Ethics and Information Technology (2019) 21:197–208
https://doi.org/10.1007/s10676-019-09502-5

ORIGINAL PAPER

Data science ethical considerations: a systematic literature review and proposed project framework

Jeffrey S. Saltz[1] · **Neil Dewar**[1]

Published online: 9 March 2019

Abstract
Data science, and the related field of big data, is an emerging discipline involving the analysis of data to solve problems and develop insights. This rapidly growing domain promises many benefits to both consumers and businesses. However, the use of big data analytics can also introduce many ethical concerns, stemming from, for example, the possible loss of privacy or the harming of a sub-category of the population via a classification algorithm. To help address these potential ethical challenges, this paper maps and describes the main ethical themes that were identified via systematic literature review. It then identifies a possible structure to integrate these themes within a data science project, thus helping to provide some structure in the on-going debate with respect to the possible ethical situations that can arise when using data science analytics.

Keywords Big data · Data science · Ethics · Code of conduct

Introduction

Data science is an emerging discipline involving the analysis of data to solve problems and develop insights. Big data is related to data science, in that for big data, the data sets are so large and/or complex that traditional data analysis techniques are typically not viable. While there are many views on what should be included in the field of big data or data science, we adopt Saltz and Stanton (2017) definition, which includes *"the collection, preparation, analysis, visualization, management, and preservation of large collections of information*". While this definition is broader than some might use, it embraces the notion that big data and data science are more than just analytics. For the rest of this paper, we will use the term data science to refer to this domain, including big data efforts.

As the field of data science grows, data scientists, just as professionals in other fields, will face pressure to deliver results. In trying to deliver results, the question of what is appropriate or ethical should arise. As an example of an ethical situation that data scientists might have to contemplate, one might ask if it is acceptable for an organization to develop a model that predicts the health care cost of a prospective employee, such as by exploring an employee's eating habits and exercise routine (Gumbus and Grodzinsky 2016). In order to address this type of question, the data science team, and the management of that organization, need to be aware of the possible ethical situations a project might encounter, so as to at least be able to consciously explore the ethical dilemma.

From a broader perspective, since ethics has been found to be a key component that can help determine the acceptance of new technologies (Stahl et al. 2016), it is important that data scientists consider the harm that might arise from their work so as to not stunt the adoption of data science. Without exploring these questions, the unethical use of data science could impact the reputational and economic well being of an organization, such as the public's well publicized reaction to Target's alleged prediction of a teenager's pregnancy (Someh et al. 2016).

However, as data science is a new domain, the full breadth and depth of data science ethical challenges has not yet been explored. In fact, it has been noted that this growing field has often excluded ethical analysis in both practice and academia (Martin 2015) and that there is no widespread agreement about what constitutes ethical versus unethical use of data science (Someh et al. 2016). The need for a focused view on ethics in data science has been reinforced by the former United States Chief Data

✉ Jeffrey S. Saltz
jsaltz@syr.edu

[1] Syracuse University, Syracuse, USA

Scientist, DJ Patil, who called on data scientists to work on developing a body of ethics (Nyes 2016). To make headway towards developing a foundational body of knowledge on data science ethics, we sought to identify the major concepts noted by practitioners and researchers running into ethical dilemmas within data science projects. Noting these key ethical dilemmas and concepts can encourage critical thinking and ethical reflection within a data science project and be a first step towards data scientists being able to systematically address the ethical impact and implications of their work using a consistent, holistic approach (Tractenberg et al. 2015).

With this goal in mind, we conducted a systematic literature review (SLR) of current data science-related scholarship that touches on ethics. We conducted an SRL since an SLR helps to identify all research related to a topic via a rigorous protocol-driven analysis and since reviewing the literature is one key to enabling the consolidation of existing knowledge and identifying gaps in current knowledge and developing research agendas (Stahl et al. 2016). Furthermore, a literature review can help move a discipline forward by clearly showing what is known (Boell and Cecez-Kecmanovic 2014). According to Rowe (2014), there are several possible goals of a literature review, such as summarizing prior research, examining contributions of past research or clarifying and/or integrating views created via previous research. As ethics within data science is such a new domain, our aim is to integrate views previously articulated, thus providing an overview of the key ethical conundrums that one might encounter within a data science project.

Hence, the key focus of this research is to create a framework of the different ethical challenges that a team might encounter when working on a data science project via the use of an SLR. Our goal was to curate the most common ethical dilemmas and challenges identified by contemporary experts in the data science field. Specifically, this paper focuses on the following research questions with respect to ethics and data science:

RQ 1: What are the key data science related ethical challenges that can be identified within the literature?

RQ 2: How might a team use these identified challenges when executing a data science project?

Section "Background" presents background information on ethics, ethics in computing and the need for ethics in data science. Section "Research method" then discusses the methodology used in our literature review. This is followed, in Section "Findings", with our findings. Section "Discussion" discusses our findings and finally, in Section "Conclusion", we present our conclusions and also provide some limitations and possible next steps.

Background

Ethics overview

We start with a brief overview of ethics. At the most basic level, it refers to the perception of something being good or right. One may speak of an "ethical use of data science" and mean that it is performed in a way that is right, proper, acceptable, or socially appropriate. Such an intuition of the ethical quality of an act is usually based on more or less explicit norms and values that are accepted within a social group or culture. Where such values and norms cease to be easily applicable or where they clash, explicit reflection on the bases and assumptions related to ethical judgments is required.

This is what ethics and the discipline of moral philosophy explores. The definition of ethics as moral guidance for behavior and principles of truth reflects the Kantian and utilitarian viewpoints as theoretically underpinning the ethical behavior of human economic actors (Mingers and Walsham 2010; Newell and Marabelli 2015). Kantian ethics argues that ethical action is based on moral values and principles, including honesty and responsibility. The Kantian perspective is therefore not concerned about the consequences of those the actions of individual actors. Conversely, the utilitarian theory focuses on consequences or outcomes. Specifically, an action is considered ethical if it is intended to maximize positive outcomes for the majority of actors (e.g. citizens in a country).

In this literature review, we focus on identifying possible ethical dilemmas, and hence, do not seek a specific Kantian or utilitarian perspective. However, it is helpful to broadly consider both perspectives, since both viewpoints offer benefits and limitations. For example, the utilitarian theory can create injustice for minority groups, since the greater good of a majority is central to the discourse. However, one could also argue that identifying the overall good in our modern and competitive world is not straightforward (Mingers and Walsham 2010).

Ethics in computing

Because data science is inextricably linked with computing, and computing has a longer history than data science, it is worth briefly reviewing ethics in computing. The potential of computing technologies to raise ethical and social issues that differ fundamentally from those raised by other technologies has been discussed since the very inception of digital computing (Wiener 1954). While there are early examples of high-level attention to the relationship between computers and ethics, a broader discourse

only started in the 1980s and 1990s. During this period, computer ethics developed into a field of applied ethics (Stahl et al. 2016). Dedicated courses on computer ethics were included in curricula, textbooks on the topics were written and academic conferences (e.g., Computer ethics philosophical enquiry and computers and philosophy) and journals (e.g., Ethics and information technology) were created.

This has led to a growing academic discourse with respect to the domain of computing ethics and on raising the awareness and interest of computing experts the in social and ethical aspects of their work, for example, by including it in standard curricula or professional accreditation. One such example of this focus on ethical challenges is within Association for Computing Machinery (ACM) code of conduct and curricula guidelines. As a consequence, most computing experts who have gone through structured training, such as through a university degree program, have an understanding of professional commitments to ethics as represented in codes and expectations of professional bodies such as the ACM, the British Computing Society (BCS), the Institution of Engineering and Technology (IET), and others (Stahl et al. 2016). Furthermore, it is not surprising that there have been several literature reviews that focus on ethics within the field of computer science for several decades. This work includes several overviews of the field (Stahl et al. 2016; Brey and Soraker 2009; Bynum 2008) and anthologies aiming to cover the main topics (Bynum and Rogerson 2003; Johnson 1985; Johnson and Nissenbaum 1995). However, none of these have explicitly focused on the field of data science and the new emerging ethical conundrums that data scientists might encounter.

The need for data science ethics

The need of ethics in data science has been frequently noted (Floridi and Taddeo 2016; Schwartz 2011; Fong 2016). It has also been noted that organizations practicing data science should provide ethical training and participative ethical assessments to analyze ethical issues (Leonelli 2016), but it is not clear that organizations have the breadth and depth of knowledge to easily offer this training.

There are many drivers for this need for ethics. For example, at a high level, Tiell and Metcalf (2016) have argued that data science introduces new classes of risk to organizations. Hence, it is not surprising that others have noted that none of the existing codes of conducts sufficiently cover the full range of potential ethical challenges a data science team might encounter (Tractenberg et al. 2015; Saltz et al. 2018). Thus, using an existing ethical framework from a software development context is not sufficient. The need for ethics has also been validated by an organized group of data scientists creating a data science code of professional conduct. However, the group, *the Data Science Association*, is not universally recognized or even known across the data science field.

Research method

While there are many approaches to a literature review, one approach, which is followed in this research, is to combine quantitative and qualitative analysis to provide deeper insights (Joseph et al. 2007). Specifically, to perform our literature review, we leveraged the guidelines for a SLR suggested by Kitchenham and Charters (2007), and hence, we structure our explanation of the methodology we used in our review by describing how we planned of the review as well as how we conducted and reported on the results of the review.

Planning the review

The plan for our SLR is summarized in Table 1. Specifically, we first defined the search space, which were the following six electronic repositories: Science Direct, Scopus, Web of Science, IEEE Xplore, ACM Digital Library and Google, which was used to explore grey literature. Next, we defined the search terms used, which where as follows: "data science" and ethics, "big data" and ethics, "data science" and ethical "big data" and ethical. We composed the search string for each database manually, based on the search functionality offered by that database's web-based user interface. The search was done on the full text of the articles, in this way we could avoid missing papers did not include our search keywords in titles or abstracts, but were relevant to the review. We kept the search to relatively recent articles since data science and the related field of big data is new,

Table 1 Search summary

Electron databases searched	ACM digital library, IEEE xplore, science direct, scopus, web of science, google
Search terms	"Big data" and ethics; "data science" and ethics; "big data" and ethical; "data science" and ethical
Publication period	2010 through 2017
Language	English
Search applied	Full text

and older articles would not capture the issues and challenges that this new domain might be creating.

Hence, to determine whether a paper should be included, in our analysis, the following inclusion criteria were defined:

- Papers published in a peer-reviewed outlet contained in ACM Digital Library, IEEE Xplore, Science Direct, Scopus, the Web of Science, or for grey literature, Google.
- Papers needed to be in English
- Papers included the relevant search terms as previously defined
- Papers that were published after 2009 (2010 or later)

In addition, the following items comprised our exclusion criteria:

- Papers that did not meet inclusion criteria;
- Papers that did not explicitly focus on ethics within a data science context, but rather, only referred to data science as a side topic
- Papers that focused on data science but only casually mentioned ethics.
- Papers that did not focus on ethical challenges a data science project might encounter, but rather, focused on high-level societal ethical considerations beyond the possible control of the organization supporting the data science effort.
- For our grey literature search, we excluded sources that had no form of review (ex. blogs)

Our exclusion of papers that discussed high-level societal ethical considerations beyond the control of the organization supporting the data science effort was driven by our desire to focus this research on enabling actionable ethics analysis within a specific data science project. This does not imply that data scientists have no role in helping to address the more overarching societal concerns, such as the impact of self-driving cars on society. In fact, data scientists can and should add their technical insight to these societal discussions (as we note in our discussion of potential next steps in our conclusion), but we view these high-level societal ethical considerations to be beyond the scope of this research.

Conducting the academic review—paper search and selection

By following the search strategy outlined in the previous section, the identified electronic databases were searched and the papers retrieved. In this initial search, 3021 papers were identified, as shown in Table 2. Note that some of these papers were duplicates, since the electronic repositories contain some overlapping sources.

Table 2 Initial search results after applying the inclusion criteria

	"Data science" + ethics	"Big data" + ethics	"Data science" + ethical	"Big data" + ethical
Science direct	187	703	158	832
Scopus	42	261	41	167
Web of science	30	256	36	142
IEEE xplore	2	31	5	30
ACM digital library	9	40	9	40

Table 3 Results after applying the exclusion criteria

	"Data science" + ethics	"Big data" + ethics	"Data science" + ethical	"Big data" + ethical
Science direct	0	8	3	2
Scopus	11	12	11	10
Web of science	7	9	7	17
IEEE xplore	0	4	0	3
ACM digital library	2	2	2	6

An extensive inspection of the studies' titles and abstracts was then made to apply the exclusion criteria. If needed, the papers were skimmed to confirm it should have been included or excluded. In total, as shown in Table 3, 116 papers were identified for further review. However, just as with the initial search results, there were duplicate papers within the count. After removing those duplicates, a total of 50 papers were identified for detailed analysis.

Conducting the grey literature review—paper search and selection

To augment the papers identified during our literature review of academic peer reviewed papers, we also searched Google for grey literature and other articles that might be useful. The internet sources were used only if the content had some sort of peer review, such as books, news items from the website of major newspapers, websites of professional bodies or professional journals. In addition, just as for the academic literature review, the publications were limited to those that were written in English after 2009 and had a focus on the topic of ethics in the field of data science.

In terms of conducting the review of the articles returned from Google, the titles and summary of the highest-ranking papers were evaluated to determine the relevance to our area of study. If there was uncertainty in this step, we screened the actual article, after which the paper was included or excluded. The analysis of the articles returned by each

Table 4 Breakdown of google search results

Source	Number of articles
Grey literature	26
Other peer reviewed	4
Article found via SLR	5
Total	35

Google search was stopped after reviewing the first 300 articles.

This review identified 35 additional articles. As shown in Table 4, of those 35 articles, five were also identified via our academic peer review described in the previous section, four were other peer reviewed journal papers that were not identified via our previous academic review (ex. a law review article that was not part of our search repository) and 26 were grey literature articles (i.e., an HBR article).

Conducting the review—data extraction and synthesis

According to the guidelines provided by Kitchenham and Charters (2007), we defined a data extraction process to identify relevant information from the 80 papers (50 from our SLR, four additional academic papers and 26 grey literature articles) that pertain to our research questions.

Our data extraction process included the following: First, we set up a form to record ideas, concepts, contributions, and findings of each of the 80 papers. Using this form ensures subsequent higher-order interpretation. The following data were extracted from each publication: (i) review date; (ii) title; (iii) authors; (iv) reference; (v) database; (vi) year of publication and (vii) an electronic link to the actual paper.

Once the extraction was completed, we used content analysis (Elo and Kyngäs 2007; Hsieh and Shannon 2005) to explore the key ethical concepts discussed within each of the papers. Each of these key concepts was also recorded as part of the data extraction. Specifically, the papers were analyzed through an iterative process of item surfacing, refinement and regrouping to generate the key themes used as our framework to describe the ethical challenges noted in the papers.

Finally, we assessed the repeatability of our data extraction and categorization by using an inter-rater analysis among the researchers (Fleiss et al. 2004). To find the inter-rater agreement among the researchers, we had two independent coders evaluate the papers. After training, the coders agreed on 89% of the coding decisions. Disagreements were discussed and agreed upon to create a final coded data set.

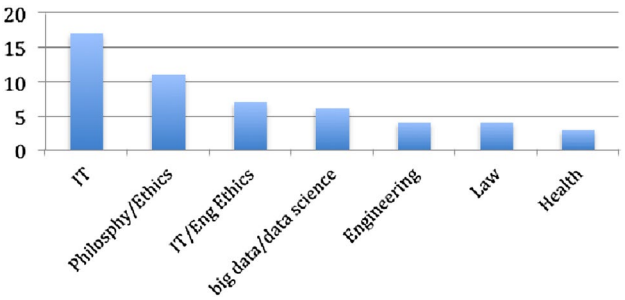

Fig. 1 Number of articles by focus of journal

Findings

We first note that the majority of the identified papers have been recently published. In fact, only eight of 80 identified articles were published prior to 2014, four were from peer reviewed journals and conferences and the other four were identified via our Google search. This is not surprising, as this coincides more broadly with the increasing use of data science across a range of contexts.

In terms of the publication outlet focus, as one can see in Fig. 1, the highest concentration of articles were published in information technology focused journals/conferences, where there were 17 relevant articles published. However, there were also several other domains that had more than five papers published, including journals focused on ethics and journals/conferences focused on data science. A special issue of *Philosophical Transactions A*, which focuses on a range of philosophical topics within the physical, mathematical and engineering sciences, generated six of the eleven articles published within a philosophy/ethics focus. This was the highest number of articles identified from any publication source. Note that since the domain focus for grey literature was often not clearly defined, we restricted our domain analysis to peer reviewed academic papers.

In our analysis of the articles, we identified four key themes (the need for an ethics framework, the newness of the field, data related challenges and model related challenges). The rest of this section describes each of these themes in more detail.

Newness of the field

One theme that was often noted was the challenge due to the newness of the field. Specifically, since the field is new, many ethical norms and regulations may not yet have been explored or defined (Metcalf et al. 2016; Sweeny 2013). This is further complicated by the fact that ethics and regulation tend to lag technology improvements (Zwitter 2014) and the fact that data science might introduce new classes of risk to an organization (Tiell and Metcalf 2016). In general, at least

partly due to the newness of the field, it was believed that it would be difficult to predict all the potential relevant ethical issues (Tractenberg et al. 2015). Hence, in an emerging field such as data science, there may be a lack of regulatory/legal clarity for certain situations. There may also be ethical implications that have not have been previously considered by others or even been highlighted as a potential ethical dilemma.

One such example is anonymity. While the need for anonymity is not new to data science, the thought process with respect to how to ensure anonymity must be re-examined with the emergence of advanced data science linking techniques. An example of one such ethical situation was when Netflix was sued by a closeted lesbian mother after University of Texas researchers demonstrated that Netflix data published for a competition, when combined with data from the IMDB website, uniquely identified customers and their viewing preferences (Drosou et al. 2017). In a different example of the current ambiguity relating to ethics and data ownership, a program such as Cisco's 'Connected Athlete' (Harkens 2016) collects vast amounts of biometric data on athletes in order to improve performance, prevent injury, and increase fan immersion. However, to work effectively, the creation of large databases of the collected information is required. The ownership of these databases, and the data contained within these databases, is unclear and may inform how the general public's health data is treated in future (Harkens 2016). Both of these examples describe situations that may have existed previously, but have become much more pressing due to the advent of data science.

The need for an ethics framework

The need for creating an ethical framework was the second theme identified. For example, it was suggested that creating an ethical framework could help establish a clear understanding of the vocabulary needed for discussing issues related to data science ethics (Voronova and Kazantsev 2015; Tractenberg et al. 2015). A framework could also enable data science teams to address the ethical impact and implications of data science and its applications using a consistent, holistic and inclusive approach (Tractenberg et al. 2015). In terms of leveraging an existing code of ethics, many noted that nothing was available that fully cover what is needed (Stoyanovich et al. 2017; Leonelli 2016; Voronova and Kazantsev 2015), and it was also noted that using a more general code of ethics would the lack the specificity to be useful (Stoyanovich et al. 2017).

In terms of what a framework or process could look like, some argued for a general framework that encourages critical thinking and ethical reflection (Leonelli 2016; Floridi and Taddeo 2016). This general framework could also help address questions concerning the responsibilities and liabilities of people in charge of data science processes, strategies and policies. Others focused on a specific aspect of a fully defined end-to-end process, such as the need for a data governance process to define how data is captured, stored and used (Dorasamy and Pomazalová 2016). Yet others suggested creating an actual information technology system to ensure ethics (Stoyanovich et al. 2017), even though they recognized that this is clearly a longer-term vision, as opposed to something that might be created in the short term. In any event, the goal of such a framework would be to help ensure ethical practices fostering both the progress of data science and the protection of the rights of individuals and groups (Floridi and Taddeo 2016).

Data related challenges

The data related challenges theme focuses on the key ethical situations that can arise relating to the collection and use of data. The growth of data science is in part due to the increasing amount of data that is generated, stored and available to data scientists to help predict future events based on past trends. As the two examples in the previous section show, data scientists often integrate multiple distinct data sources to generate novel insights. However, the previous examples also show that the collection and use of data creates many potentially challenging ethical situations. In fact, many of the ethical issues can be thought of as potential issues in the data supply chain (Martin 2015). Three key data related challenges were identified and are described below.

Privacy and anonymity

An individual's right to choose which of their activities and facts are shared with others is an important consideration that data science teams need to contemplate. In a digital age, this includes both what the individual *chooses* to publish and their ability to *control* with whom the data is shared. Privacy issues focus on who should control access to data and ownership concerns not just who owns the collected data but which rights can be transferred and what obligations collecting or receiving such data entails (Mateosian 2013; Wielki 2015).

The ability of aggregating and linking data enables one to merge multiple data sets and creates the ability for harm to arise from that linking of disparate information sources. For example, it has been noted that people can be re-identified from anonymous data using zip code, birth date and gender with 87% accuracy (Gumbus and Grodzinsky 2016). The impact of aggregating and linking data, and the ability for harm to arise from that information, has been noted as differentiators from other fields (Stevenson 2014; Fairfield and Shtein 2014).

For example, in the previously noted Netflix situation, Netflix, who was the database publisher, failed to understand either that this re-identification was possible or why this re-identification was problematic, revealing a lack of knowledge of either technical or ethical issues in their research. Due to this, Metcalf et al. (2016) point out that the phenomenon of data science has introduced "a change in the relationality, flexibility, repurposing and de-contextualization of data" requiring development of new ethical considerations.

Data misuse

Being able to access or collect data does not mean that it is ethical to use that data (Boyd et al. 2014). For example, there are many web sites that prohibit the collection and use of crawled data and that the use of a web crawler to gather that data may breach the terms of use of a website. Furthermore, there are "upstream" ethical issues, such as the privacy implications of how big data is gathered in the first place (Pascalev 2017). This is due to the fact that big data technology has introduced changes that impact how organizations collect information about individuals, as well as affect how individuals control the access, use and retention of that collected personal data. Unfortunately, that collected personal data is often used for purposes beyond its' intended purpose and many users would consider such practices a violation of their right to privacy (Pascalev 2017).

In a different but related example, access to customer data is typically achieved through a customer agreeing to a published usage policy. However, Tene and Polotensky (2012) suggest that consumers fail to read and understand these policies, which raises many questions with respect to actual consent. Complicating this challenge is that it is often unreasonable to expect consumers to read and agree to the published policy (since for many tools and apps, consumers really have no choice). Hence, even if the analytics where the data is being used is ethical, there might be issues relating to how the data was gathered in the first place or if the data is being used in a manner agreed to by the individual who provided the data. In reality, understanding if consent was given to use the data for its proposed use is still more in the grey area of feelings, opinions, and right treatment (Braun and Garriga 2018). However, there are some high level suggestions that organizations could follow, such as taking ownership of their data sources, not entering into confidentiality agreements that preclude explaining who are their data partners and making the data supply chain visible so that an organization has the ability to ensure no data misuse (Martin 2015).

Finally, an example of the ambiguity of data misuse is as follows. Suppose that an energy supply company finds a way to monetize its customers' electricity smart meter data by selling that data to an organization that wants to learn about how people live, yet has no intention of ever selling any product directly to those customers. The data would provide additional revenue to the energy supplier, yet there might be no incremental benefit to the energy supplier's customers. In this situation, it's not clear who owes the data and if that data is being misused. In other words, perhaps the customers should expect to share some of the bounty via reduced energy pricing (Grindrod 2016).

Data accuracy and validity

Understanding data accuracy is a key aspect of the data scientist's role. This theme not only covers the accuracy of the data, but also whether the data being used is appropriate for the problem being addressed. In other words, the data scientist needs to ensure the 'fitness of purpose' with respect to how the data is used. Otherwise, data can be taken out of context or might not be used in the spirit of how the data provider intended.

A simple example is that raw data is routinely cleaned prior to detailed analysis, and it has been noted that imputing missing values, excluding records with missing values, removing outliers and transforming variables could generate inaccurate results and/or be minimally documented and have a significant impact on the downstream analytical results (Fuller 2017; Boyd et al. 2014).

A more advanced example where data accuracy and validity might arise is with respect to teacher evaluations. A growing number of states use data from standardized test-scores of a teacher's students to develop teacher performance scores. The output from these models is sometimes used in decisions about teacher tenure, dismissal and compensation. However, many question the accuracy of a single student test score as input into this model. It has been noted that when "when any one student takes a math test, on any one day, there is a huge uncertainty around that score. It could be the kid got lucky this year, and guessed two or three right questions. Or the kid this morning could not have been feeling well. Consequently that score on any one day is not necessarily a good reflection of a kid's attainment level" (Butrymowicz and Garland 2012). Hence, some argue that, even though the actual database has the correct scores stored in the database, the data from one test is not accurate and should not be used as a key input for the model (Butrymowicz and Garland 2012).

Challenges when using analytical models

The model theme focuses on the ethical challenges that can arise from building and using analytical models. An analytical model is a mathematical technique used for simulating, explaining, and making predictions about future situations based on past data. In other words, an analytical model is a

set of mathematical functions that encapsulate the prediction of a certain situation based on past information. However, the use of an algorithm (analytical model) might introduce or amplify a range of ethical situations. Three key model related challenges were identified and are described below.

Personal and group harm

Based on a how a model is used, there can be a significant impact on a person or a group of people. One concern is that data science models can be built using data that records a bias, and thus, the model might also have that bias, and as such, systematically disadvantage a societal sub-group (Crawford 2013). Specifically, the identification of types of individuals that are grouped together may lead to serious ethical problems, such as group (e.g. ageism, ethnicism, sexism) discrimination (Floridi and Taddeo 2016).

A simple example discussed by Crawford (2013), is that, using accelerometer and GPS data from smartphones, an organization predicted potholes and instantly reported them to the city of Boston. However, older, poorer people was less likely to have smartphones. This means that the smartphone data was missing information from significant parts of the population—often those who have the fewest resources. Hence, there needs to be a focus on avoiding discrimination and bias, which might unknowingly occur via the use of a data science model.

More generally, analytics allows for a new type of algorithmically assembled group to be formed that does not necessarily align with classes already protected by privacy and anti-discrimination law or addressed in fairness and discrimination-aware analytics. In this situation, individuals are linked according to offline identifiers (e.g. age, ethnicity, geographical location) and shared behavioral identity tokens, allowing for predictions and decisions to be taken at a group level rather than an individual level (Mittelstadt 2017). A simplistic example of such a group is 'dog owners aged 38–40 that exercise regularly'. Being identified as a member of this group could drive a variety of automated decisions with harmful or beneficial effects for individual members, such as a preferential rate for health insurance (Mittelstadt 2017).

Subjective model design

Another concern is that while data science can bring objectivity to decision making, there is subjectivity within data science modeling, in that decisions must be made about which algorithm to use, which data sources to use, whether one data point should be used as a proxy for a missing fact, and how to interpret results (Sandvig et al. 2014). In other words, biases in the interpretation of data may lie not only in the tools a data scientist uses, but also in the data scientist themselves (Fuller 2017). This is re-enforced by Boyd and Crawford, who note that "researchers must be able to account for the biases in their interpretation of the data. To do so requires recognizing that one's identity and perspective informs one's analysis" (Boyd and Crawford 2012).

One simple example of subjective model design is in the field of sports analytics, where a model might be created that looks for a person to play a specific position on the field or court. For instance, in basketball, a player traditionally played one of five well-defined positions, and analytics were developed to identify the best possible player for each of these positions. However, these models had a subjective model design in that they incorrectly oversimplified the skill sets of basketball players and also pigeon-holed players into one of these five positions. In other words, the existing models might not accurately evaluate a player's specific skill and hence, misclassify a player's abilities to play multiple positions (Chen 2017).

Model misuse and misinterpretation

Most predictive models are statistical in nature. They provide no guarantees; rather, they tell us about areas where increased probability of an outcome might guide us to act differently. Due to this, the data scientist's ethical responsibilities do not end with the completion of a model. The data scientist also has a duty to explain their models and the implications of using a model. In particular, the model must be explained using language that non data scientists, such as managers, can understand. In other words, attention needs to be paid not only to the analysis of the data, but also to the presentation of the fruits of that analysis, and it is crucial that those who devise the analytics clearly understand and explain their impact (Fuller 2017).

Stated another way, due to their statistical nature, no model is completely accurate. Hence, it is important to explain model accuracy. With this in mind, the team must ensure that the analytical decision reflects the scale, accuracy and precision of the data that was used in creating the model (Clarke 2016). In addition, a reasoned justification should be made for the chosen levels of automation in decision-making (De Laat 2017), and this should be periodically re-evaluated for soundness via an appropriate level of oversight and governance.

Another aspect of this challenge is model transparency. Specifically, algorithmic outcomes of machine learning are often difficult to interpret, even by experts, and an explanation in understandable terms as to why a specific decision is recommended often cannot be supplied. The model is effectively a black box to everyone, layman and expert alike (De Laat 2017), which can make model transparency (or explainability) very difficult. In this situation, transparency delivers very little in terms of explanation (one can offer technical

clarifications about the accuracy of an algorithm—but not about the reasons behind its recommendations). For example, neural networks and support vector machines, which are both popular modeling techniques, have this challenge. When using a neural network, a middle layer (or more than one) is inserted connecting input and output. The weights connecting input variables to the middle variables, as well as those connecting the middle variables to the output variable are adjusted via several iterations within model development. The end model obtained displays all those weights, but cannot be interpreted as to how much the various input variables contribute to the outcome. In the situations where there is a high degree of regulation or a right of challenge, the empirical models must be simple enough to allow some explanation, such as explaining which covariate is driving a particular inference or decision. This is the case, for example, when one wants to lend money or to deny an operation (Grindrod 2016). Hence, in such situations, the choice of the possible model must be severely curtailed—perhaps even reduced to logistic regressions (Grindrod 2016).

Discussion

Many professional bodies have developed codes of conduct, as described by Tractenberg et al. (2015). In reviewing the articles identified during our literature review, we found that there was a gap between the codes of conduct's general statements such as "Do No Harm" and the many specific ethical concerns discussed in individual papers and noted in our key themes. In fact, we were unable to find a general map of the ethical considerations relevant to data science to assist the practitioner through the course of a project.

One way to explore how teams could use these themes is to focus on the data and model related themes and how to integrate these identified ethical challenges within a data science process. To integrate our themes within a data science process, we first note that current descriptions on how to execute data science projects generally adopt a task-focused approach, conveying the techniques required to analyze data. While these process models differ in details, at a high level, they are broadly similar. For example, Jagadish et al. (2014) describe a process that includes acquisition, information extraction and cleaning, data integration, modeling, analysis, interpretation and deployment. This step-by-step view is similar to CRISP-DM (*Cross Industry Standard Process for Data Mining)*, which was established in the 1990s (Shearer 2000), and is still the most widely used process (Haffar 2015) within the field of data discovery and data science. Hence, we can use that process model as a way to integrate the identified ethical challenges with the phases of the data science project life cycle. CRISP-DM mentions six high-level phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment.

Table 5 shows the mapping of the identified themes to the project phases. Not surprisingly, the data related challenges map to the data understanding and data preparation phases, and the model related challenges map to the modeling, evaluation and deployment phases. However, there was no identified ethical theme related to the business understanding phase. It makes sense that this theme was not a key area of focus in the literature, since this phase is more focused on topics such as ensuring accountability, which while important, might not be a key focus of a paper exploring new ethical issues relating to data science. Hence, for this business understanding phase, two ethical new considerations are proposed. First, at the start of the project, the team should consider, at a conceptual level, the potential personal and group harm. In addition, the team should also explore team accountability of the potential ethical situations. While this mapping is not a fully defined ethics framework for data science projects, this list of ethical considerations for each project phase could be a first step towards a structured dialog that should occur during every data science project.

However, just mapping the key ethical themes to the different project phases within a data science project might not be sufficient. For Example, Manders-Huits & Zimmer (2009) note three key challenges when inserting ethics within a project: (1) confronting competing values; (2) identifying the role of the values advocate; and (3) the justification of a value framework. These challenges suggest that as part of an effort to integrate ethics within a data science project, one needs to explore the motivation and drivers of the stakeholders within that data science project. For example, recent negative headlines with respect to ethics in a few data science projects could be used to help motivate stakeholders appreciate the importance of exploring these

Table 5 Framework to explore the key ethical considerations by phase of project

Project phase	Key ethical themes	Ethical considerations
Business understanding	Project initiation/ management challenges	Personal and group harm
		Team accountability
Data understanding/ data preparation	Data challenges	Data misuse
		Data privacy & anonymity
		Data accuracy
Modeling	Model challenges	Personal and group harm
Evaluation		Subjective model design
Deployment		Misuse/misinterpretation

key considerations and on ensuring accountability that these considerations are properly explored.

Conclusion

To help consolidate the ethical challenges a data science team might encounter and provide a basis of future research in the field of data science ethics, this work explores the discourse related to data science ethics. We note the increasing dialog on this subject, as demonstrated by the significant increase in the number of recently published articles on this subject. This review is of importance to data science professionals who need to gain insight into how ethical debates relate to their work. It is also of interest to scholars who focus on ethics and data science, and who want to contextualize their work in a broader context. We hope that this work provides a foundation for future research in the field of data science ethics.

Our analysis consolidated the discussion into the key ethical challenges a data scientist might encounter, thus addressing our first research question (what are the key ethical challenges identified within the literature). From our literature review, we identified two general paths to cause harm. First, with respect to data related challenges, the preparation, storage and dissemination of data could impinge on the privacy or anonymity of the subject, or cause bias in the resulting analytics. For example, just because data is available, it does not mean it is ethical to use that data (Boyd 2012). Second, with respect to model related challenges, a data science model might operate incorrectly, so for example, some subjects could be misclassified, resulting in harm. Furthermore, a model might operate correctly, but the objective of the model is inherently unfair to some subjects. In addition, while data science can bring objectivity to decision making, there is subjectivity within data science modeling, in that decisions must be made about which algorithm to use, which data sources to use, whether one data point should be used as a proxy for a missing fact, and how to interpret results (Sandvig et al. 2014).

In addition to describing the key ethical challenges a data science team might encounter, we also link these challenges to the phases within a project. This provides a framework that a data science team could use to help ensure that ethics have been appropriately considered within a data science project, which addresses our second research question (how might a team use these identified challenges).

Limitations

One limitation with respect to this research is the key words that were used within our SLR. This limitation is inherent in any SLR and it is possible that our key words only identified a subset of the desired literature. In other words, other papers might have used a different vocabulary to express similar ideas, and thus, there could have been articles that were missed. Based on the fact that we used broad search terms, we do not believe that there was a large swath of relevant literature that was excluded, but it is certainly possible that specific papers were missed during our SLR.

Furthermore, an article offering a survey of a large topic area such as ethics and data science can not, by necessity, go into significant depth with regard to all aspects of discourse that were surveyed. We used our framework to explain the current discourse on this topic and elaborated on some of the key ethical issues covered by each topic. However, we did not go into depth on each topic, as our main purpose was to map the topics and issues that have been discussed in the field, not fully explore all the ethical challenges relating to big data science. Another limitation is that new topics and themes may be apparent only after additional articles are published.

Next steps

One area of potential exploration is how these findings could be integrated within a data science curriculum, either at the undergraduate or graduate level. For example, these ethics concepts could be integrated within existing classes via the creation of key questions (based on our ethical considerations) that could be shared with students, thus providing students with a basic toolkit to help students think about these challenges within the context of a data science class project. However, potential barriers to introducing these concepts within data science courses would also need to be explored, and likely include barriers such as instructors that have limited knowledge with respect to these ethical considerations or instructors that might believe that ethics should be not integrated within data science courses, but rather, be an optional add-on course.

Other possible areas of future research include exploring some of the identified issues in more depth, or exploring some of the higher-level societal ethical considerations that were specifically excluded from our analysis, which could enable data scientists to help provide a technical perspective into these societal ethical challenges. Finally, a different next step could be to leverage this review in the creation of an industry accepted code of conduct.

References

Boell, S., & Cecez-Kecmanovic, D. (2014). A hermeneutic approach for conducting literature reviews and literature searches. *Communications of the Association for Information Systems, 34*, 1.

Boyd, D, & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society, 15*(5), 662–679.

Boyd, D, Levy, K., & Marwick, A. E. (2014). The networked nature of algorithmic discrimination. In *Data and discrimination: Collected essays* (pp. 43–57). Washington, DC: Open Technology Institute.

Boyd, K. (2012). Critical questions for big data. *Information, Communication & Society, 15*, 662–679.

Braun, A., & Garriga, G. (2018). Consumer journey analytics in the context of data privacy and ethics. In C. Linnhoff-Popien, R. Schneider & M. Zaddach (Eds.), *Digital marketplaces unleashed*. Berlin: Springer.

Brey, P., & Soraker, J. (2009). Philosophy of computing and information technology. In D. M. Gabbay, A. W. M. Meijers, J. Woods, & P. Thagard (Eds). *Philosophy of technology and engineering sciences* (pp. 1341–1408). North Holland: Elsevier.

Butrymowicz, S., & Garland, S. (2012). How New York city's value-added model compares to what other districts, states are doing, hechingerreport. Retrieved from http://hechingerreport.org/content/how-new-york-citys-value-added-model-compares-to-what-other-districts-states-are-doing_7757/.

Bynum, T. (2008). Computer and information ethics. In *Stanford encyclopedia of philosophy*. Retrieved from http://plato.stanford.edu/entries/ethics-computer/. Accessed 14 January 2016

Bynum, T., & Rogerson, S. (2003). *Computer ethics and professional responsibility: Introductory text*. New York: Wiley

Chen, A. (2017). Using machine learning to find the 8 types of players in the NBA, Fastbreak. http://fastbreakdata.com/classifying-the-modern-nba-player-with-machine-learning-539da03bb824.

Clarke, R. (2016). Big data, big risks. *Information Systems Journal, 26*(1), 77–90.

Crawford, K. (2013). The hidden biases in big data. Harvard Business Review Online Edn. Harvard Business Review.

De Laat, P. B. (2017). Big data and algorithmic decision-making: Can transparency restore accountability? *ACM SIGCAS Computers and Society, 47*(3), 39–53.

Dorasamy, N., & Pomazalová, N. (2016). Social impact and social media analysis relating to big data. In *Data science and big data computing* (pp. 293–313). Cham: Springer.

Drosou, M., Jagadish, H. V., Pitoura, E., & Stoyanovich, J. (2017). Diversity in big data: A review. *Big data, 5*(2), 73–84.

Elo, S., & Kyngäs, H. (2007). The qualitative content analysis process. *Journal of Advanced Nursing, 62*(1), 107–115.

Fairfield, J., & Shtein, H. (2014). Big data, big problems: Emerging issues in the ethics and data science of journalism. *Journal of Mass Media Ethics, 29*, 38–51.

Fleiss, J. L., Levin, B., & Paik, M. C. (2004). Determining sample sizes needed to detect a difference between two proportions. *Statistical Methods for Rates and Proportions, 2*, 64–85.

Floridi, L., & Taddeo, M. (2016). What is data ethics?. *Philosophical Transactions Series A, 374*, 2083.

Fong, K. (2016). The ethics conversation we're not having about analytics. Harvard Business Review Online Edn. Retrieved from http://blogs.hbr.org/2013/04/thehidden-biases-in-big-data/. Accessed 20 August 2017.

Fuller, M. (2017). Big data, ethics and religion: New questions from a new science. *Religions, 8*(5), 88.

Grindrod, P. (2016). Beyond privacy and exposure: Ethical issues within citizen-facing analytics. *Philosophical Transactions of the Royal Society A, 374*(2083), 20160132.

Gumbus, A., & Grodzinsky, F. (2016). Era of big data: Danger of descrimination. *ACM SIGCAS Computers and Society, 45*(3), 118–125.

Haffar, J. (2015). *Have you seen ASUM-DM?* Retrieved from IBM: https://developer.ibm.com/predictiveanalytics/2015/10/16/have-you-seen-asum-dm/.

Harkens, A. (2016). 'Rear window ethics' and discrimination: The darker side of big data. In *European conference on e-government* (p. 267). Academic Conferences International Limited.

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277–1288.

Jagadish, H., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. *Communications of the ACM, 57*(7), 86–94.

Johnson, D. (1985). *Computer ethics*. Upper Saddle River: Prentice-Hall.

Johnson, D., & Nissenbaum, H. (1995). *Computers, ethics and social values*. New York: Pearson.

Joseph, D., Ng, K., Koh, C., and Ang. S (2007). Turnover of information technology professionals: A narrative review, meta-analytic structural equation modeling, and model development. *MIS Quarterly, 31*(3), 547–577.

Kitchenham, B., & Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering*. UK: Keele.

Leonelli, S. (2016). Locating ethics in data science: Responsibility and accountability in global and distributed knowledge production systems. *Philosophical Transactions of the Royal Society A, 374*(2083), 20160122.

Manders-Huits, N., & Zimmer, M. (2009). Values and pragmatic action: The challenges of introducing ethical intelligence in technical design communities. *International Review of Information Ethics, 10*(2), 37–45.

Martin, K. E. (2015). Ethical issues in the big data industry. *MIS Quarterly Executive, 14*, 2.

Mateosian, R. (2013). Ethics of big data. *IEEE Micro, 33*(2), 60–61.

Metcalf, J., Keller, E., Boyd, D. (2016). Perspectives on big data, ethics and society. Council for Big Data, Ethics and Society. http://bdes.datasociety.net/council-output/perspectives-on-big-data-ethics-andsociety/.

Mingers, J., & Walsham, G. (2010). Towards ethical information systems: The contribution of discourse ethics. *MIS Quarterly, 34*(4), 833–854.

Mittelstadt, B. (2017). From individual to group privacy in big data analytics. *Philosophy & Technology, 30*, 475–494.

Newell, S., & Marabelli, M. (2015). Strategic opportunities (and challenges) of algorithmic decisionmaking: A call for action on the long-term societal effects of 'datification'. *The Journal of Strategic Information Systems*. https://doi.org/10.1016/j.jsis.2015.02.001.

Nyes, K. (2016). White house to data scientists: We need you. Computer world. Retrieved from http://www.computerworld.com/article/3125660/big-data/white-house-to-data-scientists-we-need-you.html. Accessed 20 August 2017.

Pascalev, M. (2017). Privacy exchanges: Restoring consent in privacy self-management. *Ethics and Information Technology, 19*(1), 39–48. https://doi.org/10.1007/s10676-016-9410-4.

Rowe, F. (2014). What literature review is not: Diversity, boundaries and recommendations. *European Journal of Information Systems, 23*(3), 241–255.

Saltz, J., Dewar, N., & Heckman, R. (2018). Key concepts for a data science ethics curriculum. In *Proceedings of the 49th ACM technical symposium on computer science education* (pp. 952–957). ACM.

Saltz, J., & Stanton, J. (2017). *An introduction to data science*. Thousand Oaks: SAGE Publications.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). An algorithm audit. In *Data and discrimination: Collected essays*. New York: New America, Open Technology Institute.

Schwartz, P. M. (2011). Privacy, ethics and analytics. IEEE security and privacy 9(3). IEEE.

Shearer, C. (2000). The CRISP-DM model: The new blueprint for data mining. *Journal of Data Warehousing, 5*(4), 13–22.

Someh, I. A., Breidbach, C. F., Davern, M. J., & Shanks, G. G. (2016). Ethical implications of big data analytics. In *ECIS* (pp. Research-in).

Stahl, B. C., Timmermans, J., & Mittelstadt, B. D. (2016). The ethics of computing: A survey of the computing-oriented literature. *ACM Computing Surveys (CSUR), 48*(4), 55.

Stevenson, D. (2014). *Locating discrimination in data-based systems. Data and discrimination: Collected essays* (16–20). Washington, DC: New America/Open Technology Institute

Stoyanovich, J., Howe, B., Abiteboul, S., Miklau, G., Sahuguet, A., & Weikum, G. (2017). Fides: Towards a platform for responsible data science. In *SSDBM'17-29th International Conference on Scientific and Statistical Database Management*.

Sweeney, L. (2013). Discrimination in Online Ad Delivery. ACM Queue 11(3). Association of Computing Machinery.

Tene, O., & Polotensky, J. (2012). Privacy in the age of big data. Stanford Law Review.

Tiell, S., & Metcalf, J. (2016). The Universal Principles of Data Science Ethics. Accenture Labs. https://www.accenture.com/t20160629T012639__w__/us-en/_acnmedia/PDF-24/Accenture-Universal-Principles-Data-Ethics.pdf.

Tractenberg, R. E., Russell, A. J., Morgan, G. J., FitzGerald, K. T., Collmann, J., Vinsel, L., … Dolling, L. M. (2015). Using ethical reasoning to amplify the reach and resonance of professional codes of conduct in training big data scientists. *Science and Engineering Ethics, 21*(6), 1485–1507.

Voronova, L., & Kazantsev, N. (2015). The ethics of big data: Analytical survey. In *Business informatics (CBI), 2015 IEEE 17th conference on* (Vol. 2, pp. 57–63). IEEE.

Wielki, J. (2015). The social and ethical challenges connected with the big data phenomenon. *Polish Journal of Management Studies, 11*(2), 192–202.

Wiener, N. (1954). *The human use of human beings*. New York: Doubleday.

Zwitter, A. (2014). Big data ethics. *Big Data & Society, 1*(2), 2053951714559253.