WHAT ARE THE RISKS ASSOCIATED WITH PERFORMANCE ANALYSIS FOR INDIVIDUALS?

A Research Paper submitted to the Department of Engineering and Society
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Systems Engineering

By

Aniket Chandra

March 27, 2020

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISOR
Catherine D. Baritaud, Department of Engineering and Society

Researchers believe that team and individual team sports may be viewed as dynamical systems and, thus, they should be thoroughly investigated using congruent concepts and tools (Williams et al, 1999). The combination of these concepts and tools is referred to as Performance Analytics. It is also believed by researchers that the dynamics of sports systems are similar to other systems such as workplace management, or education innovations. Thus, Performance Analytics can be applied to these fields as well. The technical research aims to produce a proposal for the University of Virginia's Board of Visitors and administrators about the launch of a Performance Analytics Center, here at the university. This center is envisaged to be integrated as a pan-university initiative, where any and all academic and athletic departments can collaborate in research regarding performance analytics as well as the introduction of academic programs in the field. The process of designing this center has proven to be a difficult task to accomplish. This is primarily due to the various restrictions that are put in place to protect the privacy of personal data. As part of the initial outscoping, the capstone team elected to engage in proof-of-concept projects with different teams in the Virginia Athletics department. The idea was to conduct various data analyses of personal athletic data, and demonstrate the value of the results. However, a majority of the groups reached a significant roadblock when they were denied access to the personal data of student-athletes, due to privacy restrictions.

These constraints inspired the STS Research to evaluate how the personal data of athletes and others have the potential to be abused, and how these individuals could come into harm's way. These inquiries lead to a further line of questioning regarding 1) What industries have the highest potential for abuse? 2) Where are the current laws lacking in providing security? 3)What needs to change, so that the risk of harm is mitigated? Loosely coupled, the STS Research aims

to provide an understanding of the possible negative impacts of enhanced performance analytics being integrated into our society. This research will use the Social Construct of Technology framework to evaluate the risks of performance analytics. In this paper, there is an argument for enhanced technology measures and regulations that enable greater security. These measures will be supported by evidence from previous forays into similar research and and studies from experts in the field of technology and public policy.

BUSINESS IS A-BOOMIN'!

How do businesses stay profitable and maintain an advantage over their competitors? How do companies dominate an industry and expand into new markets in today's society? They innovate. They think differently. They invest in technology. Big Data is today's technology hot topic. Big Data is the combination of structured, semistructured and unstructured data collected by organizations that can be mined for information and used in machine learning projects, predictive modeling and other advanced analytics applications. When it comes to industry and commerce, the technology is more aptly termed Business Intelligence & Analytics or BI&A. Now realizing the great importance of big data, many analytical companies are engaged in finding hidden information in consumer data (Uzialko, 2018).

Over the last two decades, BI&A has played a great role in the growth of many companies. Industry studies have highlighted this significant development. For example, based on a survey of over 4,000 information technology professionals from 93 countries and 25 industries, the IBM Tech Trends Report identified business analytics as one of the four major technology trends in the 2010s (2011). In a survey of the state of business analytics by Bloomberg Businessweek , 97 percent of companies with revenues exceeding $100 million were

found to use some form of business analytics (2011). A report by the McKinsey Global Institute (Manyika et al. 2011) predicted that by 2018, the United States alone will face a shortage of 140,000 to 190,000 people with deep analytical skills, as well as a shortfall of 1.5 million data-savvy managers with the know-how to analyze big data to make effective decisions.

It is estimated that the amount of data stored in the world's IT systems is doubling about every two years, and enterprises have responsibility for about 85% of that data (Surbakti et al., 2019). The need for big data has driven up demand for big data experts, as well as their salaries, while the supply is in shortage. To deal with this pressure, organizations are increasing their budgets, their recruitment and retention efforts, offering more training opportunities to current staff to develop the required talent, and buying analytics solutions that are designed for users who are not data science experts (Adrian, 2013). According to the researcher, Alexander Adrian, as this technological phenomenon grows, the number of analysts processing through scores of personal data is also likely to increase. This would mean that the accessibility to personal data will become much easier, and in turn, at greater risk of abuse.

How do these companies get access to the data and what do the analysts do with it? Big Data is created every day by the interactions of billions of people using computers, GPS devices, cell phones, sensors and medical devices, data-intensive areas such as atmospheric science, genome research and astronomical studies. This process offers a way to better understand and meet their customers' demands. By analyzing customer behavior, as well as vast troves of reviews and feedback, companies can nimbly modify their digital presence, goods or services to better suit the current marketplace (Uzialko, 2018). The sources of data for these firms are endless; any information exchange between the consumer and the firm will result in data

generation. The methods in which the data is collected, also determines what results the analysis will produce, as is illustrated in Figure 1.

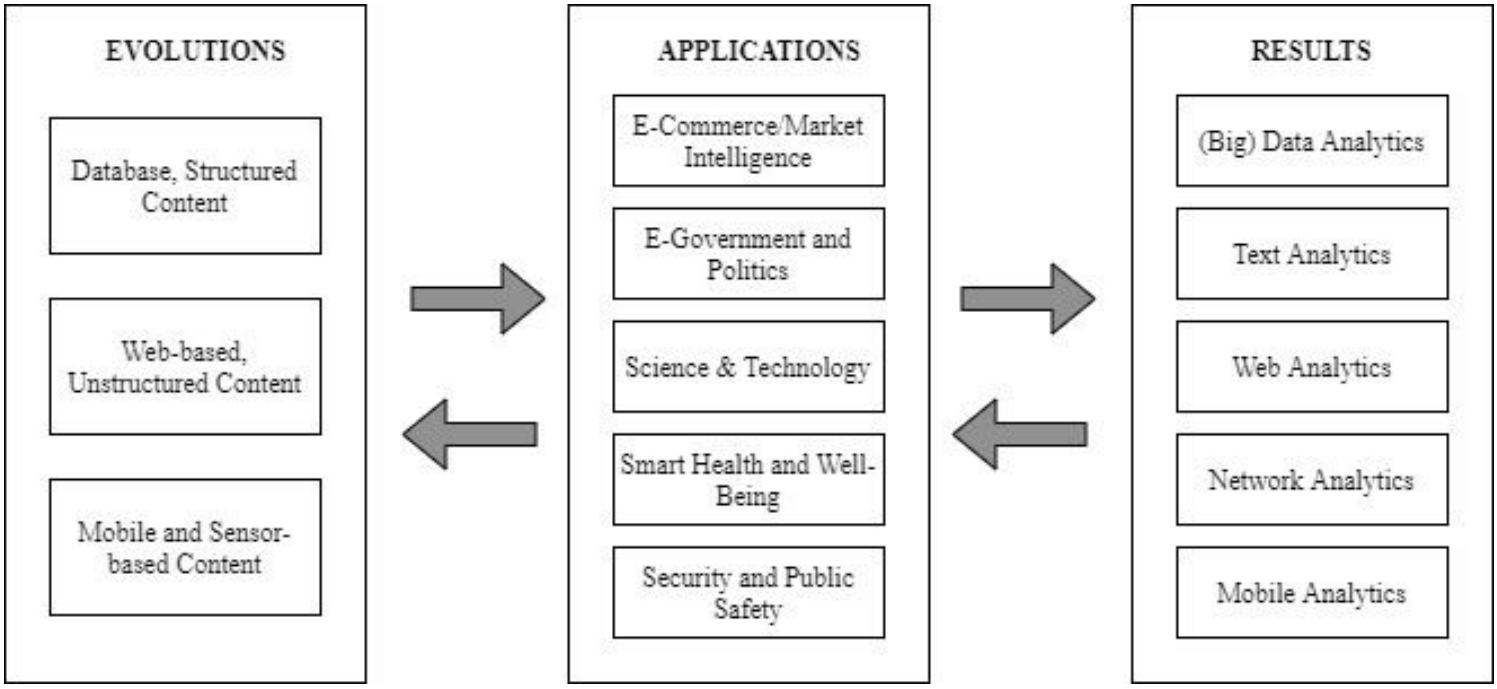

Figure 1: An Overview of Business Intelligence: The framework of how data is stored, the various application potentials for analysis, and the results. (Chandra, 2020).

When it comes to the actual analysis and data processing, each firm or entity has their own specific approach and employs their own unique business practices. However, all big data analytics strategies exhibit similar, if not the same characteristics. These traits are called The 5 V's of big data analytic and are as follows: Volume, Velocity, Value, Veracity and Variety (Ishwarappa, 2015), and are illustrated in Figure 2 on page 6. Volume refers to the sheer amount of data collected , the types of data collected and the storage capabilities of the firm. Velocity refers to the increasing speed at which data is collected, and the speed at which the data is analyzed and processed. Variety is how the data is categorized and structured; 90% of the big data is unstructured and thus poses a significant problem for analysts. Veracity covers the

probability of dirty or poor data. The quality of data collected can vary greatly; hindering accuracy of results. Value is the most important, it refers to the actionability of the results (Ishwarappa, 2015).

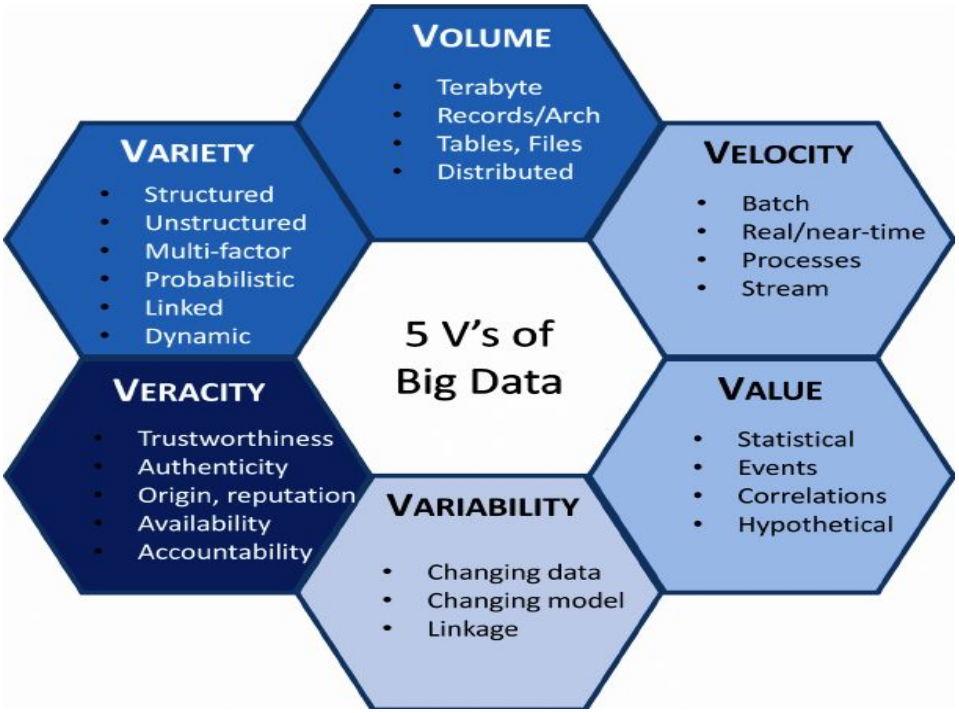

Figure 2: The Five V's of Big Data: The components of every successful analysis of consumer data. Often, Variability is considered a minor characteristic of the process. (IBM, 2014)

In the last three decades, big data has been applied to diverse fields, such as the government, international development, and education. It is only now that the US healthcare system has begun to explore its under-utilized data. Big data is not only referencing the quantity, but also the complexity, diversity, and relativity of the information. This information may be analyzed to reveal patterns, trends, and associations that may be applicable to the healthcare field. This information can be gathered through sources, such as Electronic Health Records (EHRs), Intelligent Research in Sight registry (IRIS) , and Merit-based Incentive Payment System (MIPS). Recognizing patterns would aid in predicting preventative measures for an

increased holistic and personalized patient care. Although big data proves to have endless beneficial applications, it can bring into question the ownership of this information. Additionally, big data poses a risk for security breaches (Yi Yee, S.W. et al., 2020).

YOUR DATA, YOUR LIFE!

Issues arise concerning users' data, including informed consent for use and disclosure, retention (even after the account is closed), access, and the adequacy of the consideration provided (Wigan & Clarke, 2013). The majority of the debate on Big Data's downsides today addresses threats to personal information privacy. Loss of privacy can in turn result in crimes such as identity theft or cyberstalking. The field of information security overlaps somewhat with privacy, because when security is breached, privacy can be compromised. Privacy breaches in the collection, use, and sharing of big data have affected all the major tech players, be it Facebook, Google, Apple, or Uber, and go beyond the corporate world including governments, municipalities, and educational and health institutions (Chen & Quan-Hasse, 2018).

A high profile case of data misuse occurred back in 2014 when an employee at one of the world's fastest growing companies; Uber; violated the company's policy by using its "God View" tool to track a journalist who was late for an interview with an Uber exec. If you are unfamiliar, "God View" allowed the company's staff to track both Uber vehicles and customers.)The tool was unavailable to drivers, but was, at the time, apparently "widely available" at a corporate level. Tracking the journalist obviously flies in the face of Uber's privacy policy at the time, which stated that employees are prohibited to look at customer rider

histories except for “legitimate business purposes.”(Morgan, 2017). Social network data are actually the most reliable sources of real-life big data, thanks to well-known Web social networks like Facebook and Twitter. Here, mining such data is of primary interest, but the need for privacy and security very often limits the real impact of these tasks (Cuzzocrea, 2014).

Another concern is the threat of illegal discrimination in areas such as housing, education, finance, or insurance or unacceptable classification and labeling of individuals in target marketing (Markus, 2015). The telecommunications company AT&T paid over $25 million to the Federal Communications Commission back in 2015, as a result of an investigation that discovered that employees at international call-centers illegally disclosed the personal information of upwards of 280,000 customers. The workers sold U.S. AT&T customer names and Social Security numbers to third parties who used it to unlock mobile phones, so the devices would work on networks other than AT&T’s (Ruiz, 2015). Morgan Stanley discovered in 2015 that a financial adviser downloaded account data on 10% of their wealth management clients, about 350,000 people. 900 of those client accounts later showed up on the anonymous text sharing site, Pastebin (Viswanatha, 2016). These examples highlight the lack of secure infrastructure in place, that protects vital and sensitive information of consumers.

It is somewhat unclear how breaches occur, each incident can be unique in their own way. One likely reason is the technology used to store the data: In cloud infrastructures, databases are often outsourced based on the well known DaaS (Database as a Service) paradigm. This gives rise to very problematic security issues as query processing procedures may easily access sensitive data sets and determine privacy breaches (Cuzzocrea, 2014). It is also possible for even the most restrictive of data management mechanisms to have accidental breaches. There

are two widely publicized cases of information leaks occurring through two individuals associated with the US government; Bradley Manning leaking US diplomatic cables to Wikileaks, and Edward Snowden leaking classified NSA data to media organizations (Nunan & Di Domenico, 2015).

The introduction of Big Data analytics and the approach of Internet-of-Things (IoT) has excited medical professionals since the phenomenon started gaining momentum in the early 2010s. By 2020, 40% of IoT-related technology will be health-related, more than any other category, making up a $117 billion market. The convergence of medicine and information technologies, such as medical informatics, will transform healthcare as we know it, curbing costs, reducing inefficiencies, and saving lives (Dimitrov, 2016). Healthcare organizations store, maintain and transmit huge amounts of data to support the delivery of efficient and proper care. Nevertheless, securing this data has been a daunting requirement for decades. Complicating matters, the healthcare industry continues to be one of the most susceptible to publicly disclosed data breaches. In fact, attackers can use data mining methods and procedures to find out sensitive data and release it to the public and thus data breaches happen (Abouelmehdi, et al., 2017). The current situation of big data analytics in healthcare can be described in the Social Construction of Technology model or SCOT. The SCOT model is a theory in which human actions shape technological advancements and not the other way around. Major concepts of the model include interpretive flexibility, where each technological artifact has different uses for various groups. The different interpretations often give rise to conflicts between the different social groups, over the intended use and design of the technology (Pinch & Baker, 1984) . When the model is applied to healthcare: at the center would be the data collected for healthcare purposes,

surrounded by the dimensions associated with big data analytics; The F V's. The third layer would be the various applications of healthcare data analysis, ranging from Disease Management to Drug Discovery. The following layer, would be the challenges associated with data privacy and security and lastly; the strategies that can be used to face these challenges. The SCOT framework model of big data analytics in healthcare is illustrated in Figure 3.

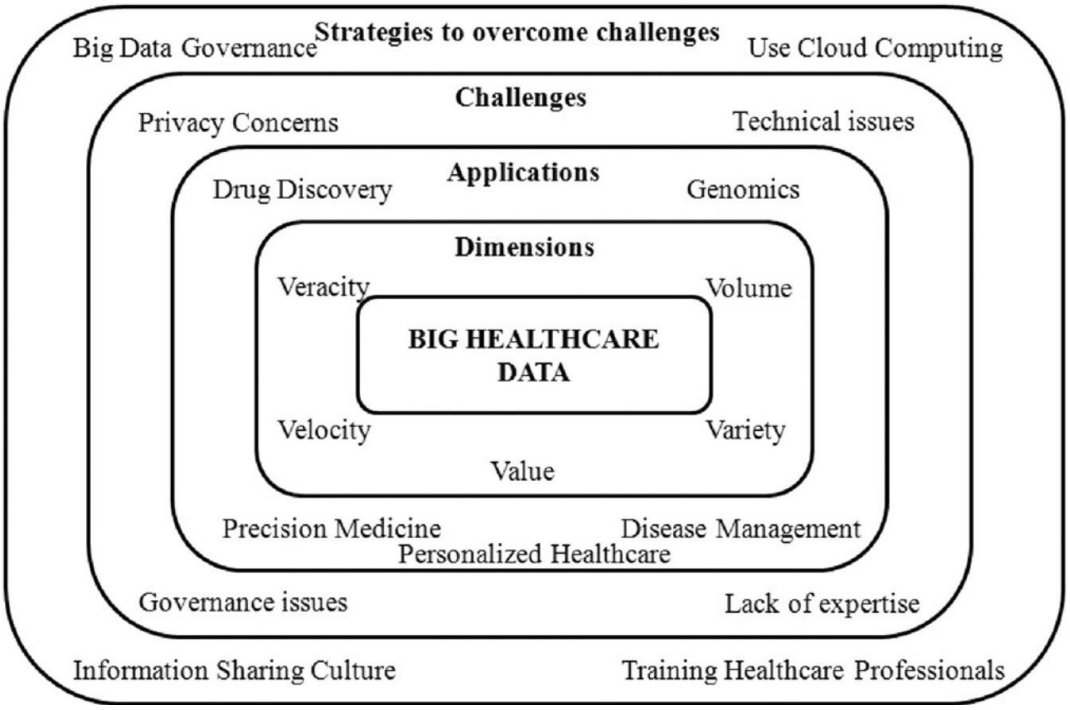

Figure 3: Healthcare Framework: Strategies and challenges of big data in healthcare (Chandra, 2020).

With the ever-changing risk environment and introduction of new emerging threats and vulnerabilities, security violations are expected to grow in the coming years. Moreover, the Affordable Care Act will lead to more enrollments for health insurance, making it an attractive focal point for hackers and opening a floodgate of healthcare breaches in the coming years. Security breaches of Electronic Health Records (EHRs) can risk patient privacy and violate the Health Insurance Portability and Accountability Act (HIPAA) of 1996 and the Health

Information Technology for Economic and Clinical Health (HITECH) Act of 2009 in the United States. Moreover, most healthcare data centers have HIPAA certification, but that certification does not guarantee patient record safety. The reason being, HIPAA is more focused on ensuring security policies and procedures than on implementing them (Patil & Seshadri, 2015).

The design of the Performance Analytics Center as part of the technical project will have to include safeguards and security measures to protect the data from the student-athletes in the sports analytics component of the design. The center will also need similar measures for the data retrieved for the academic programs. The Frank Batten School of Leadership and Public Policy, at the University of Virginia, has shown great interest in the center. The McIntire School of Commerce has shown similar excitement for the potential of Leadership and Workforce Analytics research. However, both research fields have their own risks of disadvantaging those who provide their data for analysis. For example, if research was conducted for a client of the center to analyse efficiency and workplace management in an office; those employees are at risk of reprimand or even termination, should the results display the need for such actions. It is imperative that the individuals the center receives data from, are protected from any repercussions caused by the analysis' results.

What are the safeguards that can be implemented to prevent the abuse of personal and consumer data? How can data volunteers be protected from any future disadvantages, should the analysis produce the necessary results? The aim of this STS research is to highlight the risks associated with collecting the data needed to have a fully-functional Performance Analytics Center, here at the University of Virginia. The development of this center needs to include measures to protect the student athletes and any other voluntary data donors.

One of these measures is a system of checks and balances to make sure that any data is not collected involuntarily, from unassuming or unaware subjects. A system that would identify and mitigate the risks of abuse, or the likelihood of future repercussions. Another measure is the security of the data storage facilities with high resilience; Were there to be a breach of data, what steps will be taken to contain the breach and prevent further breaches from occurring?

As more and more companies are branching out into Big Data Analysis, the likelihood of consumer abuse is likely to increase. There is currently no federal law protecting consumers after they have volunteered their data. The closest the US Government came to a suitable law to protect consumer privacy was the Fair Credit Reporting Act (FCRA) of 1970. This act applies to Consumer Reporting Agencies (CRA) , those who use consumer reports, such as a lender, and those who provide consumer-reporting information, for example a credit card company. "Consumer reports" are any communication issued by a CRA regarding a consumer's creditworthiness, credit history, credit capacity, character, and general reputation that is used to evaluate a consumer's eligibility for credit or insurance. A CRA must, "follow reasonable procedures to assure accuracy of the information . Where data is "inaccurate or incomplete or cannot be verified," a CRA must immediately correct the data (Boyne, 2017).

SECURITIZATION AND LEGISLATION

What can corporate entities such as Healthcare practices, private businesses and analytics research centers do to protect their data? One solution is Data Anonymization: technique wherein the information that discloses the identity is removed from datasets, so that the people who are

defined by the information can remain unknown i.e. sensitive data is de-identified though its format and data type is preserved (Goswami & Madan, 2017). Despite its promise, Data Anonymization has proven to be an unsuccessful technique. In the mid-90s when the Commonwealth of Massachusetts Group Insurance Commission (GIC) released the anonymous health record of its clients for research to benefit the society. The GIC hid some information like name, street address etc. so as to protect their privacy. Latanya Sweeney, then a PhD student in MIT, using the publicly available voter database and database released by GIC, successfully identified the health record by just comparing and co-relating them (Jain, et al., 2016) .

According to researchers at the Department of Computer Science at Chouaib Doukkali University in Morocco, the best solution is enhanced technology measures (Abouelmehdi, et al., 2017). These researches describe a 4-stage process to secure data in a 2017 paper published in the ScienceDirect Journal. Their plan is as follows:

1) Authentication: Authentication is the act of establishing or confirming claims made by or about the subject are true and authentic. It serves a vital function within any organization: securing access to corporate networks, protecting the identities of users, and ensuring that a user is who he claims to be. It allows only one authorized person to read or write critical data. In a healthcare system, both healthcare information offered by providers and identities of consumers should be verified at the entry of every access (Abouelmehdi, et al., 2017, para. 27).
2) Encryption: Data encryption is an efficient means of preventing unauthorized access of sensitive data. Its solutions protect and maintain ownership of data throughout its lifecycle: from the data center to the endpoint (including mobile devices used by

physicians, clinicians, and administrators) and into the cloud. Encryption is useful to avoid exposure to breaches such as packet sniffing and theft of storage devices (Abouelmehdi, et al., 2017, para. 29).

3) Data Masking: Masking replaces sensitive data elements with an unidentifiable value, but is not truly an encryption technique so the original value cannot be returned from the masked value. It uses a strategy of de-identifying the data sets or masking personal identifiers such as name, social security number and suppressing (Abouelmehdi, et al., 2017, para. 31).
4) Access Control: Once authenticated, the users can enter an information system but their access will still be governed by an access control policy which is typically based on the privilege and right of each practitioner authorized by a patient or a trusted third party. It is then a powerful and flexible mechanism to grant permissions for users (Abouelmehdi, et al., 2017, para. 33).

Despite the research team's intention to develop a system of safeguards specifically for the healthcare industry; these measures can be applied to any and all industries that rely on consumer and personal data.

The current Fair Credit Reporting Act is simply not comprehensive enough to protect consumers and individuals. Reasonable procedures could mean any and all courses of action, especially ones that would still create disadvantages. What new laws are needed to counter this phenomenon? There are some new pieces of legislation circling around in the US Congress but either they were voted down after debate, or simply did not gain enough momentum to be

introduced to committee and debating floors. According to Issak & Hanna, there are, however, some movements behind bills that are considered to be promising (2018).

There is a bill, titled the CONSENT Act (S.2639) or “Customer Online Notification for Stopping Edge-Provider Network Transgressions”, which will require the FTC to establish privacy protections for customers of online edge providers. The bill will require explicit opt-in consent from users of Facebook and other online platforms before these online platforms use, share, or sell any of their users' information, as well as explicit notification any time data is gathered, shared, or sold to a third party, in addition to adding new reporting requirements in case of a data breach involving sensitive customer proprietary information (Isaak & Hanna, 2018). The CONSENT Act was introduced in late 2018, and failed to be enacted upon by the end of the Congress session on January 17th 2019. However, Sen. Ed Markey of Massachusetts plans on “reintroducing the bill after gaining additional support from fellow law-makers in the Senate” (Isaak & Hanna, 2018).

Another bill that was introduced in 2019 was the Social Media Privacy Protection and Consumer Rights Act (SMPCR). This legislation draws similar constraints to the CONSENT Act regarding disclosure of privacy policy and obtaining initial consent and privacy preferences, but adds restrictions on modifications to privacy terms, provisions regarding withdrawal of consent, and procedures when a violation of privacy has occurred, for example: notification, data erasure, and ceasing to collect any further data (Isaak & Hanna, 2018). This legislation was introduced in the 2019-2020 session of Congress, but did not make it out of the committee floor. There is some movement to re-introduce the bill in the current session after a “re-working of the details" (Hoffman, 2019).

Only a handful states in the US, have enacted legislation and implemented laws that protect consumer data privacy. The California Consumer Privacy Act of 2018 (CCPA) was enacted in June 2018 and amended in September, and will become effective Jan. 1, 2020, with likely additional amendments in 2019. The CCPA is one of the broadest online privacy laws in the U.S., affecting companies across the country that do business with California residents. Vermont in 2018 enacted a law that requires data brokers, businesses that collect and sell or license personal information to third parties, to disclose to individuals which data is being collected and to permit them to opt out of the collection (Greenberg, 2020).

One could look internationally at countries or organizations, such as the European Union (EU), for examples on how to introduce national and supra-national policies that ensure the privacy of private residents and consumers. The EU's General Data Protection Regulation (GDPR), which took effect in May 2018, gives all EU citizens greater access to their data, a right to portability, a right to be forgotten, and right to learn when their data has been hacked (Rustad & Koenig, 2019) The GDPR is thought to be inspired by regulatory concepts initiated in the US. "The GDPR imports long-established US tort concepts for the first time into European privacy Law, including: deterrence-based fines, collective redress, wealth-based punishment and arming data subjects with the right to initiate public enforcement", (Rustad & Koenig, 2019, p. 18). The situation seems as if the ideas and solutions to counter data security issues are being presented in debate across the US but have failed to gain much traction. All the while, the same concepts and ideas have already been implemented by peer and competitor countries' governments in their own consumer protection laws.

RECOMMENDATIONS FOR SOLUTIONS

Data Privacy and Security, as well as Consumer Protection are increasingly pertinent issues, especially as data analytics becomes more integrated with business practices and our society today. The inter-connectivity of our internet enabled devices that we use daily, both professionally and socially, puts the majority of the US population at risk. The rapid digitization of personal records, ranging from financial to healthcare, eases accessibility for unauthorized individuals or malicious actors.

There are, however, many solutions to the issues that can be readily implemented or have already done so elsewhere. Data Anonymization will help to protect the identity of data volunteers, while additional technology measures such as Encryption and Access Control will allow for greater security of the overall data storage mechanisms. There has been significant research and testing of these measures, thus a space for continuous improvement and adjustment can be set in place for the future of Data Protection. Further research is required to determine the effectiveness or success of these safeguards, and whether they should be replaced or improved.

Legislation and regulation are the key measures that need to be improved, or even implemented in order to tackle these issues successfully. The US government needs to provide more effort on developing effective Data Privacy laws, ones that can pass scrutiny and voting in Congress. The EU's GDPR act of 2018, provides for a great example on how to implement large-scale legislation. One recommendation is to research and analyze the effectiveness of data

privacy laws in other countries, and how these laws can be adapted to suit the practices of and culture of the United States.

WORKS CITED

Abouelmehdi, K., Beni-Hessane, A., & Khaloufi, H. (2018). Big healthcare data: preserving security and privacy. *Journal of Big Data*, *5*(1). doi: 10.1186/s40537-017-0110-7

Adrian, A. (2013). Big Data Challenges. *Database Systems Journal, 4*(3)*,* 31-40. Retrieved from: www.dbjournal.ro/

Boyne, S. (2017). Data Protection in the United States: U.S. National Report. *Indiana University Robert H. McKinney School of Law Research Paper No. 2017, 11.* Doi: 10.2139/ssrn.3089004

Chen, W., & Quan-Haase, A. (2018). Big Data Ethics and Politics: Toward New Understandings. *Social Science Computer Review*, *38(1)*, 3–9. doi: 10.1177/0894439318810734

Chandra, A. (2020, March). An Overview of Business Intelligence: What can be produced through the analysis of consumer data. Figure 1. STS Research Paper. (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Chandra, A. (2020, March). Healthcare Framework: Strategies and challenges of big data in healthcare. Figure 3. STS Research Paper. (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Cuzzocrea, A. (2014). Privacy and Security of Big Data. *Proceedings of the First International Workshop on Privacy and Security of Big Data - PSBD, 14,* 45-47. doi:10.1145/2663715.2669614

Dimitrov, D. V. (2016). Medical Internet of Things and Big Data in Healthcare. *Healthcare Informatics Research, 22*(3)*,* 156-163. doi: 10.4258/hir.2016.22.3.156

Goswami, P., & Madan, S. (2017). Privacy preserving data publishing and data anonymization approaches: A review. *2017 International Conference on Computing, Communication and Automation (ICCCA)*, *1,* 139-142 doi: 10.1109/ccaa.2017.8229787

Greenberg, P. (2020, January 3). 2019 Consumer Data Privacy Legislation. *National Conference of State Legislatures.* Retrieved from: https://www.ncsl.org/

Isaak, J., & Hanna, M. J. (2018). User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. *Computer*, *51*(8), 56–59. doi: 10.1109/mc.2018.3191268

Ishwarappa, A. J. (2015). A Brief Introduction on Big Data's 5V's Characteristics and Hadoop Technology. *Procedia Computer Science, 48*(1)*,* 319-324. Retrieved from https://www.sciencedirect.com

Jain, P., Gyanchandani, M. & Khare, N.(2016). Big data privacy: a technological perspective and review. *Journal of Big Data, 3*(25), 3-10 . doi.: 10.1186/s40537-016-0059-y

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. & Byers, A.H. (2011, May 1). Big Data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute.* Retrieved from https://www.mckinsey.com.

Markus, M. L. (2015, January). New games, new rules, new scoreboards: the potential consequences of big data. *Journal of Information Technology, 30*, 58-59. doi:10.1057/jit.2014.28

Morgan, R. (2017, August 15). Uber settles federal probe over 'God View' spy software. *New York Post.* Retrieved from www.nypost.com

Patil, H. K., & Seshadri, R. (2014). Big Data Security and Privacy Issues in Healthcare. *2014 IEEE International Congress on Big Data*. doi: 10.1109/bigdata.congress.2014.112

Pinch, T. J., & Bijker, W. E. (1984) The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. *Social Studies of Science, 14*, 399-441.

Ruiz, R. R. (2015, April 8). F.C.C. Fines AT&T $25 Million for Privacy Breach. *The New York Times.* Retrieved from http://www.nytimes.com/

Rustad, M. L. & Koenig, T. H. (2018). Toward A Global Data Privacy Standard. *Florida Law Review, 71,* 18-16. Retrieved from https;//www.ssrn.com/

Surbakti, F. P. S., Wang, W., Indulska, M., & Sadiq, S. (2020). Factors influencing effective use of big data: A research framework. *Information & Management*, *57(1)*, 103146. doi: 10.1016/j.im.2019.02.001

The 5 V's of big data. (2019, April 16). *IBM Blogs.* Retrieved from https://www.ibm.com/blogs/watson-health/the-5-vs-of-big-data/

Uzialko, A.C. (2018, August 3). How Businesses Are Collecting Data (And What They're Doing With It). *Business New Daily*. Retrieved from: https://www.businessnewsdaily.com/

Viswanatha, A. (2016, June 8). Morgan Stanley Fined $1 Million for Client Data Breach. *The Wall Street Journal.* Retrieved from https://www.wsj.com/

Wigan, M.R. & Clarke, R. (2013). Big Data's Big Unintended Consequences. *IEEE, 46*, 46-53. doi: 10.1109/MC.2013.195

Williams, A.M., Davids K., & Williams, J.G. (1999). Visual Perception and Action in Sport. London: *Taylor and Francis, 7(1),* 1-3.

Yee, S. W. Y., Gutierrez, C., Park, C. N., Lee, D., & Lee, S. (2020). Big Data: Its Implications on Healthcare and Future Steps. *Impacts of Information Technology on Patient Care and Empowerment, 1,* 82–99. doi: 10.4018/978-1-7998-0047-7.ch005

BIBLIOGRAPHY

Abouelmehdi, K., Beni-Hessane, A., & Khaloufi, H. (2018). Big healthcare data: preserving security and privacy. *Journal of Big Data*, 5(1). doi: 10.1186/s40537-017-0110-7

Adrian, A. (2013). Big Data Challenges. *Database Systems Journal, 4(*3*),* 31-40. Retrieved from: www.dbjounral.ro

Alamar, B. & Mehrotra, V. (2011, October 7). Beyond 'Moneyball': Rapidly evolving world of sports analytics, Part I. *Analytics Magazine*. Retrieved from: http://analyticsmagazine.org/

Alamar, B. & Mehrotra, V. (2011, October 7). Figure 1: A framework of Sports Analytics: The key components in the transformation of data to knowledge, p.3. A*dapted from Beyond 'Moneyball': Rapidly evolving world of sports analytics, Part I*. *Analytics Magazine*.

Armstrong, S., Jovanov, E., & Kerwin, D. (2007, April). Wireless connectivity for health and sports monitoring: a review. *British Journal of Sports Medicine, 41*(1), 285-289. Retrieved from www.bjsm.bmj.com

Boyne, S. (2017). Data Protection in the United States: U.S. National Report. *Indiana University Robert H. McKinney School of Law Research Paper No. 2017, 11.* Doi: 10.2139/ssrn.3089004

Bruemmer, M. (2014, August 8). Diagnosing the Risks of Online Healthcare. I*nternational Association of Privacy Professionals Journal.* Retrieved from: www.iapp.org/

Chandra, A. (2019, October). Technical Project Gantt Chart: The timeframe of the technical project. Figure 2. Prospectus. (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Chandra, A. (2019, October). The Cycle of Big Data Analytics: How businesses take advantage

of consumer data. Figure 3. Prospectus. (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Chandra, A. (2019, October). Pacey's Triangle: Analysis of the cultural, technological and organizational aspects of Big Data. Figure 4. Prospectus. (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Chandra, A. (2020, March). An Overview of Business Intelligence: What can be produced through the analysis of consumer data. Figure 1. STS Research Paper. (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Chandra, A. (2020, March). Healthcare Framework: Strategies and challenges of big data in healthcare. Figure 3. STS Research Paper. (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Chen, W., & Quan-Haase, A. (2018). Big Data Ethics and Politics: Toward New Understandings. *Social Science Computer Review*, *38*(1), 3–9. doi: 10.1177/0894439318810734

Cuzzocrea, A. (2014). Privacy and Security of Big Data. *Proceedings of the First International Workshop on Privacy and Security of Big Data - PSBD, 14,* 45-47. doi:10.1145/2663715.2669614

Davis, K. (2012). Ethics of Big Data: Balancing Risk and Innovation. New York, NY: O'Reilly Media, Inc.

De Mauro, A. & Greco, M. (2015, February). What is big data? A consensual definition and a review of key research topics. *AIP Conference Proceedings, 1644(1)*, 97-99. doi:10.1063/1.4907823

Dimitrov, D. V. (2016). Medical Internet of Things and Big Data in Healthcare. *Healthcare Informatics Research, 22*(3)*,* 156-163. doi: 10.4258/hir.2016.22.3.156

Goswami, P., & Madan, S. (2017). Privacy preserving data publishing and data anonymization approaches: A review. *2017 International Conference on Computing, Communication and Automation (ICCCA)*, *1,* 139-142 doi: 10.1109/ccaa.2017.8229787

Greenberg, P. (2020, January 3). 2019 Consumer Data Privacy Legislation. *National Conference of State Legislatures.* Retrieved from: https://www.ncsl.org/

Hayhurst, C. (2019, August 8). Data Analytics Helps College Coaches and Athletes Optimize Training and Performance. *EdTech Magazine*. Retrieved from: https://www.richmond.com/

Isaak, J., & Hanna, M. J. (2018). User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. *Computer*, *51*(8), 56–59. doi: 10.1109/mc.2018.3191268

Ishwarappa, A. J. (2015). A Brief Introduction on Big Data's 5V's Characteristics and Hadoop Technology. *Procedia Computer Science, 48*(1)*,* 319-324. Retrieved from https://www.sciencedirect.com

Jain, P., Gyanchandani, M. & Khare, N.(2016). Big data privacy: a technological perspective and review. *Journal of Big Data, 3*(25), 3-10 . doi.: 10.1186/s40537-016-0059-y

Kumar, R., (2016). Impact of Big Data Analytics on Healthcare and Society. *Journal of Biometrics & Biostatistics*. Retrieved from https://www.omicsonline.org/

Lindsey, G. R. (1959). Statistical Data Useful for the Operation of a Baseball Team. *Operations Research, 7(1)*, 197-207.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. & Byers, A.H. (2011, May 1). Big Data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute.* Retrieved from https://www.mckinsey.com.

Markus, M.L. (2015). New games, new rules, new scoreboards: the potential consequences of big data. *Journal of Information Technology, 30(1)*, 58-59. doi:10.1057/jit.2014.28

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. New York, NY: Houghton Mifflin Harcourt.

Morgan, R. (2017, August 15). Uber settles federal probe over 'God View' spy software. *New York Post.* Retrieved from www.nypost.com

McElroy, W. (2019, June 1). UVA has had the best athletic year of any college sports program. *Richmond Times-Dispatch*. Retrieved from: https://www.richmond.com/

Nunan, D. & Di Domenico, M. (2017). Big Data: A Normal Accident Waiting to Happen? *Journal of Business Ethics*, *145*(3), 481-491. doi.org/10.1007/s10551-015-2904-x

Patil, H. K., & Seshadri, R. (2014). Big Data Security and Privacy Issues in Healthcare, *2014 IEEE International Congress on Big Data*,, pp. 762-765. doi: 10.1109/bigdata.congress.2014.112

Pinch, T. J., & Bijker, W. E. (1984) The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. *Social Studies of Science, 14*, 399-441.

Ruiz, R. R. (2015, April 8). F.C.C. Fines AT&T $25 Million for Privacy Breach. *The New York Times*. Retrieved from http://www.nytimes.com/

Rustad, M. L. & Koenig, T. H. (2018). Toward A Global Data Privacy Standard. *Florida Law Review, 71,* 18-16. Retrieved from https;//www.ssrn.com/

Siemens, G., Dawson, S. & Lynch, G., (2013, December 3). Improving the Quality and Productivity of the Higher Education Sector. *Society for Learning Analytics Research*. Retrieved from http://www.solaresearch.org/

Smith, B. (2019, August 6). Data Analytics Helps College Coaches and Athletes Optimize Training and Performance. *EdTech Magazine*. Retrieved from: www.edtechmagazine.com

Surbakti, F. P. S., Wang, W., Indulska, M., & Sadiq, S. (2020). Factors influencing effective use of big data: A research framework. *Information & Management*, *57*(1), 103-146. doi: 10.1016/j.im.2019.02.001

The 5 V's of big data. (2019, April). *IBM Blogs*. Retrieved from https://www.ibm.com/

University of Virginia: Office of Communications. (2019). Great And Good: The 2030 Plan. Charlottesville, VA: Author.

Uzialko, A.C. (2018, August 3). How Businesses Are Collecting Data (And What They're Doing With It). *Business New Daily*. Retrieved from: https://www.businessnewsdaily.com/

Viswanatha, A. (2016, June 8). Morgan Stanley Fined $1 Million for Client Data Breach. *The Wall Street Journal.* Retrieved from https://www.wsj.com/

Wigan, M.R. & Clarke, R. (2013, June). Big Data's Big Unintended Consequences. *IEEE*, *46*(6), 46-53. doi: 10.1109/MC.2013.195

Williams, A. M., Davids K., & Williams, J.G. (1999). Visual Perception and Action in Sport. London: *Taylor and Francis, 7(1),* 1-3.

Yee, S. W. Y., Gutierrez, C., Park, C. N., Lee, D., & Lee, S. (2020). Big Data: Its Implications on Healthcare and Future Steps. *Impacts of Information Technology on Patient Care and Empowerment, 1,* 82–99. doi: 10.4018/978-1-7998-0047-7.ch005