Data Science and Ethics

Uchenna Nzenwata*   Olugbohungbe Rotimi   Adetunji Oluwatofunmi   Adedeji Adegbenle
Johnson Hinmikaiye
Department of Computing and Engineering Sciences, Babcock University, Ogun State, Ilishan-Remo 121103, Nigeria

Abstract
Data Science today profoundly influences how business is done in fields as diverse as the life sciences, smart cities, and transportation. Ethics in this study is defined as shared values that has helped humanity differentiate right from wrong; and data science is narrated as an inter-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from many structured and unstructured data. Data ethics have really become a real issue since different organizations have different working operational standards and rules guiding resourceful data. Also, due to the spark increase in data, majority of the organizations do not understand the importance of data they produce regularly. Hence, the purpose of this study paved way to establish general acceptable and basic guidelines for the use of data amongst data scientists and data practitioners. To establish these guidelines, a systematic literature review approach was used to review some related literatures. Six (6) professional parastatals were examined, in which Four (4) of these professional parastatals were discovered to have similar guidelines with the basic guidelines established in this study. We scrutinized the existing similar guidelines and perceive the need to present it in our study. This was based on the discovery of similar guidelines from the sampled data science parastatals; the approximation of 67% similarity guidelines was observed which passes a threshold of what we termed to be an accepted guideline. The established guidelines can be used by organizations and data-related professions to develop explicit internal policies, procedures, guidelines and best practices for data ethics.
Keywords: Algorithms, Artificial intelligence, Blockchain, Data science, Ethics, Microtargeting.
DOI: 10.7176/CEIS/11-5-02
Publication date: September 30th 2020

1. Introduction
According to (Lee, 2019), "Data Science today profoundly influences how business is done in fields as diverse as the life sciences, smart cities, and transportation". As cogent as these fields have become, the dangers of data science without ethical considerations is as equally obvious; whether it be the protection of personally identifiable data, understood bias in automated decision-making, the illusion of free choice in psychographics, the social impacts of automation, or the apparent divorce of truth and trust in virtual communication.

Come to think of it, if ethics, herein this term paper, is defined as shared values that help humanity differentiate right from wrong, then the increasing digitalization of human activity shaped by the very definitions of how we evaluate the world around us using datasets should not be stressed without ethics. This means that data science shakes the perceptual foundations of values, community and equity. Therefore, our ever-increasing reliance on information technology has fundamentally transformed traditional concepts of privacy, fairness and representation, not to mention free choice, truth and trust. This is a high call for data ethics.

It should be a point to note that the awareness of data science and its algorithms have an increased and fundamental impact on society, which is vivid around the world. To our knowledge, according to (Garzcarek and Steuer, 2019), "the first data science application recognized to have a large impact on societal processes are election forecasts and polls on voting behavior". Hence, many countries have their diverse regulations on what is allowed to publish when in context of an upcoming or ongoing election. From the other definition of ethics, it is glaring that ethics majorly deals with human; how human considers what is right and wrong. The question of course is, "what has ethics got to do with data?". As we all know, data refers to any form of recorded information (Purtova, 2018).

1.1 Statement of Problem
Data ethics have really become a real issue since different organizations have different working operational standards and rules guiding resourceful data. Also, due to the spark increase in data, majority of the organizations do not understand the importance of data they produce regularly, neither do they envisage great insight from the use of data; but have gained orientation on data protection. Most of these organizations do not understand the basics of data protection other than mere avoidance of security breaches (Brown, 2016). This is a high call for generalizable ethics guiding data science that will span beyond the horizon of security.

1.2 Aim and Objectives
In this study, our soul aim is to establish a guideline that will be the best suit guidelines for the data scientists and

practitioners. This aim will be guided by specific objectives - to systematically identify, describe, review and categorize some of these ethical issues in data science. Also, we will identify some professional parastatals on data science and examine some of their data science ethical considerations against the established basic guidelines.

1.3 Methodology
To be able to achieve the set objectives, a thorough and systematic literature review method was employed. This enables the study to have an overview of what had been scripted on data ethics by some data professional parastatals. These reviews are then used to judge the codes of practices that will be suggested as well shaped norms and guidelines for ethical data practices in this study.

1.4 Significance of Study
Since there is no single, detailed code of data ethics that can be fitted to all data contexts and practitioners; this study can be used as a basic guide for organizations and data-related professions to develop explicit internal policies, procedures, guidelines and best practices for data ethics that are specifically adapted to their own activities.

2. Literature Review
While there is no single definition of data science, Data science, according to (Raja and Supriya, 2018), "is an inter-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from many structural and unstructured data".

Ethics on the other hand, encompasses the following dilemmas; how to live a good life, our rights and responsibilities, the language of right and wrong, moral decisions; in other words, ethics is seen as shared values that help humanity differentiate right from wrong. Whereas, (Richterich, 2018) definition of data ethics gave a summary of what it should be. He defined data ethics as a new branch of ethics that studies and evaluates moral problems related to data (including generation, recording, curation, processing, dissemination, sharing, and use), algorithms (including AI, artificial agents, machine learning, and robots), and corresponding practices (including responsible innovation, programming, hacking, and professional codes), in order to formulate and support morally good solutions (e.g. right conducts or right values). Could it be that data and the science of data existed before the advent of data ethics? It was a debated course on the part of the data scientists. It was obvious that data scientists and statisticians were making ends meet without the knowledge of ethics in data science, with the contribution of (Stark and Hoffman, 2019), they inspired data scientists in their work to participate in the debate on the impact that their professional work has on society, and to become active in public debates on the digital world as data science professionals. In this debate, questions of "how do ethical principles (e.g., fairness, justice, beneficence, and non-maleficence) relate to actual situations in our professional lives?" and "what lies in our responsibility as professionals by our expertise in the field?". More specifically, they made an appeal to statisticians that may consider themselves not as data scientists, nor what they do as data science, to join that debate, and to be part of the community that establishes data science as a proper profession ethics. This claim was not a deviation from what (Helbing, 2018) posited; data science is in the focal point of current societal development. Without becoming a profession with professional ethics, data science will fail in building trust in its interaction with and its significant contributions to society.

On the inspection of the importance of data ethics themed as the ethical impact of data science by (Vitak, Shilton, and Ashtorab, 2016), they emphasized the complexity of the ethical challenges posed by data science; and because of such complexity, they were of the opinion that data ethics should be developed from the start as a macroethics, that is, as an overall framework that avoids narrow, ad hoc approaches and addresses the ethical impact and implications of data science and its applications within a consistent, holistic and inclusive framework. Only as a macroethics will data ethics provide solutions that can maximize the value of data science for our societies, for all of us and for our environments.

The validity of data ethics cannot be over emphasized. Therefore, acknowledging its potency in the world driven by data is of great paramount. Also, with the existing ethics in computing and information, data ethics finds it seamless in implementation. In other words, data ethics can build on the foundation provided by computer and information ethics, which had focused for over the 30 years on the main challenges posed by digital technologies (Demy, Lucas, and Strawser, 2016; Confente, Siciliano, Gaudenzi and Eichoff, 2019). This rich legacy is most valuable. It also fruitfully grafts data ethics onto the great tradition of ethics more generally. At the same time, data ethics refines the approach endorsed so far in computer and information ethics, as it changes the level of abstraction (LoA) of ethical enquiries from an information-centric (LoA$_I$) to a data-centric one (LoA$_D$).

With the background and review to this study, the next section identifies some of the current ethical challenges as it affects data practitioners and statistics.

3 Ethical Issues in Data Science
Certainly, humanity has much to gain from the potential benefit from using data in areas such as health care; for

disease diagnosis, and work. However, the absence of a moral framework affects the picture of acceptable data use and organizational behavior. In 2019, the outcome was seen in multiple manifestations, including corporate reputation, such as Facebook's $5 billion data privacy fine and legislation, such as the California Consumer Privacy Act, which went into force January 1, 2020 (Determann and Gupta, 2019). Together these issues also call into question how organizations consider data and its ethical use. Some of these issues as enlisted by (Lee, 19) are identified in the sub-sections below.

3.1 Data as Citizens' personal property
Possibly no area of ethics in data science has received more attention today than the protection of personal data. The digital transformation of our interactions with social and economics communities reveals who we are, what we think, and what we do. In accordance with (Ball, et al., 2020), "the recent introduction of legislation in Europe, India, and California specifically recognize the rights of digital citizens, and implicitly address the dangers of the commercial use of personal and personally identifiable data". The attention given to this legislation extends far beyond concerns for data protection. As data become the new currency of the world economy, the lines between public and private, between individuals and society, and between the resource rich and the resource poor are being redrawn (Müller, 2020).

3.2 Automated decision-making
The ability to thoroughly take decisions among alternative possibilities has long been regarded as a condition that separates man the living from machines. In the ethics of Artificial Intelligence and Robotics, (Yawney, 2018), said this is furthered as innovations in data science progress in algorithmic trading, self-driving cars, and robotics, the distinction between human and artificial intelligence is becoming increasing difficult to distinguish. Current applications of machine learning cross the threshold of decision support systems and enter the field of artificial intelligence where sophisticated algorithms are designed to replace human decision-making. Beyond this logic introduces several ethical considerations. Are humans willing to accept that these applications, which by their very nature, learn from our experience, making us prisoners of our past and limiting our potential for growth and diversity? Do we really understand that the inherent logic of these applications can be modified, which creates opportunities to cheat the system? Last and but not least, who is legally responsible for the contained bias inherent in automated decision -making? These questions are worth bearing in mind ethically.

3.3 Illicit Microtargeting Application Activity
Microtargeting is a digital marketing technique that seeks to match ad campaigns (advertisement campaign using ad-ons) with the consumers or businesses most likely to be interested in your product or service. This is done using demographic, past purchasing and browsing history to pinpoint the best market for what you are selling. This tool is declared as ineffective marketing tool by (Yawney, 2018). Applications of microtargeting has since been pitched as powerful tools of influence in the fields of marketing, politics and economics (Andhoy, 2019). Moreover, micro-targeting techniques allow researchers extrapolate sensitive information and personal preferences of individuals even when such data is not specifically captured. Hence, as the customers become the target, there is a real danger that data science is used less to improve organizations' and/or social products or service offering than to turn consumers into objects of manipulation.

3.4 Distributed Ledger Technology/Block Chain Technology
The goal of Information technology has long been to provide a single version of the truth to facilitate exchanges of products, services and ideas. According to (Long, 2018), "distributed ledger technologies in general, and blockchain technologies in particular, offer their share of hope both for a more transparent and traceable source of information". Yet this vision of an Internet of Value is partially clouded by the potential societal challenges of relatively untested technologies. Can technology in and of itself be the standard for both truth and trust? To what degree will people and organizations accept the importance of transparency? On what basis can social values like the right to be forgotten be reconciled with the technical requirements of public ledgers? Freed from the conventions and the logic of financial capitalism, can human nature accept a radically different basis for the distribution of wealth?

3.5 Human VS Machine Intelligence
The impact of information technology on the organization of private and public enterprises has been largely debated over the last four decades (Healey and Wood, 2019). The impact of data science on the function of management has received considerably less attention. In the trade press, technology is often seen as ethically neutral, proving a digital mirror of the dominant managerial paradigms at any point in (Latour and Law, 2018). Being confident, I want to state clearly that in academia the relationship is subject to closer scrutiny. Which is in no deviation with the works of authors like Latour, Callon and Law, who have demonstrated how different forms

of technology influence the way that managers, employees and customers (Shadowen, 2019) perceive the reality of social and economic exchanges, markets and industries. When focusing on data science, these concerns for the context bring to light their own lot of ethical considerations. If Data is never objective, to what extent must management understand the context of how the data was collected, analyzed and transmitted? Similarly, as algorithms have become more prevalent and complex, to what extent do managers need to understand their assumptions and their limits? As applications assume larger and larger roles in key business processes, to what extent should management be defined around the coordination of human and software agents? Seen from another perspective, as artificial intelligence matures, which functions of management should be delegated to bots and which should be reserved for humanity? These are the fundamental issues that are questionable in organizations that are totally dependents on machines and less consideration on man. Hence, ethics should be defined.

3.6 Bias, Discrimination and Exclusion: Algorithms and Artificial Intelligence Challenges

Algorithms and artificial intelligence can create biases, discrimination or even exclusion towards individuals and groups of people. This issue is one where data science expertise is very important for understanding the extent of the problem. There is this one stressing point that is often overlooked, when algorithmic bias is discussed; the very nature of the most commonly applied algorithms, called pattern recognition or classification and clustering, if applied to humans, is applying bias. In statistics they form a prior belief on an individual generated by experience with other individuals assigned to the same group (Shadowen, 2019). In this case, questions of fairness and justice are touched by the use of these algorithms for judgement/prediction and decision making in general.

3.7 Quality, Quantity, Relevance: The challenges of data curated for AI

The acceptance of the existence of potential bias in datasets curated to train algorithms is of paramount importance. Even if implemented in best of mind, there may be unexpected bias in the training data. One famous example of algorithmic training going wrong was Microsoft's twitter bot Tay (Perez, 2016). Tay was implemented to act on Twitter as a regular user. The bot should learn from the comments by others how to perform common twitter conversations. In less than a day the humans had learned how to manipulate the learning algorithm in such a way that Tay started to speak out cruel and racist paroles. Microsoft decided to take Tay offline less than a day after it started learning. Another example for a similar event is an AI system at amazon. That system should help to find the most qualified applicants in their huge stream of applications. The experiment had to be stopped, when it was noted that the algorithm systematically downgraded applications of women. (Kodiyan, 2019) suggested probable causes for that behaviour are given. The training data contained mostly applications of men, so most of the successful applicants were men. There are not too many details, but as a consequence any appearance of the word woman reduced the chances of that applications.

4. Suggested Guidelines/Practices for Data Ethics

As we all know, there is no single, detailed code of data ethics that can be fitted to all data contexts and practitioners (Vallor, 2018); organizations and data-related professions should therefore be encouraged to develop explicit internal policies, procedures, guidelines and best practices for data ethics that are specifically adapted to their own activities (e.g., data science, machine learning, data security and storage, data privacy protection, medical and scientific research, etc.). However, there exist similar guidelines amongst other data science parastatals examined, which are identified to have 67% similarity shared contents, leading to specific codes of practice that can be well shaped to form the norms and guidelines for ethical data practice, and they are as listed below.

1. **Keep Data Ethics in Focus:** Data ethics is a pervasive aspect of data practice. Because of the immense social power of data, ethical issues are virtually always actively in play when we handle data. Hence, data practice must involve ethical considerations that are universal and central, not irregular and marginal, the individual and organizational efforts need to strive to keep ethics in the spotlight.

2. **Consider the Human Lives and Interests Behind the Data:** For the facts that data we handle are generated by non-human entities (for example, recordings of ocean temperatures), these data are being collected for important human purposes and interests. And much of the data under the 'big data' umbrella concern the most sensitive aspects of human lives: the condition of people's bodies, their finances, their social likes and dislikes, or their emotional and mental states. A decent human would never handle another person's body, money, or mental condition without due care; but it can be easy to forget that this is often what we are doing when we handle data.

3. **Focus on Downstream Risks and Uses of Data:** As a data practitioner, it is essential to think about what happens to or with the data later on, even after it leaves my hands. Even if, for example, I obtained explicit and informed consent to collect certain data from a subject, I should not ignore how that data might impact the subject, or others, down the road. If the data poses clear risks of harm if inappropriately used or disclosed, then I should be asking myself where that data might be five or ten years from now, in whose hands, for what purposes, and with what safeguards. I should also consider how long that data will remain

accurate and relevant, or how its sensitivity and vulnerability to abuse might increase in time. If I can't answer any of those questions or have not even asked them then I have not fully appreciated the ethical stakes of my current data practice.

4. **Envision the Data Ecosystem:** Not only is it important to keep in view where the data I handle today is going tomorrow, and for what purpose, I also need to keep in mind the full context in which it exists now. For example, if I am a university genetics researcher handling a large dataset of medical records, I might be inclined to focus narrowly on how I will collect and use the genetic data responsibly. But I also have to think about who else might have an interest in obtaining such data, and for different purposes than mine (for example, employers and insurance companies). I may have to think about the cultural and media context in which I'm collecting the data, which might embody expectations, values, and priorities concerning the collection and use of personal genetic data that conflict with those of my academic research community. I may need to think about where the server or cloud storage company I am currently using to store the data is located, and what laws and standards for data security exist there. The point here is that my data practices are never isolated from a broader data ecosystem that includes powerful social forces and instabilities not under my control; it is essential that I consider my ethical practices and obligations in light of that bigger social picture.

5. **Mind the Gap Between Expectations and Reality:** When collecting or handling personal or otherwise sensitive data, it's essential that we keep in mind how the expectations of data subjects or other stakeholders may vary from reality. That is, we should always ensure that the expectations of the researcher or a data practitioner concerning subjects' data should be demystified and made known to the subjects. There should be no room for expectation discrepancy since the data practitioners know a lot than the data subjects. Therefore, Agreements with data subjects who are 'in the dark' or subject to illusions about the nature of the data agreement should in general, be ethically legitimate.

6. **Treat Data as a Conditional Good**: Some of the most dangerous data practices involve treating data as unconditionally good. Data should be collected only as much of it as we need, when we need it, store it carefully for only as long as we need it, and purge it when we no longer need it. The second dangerous practice that treats data as an unconditional good is the flawed policy that more data is always better, regardless of data quality or the reliability of the source. Data are a conditional good, only as beneficial and useful as we take the care to make them.

7. **Avoid Dangerous Hype and Myths around 'Big Data'**: Data is powerful, but it isn't magic, and it isn't a silver bullet for complex social problems. There are, however, significant industry and media incentives to portray 'big data' as exactly that. This can lead to many harms, including unrealized hopes and expectations that can easily lead to consumer, client, and media backlash. That is, not all problems have a big data solution, and we may overlook more economical and practical solutions if we believe otherwise. We shouldn't ignore problems that might require other kinds of solutions, or employ inappropriate solutions, just because we are in the thrall of 'big data' hype.

8. **Establish Chains of Ethical Responsibility and Accountability:** It should be clear who is responsible for each aspect of ethical risk management and prevention of harm, in each of the relevant areas of risk-laden activity (data collection, use, security, analysis, disclosure, etc.) It should also be clear who is ultimately accountable for ensuring an ethically executed project or practice. Who will be expected to provide answers, explanations, and remedies if there is a failure of ethics or significant harm caused by the team's work? The essential function of chains of responsibility and accountability is to assure that members of a data-driven project or organization take explicit ownership of the work's ethical significance.

9. **Practice Data Disaster Planning and Crisis Response:** Most people don't want to anticipate failure, disaster, or crisis; they want to focus on the positive potential of a project. While this is understandable, the dangers of this attitude are well known, and have often caused failure, disaster, or crisis that could easily have been avoided. Data practitioners must begin to develop the practice of data disaster planning. Known failures should be carefully analyzed and discussed ('post-mortems') and results projected into the future. 'Pre-mortems' (imagining together how a current project could fail or produce a crisis, so that we can design to prevent that outcome) can be a great data practice.

10. **Promote Values of Transparency, Autonomy, and Trustworthiness:** The most important thing to preserve a healthy relationship between data practitioners and the public is for data practitioners to understand the importance of transparency, autonomy, and trustworthiness to that relationship. Hiding a risk or a problem behind legal language, disempowering users or data subjects, and betraying public trust are almost never good strategies in the long run. Clear and understandable data collection, use, and privacy policies, when those policies give users and data subjects actionable information and encourage them to use it, help to promote these values. Although, we can't always be completely transparent about everything we do with data: company interests, intellectual property rights, and privacy concerns of other

parties often require that we balance transparency with other legitimate goods and interests.

11. **Design for Privacy and Security:** This might seem like an obvious one, but nevertheless its importance can't be overemphasized. 'Design' here means not only technical design (of databases, algorithms, or apps), but also social and organizational design (of groups, policies, procedures, incentives, resource allocations, and techniques) that promote data privacy and data security objectives. How this is best done in each context will vary, but the essential thing is that along with other project goals, the values of data privacy and security remain at the forefront of project design, planning, execution, and oversight, and are never treated as marginal, external, or 'after-the-fact' concerns.

12. **Make Ethical Reflection & Practice Standard, Pervasive, Iterative, and Rewarding:** Ethical reflection and practice, is an essential and central part of professional excellence in data-driven applications and fields. Yet it is still in the process of being fully integrated into every data environment (Popovič, Hackney, Tassabehji, and Castelli, 2018). The work of making ethical reflection and practice standard and pervasive, that is, accepted as a necessary, constant, and central component of every data practice, must continue to be carried out through active measures taken by individual data practitioners and organizations alike. Ethical reflection and practice in data environments must also, to be effective, be instituted in iterative ways. That is, because data practice is so increasingly complex in its interactions with society, we must treat data ethics as an active and unending learning cycle in which we continually observe the outcomes of our data practice, learn from our mistakes, gather more information, acquire further ethical expertise, and then update and improve our ethical practice accordingly. Most of all, ethical practice in data environments must be made rewarding: team, project, and institutional/company incentives must be well aligned with the ethical best practices described above, so that those practices are reinforced and so that data practitioners are empowered and given the necessary resources to carry them out.

5. Outcomes and Discussion

Having identified the data science ethical issues upon which the above section guidelines were proposed, it should be established here that these guidelines have been adopted by some professional data science parastatals, which is an indication that these guidelines are not out of place in their own original statements.

This is evident when reviewed alongside with American Statistical Association (ASA) (American Statistical Association, 2018). The American Statistical Association (ASA) guidelines have eight sections, six of these sections are focused on individuals and groups of people to which the statistical work may matter. Hence, The American Statistical Association did not consider any of the data science ethical guidelines established in this study since it deals with constructive guidelines as related to human and the public for statistical research by considering data as entity that should not be worried about.

The Association of Computing Machinery (ACM) Code of ethics (Gotterbarn, 2018) has a preamble, and four sections. On a general level, the code addresses few of the ethical issues that is presented in this study. Yet, the Code is not a code for data science, and it is not providing the constructive guidance on the integrity of data.

The German Informatics Society (GIS) has a long history of its ethical guidelines (Trinitis, Class and Ullrich, 2019). These guidelines are concise and consist of a preamble and 13 very short sections. There are no data science specific sections in these guidelines, nevertheless many important aspects of the discussed ethical issues in this study are touched. Therefore, this study considers that the structure of the ethical guidelines of the German Informatics Society (GIS) can be a good tool to improve ethical guidelines for data science, which in variant case is similar to the guidelines discussed in this study.

The Accenture Data Ethics Research Initiative (ADRI), has a set of twelve principles proposed in (Accenture, 2014), which provide a baseline for those seeking such guidance and those looking to develop a group-specific code of data ethics. These guidelines is the most informed on the ethical issues discussed in this study. What does that imply? It should be noted here that what is being discussed in this document can be adopted as standard for the general guidelines on Data Science ethics. Also, the Royal Statistical Society (RSS) Data Science Section and the Institute and Faculty of Actuaries (IFoA) posited data science ethical guidelines for members working in the area of data science (Perkins, Davis and du Preez, 2020). It is intended to complement existing ethical and professional guidance and is aimed at addressing the ethical and professional challenges of working in a data science setting as identified in this study. They tried to place the guidelines on a scale of five themes, which encompass the major guidelines provided in this study. Here, we can compliantly state that the guidelines presented in this study is not out of place with the ideas presented in Royal Statistical Society (RSS) and the Institute and Faculty of Actuaries (IFoA) guidelines.

The resulting ethical guidelines established in this study found a great application in the Data Science Association (DSA) by improving the data science profession, eliminating bias and enhancing diversity, and advancing ethical data science throughout the world (Trepanier, Shiri and Samek, 2019). This study shares similar ethical guidelines with Data Science Association. At this point, it suffices place the established guidelines herein,

in this study as an adaptable guideline that serve as basis for data science activities though vary amongst different points of view.

6. Conclusion

The views shared in this document were gathered from different sources on ethics in data science. It is obvious that not every organization incorporates ethics in the use of data. Also, the understanding of data ethics is very much myopic as professionals attempt to limit data ethics as the principle of seeking permissions from data subjects. From the few comparisons and exploration done amongst other literatures, basic legal guidelines were extrapolated to be set as standard that measure the generality of other guidelines set by the examined professional parastatals.

Fundamentally, there is an understanding that the morality of the data science community is evolving and that it is a shared task to develop it, which in turn needs open discussions. As such, after the examination of six (6) selected professional parastatals and their ethical guidelines, four (4) of these professional parastatals were discovered to have similar data science ethical guidelines with those established in this study. Therefore, it implies that the guidelines set in this study can be adopted as basics to data science ethical guideline.

References

Accenture. (2014). Universal principles of data ethics: 12 guidelines for developing ethics codes. 12. https://www.accenture.com/t20160629T012639Z__w__/us-en/_acnmedia/PDF-24/Accenture-Universal-Principles-Data-Ethics.pdf

American Statistical Association. (2018). Ethical Guidelines for Statistical Practice, Approved April 2018.

Andhov, A. (2019). Corporations on Blockchain: Opportunities & Challenges. *Cornell International Law Journal, Forthcoming*.

Ball, K., Canhoto, A., Daniel, E., Dibb, S., Meadows, M., & Spiller, K. (2020). Organizational tensions arising from mandatory data exchange between the private and public sector: The case of financial services. *Technological Forecasting and Social Change*, *155*, 119996.

Brown, I. (2016). The economics of privacy, data protection and surveillance. In *Handbook on the Economics of the Internet*. Edward Elgar Publishing.

Confente, I., Siciliano, G. G., Gaudenzi, B., & Eickhoff, M. (2019). Effects of data breaches from user-generated content: A corporate reputation analysis. *European Management Journal*, *37*(4), 492-504.

Demy, T. J., Lucas Jr, G. R., & Strawser, B. J. (Eds.). (2016). *Military Ethics and Emerging Technologies*. Routledge.

Determann, L., & Gupta, C. (2019). India's Personal Data Protection Act, 2018: Comparison with the General Data Protection Regulation and the California Consumer Privacy Act of 2018. *Berkeley J. Int'l L.*, *37*, 481.

Garzcarek, U., & Steuer, D. (2019). Approaching Ethical Guidelines for Data Scientists. *Studies in Classification, Data Analysis, and Knowledge Organization*, 151–169. https://doi.org/10.1007/978-3-030-25147-5_10

Gotterbarn, D. W., Brinkman, B., Flick, C., Kirkpatrick, M. S., Miller, K., Vazansky, K., & Wolf, M. J. (2018). ACM Code of Ethics and Professional Conduct.

Healey, K., & Woods Jr, R. H. (2019). *Ethics and Religion in the Age of Social Media: Digital Proverbs for Responsible Citizens*. Routledge.

Helbing, D. (2019). Societal, economic, ethical and legal challenges of the digital revolution: From big data to deep learning, artificial intelligence, and manipulative technologies. In *Towards Digital Enlightenment* (pp. 47-72). Springer, Cham.

Kodiyan, A. A. (2019). An overview of ethical issues in using AI systems in hiring with a case study of Amazon's AI based hiring tool.

Latour, B., & Law, J. (2018). Social complexity, actor networks and the SPACE framework.

Lee, S. (2019). The Ethics of Data Science. Retrieved from https://www.baieurope.com/blog/the-ethics-of-data-science

Lee, S. (2019). The Ethics of Data Science. Retrieved from https://towardsdatascience.com/the-ethics-of-data-science-e3b1828affa2

Long, R. J. (2018). *New office information technology: Human and managerial implications* (Vol. 30). Routledge.

Müller, V. C. (2020). Ethics of artificial intelligence and robotics.

Perkins, S., Davis, H., & du Preez, V. (2020). Practical Data Science for Actuarial Tasks.

Perez, S. (2016). Microsoft silences its new AI bot Tay, after Twitter users teach it racism. *Tech Crunch*.

Popovič, A., Hackney, R., Tassabehji, R., & Castelli, M. (2018). The impact of big data analytics on firms' high value business performance. *Information Systems Frontiers*, *20*(2), 209-222.

Purtova, N. (2018). The law of everything. Broad concept of personal data and future of EU data protection law. *Law, Innovation and Technology*, *10*(1), 40-81.

Raja, M. S., & Supriya, R. (2018). Big Data Analytics and Data Science–A Review on Tools and Techniques. *International Journal of Advanced Research in Computer Science*, *9*(Special Issue 1), 107.

Richterich, A. (2018). *The Big Data Agenda. Data Ethics and Critical Data Studies*. University of Westminster Press.

Shadowen, N. (2019). Ethics and bias in machine learning: A technical study of what makes us "good". In *The Transhumanism Handbook* (pp. 247-261). Springer, Cham.

Stark, L., & Hoffmann, A. L. (2019). Data is the new what? Popular metaphors & professional ethics in emerging data culture.

Trepanier, C., Shiri, A., & Samek, T. (2019). An examination of IFLA and Data Science

Association ethical codes. *IFLA journal*, *45*(4), 289-301.

Trinitis, C., Class, C. B., & Ullrich, S. (2019). The german informatics society's new ethical guidelines: POSTER. In *Proceedings of the 16th ACM International Conference on Computing Frontiers* (pp. 302-304).

Vallor, S. (2018). An Introduction to Data Ethics. *Course module.) Santa Clara, CA: Markkula Center for Applied Ethics*.

Vitak, J., Shilton, K., & Ashktorab, Z. (2016). Beyond the Belmont principles: Ethical challenges, practices, and beliefs in the online data research community. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 941-953).

Yawney, L. (2018). *Understanding the "micro" in micro-targeting: an analysis of the 2018 Ontario provincial election* (Doctoral dissertation).