Big Data Seeks Value, Make it Ethical

Elie Chamoun
Computer Science department
American University of Culture and Education AUCE
Beirut, Lebanon
elie.chamoun@ul.edu.lb

Nour Charara
Computer Science department
American University of Culture and Education AUCE
Beirut, Lebanon
nourcharara@ul.edu.lb

Abstract— **Over decades' data sets became the lifeblood of business or the new oil of the economy and driving the change. Accordingly, nowadays data has become more critical and known as Big Data. Compared to traditional datasets (formatted with titled fields), Big Data typically includes unstructured masses of data that require real-time analysis. In addition, Big Data originates new opportunities for discovering new experiences and values, helps us gain a deeper understanding of hidden values, incurs new challenges and discover answers on how to organize and effectively manage values for these massive and complex datasets. Therefore, owners of industries and business men have become interested in the potential of Big Data; tackle challenges by leveraging the power of advanced technologies (Artificial intelligence, Healthcare, Smart City, Banking and Finance, Oil & Gas, Data Mining. Internet of Things) engaging in an ethical debate (Specialists, Individuals) to build a better world.**

Ethical debates are typically articulated within the context of ethical theories. An ethical theory is framed and reviewed in this paper. This theory, named SAS Theory, categorized the Value (ethical wise and not as money revenue) in Big Data to three functions. (Sustain, Align, Support). These functions facilitate and stimulate the ethical behavior of Big Data to turn Values into Actions. Furthermore, this theory outlines the ethical behavior of the user personas and organizations offline and online. It also examines and shows how it might be engaged to conserve the high ethical service in some industries and technologies. Finally, this theory presents three functions that group the technologies in categorized properties using a specific parameter which is the Value Personas. Categorizing means to assign a specific ethical solution for a given technology or business.

Keywords — Big Data, Ethics, SAS Theory, Value, Value Personas.

I. INTRODUCTION

Big Data is known as extreme large data sets. The features of Big Data have been classified according to five fundamental elements (5Vs), which are Volume (size of data), Variety (different types of data from several sources), Velocity (data collected in real time), Veracity (uncertainty of data) and Value (benefits to various industrial and academic fields). [1]

Clark [2] indicate that 3 elements have been added to the basic 5 elements: Visualization (interpretation of data and identification of the most relevant information for the users), Viscosity (latency data transmission between the source and destination), and Variability (context of data).

Other researchers, like Manogaran [3], introduce additional characteristics beyond the 8Vs model, such as Validity (correct processing of the data) and Virality (speed of the data sent and received from various sources) with a total of 10Vs.

IBM (International Business Machines) data scientists break Big Data into four dimensions: Volume, Variety, Velocity and Veracity [4]. As remarked, the element Value has been removed from the five original dimensions. Subsequently, Big Data is unethical without a Code of Ethics like Value and, therefore, there would be no need to argue about the right of Privacy, Security and Policies.

The frequently cited features of "volume, speed, and variety" are useful benchmarks - persistent features such as the size of the datasets, the speed at which they can be acquired and queried, and the wide range of file formats and types that generate data.

The Value element, in Big Data, is important enough to pose practical rather than theoretical problems in computer ethics. Hence, privacy breaches arise in some actions taken by businesses as an outcome of big data analytics and lead to embarrassment and even lost jobs of those involved.

The thin line dividing between new data and old data in Big Data is related to the development of technology. New developments in this space make old privacy issues and other ethical issues much more pressing. Comparing the cell phone ten years ago and now, it can record conversation, take photos and videos and stream directly online.

Moreover, Big Data is «Ethically Neutral» [5]. This means that Big Data does not include an integrated perspective on what is good data or bad data upon its creation, or where is the good or bad usage of this data without analyzing it.

However, Big Data generates a "forcing function" in our lives by its size and speed. Today, millions of people want to share the same information (flash news for example) with each other. This is a direct example of how the "forcing function" [1] of Big Data literally influences our lives.

Influence is a two-way street, just like the scientific principle, we cannot observe a system without changing it. Big Data cannot be used without impact. This impact is where ethical issues live. Big Data can amplify our values,

[1] "A forcing function" is an aspect of a design that prevents the user from taking an action without consciously considering information relevant to that action." It forces conscious attention upon something ("bringing to consciousness") and thus deliberately disrupts the efficient or automatized performance of a task. [6]

making them much more powerful and influential, especially when they are collected and focused on a specific desired outcome (Commenting and disliking a bullying act found on the net).

As the forcing function of Big Data pushes data in our organizations and in individual lives, the balance between risk and innovation will continue to be an urgent need to meet and maintain the ability of Big Data to generate benefits from values rather than prejudices.

II. BALANCING BIG DATA AND ETHICS

A. Big Data tends to be a wide-ranging category

One approach is balance between risk and innovation. An article in CBS News regarding innovation titled "Predictive Policing" program that can actually predict where crimes will happen. The crime prediction boxes come from the same kind of mathematical calculation used to predict earthquakes and aftershocks. Los Angeles Police Chief Charlie Beck says «The real measure of this is not how many people you catch it is how much crime you prevent». [7]

Massive data represents tremendous opportunities for the benefit of business, education, health care, government sector, manufacturing, and many other areas. The risk, however, for privacy, is the ability to manage our reputation and online identity. What it might mean to lose or gain ownership of our personal data is just the beginning of the ethical issues. To take advantage of the benefits of Big Data innovations, we need to understand the practical risks of implementing them.

Another approach is the anonymization of data sets before they are published, targeted advertising, and so on. As the lawyer Paul Ohm [8] points out, «the data can be useful or perfectly anonymous, but never both». So, let's suppose we know things about a person in particular: where did he eat, what did he eat? It is very unlikely that we will end up violating his privacy by broadcasting "the information" that a particular person likes Pizza House and Pizza. Suppose we have this information for about 10 million of peoples, patterns emerge that do make it possible to tie data points to particular named, "located individuals".

Big Data tends to be a wide ranging technological category like Artificial Intelligence, Data mining, globalization, medicine and others. The task is what if we can categorize these technologies using functions and parameters. Where would Big Data stand regarding ethics?

B. Ethics and Big Data

As mentioned before, Big Data itself, like any technology, is ethically neutral. The use of Big Data is not the case [9]. Although the ethics involved are abstract concepts, they can have very real implications. The goal is to develop better ways to engage in an intentional ethical inquiry to inform and align our actions with our values. How could this alignment be reflected?

There is a significant amount of effort to create a digital «Bill of Rights» [10] for the acceptable use of Big Data. The White House recently released a bill on the protection of consumer rights. The values which that bill supports include transparency, security and accountability.

On a technological level, businesses need to work in ethical frameworks - such as the IEEE P7001 - Transparency in Autonomous Systems – [11], but what about data? More machines learn to go solo. Does one have to make sure that the inner uploaded data, with which they operate, are clean and impartial? For example, could a smart computer run the defense for a given country?

The challenge is how to honor these values in our daily actions decision.

C. The ethical decision points

Davis and Patterson define 'Ethical decision points consist of a series of four activities that form a continuous loop' [12].

a) Review: the discovery and discussion in fundamental organization values is in demand. An understanding of what values actually are not what it is estimated or other thinks. No jump to solutions without first identifying the ethical issue(s) in the situation. Example: how to value transparency in the use of Big Data.

b) Analysis: review of current ethical situation and gather the facts, actual data-handling practices and an assessment of how well these facts align with core organizational values. The exploration of whether a particular use of Big Data technology aligns with the values that have been identified Example: Should they create this new product feature using Big Data?

c) Express: a simple, clear written expressions of where, when and how identified values and actions align—and where they don't—using a common vocabulary for discussing. What specific virtues are relevant in the situation? Example: This new product feature that uses big-data technology supports value of transparency?

d) Action: tactical plans to close alignment gaps that have been identified and to encourage and educate on how to maintain that alignment as conditions change over time Example: If we build this new product feature, we must explicitly share (in a transparent way) with our customers and ourselves on how that feature will use personal data.

These four activities are validated by three functions where ethics can converge with Big Data.

III. CONVERGING BIG DATA WITH ETHICS

A discussion of ethics and Big Data depends on how people define ethics. In general, ethics involves analyzing the conduct that may be beneficial or harmful to other people [13]. However, ethics is a subject that has been studied for at least 2,400 years, since then there have been a number of formulations of ethical principles (like Hippocrates 460 BCE "revered for his ethical standards in medical practice" [14]). Solid ethical theories share common property. They

allow the individual to convince using logical and reasoned arguments based on the principles defined by ethical theory. To illustrate this, the ethical theories will outline the ethical behavior of the user personas and organizations offline and online to conserve the best ethical service. Simultaneously, this theory helps to frame our understanding for values and moral issues than examined and show how it might be engaged to fix Big Data values in some industries and technologies.

While the nine elements mentioned earlier are related directly to data (Volume, Variety …etc.), Value(s) «are things that people care about» [15]. Thus Values are connected to fundamental human and personal aspects.

A. The Value Personas:

Patterson defines the Value Personas [16] as an «evolution of traditional user personas» that express how a specific value shows up and influences action within an organization or even society. Value Personas shed light on moments when the use of Big-Data technologies raises an ethical (or value-focused) decision point. A Value Personas can suggest options for how to align shared values with proposed action from various organizational role perspectives.

B. The Three Ethical Functions in Big Data:

Valuable personas provide a means to frame an explicit ethical inquiry and are very flexible and modifiable in many given ways. How? The Valuable personas play the parameter in functions presented in "Values". In this paper, the Valuable personas deliver an ethical behavior which appears in the three functions that control Big Data. This leads us to the SAS (Sustain, Align, and Support) theory:

a) Sustain: this function is generated when ethics are set upon creation of big data, where the ethical decision should be taken instantly.

- *Challenge*- Our ethical values are inherent in our actions all the time. The 'Sustain Function' shows up in online technology while surfing the net and using social media, or in some technologies like Artificial intelligence and robotics.
- *Solution*- Values can't take action; persons do. Valuable Personas can help analyze the conflicts between what you value and how you should act based on these values. The behavioral response should be spontaneous to take action. That means only when taking action online, for example, the ethical and non-ethical value shows up. The advantage of Value Personas is identifying which values are showing up in actions and how.

The Value Personas can provide a mechanism for developing a common vocabulary based on and inspired by the own personal moral codes, and aims at developing a set of common and shared values, which help to reduce barriers on line and could inspire collaboration to productivity and innovation.

b) Align: Values can be aligned intensely with Big Data. This function is applicable on technology and organization more than individuals.

- *Challenge*- While it is true that an organization may hold conflicting values; this can lead to contradictory actions. This conflicting issue could rise in some technology, like medicine. Ethics for this career have been generated before the rise of Big Data concept. For Example, changing the data-handling policy for an organization or product without notifying anyone means it is not acting in alignment with its values.
- *Solution*-The Value Personas can help make those conflicts transparent. Value Personas can help analyze the conflicts between what we value and how we should act based on those values.

c) Support: Values are not ethics. Ethics are derived from values. This function illustrates values as a "stone" holding Big Data.

- *Challenge*- The word "Ethics" is an expression of which action is valued and which action is not. Values measure whether an action is ethical or not.
- *Solution* –The Value Personas is the 'key' by which ethical alignment can be measured. In practice, the weight of this 'key' starts with ethical information literacy in education. Courses about computer ethics need to be included in education programs, workshops and conferences. In practice, ethics start in education. Additionally, public libraries can serve technical support for those who need answers regarding ethical issues for students and other attendees by offering an online service and seminars.

These functions are useful for understanding how technologies are classified in Big Data properly and the importance of the Value Personas as a parameter or 'Key'.

IV. APPLYING ETHICAL THEORIES FUNCTIONS TO BIG DATA

The usage of these ethical perspectives described above is useful for understanding how issues are revealed in Big Data ethics. However, it is possible to better understand how and why ethics help to shed light on a problem such as Big Data ethical concerns.

A. Role of Value Personas

As a key, the Value Personas help clear these values. Following organizational actions, Value Personas provide a way to facilitate discussion about organizational alignment in actions, business practices and individual behaviors based on common set of values. Value Personas provide a description of key roles, ethical decision points, alignment, actions and anticipated outcomes. Value Personas help identify shared values and create a vocabulary for explicit

dialogue, reducing the risk of misalignment (produced by some politics in an organization) and encouraging collaboration and innovation between the team works in an e-type (e-business, e-HR …) technology.

Value Personas may either evolve regularly based on changes in market conditions, new technologies, legislation, common practices, or evolutions in business model, or may stay essentially intact and unchanged for long periods. The dynamics of the conditions where business decisions are made are highly variable and subject to influences that are often difficult to expect. Value Personas can be expected to be used in the process of being able to be revised, updated, and adjusted.

B. Turn discussion into action

Technologists are constantly working to change the capabilities of Big Data. Social norms and legislation evolve more slowly. Competitive market forces, depending on the sector, can evolve at many different rates, instantaneously, quarterly or annually. There is no reason to hope that the ability to maintain alignment between values and actions can be fully expressed in advance of all commercial conditions. Indeed, one of the advantages of the innovation opportunities offered by Big Data is that it allows companies and organizations to adapt quickly to these market forces and competition.

The Google Code of ethics Conduct is an example. This code is one of the ways where Google put values into practice, with a cliché «Don't be evil». This code is built around the recognition that everything they do in connection with work at Google will be, and should be, measured against the highest possible standards of ethical business conduct [17]. In this code, Google expects from its employees and Board members to know and follow the Code of Ethics. Moreover, Google is ready to answer, thru the Ethics & Compliance Helpline, for any question from its employees regarding a concern of a suspected violation of the Code or any other Google policy. Finally, if any employee believes a violation of law has occurred he can always raise that to the Ethics & Compliance Helpline or with contact a government agency. This is an example on how organizations can “create and develop” a Value Personas.

The ability to align values with actions allows organizations to create a common and shared sense of action and purpose about any given business initiative. What matters, ethically speaking, is turning the question of "*should we do this*" into "*how can we do this*" and eventually free more thinking and collaborative work.

Value Personas, as a tool for developing this capacity, are inherently evolving. Comprehensive, multi-day workshops can give organizations enough time to develop an awareness regarding their values and articulate suggested actions at various ethical decision points. However, these methods and tools work equally well in conferences conversations or in an informal meeting where ethical issues suddenly arise and ethical discussions begin.

In general, organizations have a set of ideal core values to provide a starting point for the less formal use cases. However, even in an incident of disagreement over core values, Value Personas can generate a tool to start productive ethical conversations. This conversation could begin about A.I. that represents a totally different intelligence from ours as humans.

These conversations become productive when made transparent and explicit. Value Personas can help as a tool. The goal is clear, and divergent actions of plan are required. One must take into account each person’s role involved, and define exactly what each person will do, in what order, and highlight, predicts the results or objectives for each action. McEwan argues "A.I. becomes responsible for helping in their design and their generation” [18]. The real challenge is that these designs might have real problems understanding each other. An argument question the Value Personas have to deal with.

Therefore the products, features and services, should make the organization more beneficial for all its users with one guiding principle: Is what they’re offering useful in ensuring ethical action?

C. Standard ethical models an organization should earn

There are four critical Values that could establish standard models for Big Data in organizations:

a) Privacy Vs secrecy: Privacy doesn’t always mean secrecy. Ensuring data confidentiality means defining and enforcing information rules - not just rules for collecting data, but also for their use, maintaining and keep. Data owners must have the ability to manage the flow of their private information on massive third-party analytical systems

b) Although private information generated and shared it can still remain confidential: It is unrealistic to think about information as secret or shared, totally private or completely public. For many reasons, data (and metadata) are shared or generated by design with trusted services (e.g. address books, images, GPS, cell tower, and Wi-Fi localization of our cell phones). But it's not because we're talking about medical data, financial data, address book data, location data, reading data, or other things that we share the generated information.

c) Transparency in Big Data: Massive data is powerful when secondary uses of datasets produce new predictions and readings. Obviously, this leads to commercial data, with people like data brokers, collecting massive amounts of data about data owners (in general are clients), often without their knowledge or permission, and shared unexpectedly. For Big Data to work in ethical terms, data owners (people which create and process data) need to have a transparent view of how their data is used or sold

d) Big Data able to compromise identity. Privacy and secrecy protections are not enough anymore. Big Data analysis can compromise identity by allowing official organization and surveillance institutional

oversight and even determine who we are before we make our own opinion. Big Data owners need to start thinking about what kind of Big Data predictions and inferences they would allow and which ones they should not.

D. *Sighting the Qualitative.*

With the help of the social media technologies, the data created by users are regenerated by developers and analysts, reducing the human experience to limited set of quantitative variables. The capitalist modernity is the source that shapes the world to numbers with statistical analysis. Therefore, this paper suggested the needs for reflection on the qualitative nature of ethics in Big Data regarding quantitative heterogeneous arithmetic large datasets.

E. *Turning action to law*

There is a lot of work to be done to translate the S.A.S. theory with its three functions (Sustain, Align, and Support) into laws and rules that will lead to an ethical management of Big Data. Moreover, we must certainly develop more principles by building more powerful technological tools. Every person involve in Big Data management should engage in the ethical discussion of how Big Data is used. Database developers and administrators are at the forefront of the issue. The law is a powerful part of the Big Data ethic, but it is far from being able to handle the many use cases and nuanced scenarios that arise. Organizational principles, institutional ethics statements, self-monitoring and alternate forms of ethical guidelines are also needed. The technology itself can help provide an important element of the ethical mix.

How? This could be in form of an Intelligent Data Tracer (IDT) that would tell us how our data is being used and would allow us to decide whether or not we want our data to be used in the analysis that takes place beyond our spheres of awareness and control. We also need clear rules to determine (by default) what types of personal data processing are allowed and what types of decisions based on these data are acceptable especially when data affect people's lives.

But the important point is this; first we need a great data ethics rules second a Value Personas to be at the center of these critical ethical discussions and keep in mind Big Data ethics is for everyone.

V. CONCLUSION

The ethical contexts described in this paper examine the ethical behavior of Big Data to turn Values into Actions using functions (Sustain, Align, and Support) and parameters like Value Personas described in the SAS theory. What makes ethics actions so valuable is that it helps formulate arguments about what is right or wrong using logical and coherent opinions. This can help to evaluate and understand that the usage of Big Data is morally right.

Ethical feasible theories take all people, other than the decision maker, into consideration. It assumes that moral is good and moral principles are objective and based on reasoning facts and common maintained values. [19] The use of viable ethical theories helps to better describe our problems with Big Data based on a set of clearly articulated moral values.

In our daily lives, Big Data becomes a main force. It affects what others know about us and vice versa, as well as how we act because of the information they share with us. Besides our contribution, like it or not, the tools of this force like cameras, sensors surround us and we use it.

By examining this ethical theory, we can better recognize converging and diverging views on moral data, better understand the context and logic of the arguments presented and, in doing so, better assess how the future sequence of action is or should be justified.

This general conclusion sheds light on the use of Big Data. It also unlocks the way for finding ways to simplify ethical gaps. Big Data is here to stay, to value ethics with results facilitating advances in the given technologies, artificial intelligence, medicine, globalization, to name a few. The objective of this paper is to categorize these technologies in functions and assign the right Value Personas parameters. The positive ethical usage results provide the balance point that supports the use of Big Data. Moreover, the use of ethical theories, such as SAS, aids us better to recognize and manage how Big Bata affects in our lives.

REFERENCES

[1] J. W. B. S. S. J. Shilpa G. Kolte, "Big Data Summarization : Framework, Challenges and Possible Solutions," Advanced Computational Intellegence: An International Journal (ACII), vol. 3, no. 4, pp. 1-9, October 2016.

[2] D. Clark, "5 Things You Need to Know about Big Data," KD Nuggets, 03 March 2018. [Online]. Available: https://www.kdnuggets.com/2018/03/index.html. [Accessed 31 October 2019].

[3] Gunasekaran Manogaran et al., "Big Data Knowledge System in Healthcare," in Internet of Things and Big Data Technologies for Next Generation Healthcare, Springer International Publishing, 2017, pp. 133-157.

[4] "Infographics & Animations," IBM Big Data & Analytics Hub, 2018. [Online]. Available: https://www.ibmbigdatahub.com/infographic/four-vs-big-data. [Accessed 31 October 2019].

[5] K. D. Doug Patterson, "Big Data, Big Impact," in Ethics of Big Data, O'Reilly Media, Inc, 2012, p. 64 p..

[6] Bill Papantoniou, "The Glossary of Human Computer Interaction," Interaction Design Foundation, 2002. [Online]. Available:https://www.interaction-design.org/literature/book/the-glossary-of-human-computer-interaction/forcing-functions. [Accessed 30 Nvember 2019].

[7] B. Orr. LAPD computer program prevents crime by predicting it. April 11, 2012. https://www.cbsnews.com/news/lapd-computer-program-prevents-crime-by-predicting-it/. [Accessed 31 October 2019]

[8] P. Ohm Broken promises of privacy responding to the surprising failure of anonymization 2010 https://www.uclalawreview.org/pdf/57-6-3.pdf [Accessed 31 October 2019]

[9] D. Boyd and K. Crawford, "Critical Questions for Big Data: Provocations for a Cultural,Technological, and Scholarly Phenomenon.," Information, Communication, & Society, vol. 15, no. 5, pp. p. 662-679, 2012.

[10] D. Weitzner "We Can't Wait: Obama Administration Calls for

A Consumer Privacy Bill of Rights for the Digital Age," The White house president Black Obama, 23 February 2012. [Online].Available: https://obamawhitehouse.archives.gov/blog/2012/02/23/we-can-t-wait-obama-administration-calls-consumer-privacy-bill-rights-digital-age [Accessed 18 october 2019].

[11] IEEE Standards Association, "IEEE Announces Standards Development Project to Address Transparency of Autonomous Systems," IEEE Standards Association, 2016.

[12] K. Davis and D. Patterson, Ethics of Big Data, Beijing, Cambridge, Farnham, Koln, Sebastopol, Tokyo: O'Reilly, 2012, p. 22 p..

[13] D. Smith, "Five principles for research ethics : Cover your bases with these ethical strategies.," Monitor on Psychology, vol. 34, no. 1, p. 56, January 2003.

[14] W. D. Smith, "Encyclopedia Britanica," 2018. [Online]. Available: https://www.britannica.com/biography/Hippocrates#accordion-article-history. [Accessed 05 January 2019].

[15] Critical thinker academy, "What are Moral Values?," How to Build a Compelling Moral Argument, 2018. [Online]. Available: https://criticalthinkeracademy.com/courses/moral-arguments/lectures/659294. [Accessed 22 January 2019]..

[16] K. Davis and D. Patterson, Ethics of Big Data, Beijing, Cambridge, Farnham, Koln, Sebastopol, Tokyo: O'Reilly, 2012, p. 57 p..

[17] "Google Code of Conduct," Alphabet Ivestor Relations, 31 July 2018.[Online]. Available: https://abc.xyz/investor/other/google-code-of-conduct.html. [Accessed 26 October 2019].

[18] B. Walsh Ian McEwan on His New Novel and Ethics in the Age of A.I. May 20,2019 https://onezero.medium.com/ian-mcewan-on-his-new-novel-and-ethics-in-the-age-of-a-i-f30ec47bac72 [Accessed 31 October 2019]

[19] M. J. Quinn, Ethics for the Information Age, 7th edition ed., Pearson Education, 2016, p. 544 pages.