The Ethics of Big Data: Analytical Survey

L. Giber, N. Kazantsev
National Research University “Higher School of Economics” (HSE)
20, Myasnitskaya str., Moscow, Russia, 101000
e-mail: lbgiber@gmail.com

Abstract. The number of recent publications on the matter of ethical challenges of the implementation of Big Data has signified the growing interest to all the aspects of this issue. The proposed study specifically aims at analyzing ethical issues connected with Big Data.

Key words: Big Data, cloud computing, data security, ethics, legal regulation.

1. Introduction

An increasing amount of publications in the sphere of ethical challenges of big data signify a continuous increasing concern for this issue. Though there is a wide range of studies, there are still certain blank spots in this field to be found.

It should be admitted that with the rapidly increasing size and scope of information that big data technologies can provide businesses nowadays, maintaining an ethical framework may benefit from a clear understanding of vocabulary for discussing issues of coherent and consistent practices, as far as most fatal difficulties recently are connected with considerable confusion and uncertainty in the definitions of main concepts, their research structure, essential characteristics and components. All this indicates that this theme has not been sufficiently elaborated, particularly from a theoretical perspective.

Therefore, this paper firstly addresses current debates over data in general, and big data specifically, by examining the ethical issues arising from advances in knowledge production. Typically, ethical issues such as privacy and data protection will be discussed in the context of big data systems as a source and basis for the production of knowledge. Here we argue that this indeed is an evidence of more dramatic change of the society: there is a reason to believe that human autonomy is undermined by the advancement of scientific knowledge.

To make this argument, we first offer definitions of data and big data from various perspectives, basing on the critical review of existing scientific literature. Afterwards we analyze the reasons of current growth of data-driven analyses on the example of studies of human behavior. Next, we distinguish between the applied and scientific contexts in which big data research is used, and argue that this research has quite different implications. We conclude that the application of Big Data is both constrained and empowered by the nature of available data sources, nevertheless big data research will inevitably be-
come more pervading, and this will require awareness of the specialists in the area of data science, policymaking, moreover of a wider public about the contexts of Big Data analyses and its often unintended ethical consequences.

2. Problem statement

As theorists show continuous concern for the issue of the ethical challenges of applying Big Data with the rapidly increasing size and scope of information that big data technologies can provide businesses nowadays, an ethical framework has not been elaborated yet, particularly from a theoretical perspective.

There still remains a certain lack in a sort of universal ethical framework, which may result in unintended ethical consequences, such as loss of identity and personal data.

It becomes obvious that some further researches based on practical cases are needed to improve the current theoretical works, to allow for more accurate results of application of Big Data from ethical perspective.

It follows from what has been discussed before that, proposed study specifically aims at analyzing ethical issues connected with Big Data.

Therefore the specific tasks of proposed study will be the following:

1. To offer definitions of data and big data from various perspectives, basing on the critical review of existing scientific literature.

2. To address current debates over data in general, and big data specifically, by examining the ethical issues arising from advances in knowledge production.

3. To analyze the reasons of current growth of data-driven analyses on the example of studies of human behavior.

4. To distinguish between the applied and scientific contexts in which big data research is used and their implications.

5. To draw conclusions on ethical consequences of Big Data usage.

3. Literature Review

Many authors (see [1, 2]) believe that Big Data analyses raise major questions about a loss of human privacy and autonomy as a consequence of applying deterministic knowledge to human behavior. These questions focus on free will and human agency, as far as advancing knowledge discovery processes may seem to take these away from individuals.

A huge part of what we know about the world, especially about social and political phenomena studied in social sciences, is based on data analysis. It follows for instance from the papers “Political values,” in Comparative Politics: The Problem of Equivalence by Ronald Inglehart [3], “Quantitative analysis” in Approaches and Methodologies in the
Social Sciences: A Pluralist Perspective” by Mark Franklin [4], and “Causal Explanation” by Adrienne Heritier [5].

The fact that quantitative analysis of data is considered as the possible way to obtain a clear view seems promising. Quantitative methods, due to their basis of clear mathematical and statistical models, provide researchers with results of accurate, statistically verified numerical data, helpful for better understanding of studied phenomena and getting precise quantitative values of the studied parameters.

According to Gary King, Robert Keohane, and Sidney Verba [6], whether quantitative or qualitative study involve the dual goals of describing and explaining, both of which depend upon the rules of scientific inference. It is important that authors highlight the difference between interpretation and inference, which is a fair step of causal account and goes beyond interpretation, still there are multiple ways for acquiring the knowledge. Actually particular facts in close correlation with general knowledge generates good inference, which may help choose proper method for research.

Big data extends knowledge into new domains, it achieves greater accuracy in pinpointing individual behavior, and the capability of generating this knowledge may be carried out by new actors and more powerful tools.

It may be suggested that academic and applied research have mutual goal of producing knowledge based on large-scale data, though they differ in the ways that knowledge will be used: academics intent to generate information about human behavior, and those who are engaged in applied researches are interested in changing such behavior. It is the reason for us to suggest that, though these two kinds of researches overlap, the ethical implications of them differ significantly in the terms of privacy and data protection.

Main reason why it is still important to pay particular attention to their overlap is that both of them have their purpose in the longer-term at pansophy about human behavior, even if the applications of this knowledge remain analytically distinguishable.

Following part presents a review of the most significant theoretical approaches together with some researches in the field of implementation of Big Data that were used in analyzing and interpreting the results to be obtained. That data generally and big data more specifically are the subject of major contemporary debate in the society. An increasing amount of publications in the sphere of ethical challenges of Big Data signify a continuous increasing concern for this issue. Though there is a wide range of studies, there are still certain blank spots in this field to be found.

It should be admitted that one of the most fatal difficulties in any study of ethics of Big Data is a considerable confusion over the meanings of “Data” in general an “Big data” particularly: there is still some uncertainty in the definition of this concept and its research structure, essential characteristics and components. All this indicates It follows
from what has been said, despite many discussions of the implications of Big Data, neither data nor big data have been defined in the academic literature.

Such definitions are going to be provided further in this paper in order to explain why knowledge based on data-driven research is new and why the sources of data in this research are distinctive. Once we the groundwork is laid by establishing what is new and distinctive about data-driven knowledge, we will discuss its ethical implications. It may be argued that apart from current policy debates about privacy and data protection, data — driven research raises larger issues about the role of knowledge in society and about how knowledge can be used in relation to human behavior, with implications for how to create greater awareness among data scientists, policymakers, and a wider public in general.

The issues related to privacy and data protection are well-known nowadays and recently, there has been a shift from talking about privacy in general to discussions about data, and specifically digital data. In brief, privacy and data protection law is established in order to safeguard and guarantee individuality and autonomy in society. There are currently many debates about data (for example, the “right to be forgotten” in Europe, and the “White House review on BigData”) and privacy and data protection laws are spreading around the world (currently they are being adopted in more than 100 countries). Greenleaf [7] discusses that the efficiency of the data privacy principles is consequence of their ideological effect and global nature, and to the same extent of their enforcement, which is often lacking.

First we should draw our attention to the term “ethics”: this term may be seen is ambiguous, as far as the given by different people meanings to this term vary significantly, for instance, if one sees the sense of “ethics” in religion, other may see it in being moral or in following the laws. However, ethics cannot be confined to following laws, doing what so ever that is accepted in the society or even religion (though religion itself can set ethical standards). How can one define ethics? First and foremost, we may suggest that ethics refers to what is called “right and wrong”, and prescribe what people should and should not do, usually in terms of rights, morals, and compiles honesty, compassion and loyalty.

If seeing it from another perspective, we may also claim that ethics is the framework of ethical standards: their study, development, modification and implementation.

According to the Business dictionary, ethical standards “principles that when followed, promote values such as trust, good behavior, fairness, and/or kindness”, in its simplest form they include the right to life, the right to freedom from injury, and the right to privacy.

As it was mentioned in “Thinking ethically: A framework for moral decision making” [8] by Velasquez, M., Andre, C., Shanks, T. and Meyer, M. J., feelings, “laws, and
social norms can deviate from what is ethical. So it is necessary to constantly examine one's standards to ensure that they are reasonable and well-founded.”

It follows from what has been discussed that ethics means continuous studying of changing moral beliefs of the society we are living in, in order to ensure that we live under the standards that are favorable for lives.

Now when our society is rapidly changing, and even the interaction between two people can be held on-line, we should turn to the ethics of what is connected with ethics of IT. For instance, cloud computing which has become one of the fastest growing segments of the IT industry, and as increasingly amounts of information about individuals and companies are placed in the cloud, the concerns about how safe it is in sense of dealing with private data are beginning to grow. Consequently, security issues, ranked first as the greatest challenge issue of cloud computing, has played a major role in slowing down its acceptance.

Gartner 2008 discussed seven security issues that need to be addressed before businesses consider switching to the cloud computing, they were further discussed in “Cloud computing security issues and challenges” (see [9]). According to [9] they are as follows:

1) privileged user access;
2) regulatory compliance;
3) data location;
4) data segregation;
5) recovery;
6) investigative support;
7) long-term viability.

There are various security issues in the area of cloud computing such as data loss, phishing, botnet, which pose serious threats to data and software run in the Cloud. Several studies have been carried out recently relating to security issues in cloud computing, the results can be found in the following table.

There are several concepts that are closely correlated with the ethical framework of the implementations of Big Data, which have to be kept in mind. According to the book Ethics of Big Data by Kord Davis and Doug Patterson there is a framework of big data ethics for both individuals and organizations, which consists of four common elements: first, identity answering the question “What is the relationship between our offline identity and our online identity?”, next privacy answering the question of “Who should control access to data?”, third ownership — “Who owns data, has rights to transfer the data, and what are the obligations of people who generate and use that data?”, and finally reputation — “How can we determine what data is trustworthy?” [10].
Table 1.

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Publication</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENISA</td>
<td>2009</td>
<td>Cloud computing: benefits, risks and recommendations for information security</td>
<td>investigated different security risks related to adopting cloud computing along with the affected assets, the risks likelihood, impacts, and vulnerabilities in the cloud computing may lead to such risks</td>
</tr>
<tr>
<td>R. K. Balachandra, P. V. Ramakrishna and A. Rakshit</td>
<td>2009</td>
<td>Cloud Security Issues</td>
<td>discussed the security SLA’s specification and objectives related to data locations, segregation and data recovery</td>
</tr>
<tr>
<td>P. Kresimir and H. Zeljko</td>
<td>2010</td>
<td>Cloud computing security issues and challenges</td>
<td>discussed high level security concerns in the cloud computing model such as data integrity, payment and privacy of sensitive information</td>
</tr>
<tr>
<td>B. Grobauer, T. Walloschek and E. Stöcker</td>
<td>2010</td>
<td>Understanding Cloud Computing Vulnerabilities</td>
<td>discussed the security vulnerabilities existing in the cloud platform. The authors grouped the possible vulnerabilities into technology-related, cloud characteristics-related, security controls related</td>
</tr>
<tr>
<td>S. Subashini and V. Kavitha</td>
<td>2010</td>
<td>A survey on security issues in service delivery models of cloud computing</td>
<td>discussed the security challenges of the cloud service delivery model, focusing on the SaaS model</td>
</tr>
<tr>
<td>S. Ramgovind, M. M. Eloff and E. Smith</td>
<td>2010</td>
<td>The Management of Security in Cloud Computing</td>
<td>discussed the management of security in Cloud computing focusing on Gartner’s list on cloud security issues and the findings from the International Data Corporation enterprise</td>
</tr>
<tr>
<td>M. A. Morsy, J. Grundy and I. Müller</td>
<td>2010</td>
<td>An Analysis of the Cloud Computing Security Problem</td>
<td>investigated cloud computing problems from the cloud architecture, cloud offered characteristics, cloud stakeholders, and cloud service delivery models perspectives</td>
</tr>
<tr>
<td>Cloud Security Alliance (CSA) &amp; IEEE</td>
<td>2010</td>
<td>Cloud Security Alliance</td>
<td>indicated that enterprises across sectors are eager to adopt cloud computing but that security are needed both to accelerate cloud adoption on a wide scale and to respond to regulatory drivers; discussed that cloud computing is shaping the future of IT but the absence of a compliance environment is having dramatic impact on cloud computing growth</td>
</tr>
</tbody>
</table>

3.1. Defining Data and Big Data

This part presents a review of the most significant theoretical approaches together with some researches in the field of ethics of Big Data that were used in analyzing and interpreting the results to be obtained. Currently, the phenomenon of Big Data has be-
come a subject of major interest of specialists in various fields of science, consequently the public policy issues of Big Data becomes one urgent problem in this area. However, despite the diversity of works that reveal the basics of Big Data, there still seems to exist some uncertainty regarding the meaning of the term, its essential characteristics and components.

The interpretation of the concept of “Big Data” is far from being unambiguous and clear and the concept of Data is also multidimensional in nature, has many meanings. There are no universal definitions of data and big data, but since specifying what is novel about data-driven research is of major importance for understanding its implications, we will provide possible definitions here.

There is apparently no segment of activity in the world, which attracts as much attention recently as the knowledge management. According to Russell Ackoff, a professor of organizational change and systems theorist, the contents of the human mind can be arranged into five categories: data, information, knowledge and wisdom, they are closely correlated with each other; many people confuse data, information, and knowledge however they cannot be treated as synonyms. These concepts may be seen as the building blocks of information science, many theorists were elaborating on the links between these terms [11, 12], but first who put them into the single formula was Russell Lincoln Ackoff. In 1989 Russel Ackoff [13] posited a hierarchy at the top of which lays wisdom, and below understanding, knowledge, information and data.

Data in his terms are the product of observations, and are of no value until they are processed into a usable form to become information.

Information is contained in answers to “where”, “who”, “what”, and “when” questions.

Knowledge is the next layer, further refines information by making “possible the transformation of information into instructions. It makes control of a system possible” [13], and that enables one to make it work efficiently; answers “how” questions. “Understanding” for him connotes an ability to assess and correct for errors, appreciation of "why" question, while “wisdom” means evaluated understanding, an ability to see the long-term consequences of any act and evaluate them relative to the ideal of total control.

It is important to mention that Ackoff indicated that the first four categories they deal with what is known or what has been, they relate to the past. Only wisdom — fifth category deals with the future because it may be treated as a tool for creation of the future rather than just grasping ideas about the present and past.

The work of Russel Ackoff was critiqued and reviewed by many authors, though the paper by Bellinger, G., Castro, D., & Mills, A. [14] can be seen as most useful, as it is first brief and second it is really close to the original text. In the paper “Data, information, knowledge, and wisdom” its authors G. Bellinger, D. Castro D., and A. Mills [14] pro-
vide the definitions of terms, which were suggested by Russel Ackoff: first, data — raw symbols, that have no significance beyond their existence; can exist in any form; do not have meaning of themselves; next level is information, which represents “data that has been given meaning by way of relational connection”; thus information can be useful or not. Knowledge is higher level of proposed by Russel Ackoff hierarchy, which constitutes appropriate collection of information, intended to be useful; a deterministic process. Knowledge can be obtained via experience or education. Knowledge gained through experience is the most memorable, but it is also the most time consuming and costly to obtain, as experience refers to what we have done and what has happened to us in the past. Since time and money are limited resources, the amount of knowledge we can obtain from experience is also limited. Knowledge obtained through experience is also high-risk because we learn through our mistakes: we identify our errors and make corrections. Understanding, according to the “Data, information, knowledge, and wisdom” [14], is an interpolative and probabilistic process; cognitive and analytical; process through which one can synthesize novel knowledge from what has been previously held. Wisdom, the highest level of the hierarchy, may be characterized as “an extrapolative and non-deterministic, non-probabilistic process” [14]; may be treated as a uniquely human state, as it recalls to moral and ethics.

How can one define Big Data? First we may say that Big Data can be defined as research that implies an alteration and modification in the scale, intensity and scope of knowledge about the studied problem. Big data can be characterized as data sets, which sizes are beyond the ability of commonly applied software tools to analyze, process, gather,
er, store, and manage data within a tolerable amount of time [2]. META Group (Gartner) proposed to use the "3Vs" model for defining Big Data: thus, Big Data can be defined as “high-volume, high-velocity and high-variety information assets” [15]. Additionally, a new V- "Veracity" may be added to describe it.

Once again, it is emphasized in recent definitions that Big Data can require for specific “Technology and Analytical Methods for its transformation into Value” [16].

4. The ethics of big data: discussion

It can be admitted that quite different possibilities are attached to academic and commercial research. If those who are engaged in academic social research intend to generate and generalize information about human behavior and not to change it; those who are conducting applied researches (e.g. specialists in marketing) usually are interested in changing human behavior. Thus the usage of big data in the sector of applied researches stops being neutral in its nature.

It follows from what has been said, that there is significant difference in the usage of Big Data in academic and applied researches: in academic research big data is treated as a tool to generate abstract knowledge, without the need, or even will to use this knowledge to change behavior; from the other hand in applied researches the knowledge, which is generated on the basis of using Big Dara, can be used to change studied behavior.

It seems to me logical now to define the term “marketing”, to understand how Big Data is correlated with it.

Marketing is defined by the American Marketing Association as “the activity, set of institutions, and processes for creating, communicating, delivering, and exchanging offerings that have value for customers, clients, partners, and society at large.”

Now it is important to turn to the term “Social media marketing” and define it. The paper “Redefining social marketing with contemporary commercial marketing definitions” [17] defines social marketing as:

“adaptation and adoption of commercial marketing activities, institutions and processes as a means to induce behavioral change in a targeted audience on a temporary or permanent basis to achieve a social goal”.

With this definition, the connections between marketing and Big Data can be easily found: for instance, information can yield product insights allowing marketers to make suggestions for companies to create products people want, information may also help to understand how to efficiently set the value of those products, consequently the distribution (and production) strategies will be optimized in order to deliver the product to the consumer, and may help to determine the adequate rate of exchange (price) in order to assure a healthy profit.
To wrap it all up, the finding can be drawn on the account that the more information you have the more advanced marketing decisions one can make.

It follows from what has been discussed above that, there is an obvious potential between the concept and the process.

Now I am going to answer the question, how can Big Data be used to achieve the best strategic marketing objectives?

The term "big data" does not just refer to data itself; it refers to the data and to its challenges, capabilities and competencies linked to the storing and analyzing given huge data sets to sustain the level of decision making, which is more accurate and well-timed than anything was previously attempted — so called “Big-Data-driven decision making”.

Because Big Data holds the potential to describe target customers with an accuracy and level of detail, which was not even imagined even 10 years ago.

For example, marketers nowadays can gain data on people’s consumption habits, digital clicking behavior, time spent and activities (for instance sharings — reposts or re-tweets; comments; likes etc.) on various sites, purchasing history, personal preferences based on posts in social media sites, to be short almost every piece of information that can be imagined can be found my marketers (or social media marketers) nowadays on-line. It follows from what has been just discussed that it makes much sense that marketers nowadays are more and more turning to Big Data.

Now we are going to analyze and try to answer the question whether marketing can benefit from using Big Data and which particular benefits can be seen. By conjoining big data with the integrated marketing strategy, organizations may have numerous benefits.

It can be admitted that there is a growing demand for processing the existing big data to collect valuable information all over the world, and every and each business is recently also in demand for specific tools for processing and analyzing big data, in order to fulfill the needs of formulating marketing strategies and analysis of various industries.

The usage of Big Date in social media marketing researches may result in for example creation a more accurate profile of the target consumers; secondly it may help to forecast consumer reaction to product offerings and marketing messages; thirdly personalize those marketing messages and product offerings and optimize production and distribution strategy. Moreover, Big Data can be used in SMM in order to create and use more accurate assessment measures. Finally, retain more customers less expensively. Consequently, it may be suggested that the usage of Big Data may result in enhancement of the long-held SMM capabilities and give rise to a set of novel impressive capabilities.

As it was already mentioned before social media sites can be treated as a major source of Big Data for those involved in marketing researches and every piece of information can be easily aggregated, analyzed, and interpreted nowadays — what is of prime
interest for such areas as online or digital marketing, which collect and aggregate data, that consequently may be analyzed for patterns and insights.

After the insights are found, they can be integrated into any kind of digital marketing strategy: for simplest example, an insight that major part of followers of brand’s public page are involved in discussions of the posts with beautiful photos in them, may result in formulating advices for the PR department of the company to accompany all messages from Brand’s page with a photo or picture.

First, Big Data can help SMM in the process of more accurate identification of the target audience — probable byers.

In the past marketers used to make assumptions regarding socio-demographic characteristics of possible consumer of the goods: for instance about their age, gender, education, income etc.

One (SMM organization) may use all the information from social media to get more advanced and detailed information about the possible consumers: we can get all the information on socio-demographic profiles, and moreover we may get the data about which web sites consumers visit on a regular basis, which social media profiles they have and use etc. Moreover, Big Data provide the timeliest insights into who is interested or engaging with their product or content in real time.

Secondly, one can identify the drawbacks of company or brand through analyzing the data from social media.

It has to be mentioned here that the information on the quality of production can be published not only on the specific sites of recalls and comments (such as Otzovik for instance), more often this data consists of postings — almost every time emotional — on the quality of production or service.

In addition, analyzing this information may shed light on obvious minuses of the production of the company, and may result in the improvement and correcting deficiencies, consequently in the increase in sales due to the increased number of positive reviews.

Thirdly, one may answer the question whether the posts on the public page of brand (for instance, brands’ public pages in social media sites — Facebook, Vkontakte etc.) are “successful” or not.

We can evaluate and analyze the effectiveness of current smm strategies by analyzing and evaluating the involvement of users of social media sites in the posts of public page of brand: by analyzing the number of comments and sharings under the postings we may draw conclusions regarding the effectiveness of current marketing strategy in social media and formulate advices for improving it.

Finally, one can easily change human behavior: there is no secret that marketers nowadays more and more often resort to the help of so-called “agents of influence”, and
there are two types of them. First type — are regular bloggers (or users of Twitter, Facebook etc.) with biggest numbers of followers, that means the possible audience of their post — in other words, people who will see the posting — is big enough to be of prime interest for marketers (the required number of followers depends on the brand). Second type, those who can be called professional “agents of influence” — those who were specifically educated (by company or agency) to write specific types of postings.

The main aim of collaborating with agents of influence is to change consumer behavior of byers — through dissemination of positive reviews written by “agents of influence” marketers may increase the loyalty of consumers and increase the rates of sells.

It cannot be treated as an exhaustive list of ways to apply Big Data analysis in social media marketing and there is a far greater number of such implementations followed by benefits from them.

It has to be emphasized once again that by combining big data with an integrated marketing strategy, Social Media Marketing organizations can make a significant positive impact on customer engagement, customer retention and loyalty and marketing performance of the company.

Recently we may clearly see that data-driven knowledge is an advancing research area because of the availability of various novel digital data sources.

Nonetheless, it is important to bear in mind that there are also limits of this knowledge: there is a tendency to have very narrow aim in case of applied big data research, while academic big data knowledge aims at obtaining the most generalizable or broadest knowledge.

The process of generating such a powerful knowledge inalterably produces de-personalization, a more deterministic approach to the world. As Mayer-Schoenberger and Cukier (2013) point out that Big Data can “undermine the idea of personal responsibility, particularly as one of the cornerstones of the modern worldview is the idea of free will” [1].

It should be admitted that in this case they were referring to the question of law, but the issue is much wider than law, as far as big data research also challenges the notions of individuality and self-determination beyond the pale of the legal context. Similarly, the very idea of technological determinism — that one’s behavior may be not only prognosticated but also manipulated by a particular technology — goes against fundamental theories of how society functionates according to individual and collective decision-making. Besides, it may be added that even though deterministic knowledge of human behavior may seem threatening, for a set of certain social purposes, such knowledge will inevitably be needed — if we think, for example, about peoples’ energy consumption, but will it diminish the privacy of peoples’ data? It is the question that worth thinking about.
It follows from what has been discussed before, that nevertheless identifying new data sources highlights novel opportunities deriving from these sources, it also indicates the limits and restrictions of big data approaches. New sources of big data (for instance, social media) have recently become widely available and used as in the commercial researches as to a lesser extent, in the non-profit sector and in government. Consequently, it can be suggested that majority of data-driven researches nowadays are carried with narrow goals, of prescribing and changing human behavior (e.g. Consumption habits).

Brad Peters has written in Forbes that Big Data “changes the social contract” [18].

How can these changes be seen? First, we suggest that the nature of that change is complex, and there is a need with no doubts to formulate a more explicit and transparent ethical framework, that inherently contains all ethical components. Though one may claim that ethical topics connected with Big Data are centered on individuals, by raising personal privacy concerns, at the same time it generates new questions about personal identity, for instance who owns the personal data and how the increased availability and presence of more data influence individuals’ reputations.

It follows from what has been discussed before, it may be suggested that both individuals and organizations nowadays are interested in the understanding how data (e.g. personal data) is being handled. It may be stated that almost everyone nowadays are correlated with Big Data, consequently that means the discussion of ethics of Big Data has the potential to inform both individuals and organization on the resort of the benefits provided by big data and the potential risks from the unintended consequences of the unfair usage of Big Data.

As far as it may be stated, that Big Data may frequently be seen as aggregated data about people and their characteristics (e.g. consumer behavior), there is much of potential for use and abuse of this acquired data. Though direct benefits are now being realized in many areas (and most obviously by marketers), the concerns about consequences of having personal data being captured, aggregated, mined, sold, re-sold, and even linked to other data are recently just beginning to become an issue to think about. It has to be mentioned that the type and the importance of their impact of the risks are extremely difficult to determine in advance, and we are just starting to realize the implications and risks of the Big Data.

Moreover, the risks of consequences of the unfair usage of the Big Data are not just limited on the risks for individuals, much risks for the organizations can be also found, as far as the businesses are not aiming on harming their customers.

Big-data technologies may affect the meaning of such concepts as privacy, ownership, and identity for both individuals and organizations: as information is correlated and aggregated by not only the entity of origin, but additionally by those who may seek to
innovative further use and change it, it is hardly controlled how the information is used once this information out of one's hands.

We may see that society, government, and what is more important the legal system has not yet adapted to the upcoming age of big-data impacts, for instance new legislation on the resort of how big data should be handled is recently being debated by governments of all countries, but it will take much time until such legislations will be ratified. That may become an evidence of the major importance of formulating an ethical framework for applying Big Data nowadays.

The novel ethical framework for the implementation of Big Data, which is presented in this paper, is based on the critical analysis of the recent publications in the sphere of IT, cloud computing, Big Data and more particularly ethical consequences of its implementation. All ethical issues, which are needed to be thought of when generating, applying, transforming or analyzing Big Data, were divided into two groups:

2. Technology related issues.

It was proposed to allocate a total of fourteen elements of the ethical framework for the implementation of Big Data, that can be found in the following table. 2.

5. Conclusion

The ethics of big data are typically discussed in the relation to current issues that call for urgent policy and regulatory responses.

It should be admitted that with the rapidly increasing size and scope of information that big data technologies can provide businesses and academics nowadays, the consequences and risks of its unfair use are recently just being understood and thought of.

Therefore, this paper addressed current debates over big data specifically, by examining the ethical issues arising from advances in knowledge production: such issues such as privacy and data protection were discussed in the context of big data systems.

It follows from what has been discussed in the paper, that this indeed, may be seen as an evidence of more dramatic change of the society and more specifically an evidence of the process of continuing weakening of individuals’ autonomy.

Though it is hard to observe ethical implication of data-driven knowledge at an aggregated level where data are impersonal and anonymous, data about individuals are by its nature personal and often sensitive, moreover it may be suggested that there are direct influence of the applied data-driven knowledge on individuals. For instance, there is no secret that marketers nowadays more and more often resort to the help of so-called “agents of influence” who may change human consumer behavior.
<table>
<thead>
<tr>
<th>Category</th>
<th>Element</th>
<th>Operationalization of the element</th>
<th>Example of ethical implementation of Big Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Security controls related</td>
<td>Identity</td>
<td>What is the relationship between one’s offline identity and one’s online identity?</td>
<td>If online and offline identities of a person match, the information has to be stored with caution</td>
</tr>
<tr>
<td></td>
<td>Privacy</td>
<td>Who should control access to data?</td>
<td>The access to the data about individuals has to be controlled by individual or responsible organization</td>
</tr>
<tr>
<td></td>
<td>Transparency of the access</td>
<td></td>
<td>The access to the data has to be transparent</td>
</tr>
<tr>
<td></td>
<td>Ownership</td>
<td>Who owns data, has rights to transfer the data, and what are the obligations of people who generate and use that data?</td>
<td>Only person (and the responsible organization with the permission from the person) owns the rights to transfer the data</td>
</tr>
<tr>
<td></td>
<td>Reputation</td>
<td>How can we determine what data is trustworthy?</td>
<td>The data is trustworthy enough</td>
</tr>
<tr>
<td></td>
<td></td>
<td>What are the possible ways to use data?</td>
<td>The data should not be used to take advantage of the person</td>
</tr>
<tr>
<td></td>
<td></td>
<td>How are we perceived and judged by using data</td>
<td>The responsible organization should not ever be accused of usage of the data harmed the person</td>
</tr>
<tr>
<td></td>
<td>Data confidentiality</td>
<td>Is the information protected against unintended or unauthorized access?</td>
<td>The information has to be protected from being accessed by unintended or unauthorized users</td>
</tr>
<tr>
<td></td>
<td>Availability</td>
<td>Is (or how often) information available for use by its intended users?</td>
<td>The data has to be always available for the intended users access</td>
</tr>
<tr>
<td></td>
<td>Data lifecycle</td>
<td>Right to be forgotten and to erasure</td>
<td>The information about the person has to be deleted if the person intends to do so</td>
</tr>
<tr>
<td></td>
<td>Sensitivity of the data</td>
<td>Is the data sensitive (medical patient records)?</td>
<td>The data in open sources has to be not sensitive, all the sensitive information has to be stored in private storages</td>
</tr>
<tr>
<td>Technology related</td>
<td>Privileged user access</td>
<td>Who manage the data?</td>
<td>The information should be managed by the person himself or responsible authorities</td>
</tr>
<tr>
<td></td>
<td>Regulatory compliance</td>
<td>Does one have permission to use?</td>
<td>The access to the data has to be regulated with permission</td>
</tr>
<tr>
<td></td>
<td>Data location</td>
<td>Is the data store in private storage?</td>
<td>Personal data has to be stored in private storages, not in open sources</td>
</tr>
<tr>
<td></td>
<td>Data segregation</td>
<td>Are the encryption schemes in place and tested?</td>
<td>The encryption schemes have to be in place and tested</td>
</tr>
<tr>
<td></td>
<td>Recovery</td>
<td>Will one have an ability to do complete restoration in case of disaster?</td>
<td>The option of restoration (in case if the data was accidentally deleted or lost) has to be available</td>
</tr>
<tr>
<td></td>
<td>Long-term viability</td>
<td>Will the data be transferred in the event your provider ceases to exist, is acquired or contract ends?</td>
<td>The data has to be erased or given back to the person, if the contract ends</td>
</tr>
</tbody>
</table>
As it was already mentioned before, with the rapidly increasing size and scope of information that big data technologies can provide businesses nowadays, it is of major importance to maintain an ethical framework of implementation of Big Data.

The study proposes a novel framework for the implementation of Big Data, which is based on the critical analysis and comparison of the recent researches and publications conducted by various scientists in the sphere of Big Data and Cloud Computing. For that reason this paper is likely to be of major interest for the specialists involved in the areas of data science, applied researches, policymaking, moreover for a wider public, in order to aware about the contexts of Big Data analysis and its ethical consequences.

References


Этика больших данных: аналитический обзор

Л. Б. Гибер, Н. С. Казанцев
Национальный исследовательский университет — Высшая школа экономики
125413, Москва, ул. Флотская, 78/1, 42
e-mail: lbgiber@gmail.com

Аннотация. Применение больших данных (Big Data) дает беспрецедентные возможности для внедрения инноваций в экономику, здравоохранение, общественную безопасность, образование и почти в каждую сферу человеческой деятельности. В настоящее время не можем больше утверждать полную осведомленность о том, как повседневная жизнь подвергается влиянию со стороны процессов обработки больших объемов данных, что также создает риски для индивидуумов и общества в целом, если не осуществляется эффективное управление над подобными процессами. Количество последних публикаций в сфере этических проблем применения Big Data могут стать свидетельством постоянно растущего интереса ко всем аспектам этого вопроса. Исследование посвящено анализу этических рисков применения больших данных, в контексте изменения данных рисков в разных странах, в зависимости от правовых аспектов и особенностей конфиденциальности информации физических лиц.
Ключевые слова: большие данные, облачные вычисления, безопасность данных, этика, правовое регулирование.

Литература


Авторы:
Лариса Борисовна Гибер — менеджер Школы бизнеса и делового администрирования, магистрант Факультета бизнеса и менеджмента, Национальный исследовательский университет — Высшая школа экономики
Николай Сергеевич Казанцев — преподаватель кафедры моделирования и оптимизации бизнес-процессов Факультета бизнеса и менеджмента, Национальный исследовательский университет — Высшая школа экономики

418