology Review (2013), http://www.technologyreview.com/view/519851/the-big-data-conundrum-how-to-define-it/ , Accessed October, 3, 2013.

[3]McGuire,T., Manyika, J. and Chui, M. (2012)"Why Big Data is the New Competitive Advantage". Ivey Business Journal, Jul/Aug, Vol.76 Issue 4, pp. 1-4.

[4]Mayer-Schonberger, V. and Cukier, K. Big Data: A revolution that will transform how we live,work, and think. (2013) Houghton Mifflin Harcourt, Boston, NY.

[5]OECD Digital Economy Papers 222 (2013) Exploring Data-Driven Innovation as a New Sources of Growth: Mapping the Policy Issues Raised by Big Data, April.

[6]Zwitter, Andrej (2014) Big Data Ethics. Big Data & Society July – December, 2014: 1 – 6.

[7]Yu, Persis, McLaughlin,Jillian and Levy, Marina. (2014) Big Data: A big Disappointment for Scoring Consumer Credit Risk. National Consumer Law Center March 2014.

[8]Pasquale, Frank and Citron, Danielee Keats (2014) Promoting Innovation While Preventing Discrimination: Policy Goals for the Scored Society . Washington Law Review 89:1413.

[9]Murphy, Michael and Barton, Jophn. (2014) From a Sea of Data to Actionable Insights: Big Data and What it Means for Lawyers. Intellectual property & Technology Law Journal March, 26.3: 8 – 17.

[10]Newman, Nathan. (2014) How Big Data Enables Economic Harm to Consumers, Especially Low Income and Other Vulnerable Sectors of the Population. Journal of Internet law December, 18.6: 11 – 23.

[11]Curran, John (2014) FTC Chief Sounds Note of Caution on Development of Big Data. Cybersecurity Policy Report, September.

[12]Dwoskin, Elizabeth (2014) White House Takes Aim at Big Data Discrimination; Report recommends More Privacy laws. Wall Street Journal (Online) . May 1, 2014.

[13]Helbing, Dirk. (2014) Big Data Society: Age of Reputation or Age of Discrimination? http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2501356 Accessed May 1, 2014.

[14]Wireless News (2014) Consumer Watchdog Supports 6 Policy Recommendations in White House Big Data Report. May 7.

[15]Baker, Pam. (2015) Data Divination: Big Data Strategies. Cengage Learning PTR, Boston, MA

[16]Lohr, Steve (2015) Maintaining A Human Touch As the Algorithms Get to Work. New York Times, April 7, p. A3

[17]Hammond, Brian. (2014) Industry Groups Stress Need to Protect Innovation in Big Data privacy Effort. Telecommunications Reports, Sept 1,2014 80.17: 29 -32

[18]Kapoor, Bhushan.(2011), Impact of Globilization on Human Resource Management. Journal of International Management Studies 6.1 (Feb) : 1 – 8.

[19]Lohr, S. (2013). Big Data, Trying to Build Better Workers. New York Times, April 21, p. 5.

[20]Zarsky, Tal Z. (2014) Understanding Discrimination in the Scored Society, Washington Law Review, 89:1375.

[21]Kettleborough, Jonathan. (2014), Big Data. Training Journal. June. 14 – 19.

[22] Grossman, K (2014) "System-integration drives talent acquisition". http://www.peoplefluent.com/blog/hr-system-integration-drives-talent-acquisition, Accessed June 12, 2015.

[23]www.peoplefluent.com (2014) Make Your HR Data Actionable Now! Unlock the Value Trapped in Your Company's Data by using Role - Based Analytics. A Peoplefluent White Paper. Accessed May 14, 2014.

[24]Sanger, D. and Lohr, S. (2014). Call for Limits on Web Data of Customers. NYT May 2, 2014. P. A1 and B6.

[25]Rubinstein, Ira S. (2013) Big Data: The End of Privacy or a New Beginning? International Data privacy Law, Vol 3, No.2: 74 - 87

[26]Lohr, Steve (2015) Dataism: The Revolution Transforming Decision Making, Consumer Behavior, and Almost Everything Else. HarperCollins, NY,NY.

[27]Gurin, Joel (2015) Opening Business Innovation With Open Data. Business Horizon Quarterly. Issue 12, pp. 42 – 49.

[28]Citron,Danielle Keats and Pasquale, Frank (2014) The Scored Society: Due Process for Automated Predictions. March, Washington Law Review. 1 – 33.

.