Bridging the Gap: The Potential in Performance Data Coupled with Lack of Regulation

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Peter L. Worcester

Spring 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

Kathryn A. Neeley, Associate Professor of STS, Department of Engineering and Society

Introduction

In recent years, technological breakthroughs in wearable devices have given consumers the ability to track almost every facet of their lives. According to CCS Insight, 411 million smart wearables worth 34 billion dollars will be sold in 2020. (Lamkin, 2016). Fitness, activity, and sports trackers represent roughly 50 percent of this market, as seen in Figure 1 below. This rapidly growing market of wearables can collect data on all aspects of an athlete's life. (Hughes, 2017). The data collected by wearable smart devices can be defined as biometric performance data, or simply performance data. (Arnold, 2017).

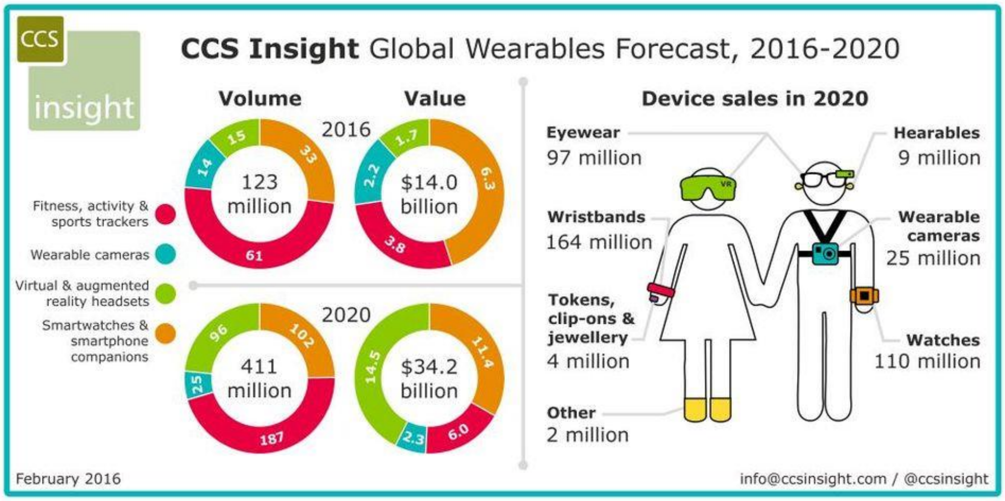

Figure 1: Global Wearable Tech Market Trends. This graphic shows that there will be roughly four hundred million wearable smart devices to be sold in 2020. (Lamkin, 2016).

Team and individual team sports should be investigated using congruent concepts and tools since they are dynamic systems. (Williams et al, 1999). Applying these concepts and tools to performance data leads to what is called Performance Analytics. Performance Analytics isn’t confined to sports, however. Researchers believe 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 project aims to produce a proposal for the University of Virginia’s administrators about the launch of a Performance Analytics

Center. This center is envisioned 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. There have been many pitfalls in when designing this center, however. This is primarily due to the various restrictions that are put in place to protect the privacy of personal data. Initially, 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 privacy constraints inspired my 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. Currently, there are a plurality of privacy and security risks surrounding personal data for both student athletes and individuals alike. Investigation into these risks leads to many questions including which industries have the highest potential for abuse, what are the current laws that are lacking in providing security, and what needs to change so that the risk of harm is mitigated. Data analytics offers great potential benefits to society but has plenty of shortcomings. Thus, this 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 Emmanuel Mesthene's model for the relationship between technological and social change to evaluate the risks of performance analytics. In this paper, I argue that implementing greater regulation will lead to a balance of protecting data privacy while also maintaining the benefits of performance analytics. This claim will be supported by evidence from previous developments into similar research and studies from experts in the field of technology and public policy.

Part 1: The Ability for Big Data to Do Good and Bad

Big Data has existed for decades but has just now recently gained public attention in a "big" way. This boom in attention is largely attributed to the exponential rise in data volume in both how it is collected and accessed. Data analytics, and its extension into Big Data, have become one of the premier developments in 21st-century sport. Athletes' performance data comprises a new and valuable category of big data. Historically, teams have used a wide variety of performance data such as heart rate, biomechanical measurements, reaction time, and self-reported wellness information. (Osborne, 2017). Whether athletes are ready for Big Data, it does appear to stay. The NCAA recently approved the use of some wearable technologies in football games for the purpose of health and safety. (Dodd, 2014). The rapidly increasing rate of collection in a wide variety of athletes' performance data will inevitably lead to issues of legality and ethical consequences in the near future.

Over the last two decades, Big Data has also 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).

Researchers estimate that the amount of data stored in the world's data systems is doubling roughly every two years, and corporations have responsibility for about 85% of that data (Surbakti et al., 2019). 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, and various sensors. By analyzing customer behavior, as well as vast amounts 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 and any information exchange between the consumer and the firm will result in data generation.

Data generation can also come from wearable devices in sports. Researchers forecast the sports analytics industry to grow into $5.2 billion by 2024. (Sports, 2020). Corporations have recognized this and have looked to university athletic programs as a potential gold mine. Performance data from student athletes can be collected in endless formats, ranging from cardiac health to the hormonal system. Sports trainers and coaches at the collegiate level can collect any of this performance data, which puts athletes at risk of unnecessary intrusions. One of the leading performance data collection companies for college and professional sports is Catapult Sports. Catapult Sports is an Australian company described as a "wearable data juggernaut." (Glaeser, 2020). Catapult Sports devices can collect up to one hundred different data metrics on athletes while the tracking device is being worn. The devices operate on a cloud-based analytics platform that allows a team's training staff to report and present data in the style that best suits the needs of the team using the wearable devices. (Wearable, 2019). However, a private cloud-based platform, while beneficial, means that the data and programs are being filtered through a "central software control point." (Knorr, 2014). As a result, this means that all teams using Catapult Sports' analytics technology may have their athletes' data stored in a centralized location. Because it appears most college teams use Catapult Sports, a large amount of player data is presumably stored by this single company, likely at either a single location or a small number of locations. Thus, Catapult Sports has the capability of collecting performance data on college athletes at all levels, raising a major area of concern in the event of a data breach.

Another recent example of data insecurity is the University of Michigan's apparel contract with Nike. The contract, worth approximately $170 million, grants broad rights that allows Nike to utilize Michigan players' performance data. (Snyder, 2016). Hence, performance data can be collected while athletes are on and off the field, which may raise data security risks for the athletes. Some college athletes may not trust the data produced by wearable sensors and what is done with it. A hidden clause grants Nike “the right to utilize … Activity Based Information … in any and all media…” (Snyder, 2016). While analytics can enhance performance, it can also threaten privacy and damage relationships. Requiring student-athletes to wear sensors, such as Nike’s, may break the trust between athletes and coaches.

In the last few decades, big data has been applied to diverse fields, such as the government, international development, and education. It is only now that the sports industry 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 sports analytics field. This information can be gathered through sources such as Catapult and Nike wearable technologies. Recognizing patterns would aid in predicting preventative measures for an increased holistic and personalized analysis. 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).

Part 2: Mesthene's Framework Applied to Performance Analytics

Issues arise concerning users' data, including informed consent for use and disclosure, retention, 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 have affected all facets of the corporate world including governments, municipalities, and education and sports (Chen & Quan-Hasse, 2018).

Emmanuel Mesthene's *Technological Change: Its Impact on Man and Society* presents a model of the relationship between technological and social change (Mesthene, 1970). In the preface, Mesthene states that modern technologies "bring about changes in institutions and individual life styles; they generate strains for our values and beliefs; and they create problems – and opportunities – for our economic and political organizations." (Mesthene, 1970). While this model is fundamentally cyclical in nature, it can be summarized in a flowchart that was handed out in class below in Figure 2. This model can be applied to any domain, including Big Data. In this supporting argument, I will analyze the extent to which Mesthene's model is well suited to this situation.

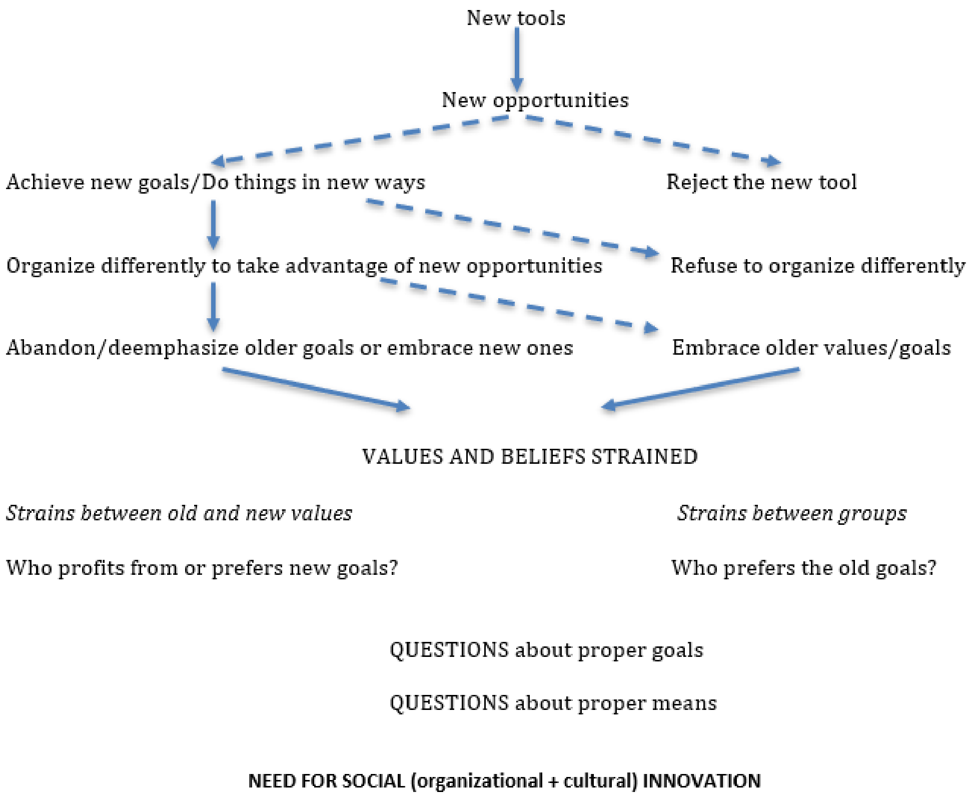
New tools
New opportunities
Achieve new goals/Do things in new ways
Reject the new tool
Organize differently to take advantage of new opportunities
Refuse to organize differently
Abandon/deemphasize older goals or embrace new ones
Embrace older values/goals
VALUES AND BELIEFS STRAINED
Strains between old and new values
Strains between groups
Who profits from or prefers new goals?
Who prefers the old goals?
QUESTIONS about proper goals
QUESTIONS about proper means
NEED FOR SOCIAL (organizational + cultural) INNOVATION

Figure 2: A Visual Summary of Mesthene's Model of the Technological Change-Social Change Relationship. (Neeley, 2019).

New technologies have been bringing about social changes since the dawn of time and performance analytics in Big Data is no different. What differentiates our time is that our society has widespread awareness of the social changes that new technologies bring. According to Mesthene, technology appears to induce change in two distinct ways: by creating new opportunities and by generating new problems for individuals and for societies. He provides a mechanism for how technological development leads to both positive and negative effects. Firstly, technological advancements provide new opportunities to a certain goal. Taking advantage of these opportunities leads to required changed in a social organization. These changes will interfere with the functions in existing social structures. Finally, the initial goals that were achieved in prior social structures can no longer by adequately achieved.

We can also break down Mesthene’s model in terms of performance data from a sociotechnical perspective. Looking at the impact of performance data from a sociotechnical perspective, one begins to see why performance data is of ethical concern. Technically, data analytics is advancing at an unprecedented rate, with the amount of available data nearly doubling every two years. Organizationally, there are huge economic benefits that businesses stand to benefit from consumer analytics. Finally, culturally, consumer data manipulation has several potential negative impacts and ethical issues.

In this chapter, Mesthene applies his methodology to a recent study of computer-based educational technology by Anthony G. Oettinger. (Mesthene, 1970). Oettinger recognizes the vast benefits and exciting possibilities with the use of computers, assorted teaching machines, and systems analysis in secondary education. However, when Oettinger looks to turn this experimental possibility into a large-scale practical application he finds that his quick technological fix will not impose educational innovation in the current social structure. There are a number of reasons for the lack of success here. First, the hardware itself is primitive, unreliable, and expensive. Second, teachers are unaware of how to apply the technology and as a result the hardware collects dust in the classroom. However, the greatest resistance in the introduction of technology into schools is the structure of the American school system itself. Positive change won’t automatically happen without organizational and cultural innovation. Mesthene ends with a powerful metaphor: “Technology may be the motor of all progress, but institutional sluggishness will most often turn out to be an effective brake.” (Mesthene, 1970).

Mesthene’s example of the introduction of educational technology can also be applied to Big Data. 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. “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 violates 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). In this example, Uber’s GPS tracking tools offered new opportunities to harvest location data. Uber organized this data such that many of their employees had wide access. By doing this, they abandoned their values of privacy and succumbed to corporate greed. This application of Mesthene’s model highlights that lack of regulation leads to data misuse and privacy risks for our society and calls for social innovation as a result.

As discussed earlier, college athletics teams collect a vast amount of performance data on their respective players. Yet, even though this data is often used for player safety and to increase player performance, the collected and stored performance data is arguably not protected. Moreover, the NCAA does not currently address the use or collection of performance data in its bylaws. As such, college athletes likely have no regulatory protection for their collected performance data.

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, and The McIntire School of Commerce has shown similar excitement for the potential of Performance 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 analyze 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.

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? Each of these solutions are potential regulatory actions that can be taken such that Performance Analytics can be adopted at the school responsibly.

Part 3: Solutions for Organizational Regulation

What can corporate entities such as private businesses and sports analytics research centers do to protect their data? One solution is data anonymization. This is a 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). In addition, the US Government came to a suitable law to protect consumer privacy called the Fair Credit Reporting Act (FCRA) of 1970. This act applies to Consumer Reporting Agencies (CRA). 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).

However, 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).

One bill that was introduced in 2019 was the Social Media Privacy Protection and Consumer Rights Act. This legislation has constraints 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.

Conclusion

Data privacy and security are increasingly relevant issues in all domains of society today, including the sports realm. 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 sports, eases accessibility for unauthorized individuals or malicious actors. Using Mesthene's model for the relationship between technological change and social change and applying it to performance analytics showed that the health and safety of student athletes and others are compromised for corporate greed. Mesthene's framework highlights that the benefits of Big Data won't be fully realized until there is social and organizational change to remedy this issue.

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. 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.

In order to maximize the benefits of Big Data, legislation and regulation are the key measures that need to be improved and instituted. The US government needs to provide more effort on developing effective data privacy laws. Unfortunately, the legistlative and administrative process takes time, and until the federal government takes action, athletes and individuals will remain unprotected. 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.

References

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

Arnold JF, Sade RM. Wearable Technologies in Collegiate Sports: The Ethics of Collecting Biometric Data From Student-Athletes. Am J Bioeth. 2017;17(1):67–70. doi:10.1080/15265161.2016.1251648

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

Dodd, D. (2014). NCAA denies ACC use of helmet cams, sideline communications. *CBS Sports*. Retrieved December 9, 2019.

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

Glaeser, C. (2020, January 3). A Buyer's Guide to Athlete Tracking Systems for Coaches. Retrieved from https://simplifaster.com/articles/athlete-tracking-systems/

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

Hughes, B. (2017, December 6). The New Wave of Sports Wearables. *HUFFPOST*. Retrieved from: https://www.huffingtonpost.com/brian-hughes/the-new-wave-of-sports-we b 12449566.html.

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

Karkazis, K., & Fishman, J. R. (2017). Tracking US professional athletes: The ethics of biometric technologies. The American Journal of Bioethics, 17(1), 45-60.

Knorr, E. (2014, August 4). Build Your Own Private Cloud. *INFOWORLD*. https://www.infoworld.com/article/2608305/cloud-computingbuild-your-ownprivate-cloud.html.

Lamkin, P. (2016, February 17). Wearable Tech Market To Be Worth $34 Billion By 2020. Retrieved from https://www.forbes.com/sites/paullamkin/2016/02/17/wearable-tech-market-to-be-worth-34-billion-by-2020/#2b5490743cb5

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.

Mesthene, E. G. (1970). Technological change: Its impact on man and society.

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

Neeley, K. (2019). Mesthene's Model of the Technological Change-Social Change Relationship

Osborne, B. (2017). Legal and Ethical Implications of Athletes' Biometric Data Collection in Professional Sport. Marq. Sports L. Rev., 28, 37.

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/

Sports Analytics Market. *Markets and Markets*, January 2020. Retrieved from: https://www.marketsandmarkets.com/Market-Reports/sports-analytics-market-35276513.html

Snyder, M. (2016, April 26). Michigan finalizes Nike contract for up to $173.8 million. *Detroit Free Press.* Retrieved from: https://www.freep.com/story/sports/college/university-michigan/wolverines/2016/04/26/michigan-nike-contract/83533954/

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

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/

Wearable GPS Sports Performance Trackers: Catapult Sports. (2019, November 11). Retrieved from https://www.catapultsports.com/products

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