International Journals of Advanced Research in Computer Science and Software Engineering
ISSN: 2277-128X (Volume-8, Issue-12)

Research Article | **December 2018**

Big Data Ethics

Matthew N. O. Sadiku, Chandra M. M. Kotteti, and Sarhan M. Musa
Roy G. Perry College of Engineering, Prairie View A&M University, Prairie View, TX 77446, United States
Email: sadiku@ieee.org; ckotteti@student.pvamu.edu; smmusa@pvamu.edu

Abstract: ***Since the dawn of the Internet, the quantity of data has exponentially increased. Data have become so readily available that the battle to protect individual liberties seems increasingly challenging accepted social and ethical norms. It is time for academics and researchers to reflect on the ethical, legal, social, and political implications of big data. The ethics of big data requires urgent regulatory and policy responses. This paper provides a brief coverage of big data ethics.***

Key words: big data, big data ethics, data ethics, research ethics

I. INTRODUCTION

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.

A good starting point is examining the some qualities of data which are central to many ethical considerations. The idea of open data is of the idea that data should be freely available and should not be restricted in its use since its restriction can impede scientific discovery. Open data yields many benefits including promoting scientific discoveries, cost savings, greater government transparency, and public education. It can fuel improvements in healthcare quality and policies. However, open data raise several complex ethical concerns related to privacy, discrimination, erroneous research findings, and litigation [2].

Data sovereignty refers to a government's control over the data that is generated within a given country. This is important to big data ethics. Data may be regarded as a commodity. Once generated and collected, data can be stored indefinitely, passing into new management and new uses.

II. BIG DATA

Big data refers to huge amounts of data, such as that generated by social networks and mobile devices. The term big data is now used in almost any field, including healthcare, manufacturing, engineering, business, education, the physical sciences, and social sciences. Big data is shaping modern human existence, and its influence is growing with time. Technologies such as artificial intelligence, machine learning, cloud computing, and the Internet of things further increase the immense power of big data.

Big data is characterized by 5Vs [3]:

- Volume, referring to the huge amount of data generated every second;
- Variety, indicating that data is produced by different sources with different formats;
- Velocity, expressing the speed at which new data is being generated;
- Veracity, referring to the truthfulness of data, whether it comes from a reputable source;
- Value, which is the desired outcome for analyzing big data.

As data has grown in size and complexity, so too have the ethical questions associated with the big data. These questions become urgent as big data research focuses on sensitive aspects of human behavior, interaction, and health. However, understanding of the ethical implications of big data lags behind.

III. ETHICAL PRINCIPLES

Ethics governs a person's behavior. Sharing data is an expectation of the human participants involved and thus a key part of ethical research. Data ethics is of increasing relevance as the quantity of data increases. Ethical questions may be considered in terms of accuracy, humane treatment, informed participants, and the necessity and applicability of the work.

Companies and professional organizations build codes of conduct and ethic. For example, professional organizations such as the Association for Computer Machinery (ACM), the Institute of Electrical and Electronics Engineers (IEEE), and the British Computer Society, have developed codes of ethics. Such professional codes of ethics set the standard for acceptable behavior within a profession and provide a signal to those that interact with the relevant group as to what to expect of the group members.

Ethics helps us to frame our ideas about what is right or wrong using rational argument. It can be applied in determining whether the use of big data is ethical. Five major ethical themes are identified [4]: (1) informed consent is taken for participating in a study, (2) privacy, which is the non-disclosure of personal information to the public, (3) ownership refers to rights regarding the modification of data, along with benefiting from intellectual property, (4) epistemology and objectivity, a number of sources reveal a connection between the ethics and epistemology of big data, and (5) big data divides between those who have or lack the necessary resources to analyze large datasets. Companies must ensure that the data they own is not used against their own subjects without consent.

Perhaps the most important rule is similar to the Golden Rule: "Do unto the data of others as you would have them do unto yours." These principles are by no means exhaustive,

IV. ETHICAL USES OF DATA

Big data ethics are integrated into a range of disciplines, such as computing, statistics, and data sciences, and user behavior using codes of conduct. Examples of ethical uses of data transaction include [5]:

- *Statutory purposes*: All collection and use of personal data by the state should be completely transparent and covered by a formal license negotiated prior to any data collection.
- *Legal purposes*: Legal scholars have started to propose means of "due process" for big data. Even assuming that their use of data is entirely legal, companies that use data in ways that are inconsistent with their values risk alienating their base and damaging their reputation.
- *Social purposes*: All uses of individual data for social purposes should comply with the transparency principles.
- *Health*: Big data will be useful to sustain and improve health care. Various forms of health sensitive information are being easily created, stored, and accessed.
- *Crime*: For crime prevention an explicit set of general principles for the harvesting and use of persona data should be established and widely publicized.
- *Commerce*: Personal data used for commercial purposes belongs to the individual. Individuals have the right to decide how and where and if their personal data is used for commercial purposes. For example, targeted advertising services combine data from many sources to customize online advertising.
- *Research*: Personal data used for research purposes belongs to the individual. The burden of ethical use and sharing is placed on the researcher.

V. CHALLENGES

Although big data has many benefits, such as efficient use of resources, scientific advances, bringing personalized service, detection of fraud and abuse, and prevention of failure or accident, it presents a number of complex, ethical challenges. The growth of unregulated big data poses a threat to social and ethical values [6]. Humans place too much faith in technology and this has the potential for bad decisions. The risk that health information will be used inappropriately can never be fully eliminated, but it can be minimized. Patients whose data were used for research purposes have the right to initiate litigation. The lack of specificity in general ethics for big data suggests the need for big data codes of ethics. Such codes will come from multiple disciplines involving big data.

VI. CONCLUSION

Ethics has been a long time concern in human computer interaction. The rapid growth of big data necessitates that all parties involved in research should understand the legal, ethical, and cultural challenges facing us. With opportunities also come the responsibility to consider the ethics of our choices in the everyday practices.

Use of big data has increased and we should watch for ethics violations. Providing ethical training in the use of big data for students and practitioners is an important part of the solution. More information on big data ethics can be found in book in [7] and a relevant journal: *Journal of Information Ethics*.

REFERENCES

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

[2] S. Hoffman, "Citizen science: The law and ethics of public access to medical big data," *Berkeley Technology Journal*, vo. 30, no. 2, 2015, pp. 1741-1806.

[3] 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.

[4] B. D. Mittelstadt and L. Floridi, "The ethics of big data: Current and foreseeable issues in biomedical contexts," *Science and Engineering Ethics*, vol 22, no. 2, April 2016, pp. 303–341.

[5] "Big data ethics," *Wikipedia*, the free encyclopedia https://en.wikipedia.org/wiki/Big_data_ethics

[6] C. L. Jurkiewicz, "Big data, big concerns: Ethics in the digital age," *Public Integrity*,vol. 20, 2018.

[7] K. Davis, *Ethics of Big Data*. Sebastopol, CA: O'Reilly, 2012.

ABOUT AUTHORS

Matthew N.O. Sadiku (sadiku@iee.org) is a professor at Prairie View A&M University, Texas. He is the author of several books and papers. He is an IEEE fellow. His research interests include computational electromagnetics and computer networks.

Chandra M. M. Kotteti (ckotteti@student.pvamu.edu) is currently a doctoral student at Prairie View A&M University, Texas. His research interests include fake news detection using machine learning and deep learning, natural language processing, big data analytics, and wireless networks.

Sarhan M. Musa (smmusa@pvamu.edu) is a professor in the Department of Engineering Technology at Prairie View A&M University, Texas. He has been the director of Prairie View Networking Academy, Texas, since 2004. He is an LTD Sprint and Boeing Welliver Fellow.

