Big Data Law

[1]Matthew N. O. Sadiku, [1]Philip O. Adebo and [2]Janet O. Sadiku,
[1]Roy G. Perry College of Engineering, Prairie View A&M University, Prairie View, TX, USA
[2]Juliana King University, Houston, TX, USA

Abstract: Big data is one of the most valuable assets in today's digital world. It is regarded as the oil of the digital era. With data at the forefront of many modern lives, understanding how to properly manage, store, and process it is vital. Data can greatly influence litigation processes. There is no uniform state, federal, or international law that governs big data. The law of big data does not exist as a concrete sets of rules, but as a myriad of regulations that are changing with the development of technology. In this paper, we discuss the legal implications of big data.

Keywords: *Big Data, Law, Big Data Law*

I. INTRODUCTION

There is data everywhere. Each day, large amounts of data are being generated. Organizations and companies now have very large data stored in their files, databases, and data warehouses. The amount of data generated by the Internet and mobile technologies is unprecedented. Data comes from a variety of sources such as sensors, social media sites, transaction records, medical records, RFID devices, video sharing, etc. This huge amount of data is collectively called big data. Big data enables the collection and use of massive amount of data generated by man and machine [1]. Big data usage covers a wide range of fields ranging from healthcare, education, research, government, commerce, and politics. It is revolutionizing virtually all aspects of our society.

There is no doubt that information is the most precious commodity for any business. Data is information in raw format, while information represents data after processing and analysis. There is data everywhere. Data comes from a variety of sources such as sensors, social media sites, smart phones, Internet, emails, ecommerce transactions, weather data, medical records, insurance records, RFID devices, video sharing, etc.

Figure 1 highlights several sources of the big data deluge [2]. The ability to collect and analyze huge amounts of data is a growing problem within the engineering community.
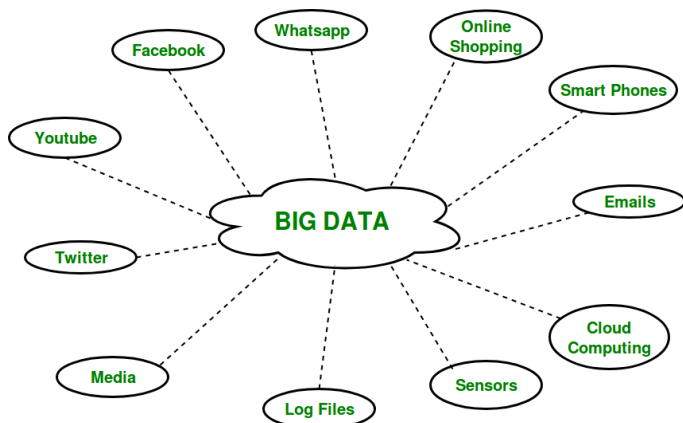

Figure 1: Sources of the big data deluge [2].

This huge amount of data is collectively called big data. Big data refers to massive amount of data that are so large that traditional processing tools cannot cope. It is a high-volume, high-velocity, and high-variety information that requires special information processing tools. Because of these characteristics, big data requires new technologies and techniques to capture, store, and analyze.

Big data is currently a major topic of discussion across a number of fields, including business, management and marketing, education, scientific research, national security, and government. While there are laws that protect certain types of personal data, such as the European Union's General Data Protection Regulation (GDPR) and the California Consumer Protection Act (CCPA), these laws do not recognize data itself as property. In the United States today, there is no single law to address the privacy concerns associated with the collection of consumer data.

WHAT IS BIG DATA?

Our lives revolve around huge data sets**.** With the advent of various social media platforms and multinational companies, the generation of data has increased drastically. As its name implies, big data is a structured, semi-structured, and unstructured data, which is very big, fast, and comes in many forms. These three types of data are shown in Figure 2 [3]. Big data may be regarded as a phenomenon since we can observe its effects like growing volume and variety. It has become the fuel that every industry needs today to flourish.

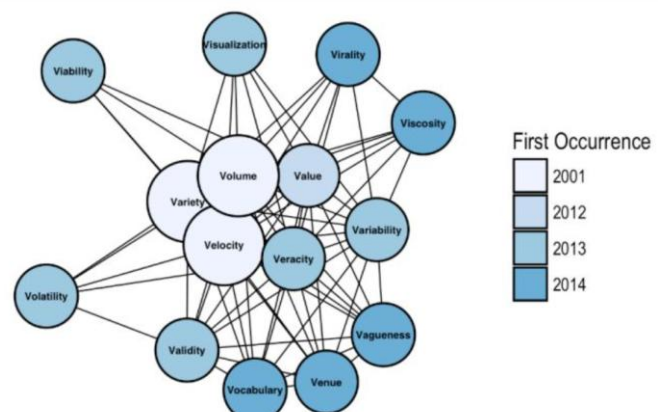Figure 2: Types of big data [3].

Figure 3: The 42 V's of big data [4].

One of the major strengths of big data is its flexibility and universal applicability to so many industries. Big data is used in several areas such as education, business, finance, government, healthcare, engineering, manufacturing, agriculture, social media, tourism, industry, entertainment, sports, construction, transportation, defense, etc. Increasingly, big data is regarded as the most strategic resource of the 21st century, similar in importance to gold and oil. It may also be regarded as the new form of currency.

Big data (BD) is a relatively newer technology that can make use of smart city services. The three main sources of big data are machines, people, and companies. As shown in Figure 3 [4], big data can be described with 42 Vs. The first five Vs are volume, velocity, variety, veracity, and value [5].

- *Volume*: This refers to the size of the data being generated both inside and outside organizations and is increasing annually. Some regard big data as data over one petabyte in volume.
- *Velocity*: This depicts the unprecedented speed at which data are generated by Internet users, mobile users, social media, etc. Data are generated and processed in a fast way to extract useful, relevant information. Big data could be analyzed in real time, and it has movement and velocity.
- *Variety*: This refers to the data types since big data may originate from heterogeneous sources and is in different formats (e.g., videos, images, audio, text, logs). BD comprises of structured, semi-structured or unstructured data.
- *Veracity*: By this, we mean the truthfulness of data, i.e. weather the data comes from a reputable, trustworthy, authentic, and accountable source. It suggests the inconsistency in the quality of different sources of big data. The data may not be 100% correct.
- *Value*: This is the most important aspect of the big data. It is the desired outcome of big data processing. It refers to the process of discovering hidden values from large datasets. It denotes the value derived from the analysis of the existing data. If one cannot extract some business value from the data, there is no use managing and storing it.

Industries that benefit from big data include the healthcare, financial, airline, travel, restaurants, automobile, sports, agriculture, and hospitality industries. Big data technologies are playing an essential role in farming: machines are equipped with sensors that measure data in their environment. Individuals with knowledge of Big Data are referred to as Big Data Specialists and they specialize in Hadoop, Map Reduce, Spark, NO SQL and DB tools like HBase, Cassandra, and MongoDB, etc.

II. BIG DATA LAW

Law is an indispensable aspect of life. There is always a set of rules and regulations attached to any activity we perform. The law industry comprises the mechanisms and processes to deal with legal issues, the courts, legal and law firms, and even the police can all be considered as the pillars of the law industry. The law industry as well as several other industries are making the use of big data available to them [6]. The main reasons many lawyers give for a reluctance to embrace big data are ethical considerations and concerns for client privacy. As a lawyer, you only have a choice: you wither compete with these systems or help build them. Figure 4 shows a lawyer analyzing big data [7].

Figure 4: A lawyer analyzing big data [7].

The Common Rule protects most human subjects involved in federally funded research and generally requires patient consent. States define their own privacy framework, which typically includes laws governing the same entities, activities, and/or types of information as the all federal laws. State laws and regulations are relevant to the extent that they restrict disclosure of identifiable information more than federal laws. Big data fits into what intellectual property rights seek to regulate.
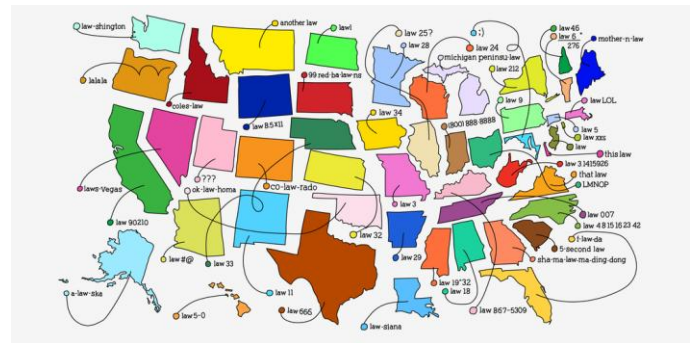

Figure 5: The state of consumer data privacy laws in the US [8].

The United States does not have a singular law that covers the privacy of all types of data. Instead, it has a mix of laws that go by acronyms like HIPAA, FCRA, FERPA, GLBA, ECPA, COPPA, and VPPA. Figure 5 shows the state of consumer data privacy laws in the US [8]. Federal regulation of big data is done through a variety of statutes, but there is no one unified federal privacy law. The federal law surrounding big data goes further than mere collection, although data collection of big data can also have potential federal legal issues. Typically data collection involves the realm of privacy and proper disclosure. Big data is affected by state privacy laws, especially privacy laws to directly address online disclosures and record keeping.

The Committee on House Administration (CHA) held a hearing to discuss data privacy risks and reforms needed in the public and private sectors. Lawmakers from both parties agree that a uniform national law is needed. As Congress members ponder different models, federal action on data privacy is unlikely to go anywhere soon. Since 2018, 34 states have passed or introduced seventy-two privacy bills regulating the commercial collection and use of personal data. To date, California, Colorado, and Virginia are the only states that have passed comprehensive data privacy legislation [9].

III. LEGAL ISSUES IN BIG DATA

How we use big data is a cause for legal concern. The law is gradually covering up the exposed parts of the vulnerabilities that big data pose. In law, utilizing big data with advanced analytics tools can speed up the justice process, present new evidence, and streamline document processing. Some of the challenges legal professionals face when dealing with big data include the following [10]:

- *Data Security:* Big data naturally brings more risk of security and privacy issues. It also bring a higher threat of a data breach. With more scrutiny placed on data protection, legal professionals must consider privacy, confidentiality, and data management more than ever. The stakes are high for protecting and securing sensitive data. Organizations handling regulated data must implement rigorous security measures to protect the sensitive and confidential data they store, access, and work with daily. Security standards vary from industry to industry and change based on the location in which they were created.

- *Consumer Privacy:* Protecting personal data is a legal requirement. When dealing with big data, legal professionals need to remain compliant with General Data Protection Regulation (GDPR). Some scholars have argued that existing information privacy laws may be inadequate to confront big data's privacy risks. Reducing your risk of sensitive data exposure is imperative. It is important to obtain consent from the users when collecting that information.

- *Volume of Data:* One of the main challenge of big data is the sheer volume of electronically stored information that sometimes has to be analyzed. The amount of data being created can be overwhelming. Legal professionals can utilize efficient ways to analyze data to overcome the challenge of big data.

- *Review Process:* Reviewing the data you have processed can be time-consuming, especially if the review is manual. This is why modern analytics tools are becoming increasingly important in litigation.

- *Variety of Data:* The variety of data can compound the challenge of volume. The variety includes active data, mobile data, cloud data, etc.

IV. APPLICATIONS

Big data may be regarded as a problem-solving philosophy that leverages massive datasets and algorithmic analysis to extract "hidden information and surprising correlations." We live in the "big data era"–an era in which companies collect vast amounts of consumer data to work more efficiently and productively. Researchers and decision-makers have realized that big data is beneficial for understanding consumer needs, improving service quality, and predicting and preventing risks [11]. Big data offers a number of benefits across different industries and practice areas. Here we provide a number of examples of applications of regulation on big data.

- *Public Health:* Big data has become the ubiquitous watch word of medical innovation. It has long been promised to substantially improve health care. Big data analytics can be used to predict critical healthcare-related information. It can be used to predict critical healthcare-related information, develop treatments in areas without specialty providers, and detect and diagnose disease. Big data holds great promise for public health, given the nature of services such as monitoring population health status,

evaluating population-based health service quality, and conducting research for innovative solutions. Some of the most well-known privacy laws are Health Insurance Portability and Accountability Act of 1996, commonly referred to as HIPPA. This statute protects patient information and controls the type of patient data that can be released. The health information privacy framework is a patchwork of often-overlapping federal and state laws that regulate specific types of information, individuals, and organizations. The HIPAA Privacy Rule governs "protected health information," which is individually identifiable information about an individual's care, health condition, or payment for care. The legal framework governing health information does not impede or prohibit many big data uses that support improvements in public health [12].

- *Business:* Big data is a powerful tool for businesses, but it also needs to be protected from misuse. It allows businesses to provide better services to customers based on their individual needs. It can help organizations identify and react to problems in near real time. It can also help avoid security threats and fraud by allowing organizations to detect anomalies in their data or on their networks [13]. Although the basic privacy laws will not fix everything, new laws could at least encourage less privacy-hostile products and services. Civil law plays an important role in determining the marketability of big data within a company and across different industries.

- *Research:* This is one of the cornerstones of good public relations practice. Currently research in public relations is being affected by the popular phenomenon of big data, a research practice that uses large amounts of information to gain insight into human behavior and beliefs. Big data allows for in-depth analysis of patterns within behavior, and it provides a level of insight that is valued by academia and industry alike. Big data is an important form of research and it will be part of research for a long time [14].

- *Surveillance:* The surveillance technology is a civil rights issue and a democracy issue. Big data technologies are revolutionizing policing. The rise of big data policing explores how data-driven surveillance technologies impact the "who," "where", "when," and "how" of everyday policing. Big data policing has a "black data" problem as the legacy issues of racial discrimination, opacity, and distortions to constitutional protections threaten to undermine the legitimacy of these law enforcement innovations [15].

- *Legal System:* Big data is driving a trend towards behavioral optimization and "personalized law," in which legal decisions and rules are optimized for best outcomes and where law is tailored to individual consumers based on analysis of past data. The use of big data will likely generate unintended consequences in the legal system. Law is abstract, values-based, and built on compromise. Big data is empirical, algorithmic, and deterministic [16].

V. BENEFITS

Big data can have big benefits. At least in the industrialized world, most people routinely contribute to and experience the effects of big data. Although big data strives to be objective, law and data have multiple possible meanings and uses. With appropriate analysis, big data can be used to determine the reasons for business failure, cost reduction, time savings, better decision making, and new product creation. Big data is driving a trend towards behavioral optimization and "personalized

law," in which law is tailored to individual consumers based on analysis of past data [16].

Big data is quickly making the leap from politics to policy. Just as US democracy depends on the accuracy of census data, so too will policymaking increasingly depend on the accuracy of big data and advanced analytics.

VI. CHALLENGES

In addition to the benefits of big data, there are many legal and ethical implications associated with its use. Big data is considered the greatest innovation or the greatest peril of our times, depending on whom you ask. The problem with big data is that there is not a uniform set of rules or regulations that would govern the collection of electronic information. The use of big data will likely generate unintended consequences in the legal system. Large-scale use of big data will create distortions that adversely influence legal decision-making, causing irrational herding behaviors in the law.

The collection of new types of data is essential for documenting persistent inequality and discrimination. Personal information that is known to a corporation can be used by companies and the government against vulnerable populations, including women, the formerly incarcerated, immigrants, religious minorities, and young people. People who have access to government databases have often used them for improper purposes. Unscrupulous companies can find vulnerable customers through a new industry of highly targeted marketing lists. Location-aware social media tools have allowed abusive spouses and partners to learn the whereabouts of their victims in real time [17].

Legal debates over the "big data" revolution currently focus on the risks of inclusion: the privacy and civil liberties consequences of being swept up in big data's net. The consequences of exclusion could be much more profound. Billions of people worldwide remain on big data's periphery because they do not routinely engage in activities that big data is designed to capture. Critics worry the world's increasing "datafication" ignores or even smothers the unquantifiable, immeasurable, ineffable parts of human experience. Since emerging technologies pose unique dangers to equality, and not just privacy, a new legal doctrine may be needed to protect those persons whom the big data revolution risks sidelining. The big data revolution may create new forms of inequality and subordination, and thus raises broad democracy concerns [18].

CONCLUSION

In this digital age, data is being collected and processed in ways never seen before. Data is critical to rapid, informed decision making. It must relate to relevant information and be collected and analyzed in ways that help drive rapid, fact-based decisions. It acquires enormous significance when it is mined and analyzed. Big data analytics is driving innovation across all industries, and there are many benefits to be gained from its analysis. Legal providers must adopt, embrace, and utilize big data to service clients in the digital age.

Legalities are an inevitable part of life. There are numerous legal frameworks designed to protect big data assets from misuse or unauthorized access by third parties. As new technologies allow companies and government to gain greater insight into our lives, it is vitally important that these technologies be designed and used in ways that respect the values of equal opportunity and equal justice.

Data is used to inform, but what the ethical dilemmas in data analysis show is that it is up to individuals, not the data, to make decisions. The legal landscape surrounding big data analytics is uncertain and the basic privacy laws would not fix everything. More information about bid data law can be found in the books in [19-33] and the following related journals:

- *Richmond Journal of Law and Technology*
- *Northwestern Journal of Technology and Intellectual Property*

References

[1] R. Herschel and V. M. Miori, "Ethics & big data," *Technology in Society*, vol. 49, 2017, pp. 31-36.

[2] "The world of big data," August 2019, https://www.geeksforgeeks.org/world-big-data/

[3] "Where the law comes in for big data?" https://blog.sajidiqubal.com/where-the-law-comes-in-for-big-data/

[4] "The 42 V's of big data and data science," https://www.kdnuggets.com/2017/04/42-vs-big-data-data-science.html

[5] M. N. O. Sadiku, M. Tembely, and S.M. Musa, "Big data: An introduction for engineers," *Journal of Scientific and Engineering Research*, vol. 3, no. 2, 2016, pp. 106-108.

[6] "5 Applications of big data in law industry," June 2021, https://www.analyticssteps.com/blogs/5-applications-big-data-law-industry

[7] M. A. Cohen, "Why is law so slow to use data?" June 2019, https://www.forbes.com/sites/markcohen1/2019/06/24/why-is-law-so-slow-to-use-data/?sh=68fed55eb8eb

[8] T. Klosowski, "The state of consumer data privacy laws in the us (and why it matters)," September 2021, https://www.nytimes.com/wirecutter/blog/state-of-privacy-laws-in-us/

[9] S. Buka, "Big data: An update on federal data privacy legislation," March 2022, https://www.piworld.com/post/big-data-update-data-privacy-legislation/

[10] "5 Legal challenges of big data: How large volumes of data lead to legal discovery," August 2022, https://www.altlaw.co.uk/blog/legal-challenges-of-big-data

[11] M. Valent, "The United States: Big data, little regulation," *University of Miami Business Law Review*, vol. 28, no. 2, September 2020, pp. 434-455.

[12] J. H. Thorpe and E. A. Gray, "Big data and public health: Navigating privacy laws to maximize potential," *Public Health Reports*, vol. 130, no. 2, March-April 2015, pp. 171-175.

[13] K. J. Nahra et al., "United States: Big data analytics privacy law considerations," may 2023, https://www.mondaq.com/unitedstates/data-protection/1320764/big-data-analytics-privacy-law-considerations

[14] C. Myers, "Big data, privacy, and the law: How legal regulations may affect PR research," June 2018, https://instituteforpr.org/big-data-privacy-and-the-law-how-legal-regulations-may-affect-pr-research/#:~:text=However%2C%20big%20data%20research%20is,affect%20research%20in%20public%20relations.

[15] "The rise of big data policing- Author talk with Andrew Ferguson," February 2018, https://pottershousedc.org/events/2018/2/3/the-rise-of-big-data-policing-author-talk-with-andrew-ferguson

[16] C. Devins et al., "The law and big data," *Cornell Journal of Law and Public Policy*, vol. 27, no. 2, March 2017, pp. 357-413,

[17] "Civil rights principles for the era of big data," https://www.law.berkeley.edu/files/Civil_Rights_Principles_for_the_Era_of_Big_Data_FINAL(1).pdf

[18] "Big data and its exclusions," September 2013, https://www.stanfordlawreview.org/online/privacy-and-big-data-big-data-and-its-exclusions/

[19] R. Vogl, *Research Handbooks on Big Data Law*. Edward Elgar Publishing, 2021.

[20] L. A. Dickinson and E. W. Berg (eds.), *Big Data and Armed Conflict: Legal Issues Above and Below the Armed Conflict Threshold*. Oxford University Press, 2023.

[21] I. G. Cohen et al. (eds.), *Big Data, Health Law, and Bioethics*. Cambridge University Press, 2018.

[22] M. Corrales, M. Fenwick, and N. Forgo (eds.), *New Technology, Big Data and the Law (Perspectives in Law, Business and Innovation)*. Springer, 2017.

[23] T. Hoeren and B. Kolany-Raiser, *Big Data in Context: Legal, Social and Technological Insights*. Springer, 2018.

[24] M. Burri (ed.), *Big Data and Global Trade Law*. Cambridge University Press, April 2024

[25] M. C. Compagnucci, *Big Data, Databases and "Ownership" Rights in the Cloud*. Springer, 2020.

[26] L. Thomas, *Thomas On Big Data: A Practical Guide To Global Privacy Laws*. LegalWorks, 2023.

[27] S. Faro and G. Peruginelli (eds.), *Knowledge of the Law in the Big Data Age*. Sage, 2019.

[28] N. Witzleb, M. Paterson, and J. Richardson (eds.), *Big Data, Political Campaigning and the Law: Democracy and Privacy in the Age of Micro-Targeting*. Routledge, 2019.

[29] A. Koksal, *Big Data and Competition Law: Market Power Assessment in the Data-Driven Economy*. Routledge, 2023.

[30] M. C. Compagnucci et al. (eds.), *The Law and Ethics of Data Sharing in Health Sciences*. Springer, 2024.

[31] C. Phull, *Big Data Law in Canada*. Smartblock Law, 2019

[32] A. G. Ferguson, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. New York University Press, 2017.

[33] A. Zwitter and O. J. Gstrein, *Handbook on the Politics and Governance of Big Data and Artificial Intelligence*. Edward Elgar Publishing, 2023.

ABOUT THE AUTHORS

Matthew N.O. Sadiku is a professor emeritus in the Department of Electrical and Computer Engineering at Prairie View A&M University, Prairie View, Texas. He is the author of several books and papers. His areas of research interest include computational electromagnetics, computer networks, and marriage counseling. He is a life fellow of IEEE.

Philip O. Adebo is an instructor at Texas Southern University. He completed his PhD in Electrical and Computer Engineering, Prairie View A&M University with emphasis on power systems. His research interests include power systems, renewable energy, microgrids, smart-grid systems, restructuring power system, and optimization of power systems.

Janet O. Sadiku holds bachelor degree in Nursing Science in 1980 at the University of Ife, now known as Obafemi Awolowo University, Nigeria and Master's degree from Juliana King University, Houston, TX in December 2022. She has worked as a nurse, educator, and church minister in Nigeria, United Kingdom, Canada, and United States. She is a co-author of some papers and books.