Journal of Business Ecosystems
Volume 2 • Issue 1 • Semi-Annually 2021

A Study on the Wide-Ranging Ethical Implications of Big Data Technology in a Digital Society:

How Likely Are Data Accidents During COVID-19?

Izabella V. Lokshina, SUNY Oneonta, USA

Cees J. M. Lanting, DATSA, Belgium

ABSTRACT

Exponential growth in the commercial use of the internet has dramatically increased the volume and scope of data gathered and analyzed by datacentric business organizations. Big Data emerged as a term to summarize both the technical and commercial aspects of these growing data collection and analysis processes. Formerly, much discussion of Big Data was focused on its transformational potential for technological innovation and efficiency; however, less attention was given to its ethical implications beyond the generation of commercial value. In this paper, the authors investigate the wide-ranging ethical implications of Big Data technology in a digital society. They inform that strategies behind Big Data technology require organizational systems, or business ecosystems, that also leave them vulnerable to accidents associated with its commercial value and known as data accidents. These data accidents have distinct features and raise important concerns, including data privacy during COVID-19. The authors suggest successful risk mitigation strategies.

KEYWORDS

Big Data Technology, Business Ecosystems, COVID-19, Data Accidents, Datacentric Business Organizations, Ethical Implications, Risk Mitigation Strategies

INTRODUCTION

Recently, as the use of the Internet has grown, so too has an interest in the use of data gathered and analyzed by arriving at a more digitized and technologically connected society (George et al., 2014; Lokshina et al., 2017). In the 1998 film "Enemy of the State", a rogue government agency is shown as having unlimited access to private data from a variety of data sources. At that point, the scenario was so disturbing to the Federal Bureau of Investigation (FBI) that a public relations campaign was launched to assure society that the plot was pure fiction (Miller, 2013).

Nevertheless, as recent exposures about data privacy have made clear, the surveillance shown in the film is very real at this time (Jennex, 2017; Nunan & Di Domenico, 2017). What has changed is the role of governments as collectors and users of data, and the increasing importance of commercial entities as the drivers of both data gathering and analysis, i.e. data analytics (Committee on Commerce, Science, and Transportation, 2020; Marr, 2015; Molok et al., 2012; Sedayao & Bhardwaj, 2014; Zaslavsky et al., 2012).

DOI: 10.4018/JBE.2021010103

The COVID-19 emergency has taken place in an already digital society. The amount of pandemic-related data gathered and processed globally has been enormous.

Additionally, advanced computational models, some based on machine learning, have shown tremendous potential in tracing sources or predicting the future spread of COVID-19, necessary for the planning of resources. Without correct data and models, the risk to public health cannot be assessed correctly, causing a problem that the authorities cannot allocate time, assets, and resources appropriately, leading to both penury, and oversupply, and waste.

Therefore, it is essential to leverage Big Data technology and intelligent analytics and put them to effective use for the benefit of public health. Reliance on digital data sources has been of great value in outbreaks caused by new pathogens significantly improving data collection, however, precise data is still rare and hence forecasts are still less reliable and effective (Scarpino & Petri, 2019).

The needs and conditions for responsible data gathering and analysis at a global scale must be clear as Big Data technology has become critical for managing the COVID-19 pandemic in a digital society. The use of data that has been collected from digital sources for prediction and surveillance is very important in the fight against the COVID-19 pandemic, but it is equally important to use this data in compliance with data protection regulations and with due respect for privacy and confidentiality and recognizing its possibly limited validity or bias.

Previous generations of information technology were dominated by technology companies with commercial strategies based on their expertise in hardware or software. Currently, many leading Internet companies have commercial strategies built around the collection and analysis of data.

For datacentric business organizations like Google, Facebook, and others, the collection of data has become a target instead of a way to achieve any additional business goals. This comes at a point when there is a growing interest from researchers in wide-ranging ethical implications of the increasing use of data in different areas, including data privacy (Jennex, 2017; Hong & Thong, 2013; Lokshina et al., 2019a; Lu et al., 2014; Nunan & Di Domenico, 2017); cyber-hacking (Bambauer, 2014; Jain et al., 2016; Perera et al., 2015; Zaslavsky et al., 2012); government regulation (Committee on Commerce, Science, and Transportation, 2020; Fink et al., 2012; Lokshina et al., 2019b); and intellectual property (Bateman et al., 2013; Jennex, 2017; Marr, 2015).

Besides, there are concerns about the level of government surveillance of commercial social networks, for instance, revealed due to leaks by Edward Snowden (Witte, 2013). In these circumstances, the term “Big Data” has obtained popularity in both business and public policy circles as summarizing industrial and commercial aspects of these state-of-the-art data gathering and analysis processes which involve also private data collection (Marr, 2015; Nunan & Di Domenico, 2017; Perera et al., 2015; Sedayao & Bhardwaj, 2014; Zaslavsky et al., 2012).

Until now, there has been much discussion of Big Data technology focused on its potential positive effects for both business and society (George et al., 2014). This extends beyond improvements to commercial efficiency and expands to claims about its transformational effect on such areas as healthcare and delivery of public services (Manyika et al., 2011; Lokshina et al., 2017).

Nevertheless, there has been less discussion on Big Data technology devoted to ethical issues. This is not surprising given the limited research on Internet ethics (Schlegelmilch & Oberseder, 2010) and limited understanding of wide-ranging ethical implications of new technologies started being deployed (De George, 2003; Nunan & Di Domenico, 2017). Therefore, in this paper, the authors have given thought to wide-ranging ethical implications of Big Data in a digital society, including during the COVID-19 pandemic, and evaluated ethical issues related to Big Data technology by applying the normal accident theory (Perrow, 1999; Nunan & Di Domenico, 2017).

Normal accidents are “normal” in the sense that the occurrence of negative developments is inevitable and taking place unexpectedly in organizational systems that are complex, interactive, and tightly coupled (Perrow, 1999). The normal accident theory has been applied to investigate nuclear power plant accidents (Pidgeon, 2011), underground coal mine devastations (Lokshina, 2001), plane crashes (Helmreich, 1997), structural collapses (Likhterman et al., 1998), failures in hospitals, anesthesia systems, the practice of medicine and perfusion producing shocking medication reactions (Dain, 2002), as well as most recent financial system fiascos (Lokshina, 2002; Guillén & Suárez, 2010; Palmer & Maher, 2010).

The authors believe that in high-risk systems distinguished by multiple and unexpected interactions, no matter how effective safety mechanisms are, normal accidents are inevitable (Dain, 2002). These negative events take place regardless of the number of safety mechanisms, the quality of the care provided, or the awareness of operators. In other words, in complex systems, errors are made by humans (Dain, 2002; Lokshina & Insinga, 2003).

The authors suggest that emerging datacentric business organizations, which can enable or be enabled by Big Data, have system features as identified in the normal accident theory. However, the consequences of normal accidents in these datacentric business organizations are less obvious than with natural disasters, which makes identification and remedy more difficult (Nunan & Di Domenico, 2017).

In this paper, the authors discuss data accidents, which extend the normal accident theory by exploring the scale and interconnection, as well as uncertainty and distrust created by Big Data technology in a digital society. The authors investigate mostly wide-ranging ethical implications of Big Data since the consequences of data accidents must be weighed against the effectiveness and value achieved by society with the deployment of Big Data technology (Lokshina & Lanting, 2018a; Nunan & Di Domenico, 2017).

This paper is comprised of eight sections and organized as follows. Section two identifies key features of Big Data technology and datacentric business organizations that enable and are enabled by an exponential growth in data collection and analysis processes. This section also shows that large-scale data collection can help curb the COVID-19 pandemic; however, this process should respect data privacy and gain public trust. Section three analyzes normal accidents associated with state-of-the-art and emerging technologies, outlines the normal accident theory, and explains specifics of datacentric organizational systems, or business ecosystems, that enable data accidents. Section four extends the normal accident theory by investigating distinct features and potential consequences of data accidents. This section also provides a classification of data accidents and analyzes scenarios of systemic, non-commercial, and commercial data accidents including examples related to COVID-19. Section five explains what should be done to mitigate the risks and potential implications of Big Data technology given the growth in the industrial and commercial value of state-of-the-art and emerging technologies. This section also recommends successful strategies and best practices to maintain responsible data collection and processing standards on a global scale to safeguard business ecosystems, followed by the conclusion, acknowledgment, and references.

VIEWS ON BIG DATA TECHNOLOGY

In this section, the authors present their views on Big Data technology. Although data has become strongly associated with information technology, managing and making sense of data remains an old problem. Historians, politicians, and military leaders have relied on information as a source of power, as well as a driving force, for centuries, while denying access to information has always been a lever to diminish power.

However, what has changed is the speed and volume of data gathered to analyze. When, in the 1850s, managers of the United States railroads were searching for new organizational designs to help overcome the challenges of data overflow in the growing business, they were talking about the volume of data below a single megabyte in modern terms (Rosenthal, 2014).

On the other hand, an important feature of data in modern business is a constant exponential growth in volume. To use just one contemporary statistic, 90% of all present data has been generated in the last two years (IBM, 2015; Jennex, 2017).

Under these circumstances, Big Data first appeared as a term to describe the technological innovation supporting this tremendous increase in the volume of data that is collected and analyzed (Jacobs, 2009). Beyond the volume of data, enabled to be collected, Big Data technology has also changed the complexity and velocity of data to be gathered and analyzed (IBM, 2015).

Velocity refers to processing power, i.e. speed at which data may be collected and analyzed, with near real-time analysis becoming possible for very large datasets and complex analysis. Complexity, due to the volume and variety of data, is also very significant as it marks a transition from gathering data in only text format towards collecting data in metadata, video, audio, and image formats (Kuechler, 2007).

More recently, Big Data went far beyond its industrial roots to include wide-ranging commercial opportunities enabled by the analysis of data, i.e. data analytics (Manyika et al. 2011; Lokshina et al., 2017). Therefore, Big Data as a term has been identified by politicians as the means for achieving economic growth. In the UK, Big Data has become one of eight key government priorities (HM Government, 2013; Nunan & Di Domenico, 2017). In the United States, Barrack Obama's use of campaign data for the analysis has led to nicknaming him as the Big Data President (Hurwitz, 2012).

This technological innovation has been supported by several economic factors. First, the cost of storing data has been reduced to the point that it might be economically feasible to store all forms of data even while there is no direct use for it. Second, the required hardware and software have now become readily available in the marketplace (Jennex, 2017).

Much of the software that currently supports Big Data have first appeared not from traditional technological companies with commercial strategies built around their expertise in hardware or software; but from datacentric business organizations where their specific demands have forced them to resolve their own Big Data technology issues. Particularly, to support the use of Big Data technology on a wide scale, much of this technology has become available through open-source modules, which helped datacentric business organizations to adopt Big Data technology for their specific needs.

The Use of Big Data Technology on a Large Scale and Associated Commercial Opportunities

In this subsection, the authors examine the use of Big Data technology on a large scale and related commercial opportunities. Big Data technology creates the potential for numerous commercial opportunities and innovations. The most common examples come from the major Internet companies for which data gathering and analysis, i.e. data analytics, are core competencies.

For instance, Marcus (2012) specified that Google has managed to develop an advanced spell-checker, not because of its competence in natural language processing or a unique approach to spelling, but due to collecting and analyzing a massive database of real spelling corrections. Other examples of commercial opportunities based upon similar technology usage scenarios include recommendation engines used by Amazon and Netflix that benefit from vast databases of user preferences on books and films to provide recommendations for future reference.

However, the commercial benefits of Big Data technology are not limited to Internet companies. Indeed, the McKinsey report that served to promote Big Data as a commercial strategy indicated the potential in healthcare, finance, and public services (Manyika et al., 2011; Lokshina et al., 2017).

For instance, in healthcare services distinguished by using fragmented datasets, Big Data technology has been proposed to help control increasing costs and accelerate research and development (R&D) processes in new pharmaceutical systems (Groves et al., 2013; Lokshina & Lanting, 2018b).

Specific features of commercial Big Data usage scenarios that have resulted in many advantages being less publicized as they are continuous improvements to existing processes. In other words, many Big Data usage scenarios are to continually improve existing processes rather than invent new procedures.

A good example derives from the transition in airlines from relying on pilots to provide estimated time of arrival (ETA) information to employing a data-driven system, which combines several data sources including weather, radar, and flight schedules (McAffee & Brynjolfsson, 2012). Formerly, busy pilots could make generally accurate estimations; however, 30% of their estimations have been over 5 minutes off. Currently, by combining multiple data points with the use of automated algorithms, airlines obtained the power to virtually eliminate gaps between estimated and actual arrival times, saving millions of dollars per year for the airports (McAffee and Brynjolfsson, 2012; Lokshina et al., 2019a).

What brings all these examples together and makes them Big Data technology is a commercial value created by large-scale data analysis, i.e. data analytics, and not only by data gathering, together with subsequent merging data in multiple datasets. While these examples can illustrate commercial benefits from Big Data, the wide-ranging ethical implications of Big Data technology that may result from commercial manipulation remain unclear. Although the focus is often placed on the volume of data gathered, Big Data technology is less about the large volume of data and more on the capacity to search, aggregate, and analyze numerous large datasets (Boyd & Crawford, 2012).

Up to this point, there are still not many concerns about the wide-ranging ethical implications of Big Data technology. Specifically, there are only a few concerns documented about the privacy implications of this technology in a digital society.

There is nothing new about privacy concerns in the public domain related to new technologies be it photography (Warren & Brandeis, 1890), personal computers (Zuboff, 1988), or the Internet (Nissenbaum, 2004). However, typically, the nature of these concerns has masked the reality of how new technologies have used and analyzed data (Mundie 2014; Marr, 2015).

Specifically, concerns about data privacy implications have been driven by fear over the active and visible data gathering process before passive and autonomous data gathering process, associated with Big Data technology (Jennex, 2017; Nunan & Di Domenico, 2017). Therefore, concerns over data privacy implications directly related to Big Data have been less discussed among the users of Big Data technology or described in the literature because the users of this technology have less insight into running data gathering and analysis scenarios (Lokshina et al., 2019a).

On the other hand, in this paper, the authors investigate the lack of perspective when the nature of privacy risk is unknown. Returning to the main notion of Big Data technology, the authors restate that large volumes of data might be gathered and stored because it is economical and there is a potential for future analysis of this data, i.e. intelligent analytics, to create commercial value. However, the data use scenarios are undetermined, therefore wide-ranging societal implications of Big Data technology remain unknown and nontransparent.

The Use of Big Data Technology for Managing COVID-19 Pandemic in a Digital Society

In this subsection, the authors consider the use of Big Data technology to manage the COVID-19 pandemic. On January 30, 2020, the World Health Organization (WHO) director-general declared the coronavirus disease 2019 (COVID-19) outbreak a public health emergency of international concern (PHEIC). Several weeks later, the outbreak has been categorized as a pandemic.

COVID-19 has already caused a hundred times more cases than the previous coronavirus-induced PHEIC, i.e. the 2002–2003 severe acute respiratory syndrome (SARS) outbreak, and the COVID-19 numbers are expected to grow. However, compared with the 2002–2003 outbreak, the COVID-19 disaster has taken place in a much more digitized and connected ecosystem.

The amount of all data produced by 2003 is generated today within a few minutes. Furthermore, advanced computational models like those using machine learning have shown great potential in tracing sources or predicting the future spread of infectious diseases (Scarpino & Petri 2019; Wheeler, 2019).

Therefore, it is imperative to leverage Big Data technology and intelligent analytics and put them to good use for the benefit of public health. Relying on digital data sources like data from mobile phones and other digital devices is of great value in outbreaks caused by newly discovered pathogens since precise data is still rare and hence forecasts less reliable and ineffective (Scarpino & Petri, 2019).

A recent study has shown the possibility of forecasting the spread of the COVID-19 outbreak by combining data from the Official Aviation Guide with data on human mobility from the WeChat app and other digital services owned by Chinese tech giant Tencent (Wu et al., 2020).

Mobile phone data have shown potential in predicting the spatial spread of cholera during the 2010 Haiti cholera epidemic while leveraging Big Data analytics demonstrated effectiveness during the 2014–2016 Western African Ebola crisis (Bates, 2017). However, during the recent epidemics, the large-scale collection of mobile data from millions of users, especially call data records and social media reports, have raised privacy and data protection concerns as well.

In 2014, privacy concerns urged the GSM Association, i.e. an industry organization that represents the interests of mobile network operators worldwide, to issue guidelines on the protection of privacy in the use of mobile phone data for responding to the Ebola outbreak (GSMA, 2014). In the information-intensive reality of 2020, ubiquitous data points and digital surveillance tools can easily worsen those concerns.

China, the country first and heavily affected by COVID-19, has reportedly used ubiquitous sensor data and health check apps to reduce the disease spread (The Economist, 2020). According to a New York Times report, there has been little transparency in whether and how this data was cross-checked and reused for surveillance purposes (Mozur et al., 2020). For instance, the report has stated that Alipay Health Code, an Alibaba-backed government-run app that supported decisions about who should be quarantined for COVID-19, also seemed to have shared information with the police (Mozur et al., 2020).

In Italy, the European country recording one of the largest numbers of the COVID-19 cases, a local data protection authority has been urged in March 2020 to issue a statement to clarify the conditions of lawful data use for mitigation and containment purposes. In its statement, the authority warned against the privacy-infringing collection and processing of data by non-institutional actors, i.e. private employers.

Two weeks later, the European Data Protection Board issued a statement on the importance of protecting personal data when used in the fight against COVID-19 and flagged specific articles of the General Data Protection Regulation (GDPR) that provide the legal grounds for processing personal data in the context of epidemics (European Data Protection Board, 2020). For instance, Article 9 allowed processing personal data "for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health," provided such processing is proportionate to the aim pursued, respects the essence of the right to data protection and safeguards the rights and freedom of the data subject.

As Big Data technology has become critical for managing the COVID-19 pandemic in the modern digital world, the conditions for responsible data collection and processing at a global scale must be clear. The authors consider that the use of digitally available data and algorithms for prediction and surveillance, i.e. identifying people who have traveled to areas where the disease has spread or tracing and isolating the contacts with infected people, has been of great importance in the fight against the COVID-19 pandemic.

However, it is equally important to use these data and algorithms responsibly, in compliance with data protection regulations, and with due respect for privacy and confidentiality. Failing to do so could undermine public trust, which would make people less likely to cooperate in providing, when necessary in the public interest, travel and contact data, follow public health guidance or recommendations, and eventually, would result more likely in worse health outcomes (Ward, 2017).

Careful data management practices should govern both data collection and processing levels. At the data collection level, the principle of proportionality regarding affected people should apply, which means that the data collection intensity must be proportional to the seriousness of the public health threat. This data collection intensity must be limited to only what is necessary to achieve a specific public health objective and be justified on a scientific basis (International Medical Informatics Association, 2011; Nuffield Council on Bioethics, 2015; PHG Foundation, 2015).

For instance, gaining access to data from personal devices for contact tracing purposes can be justified if it takes place within specific bounds, has a clear purpose, i.e. warning and isolating people who may have been exposed to the virus, and has no less invasive alternative, i.e. using anonymized mobile positioning data, is suitable for that purpose. On the other hand, “do it yourself” health surveillance like it was termed by the Italian data protection authority should be avoided as its contribution to public health safety is doubtful (Carr, 2020).

At the data processing level, data quality and security controls are required. Data integrity weaknesses that are common when data from personal digital devices are used can introduce minor errors in one or multiple factors, which in turn can have an outsized effect on large-scale predictive models (Wheeler, 2019).

Besides, data breaches, insufficient or ineffective anonymization, and biases in datasets can become major causes of distrust in public health services. Data privacy challenges not only are technological but also depend on political and judicial decisions (Carr, 2020; Holm et al., 2020).

Requesting or warranting access to personal devices for purposes like contact tracing can be more effective than simply leveraging anonymized mobile positioning data. However, convincing providers to allow access to or even assist in decrypting cryptographically protected data, similar to what has occurred in the 2016 United States Federal Bureau of Investigation (FBI) - Apple encryption dispute, may be counterproductive, especially when the agreements between international authorities and service providers would lack transparency and proportionality (Carr, 2020).

Similar trade-offs could apply to health apps that require users to register with their names or national identification numbers. National authorities should realize that because personal data might contain valuable and sensitive information about the social interactions and recent movements of infected people, these must be handled responsibly.

Overriding consent and privacy rights in the name of disease surveillance may fuel distrust and ultimately turn out to be disadvantageous. There have been reports that China’s digital epidemic control might have worsened stigmatization and public distrust.

This risk of distrust is even greater in countries where citizens place a much lower level of trust in their government like Italy, France, and the United States (OECD, 2019). Therefore, whenever access to these data sources is required and considered to be proportional, a society should be adequately informed (Holm et al., 2020).

Secrecy about data access and use should be avoided. Transparent public communication about data processing for the common good must be pursued.

For instance, data processing agreements should disclose what data has been transmitted to third parties and for what purpose. Reports from Taiwan have shown a promising way to leverage Big Data analytics to respond to the COVID-19 crisis without fueling public distrust (Wang et al., 2020; WMA, 2016). Taiwanese authorities integrated their national health insurance database with travel history data from customs databases to assist in case identification. Other technologies like QR-code scanning and online reporting have been used also for containment purposes (Wang et al., 2020).

These measures have been combined with public communication strategies involving frequent health checks and support for those under quarantine (Wang et al., 2020). The authors consider as more countries prepare to use digital technologies in the fight against the COVID-19 pandemic, Big Data technology, represented by large-scale data collection and intelligent analytics, is among the best methods, provided applied appropriately (Committee on Commerce, Science, and Transportation, 2020; Wheeler, 2019).

VIEWS ON ACCIDENTS ASSOCIATED WITH STATE-OF-THE-ART AND EMERGING TECHNOLOGIES

In this section, the authors present their views on accidents associated with state-of-the-art and emerging technologies. The authors note that technological development has always produced a divide between the economic advantages and societal implications. The question is not whether the technological progress in certain ways defines society, but the extent to which it does define society (Heilbroner, 1967; Lokshina & Lanting, 2018a).

Although currently, data appears to be the main element in the discussion about wide-ranging ethical implications of information, communications, and computing technologies, this was not always the case. By defining a notion of ethical overload, Coates (1982) identified wide-ranging ethical issues that could relate to computer technology. The concerns were associated with speech recognition, biotechnology, and the effect on the workplace structure from working at home that currently has become an important ethical overload in the COVID-19 reality, while there has been little reference to the role of data itself (Coates, 1982).

This example demonstrates the challenges related to predicting the consequences of state-of-the-art and emerging technologies, even when these technologies are well-understood themselves. The researchers have described these challenges by stating that under similar circumstances “the new immorality is to act in ignorance of future consequences” (Coates, 1982; Lokshina & Lanting, 2018a; Nunan & Di Domenico, 2017).

Although physical accidents and natural disasters have regularly taken place in society, the latest complex, embedded technologies require special attention. The authors consider that normal accident theory could help explain the consequences of these latest technologies, as well as the need for a better understanding of why disastrous situations and catastrophic events are likely to take place.

The Normal Accident Theory and Normal Accidents

In this subsection, the authors focus on the normal accident theory and associated normal accidents. An introduction of the normal accident theory has been launched from the report for the President’s Commission on The Accident at Three Mile Island after the 1979 nuclear accident at a power plant on that site (Perrow, 1981). The report concluded that the accident was not caused by discrete technical faults or human errors as it could have been expected but by a sequence of organizational failures that take place within systems like nuclear power plants that are complex, highly interactive, and tightly coupled (Perrow, 1981).

Therefore, the focus of the document was not on identifying who to blame, but rather on investigating how organizational failures could develop causing such an accident to take place (Perrow, 1999). Generally, systems in which normal accidents can take place have two key features: complex interactivity and tight coupling (Perrow, 1999; Dain, 2002; Guillén & Suárez, 2010; Palmer & Maher, 2010; Pidgeon, 2011).

Complex interactivity refers to the possibility of a chain of unknown and unplanned failures that take place in a sequence. This is contrary to linear interactivity which refers to a series of obvious, anticipated, and planned events that take place according to a known order (Perrow, 1999).

Linearity does not suggest that a linear system is simple, i.e. missing complexity, but rather that events are well-understood and linearly take place. For instance, a production line for pharmaceuticals or the flight of airplanes are linear systems as these processes can be easily explained; however, they are not simple at all.

An accident could also take place in these circumstances; however, based on its linear nature, consequences can be easily identified and fixed. For instance, when an unexpected situation takes place in a linear system like a production line, it can be simply identified and corrected by the staff. On the other hand, complex interactivity takes place when interactions are not fully understood by the staff who must take time-constrained decisions that are necessary to mitigate an accident (Nunan & Di Domenico, 2017).

Tight coupling refers to interactions that are discrete and take place very quickly. Therefore, the parts within an organizational system that is tightly coupled impact each other and trigger failures which transform into an accident (Nunan & Di Domenico, 2017).

In these circumstances, an accident is defined as a major, abnormal system failure, contrary to an incident defined as a minor, routine system failure. Therefore, an accident is a failure of the entire system and not of a certain part of the system; the outcome is that a chain of discrete system failures needs to arise before a system accident can take place.

A challenge for the normal accident theory is that application of the term "system" is not properly defined despite its notion is placed in a center of the theory (Shrivastava et al., 2009). However, applying this term, the normal accident theory can change its focus from a discrete failure, i.e. a failure of a certain part of the system including human error, to a failure of the entire system. This means that now an organizational system where the accident can take place is the primary cause of a normal accident, and not the staff (Cummings, 1984).

The term "normal accident" means that under specific circumstances an accident can become inevitable. This definition created an important disagreement between normal accident theorists and high-reliability academics since high-reliability thinkers suggested an option to design organizational systems that are complex, highly interactive and tightly coupled, and additionally, able to survive a normal accident (Roberts, 1990; Sagan, 1993; Rijpma, 1997; Perrow, 2008; Shrivastava et al., 2009).

Normal accident theorists responded that numerous incidents, even complex incidents like multiple system failures, can be prevented while a normal accident is inevitable due to a failure of the entire system (Perrow, 2009; Shrivastava et al., 2009). Therefore, the main message of normal accident theorists is not about avoiding the risk but on how to deal with the consequences of a normal accident.

This suggestion has put in question the overall relationship between technology and society; however, the notion of the normal accident theory remained the same. For instance, disasters caused by natural sources like pandemics known in the past and COVID-19; catastrophes prompted by industrial and technological sources like former accidents at nuclear power plants; and devastations caused by deliberate sources like terrorism and malicious acts of violence, have been inevitable and even growing: some of them could have been identified and narrowed, but never escaped (Perrow, 1999; Lokshina, 2001; Nunan & Di Domenico, 2017; Bloomberg, 2020; Pillinger, 2020; Science, 2020).

The notion that a normal accident is likely to take place in an organizational system has marked the normal accident theory against all other theories concerning technology risk. Therefore, in this paper, the authors apply this view to the wide-ranging ethical implications of Big Data technology.

Normal Accidents Associated with Big Data Technology

In this subsection, the authors define normal accidents associated with Big Data technology. The authors do not suggest that Big Data technology can create normal accidents. Instead, they assume that Big Data can play a critical role in shaping organizational systems where normal accidents are possible to take place. To demonstrate this assumption, they analyze Big Data technology against the fundamentals of the normal accident theory, i.e. complex interactivity and tight coupling.

Besides, the authors expand the normal accident theory by considering specific circumstances in emerging datacentric business organizations that can enable or be enabled by Big Data technology where normal accidents are likely to take place. They also investigate uncertainty and distrust created by Big Data technology in a digital society since the consequences of normal accidents would have to be weighed against the effectiveness and value achieved by society from the deployment of this technology.

Therefore, first, the authors state that Big Data technology displays features of a tightly coupled system. Such an argument seems like a counter-intuitive statement considering that the Internet is a prototype of a loosely coupled system that has been engineered to survive the destruction of some of its parts in a Cold war era.

However, there is a tightly coupled infrastructure, independent of the Internet but only using it, which stands behind Big Data technology and companies that tie together datacentric business organizations, dependent on this technology. The authors also note there is an ongoing transition from datacentric business organizations that could control their technology stack toward cloud computing where storage and processing power are provided as service.

While datacentric business organizations could develop software to gather data, to maximize efficiency they become more dependent on the deployment of large data centers operated by companies providing cloud services like Amazon, Microsoft, or Google. For all but very large business organizations, both the efficiency and flexibility offered by these multibillion-dollar data centers make it a prerequisite to apply third-party storage and likely computing technology (Lokshina et al., 2018a; Nunan & Di Domenico, 2017).

However, when there is a failure in data centers, the consequences are far more wide-ranging, unexpected, unpredictable, and dramatic than when datacentric business organizations use their own, distributed infrastructure that also, they can control. For instance, an accidental loss of data by a software developer who balanced traffic between many servers using Amazon Web Services on Christmas Eve 2012, resulted in a sequence of failures that made online services like Netflix unavailable on Christmas morning (Cockcroft, 2012).

In their explanation of the failure, Amazon has highlighted that only a small team of software developers had access to this data; at first, the software developers did not recognize an error; and initially, the entire software development team was mystified by an error message generated by the system (Amazon, 2012). However, the problem was resolved not with the use of technical resources, but by implementing a new process to ensure that changes to the system are double-checked to avoid accidental loss of data in the future.

Second, the authors note concerning loose coupling that a distributed method by which Big Data has been collected and stored created intrinsic complexity among numerous datacentric business organizations and various technologies. Returning to a previous example of Netflix's failure, the authors note that Netflix was dependent on the infrastructure of another datacentric business organization to deliver much of its content. Besides, Netflix was reliant on the business organization with whom it competed directly, i.e. Amazon's Instant Video service.

While a few datacentric business organizations can achieve vertical integration, most of them have become dependent on a growing network of third-party technology (Lokshina et al., 2019a; Nunan & Di Domenico, 2017). Taking a hypothetical example of a mid-size online retailing company, besides hosting its website, its dependence on other websites, i.e. third-party technology, could include payment services like credit card companies, Paypal and Apple Pay; integration with social media; content delivery networks to provide video content; integration with a third-party customer relation management (CRM) services; courier and delivery services; web analytics tools and advertising servers; etc. (Lokshina et al., 2019a; Lu et al., 2014). Besides complexity that is essential for data collection, there is also a growing concern over efforts to find qualified employees who can effectively analyze data (Brown et al., 2014).

Third, the authors state that a distinct feature of a normal accident and Big Data technology is a lack of shared understanding over risks. Datacentric business organizations that have been previously used as examples of a normal accident may be designated as systems that have understood what can create an accident, although they would likely disagree on the extent to which it can be considered as a normal accident.

Besides, datacentric business organizations that frequently collected large amounts of data were governments that are expected to be motivated to respect data privacy. Both the lack of commercial incentives and restricted access to data has implied an increased level of control over data by datacentric business organizations. Instead, commercial requirements that enable or are enabled by Big Data technology have been developed around two fundamental processes: data collection with consequent analysis, i.e. data analytics, and data sharing (Lokshina & Lanting, 2018b; Nunan & Di Domenico, 2017; Marr, 2015; Sedayao & Bhardwaj, 2014).

The authors consider this transformation has established a different culture with risks for data privacy. Besides, it has put pressure on processes that generate a commercial value which can likely trigger normal accidents that the authors investigate in this paper.

VIEWS ON DATA ACCIDENTS ASSOCIATED WITH BIG DATA TECHNOLOGY

In this section, the authors suggest that normal accidents associated with Big Data technology have particular features that make them different from physical accidents that have been discussed in the normal accident theory (Perrow, 1999; Snook, 2000). The term "data" is used on purpose instead of the term "information" as Big Data has been associated with unstructured data, i.e. data as opposed to structured data, i.e. information.

From a technical viewpoint, applying structure to data requires defining particular features of data to gather before a data collection process is started. For instance, it must be defined in advance that certain data is either an image, sound, or text; or it is a name, currency, time, etc.

The process of structuring data before its collection indicates that some consideration was granted to the final use of data. This may be a semantic viewpoint, but really helpful position to reinforce the main standard of Big Data technology informing that data can be gathered despite, and potentially, without knowledge of, the aim for its final use (Anjaria & Mishra, 2017; Lokshina & Lanting, 2018a; Nunan & Di Domenico, 2017).

Additionally, the authors state that data accidents associated with large scale data collection should be considered based on the following classification:

- Systemic data accident involving a systemic incomplete or unreliable collection of data and limitations in the interpretation of data, such as
  - Systemic data accident involving a systemic incomplete or unreliable collection of data, i.e.
    - Systemic data accident resulting from the systemic incomplete collection of data, and
    - Systemic data accident resulting from the systemic unreliable collection of data.
  - Systemic data accident resulting from systemic limitations in the interpretation of data.
- Data accident involving reliably collected and correctly interpretable data, such as
  - Data accident resulting from violations of data security compromising authorized access to data, i.e. "denial of service".
  - Data accident resulting from violations of data security compromising data privacy by an unauthorized read access to data, i.e.
    - Data unintended disclosure, violating data privacy by exposing confidential data, where disclosure has limited consequences for the organization and others, and
    - Data unintended disclosure, violating data privacy by exposing confidential data where disclosure has serious consequences for the organization and others.
  - Data theft compromising data privacy by unauthorized access to data, i.e.
    - Data leak, undetected unauthorized access violating data privacy,
    - Data disclosure blackmail, where violating data privacy is used as a lever to unlawfully obtain funds or other benefits or concessions, and
    - Data exposure, where violating data privacy is a tool to intentionally expose data.
  - Data accident resulting from violations of data security compromising data validity by unauthorized manipulative access to data, i.e.
    - Data integrity loss, undetected unauthorized manipulative access violating both data privacy and integrity,
    - Data destruction blackmail, where unauthorized manipulative access to data is used as a lever to unlawfully obtain funds or other benefits or concessions, and
      - Data integrity loss, unauthorized, irrecoverable manipulative access violating both data privacy and integrity.

Several important examples of data accidents associated with large scale data collection that demonstrate the relevance and use of the proposed classification are considered in the following subsections. These examples include systemic data accidents involving a systemic incomplete or unreliable collection of data and limitations in the interpretation of data corresponding to the COVID-19 data collection, interpretations, and decisions that are considered as special case data accidents associated with Big Data technology; as well as data accidents involving reliably collected and correctly interpretable data, containing undetected unauthorized access violating data privacy like in the Manning and Snowden data leak scenarios, along with violations of data security compromising both data privacy and integrity by unauthorized manipulative access to data like in the Johnson & Johnson and the World Anti-Doping Agency (WADA) data manipulation scenarios.

Systemic Data Accidents

In this subsection, the authors suggest the COVID-19 data, the following interpretations, and decisions derived from these interpretations should be defined as systemic data accidents associated with Big Data technology. These should be considered as special case data accidents because the problems in large scale data collection and interpretations are known, but considered inevitable; and the resulting interpretations and extrapolations are considered “best-effort”, “the best we have”, or “the all we get”.

Additionally, the COVID-19 data is mostly gathered voluntarily; therefore, it is possibly or even likely incomplete. Also, the COVID-19 data can be incomplete due to errors before data is collected, i.e. some people dying at home or in an elderly home which may be considered rather as “natural death” than as the COVID-19 fatalities. Additionally, the COVID-19 data can be incomplete because of the initial anonymization of data which may suppress both severity and historical data, i.e. the number of people tested may be unknown when only the number of tests is recorded.

Besides, there is feedback between collected data, interpretations, extrapolations, resulting actions, and the effect on the contamination and, consequently, on the causes behind the sources of properly collected data, which constitutes a difficulty to assess the challenge. This effect on the data sources makes it similar to non-systemic data accidents like scenarios of suspected silent manipulation, i.e. when manipulation is suspected but both nature and scope are unknown. Nevertheless, it is critically important to define the COVID-19 data, the following interpretations and decisions derived from these interpretations as data accidents since implications of problems with large scale data collection and interpretations are neither preventing local, regional, and national authorities nor stopping datacentric business organizations from making crucial decisions such as numbers of required hospital beds or intensive care unit (ICU) beds based on extrapolations from this data.

The feedback between extrapolations, strategy, actions, reduction in COVID-19 spreading, and for instance, the number of required ventilators is also a problem. The number of ventilators, being built in the United States is likely based on a gross over-estimation of real need, without considering changes in both the therapy and spread due to measures to reduce the spreading of the COVID-19 pandemic.

Besides, in assessing the COVID-19 fatality, certain confusion is caused by failing to distinguish COVID-19 cases from COVID-19 infections. The case fatality rate is computed by dividing the number of fatalities by the total number of confirmed cases. But to obtain an accurate COVID-19 fatality rate, the number in the denominator should be the number of infections instead of the number of confirmed cases. In the early days, only part of infected people was identified as confirmed cases that resulted in an inaccurate fatality rate which drove public policy, and sowed fear, and triggered the widespread lockdowns.

The lockdowns have serious effects on public health. The United Nations (U.N.) estimated that 130 million additional people will starve this year because of economic damage resulting from the lockdowns (Givetash, 2020; U.N., 2020). Also, the parents stopped bringing their children in for immunization against diseases like diphtheria, pertussis, and polio. Many patients who have had cancer and needed chemotherapy stopped coming in for treatment, and the others skipped recommended screening or diabetic monitoring. Additionally, the lockdowns have severe psychological effects, especially among young adults and children, who are denied much-needed socialization.

In these examples, the local, regional, and national authorities are aware of problems in the COVID-19 data collection and resulting interpretations and extrapolations. Therefore, the main issue raised by these instances is that since authorities have approved the COVID-19 data manipulation given incomplete data collection and limitations in its interpretation that impact both the derived decisions and public opinion, then how society can expect that commercial applications of Big Data technology will not suffer from similar situations.

Noncommercial Data Accidents

In this subsection, the authors consider noncommercial data accidents associated with Big Data technology. First, keeping in mind (Perrow, 1999), the authors examine two recent examples of data loss accidents that have been described in the literature. Both examples are instances of information leaks that took place through two individuals associated with United States government organizations.

One case is associated with Bradley (Chelsea) Manning who has been leaking United States diplomatic cables to Wikileaks. The other case is associated with Edward Snowden who has been leaking classified National Security Agency (NSA) data and procedures to various media organizations (Nunan & Di Domenico, 2017; Shane et al., 2017). These instances do not illustrate direct commercial use of Big Data; however, they can help investigate issues around key features of data accidents.

The authors note that in the Manning and Snowden scenarios the term "accident" is used to describe outcomes when original data losses have been caused not by "accidental" system failures but unexpected interactions in organizational systems. Nevertheless, the authors consider these scenarios were accidents as they were caused by the failures of entire systems designed to prevent data losses, not by discrete events (Szoldra, 2016; Shane, 2019; Shane et al., 2017).

Like with a previous scenario given for Amazon (Cockcroft, 2012), data accidents have taken place by a combination of unexpected failures. In the Manning case, the Wikileaks incident has taken place over a junior staff member, Bradley (Chelsea) Manning, who has copied data, primarily diplomatic cables, to a CD-ROM and then contacted media (Leigh, 2010; Goel, 2017; Shane et al., 2017).

In the Snowden case, the precise method by which Edward Snowden has retrieved data from the NSA databases remains unknown, though it included copying data from a computer to a USB memory drive (Waterman, 2013; Shane, 2019; Shane et al., 2017). This simple method has not been taken into consideration because computers used to store sensitive data and confidential information are expected to have USB ports disabled.

What is important, the fact that any of the Manning and Snowden data accidents took place unidentified until data was leaked to media. While accessing and downloading data, the full meaning of Snowden leaks has been unclear due to uncertainty in scope and Snowden's intention to publicly distribute.

In both examples, authorities have been unaware of data loss until it was leaked to the media. Besides, it has not been performed by a foreign government or a criminal organization that involves multiple actors. Instead, it has been performed by a single junior staff member who conducted regular organizational tasks; however, could quickly access and transfer large data volumes. Therefore, the authors suggest the main issue raised by these instances is that if single junior staff members in high-security organizations could access, download and circulate large data volumes, with much of it, like in the Snowden scenario, classified "Top Secret", then how society can expect that commercial applications of Big Data technology will not suffer from similar situations.

Next, the authors examine two recent examples of data manipulation accidents that have been described in the literature. Both examples are instances of database breaches when data has been accessed and altered in place, posing a serious threat for involved datacentric business organizations and concerned individuals (Carr, 2020; Lokshina & Lanting, 2018b).

One scenario is associated with data manipulation in one of Johnson & Johnson's insulin pumps in 2015, which enabled hackers to administer an overdose on users with insulin. The other scenario concerns Russian hackers who breached the World Anti-Doping Agency (WADA) systems in 2016, changed in place medical data of many athletes, and then circulated this altered data damaging their reputation (Carr, 2020). These instances cannot illustrate direct commercial use of Big Data; however, they can help investigate issues around key features of data accidents.

Healthcare and pharmaceutical industries are at risk of data manipulation accidents when patients' lives are affected by tampering with data on what medications the patients are prescribed, how often they should take the medications, and what allergies the patients have. When healthcare data is manipulated on a large scale causing a great deal of harm, this is considered cyberterrorism (Carr, 2020).

Data manipulation accidents not only prompt datacentric business organizations to lose their profits and customers, or patients, or users, but also show how cybersecurity and public health and safety are increasingly connected. In the Johnson & Johnson scenario, a security vulnerability in one of its insulin pumps enabled malicious actors to overdose several patients with insulin (Carr, 2020).

Data manipulation accidents create opportunities to change reality for smear campaigns. By accessing data in organizational systems and altering it in place, data manipulation accidents are used to destroy the personal and professional reputations of various individuals. In the Russian hackers' scenario, the WADA systems have been breached and the medical data of many famous athletes have been changed in place and then released (Carr, 2020), with a clear objective, masking usage of forbidden enhancing products and procedures by others.

Additionally, data manipulation accidents are intended to influence public opinions and impact decisions. When data is changed, inevitably, choices made based on this data are also manipulated. By performing calculated alterations to data, the decision-making process of people who use this data can be strategically controlled by malicious actors (Carr, 2020).

In the global race for the development of a vaccine, understanding vaccine research and other specifics about the pandemic has become a main target for intelligence agencies around the world. Currently, "state-backed hackers" more likely to focus on highly sensitive information such as the COVID-19 vaccine research rather than just organizational data. For instance, throughout 2020, a Russian state-backed hacking group known as APT29 targeted various medical organizations involved in the COVID-19 vaccine development in Canada, the United States, and the United Kingdom, to access data in their organizational systems to use, leak, or manipulate (Fox & Kelion, 2020). In September 2020, the international vaccine supply chain was targeted by cyber-espionage to access data in its organizational systems to use, leak, or manipulate. The attackers' identity was unclear, but the sophistication of their methods indicated a nation-state, according to IBM. It followed warnings from governments, including the United Kingdom, of countries targeting aspects of vaccine research (Corera, 2020).

In these instances, data manipulation has been performed by a foreign government or a criminal organization, involving multiple actors. Therefore, the authors suggest the main issue raised by such instances is that if malicious actors could access data in organizational systems, alter data in place and then circulate, then how society can expect that commercial applications of Big Data technology will not suffer from similar situations.

Commercial Data Accidents

In this subsection, the authors consider commercial data accidents associated with Big Data technology. The authors note this is difficult to directly estimate the potential for commercial data accidents because firstly, unlike in the Manning and Snowden scenarios discussed earlier, most commercial data accidents are not made public. Secondly, commercial data accidents have the key features that distinguish them from traditional "physical accidents" discussed earlier as well. These features make data accidents very difficult to identify, assess, and remedy.

The first difference is the absence of physical artifacts that makes data accidents extremely difficult to identify when they take place. In prior noncommercial data accident scenarios, the first indication that data accidents took place became a fact that data was used at the time it was used. Therefore, if a sequence of incidents, i.e. minor failures involved in a nuclear accident, could be identified at a physical location, data accidents can be only discovered by the following data uses.

The second difference is that the consequences of data accidents are neither geographically specific nor geographically positioned. Datacentric business organizations that collect data can be nationwide, at least in standard legal terms, but the Internet and Big Data technology enable them to collect data on people from around the world with very few limitations.

Once data is lost, legally, or illegitimately, it can spread all over the world very quickly. An absence of geographical borders has some concerns in terms of authority when it applies to both reducing the consequences of data accidents and stopping the spreading (Allen and Overy, 2013).

The third difference is that a specific time when the consequences of data accidents can be identified is very difficult to foresee. After physical accidents took place, the consequences can be estimated at a certain level. However, after data accidents took place, the consequences can be only noticed due to the following spreading of this data and its analysis (Lokshina & Lanting, 2018a; Nunan & Di Domenico, 2017). And even then, the consequences could last for years in the future as new methods of data analysis can become available.

Application of Normal Accident Theory to Real Data Accident Scenarios

In this subsection, the authors inform that certain researchers believe that the normal accident theory has potential limits to be effectively applied to real data accident scenarios associated with Big Data technology. These researchers mention that the main limitation is the lack of real normal accidents proposed by the normal accident theory. They consider that many datacentric business organizations are highly reliable and can avoid normal accidents (La Porte & Consolini, 1991; Leveson et al., 2009; Leveson, 2016). For instance, Shrivastava et al. (2009) indicated that the absence of system accidents foreseen by Perrow (1999) demonstrated that some potential scenarios of system accidents could be at some level attempts to manipulate data to promote the normal accident theory (Shrivastava et al., 2009; Leveson, 2016).

Many researchers suggest that the Y2K scenario can be used to illustrate the potential limits for the application of normal accident theory to real data accidents. The Y2K problem has been a normal accident, which in many cases was designed out before it took place. Additionally, the Y2K scenario was a normal accident that demonstrated key features comparable to Big Data technology. Therefore, for many researchers, the Y2K scenario became the first example of considering the moral and societal implications of state-of-the-art and emerging technologies (Nunan & Di Domenico, 2017).

However, the real implications of the Y2K scenario were rather insignificant when compared with the anticipated extreme outcomes (MacGregor, 2003). Therefore, the authors can agree that the case for data accidents is at some level compromised by the shortage of real data accidents.

Perrow (2008) evaluated analogous consequences in the context of the nuclear industry (Perrow, 2008). The interpretation was that societal pressure despite only a few accidents that took place, helped develop a regulatory system that made it hard to establish nuclear power as an energy source. Therefore, apart from France and Japan, nuclear power continues to remain a restricted power source. Additionally, this regulatory system served to create a very different management structure with high costs, a moratorium on new nuclear power plant construction, and government ownership (Perrow, 2008). Besides, this regulatory system revealed considerable costs to decommission nuclear power plants and sites.

Therefore, keeping in mind the Fukushima nuclear accident in Japan, the authors assume that Perrow (1999) simply suggested insufficient timeframes. The authors inform that the notion remains the same, i.e. the normal accident theory can be effectively applied to real data accidents associated with Big Data technology. Additionally, the authors notify that underestimating the potential consequences of state-of-the-art and emerging technologies due to unexpected and inevitable future is unwise and risky.

Successful Risk Mitigation Strategies

In this subsection, the authors discuss what should be done if data accidents would take place. For physical systems where the risk is catastrophic and the consequences of failures can greatly offset the potential advantages, Perrow (1999) suggested that such systems must be terminated (Perrow, 1999). However, Perrow (1999) also admitted that in practice this recommendation could be unfeasible and less productive (Perrow, 1999; Lokshina, 2001; Nunan & Di Domenico, 2017).

In his theory, Perrow (1999) assumed that the risk of emerging technologies can be assessed (Perrow, 1999; Lokshina, 2001). However, given the abstraction of data accidents, a comprehensive and sensible risk analysis of data accidents is from very difficult to impossible. Additionally, the greater integration of data gathering and analysis processes in the social life indicates that the future direction of Big Data technology moves towards even bigger data. However, the authors inform that the volume and scope of Big Data and the presence of technology that potentially enables more effective data analysis and storage will not increase the risk of data accidents.

Instead, the risk of data accidents can arise from uncertainty and distrust that surround the data gathering process creating a modern “gold-rush” situation when datacentric business organizations with the largest data collections can win, even when the commercial value of volume data remains unknown.

One risk mitigation strategy can be to accept the costs derived from reducing data privacy and advantages obtained from reducing public information asymmetry. However, it cannot be achieved without compromising the legal and regulatory system that governs customer data collection and usage (Committee on Commerce, Science, and Transportation, 2020; Lokshina et al., 2019a; Nunan & Di Domenico, 2017). The entire history of governing commercial activities associated with consumers demonstrates that the regulation takes place “with a rear-view mirror” and with anticipated retroactive consequences like the regulation of pharmaceuticals and tobacco. Therefore, changes made to the regulatory system may take place retrospectively.

Another, more challenging risk mitigation strategy can be to establish a relationship between data gathering and regulation, although the pace of regulation is often pushed by the lobbying power of concerned industries and effective enforcement. First, there is an understanding among policymakers that the current regulations are inadequate given that consumers themselves are much involved in data collection and use (Lokshina & Lanting, 2018a). Second, there are concerns about the enforcement of regulation in the setting when data show increasingly cross-border features and the relevance of regulation is unclear (Nunan & Di Domenico, 2017).

Third, due to social networks and other online services operating like service organizations and presenting themselves as public services while trying to avoid appropriate regulation and legislation, they should be possibly subject to regulation like regular business organizations (Bygrave, 2014). Finally, the regulation itself is going through a transformation process when the regulation becomes not only more common but also the spirit of regulation changes (Bygrave, 2014; Lokshina et al, 2019a; Nunan & Di Domenico, 2017).

DISCUSSION AND CONTRIBUTIONS

In this section, the authors provide a discussion of the wide-ranging ethical implications of Big Data technology in a digital society, followed by the major contributions of this paper in the literature as well as in general information technology (IT) and business domains. Therefore, in attempts to foresee future organizational environments or business ecosystems, the authors accept the risk as when discussing potential data accidents, the authors use hypothetical scenarios and potential implications.

The authors do not intend to predict future data accidents because attempts to predict would be misinterpreting the inevitable and unexpected nature of data accidents. Besides, the authors consider that potential data accidents that can be foreseen can be monitored and mitigated as well.

Additionally, the authors emphasize that the term "normal" is used for data accidents to indicate that data accidents are inevitable. Therefore, as the normal accident theory suggests, data accidents have the potential to take place. However, the normal accident theory does not determine when and how these data accidents can take place.

The term "mitigation" does not refer to specific instances, i.e. to preventing another data leak, data manipulation, or financial meltdown; but instead, it suggests reassessing basic organizational environments, where normal accidents have the potential to take place. The mitigation strategies must address essential requirements to simplify and decouple these organizational systems, or business ecosystems (Leveson, 2017; Nunan & Di Domenico, 2017).

The authors focus on the notion of normal accident theory and Big Data technology concerning mitigation strategies. The mitigation strategies require to change both the acceptance measure and the degree of organizational understanding that creates tolerance only when potential consequences of data accidents associated with Big Data technology have been addressed, also keeping in mind that the features of potential data accidents are distinct and uncertain.

Therefore, the authors suggest that Big Data technology creates externalities that generally can be influenced by relevant regulatory processes. However, the distinct and uncertain features of data accidents together with technical and commercial limitations associated with Big Data technology, develop challenges for effective regulation (Lokshina et al., 2019a). Accordingly, the authors demonstrate the notion by using the entire process of updating data protection legislation in the European Union to replace the currently outdated regulation from the 1990s and to standardize legislative approaches to data privacy across the European Union and associated countries (Ashford, 2014; Jain et al., 2016; Lokshina et al., 2019a).

First, a challenge of regulation across borders has led to attempts to apply regulation to the major, United States-based, social networks. However, this resulted in problems ranging from the practical concerns over the limitations of the remit of national authorities to the threats of a trade war (Farivar, 2014; Molok et al., 2012; Nunan & Di Domenico, 2017).

Second, the attempts to regulate the ability of datacentric business organizations to gather data in a format that enables Big Data analysis, i.e. data analytics, have raised issues about the level at which society is supportive or actively opposed to such data collection. For instance, if the new European Union regulation could result in Google and Facebook limiting their features or even withdrawing from the European market, how likely the consumers would appreciate improved data privacy for themselves and others or rather criticize the European Union for limiting the scope of their online activities? (Lokshina & Lanting, 2018a, b).

Accordingly, the authors suggest that datacentric business organizations may have to change their privacy-related practices as these look incomplete and, maybe, insufficient to meet the requirements of the new European General Data Protection Regulation (GDPR) - Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons about the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (Official Journal of the European Union, 2016; European Commission, 2018; Lokshina et al., 2019a).

The authors note that there are proposals for different approaches that put collected data at and under the control of the users. For instance, the approach proposed by Tim Berners-Lee with the SOLID project introduces the concept of personal online data stores (pods), hosted anywhere the user desires (CSAIL, 2015). Applications that are authenticated by Solid, a decentralized data-linked system, are permitted to request data if the user has given the application permission.

The authors also inform that during the COVID-19 pandemic, the large-scale data collection and intelligent analysis must respect data privacy and public trust. Best practices should be identified to maintain responsible data collection and data processing standards at a global scale, and need to be applied to all types of related data collection (International Medical Informatics Association, 2011; Nuffield Council on Bioethics, 2015; PHG Foundation, 2015; WMA, 2016).

For instance, in the COVID-19 reality when contact tracing of people is implemented to trace contacts with people found positive or suspected of having the virus, which is among key strategies to contain the spreading, it is important to maintain data privacy principles while establishing contact tracing procedures as part of COVID-19 detection. Therefore, the main data privacy aspects should be consent, transparency, purpose limitation, proportionality, and security safeguards (International Medical Informatics Association, 2011; Nuffield Council on Bioethics, 2015).

The term "consent" is used to obtain agreement on providing, when required, contact tracing data to collection points, in the same way as forms or employee declarations. The term "transparency" is used to inform of how data can be used as part of the contact tracing process. The term "purpose limitation" is used to notify that personal information will be treated in strict confidence and can be used only for the purposes to identify people to who the virus may have spread. The term "security safeguards" is used to ensure that security measures are taken to protect personal data from being shared unnecessarily or leaked. Implementing these measures would support a secure contact tracing process and help maintain data privacy on a global scale (International Medical Informatics Association, 2011; Nuffield Council on Bioethics, 2015).

However, the authors are worried about data security problems when private data could be exploited by malicious actors and stalkers. Private data can be anonymized and encrypted, but these are not sufficiently secure measures to protect private data that is stored in a centralized location and accessed remotely. Additionally, there is a chance of spoofed versions of contact-tracing software to be distributed. For instance, agreements with Apple and Google can mitigate the issue but cannot eliminate the problem (Carr, 2020; Holm et al., 2020).

Given the current understanding of COVID-19, contamination can occur up to two weeks before people become symptomatic and can be transmitted via objects and air. To preserve privacy, not only the access to but also the retention time of contact data should be respected; contact data must be deleted after a recommended time keeping in mind the incubation time and the resulting delay in discovery of possible virus transfer.

Moreover, the private data can be used for other purposes, for instance, by intelligence agencies. In normal circumstances, such invasive contact tracing would be unacceptable. However, a temporary increase in surveillance should be tolerated representing a compromise between privacy and public health to safeguard public health, public health infrastructure, businesses, and business ecosystems (Carr, 2020; Holm et al., 2020).

Third, the insight into implications of Big Data technology has become more challenging due to the issues of data ownership, i.e. the rights to use data after datacentric business organizations have it collected (Lokshina & Lanting, 2018a, b). The ownership typically depends on the method by which data has been collected, and the on-going discussion in terms of regulation aims to address asymmetries which are present here (Committee on Commerce, Science, and Transportation, 2020; Official Journal of the European Union, 2016; European Commission, 2018; Nunan & Di Domenico, 2017).

For instance, when people sign up for online services like social media, they provide a "blanket consent" that their private data can be used for a frequently unnamed range of purposes for an unknown length of time. Such sign-ups represent a legal consent that permits the use of data and, basically, the ownership or at least the right to use and trade for the entire length of time as these people remain the members of these online services. However, the use of data by these online services unlikely meets the requirements of traditional academic research that is based on informed consent (Lokshina & Lanting, 2018a). For instance, the continued use of collected user data after ending the membership is an issue that should be properly resolved.

Additionally, while people expect to have at least limited ownership rights on personal data directly collected by datacentric business organizations, another important issue with Big Data technology is autonomous data collection. For instance, the commercial value of data gathered autonomously by sensors in the cars, houses, and public buildings cannot be promptly understood even by datacentric business organizations themselves (Lokshina et al., 2019a, b). Additionally, the ownership discussion over commercial data can be unrealistic as people may not know that personal data has been collected.

Fourth, the authors underline the need to know if the collected data is personal data or not. This issue is very important as data protection legislation typically does not provide the rights over personal data to people after it was aggregated and anonymized. (Committee on Commerce, Science, and Transportation, 2020; Lokshina et al., 2019a; Nunan & Di Domenico, 2017).

Fifth, the authors note that a paradox of Big Data technology is that the methods to anonymize datasets, can be also used to deanonymize the same datasets. This paradox creates further ethical issues associated with both data privacy and data ownership that earlier did not exist. Despite data anonymization and unlike in traditional data analysis that aims to produce aggregate insights, the commercial benefits of Big Data technology are heavily driven by the need to analyze personal data (Committee on Commerce, Science, and Transportation, 2020; Lokshina & Lanting, 2018b; Nunan & Di Domenico, 2017).

Besides, the authors recall the ethical issues related to the COVID-19 pandemic data collection and resulting interpretations and extrapolations. The concern is that since authorities approved the COVID-19 data manipulation given incomplete data collection and limitations in its interpretation that impact both the derived decisions and public opinion, then how society can expect that commercial applications of Big Data technology would not suffer from similar situations.

Sixth, the authors suggest that the following two mitigation strategies should be considered. One approach is a behavioral transformation in datacentric business organizations addressing the wide-ranging ethical implications of Big Data technology discussed earlier in this section. Additionally, the authors remind about a difference between the major industrial companies with commercial strategies in general and those datacentric business organizations that have data collection and analysis as a goal. Therefore, another approach is to change the focus of the industrial companies that do not plan to become data conglomerates away from Big Data technology.

Seventh, the authors inform about an issue related to the role of consumer behavior itself. For instance, datacentric business organizations like Google and Facebook have already transitioned to the philosophy that at some level respects the value of data privacy in consumer decision-making (Lokshina & Lanting, 2018a; Nunan & Di Domenico, 2017). Additionally, the literature suggests that consumers consider data privacy and security before they share personal data online, especially with fishy strangers (Johnson et al., 2012). Therefore, given the commercial drivers behind Big Data technology, this mentioned in the literature decreased consumer tolerance to online data collection indicates the potential to move away from the Big Data economy (Lokshina & Lanting, 2018a; Nunan & Di Domenico, 2017).

The authors recap that the normal accident theory has potential limits for its effective application to data accidents associated with state-of-the-art and emerging technologies discussed earlier. The first limit is that normal accident theorists seemed to be partially wrong in the past, especially with the Y2K scenario (MacGregor, 2003). However, this was not a weakness of the normal accident theory. On the contrary, this was a result of attempts to use the theory to predict future data accidents and prevent potential consequences instead of an effort to explain.

In his defense, Perrow (2009) underlined that many researchers did their analysis of emerging technologies wrong. Therefore, the Y2K scenario demonstrated challenges introduced by the risk analysis produced by researchers without a sufficient understanding of the distinct features of these emerging technologies (Perrow, 2009; Nunan & Di Domenico, 2017).

The second limit is related to a view that construct "Big Data" should be used only for reference purposes. The literature also suggests a different view that construct "Big Data" is an ill-defined phrase with an unclear and confusing application that creates an association with technological advancement without the need to properly understand its nature (De Mauro et al, 2016; Ylijoki & Porras, 2016).

The authors consider the use of the term “Big Data” has currently surpassed its understanding (Lokshina & Lanting, 2018a; Nunan & Di Domenico, 2017). However, the term emerged from a practical need to recognize a transformation in data collection and analysis processes that already took place in a digital society. Therefore, in this paper, the authors used the term when they discussed the organizational strategies and new organizational behaviors in business ecosystems to address the wide-ranging ethical implications of Big Data technology in the digital society instead of inventing new terms.

Finally, this paper makes the following major contributions in the literature as well as in general information technology (IT) and business domains. First, the authors investigated the wide-ranging ethical implications of Big Data technology in a digital society. Second, the authors determined that strategies behind Big Data technology require organizational systems, or business ecosystems, that leave them vulnerable to accidents associated with its commercial value and known as data accidents. Third, the authors established that data accidents have distinct features and raise important concerns about data privacy in a digital society, particularly, in the time of the COVID-19 pandemic. Fourth, the authors created a classification of potential data accidents and proposed successful risk mitigation strategies.

CONCLUSION

In conclusion, the authors note that exponential growth in the commercial use of the Internet has dramatically increased the volume and scope of data gathered by datacentric business organizations. Big Data emerged as a term to summarize both the technical and commercial aspects of this growing data collection and analysis processes. Previously, much discussion of Big Data was focused on its transformational potential for technological innovation and efficiency. However, less attention was devoted to its wide-ranging ethical implications beyond generating commercial value.

In this paper, the authors investigated the wide-ranging ethical implications of Big Data technology in a digital society. The authors informed that strategies behind Big Data technology require organizational systems, or business ecosystems, that leave them vulnerable to accidents associated with its commercial value and known as data accidents. These data accidents have distinct features and raise important concerns about data privacy in a digital society, especially during the COVID-19 pandemic. In this paper, the authors suggested classification and methods mitigate the risk of potential data accidents.

ACKNOWLEDGMENT

The authors would like to extend their gratitude to the anonymous reviewers for their time, insights and review, assistance, and patience during the preparation of this paper.

REFERENCES

Allen & Overy. (2013). *Big Data—Annual review 2013—Allen & Overy*. Retrieved from https://www.allenovery.com/publications/en-gb/annualreview2013/global-local/Pages/Big-data.aspx.

Amazon. (2012). *Summary of December 24, 2012, Amazon ELB service event in the US-East Region*. Retrieved from https://aws.amazon.com/message/680587/

Anjaria, K., & Mishra, A. (2017). Information leakage analysis of software: How to make it useful to IT industries? *Future Computing and Informatics Journal*, *2*(1), 10–18. doi:10.1016/j.fcij.2017.04.002

Ashford, W. (2014). *Infosec 2014: Act now, but no new EU data protection law before 2017, says ICO*. Retrieved from https://www.computerweekly.com/news/2240219908/Infosec-2014-Act-now-but-no-new-EU-data-protection-law-before2017-says-ICO

Bambauer, D. (2014). Ghost in the network. *University of Pennsylvania Law Review*, *162*(5), 1050.

Bateman, C., Valentine, S., & Rittenburg, T. (2013). Ethical decision making in a peer-to-peer file-sharing situation: The role of moral absolutes and social consensus. *Journal of Business Ethics*, *115*(2), 229–240. doi:10.1007/s10551-012-1388-1

Bates, M. (2017). Tracking Disease: Digital Epidemiology Offers New Promise in Predicting Outbreaks. *IEEE Pulse*, *8*(1), 18–22. doi:10.1109/MPUL.2016.2627238 PMID:28129137

Bloomberg. (2020). *The Coronavirus May Be "Disease X" Health Experts Warned About*. Retrieved from https://www.bloombergquint.com/business/coronavirus-may-be-the-disease-x-health-agency-warned-about

Boyd, D., & Crawford, K. (2012). Critical questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information Communication and Society*, *15*(5), 662–679. doi:10.1080/1369118X.2012.678878

Brown, B., Court, D., & McGuire, T. (2014). *Views from the front lines of the data-analytics revolution.* Mckinsey.com. Retrieved from https://www.mckinsey.com/insights/business_technology/views_from_the_front_lines_of_the_data_analytics_revolution

Bygrave, L. (2014). *Data privacy law: An international perspective*. Oxford University Press. doi:10.1093/acprof:oso/9780199675555.001.0001

Carr, R. (2020). *Data Integrity Attacks: Is Data Manipulation More Dangerous Than Theft?* Zettaset. Retrieved from https://www.zettaset.com/blog/data-integrity-attacks-data-manipulation-more-dangerous/

Coates, J. (1982). Computers and business? A case of ethical overload. *Journal of Business Ethics*, *1*(3), 239–248. doi:10.1007/BF00382776

Cockcroft, A. (2012). *The Netflix tech blog: A closer look at the Christmas Eve outage*. Retrieved from http://techblog.netflix.com/2012/12/a-closer-look-at-christmas-eve-outage.html

Committee on Commerce, Science, and Transportation. (2020). *Committee Announces Paper Hearing on Big Data and the Coronavirus*. Press Release. Retrieved from: https://www.commerce.senate.gov/2020/4/committee-announces-paper-hearing-on-big-data-and-the-coronavirus

Corera, G. (2020). *Coronavirus: Hackers targeted Covid vaccine supply 'cold chain'*. Retrieved from https://www.bbc.com/news/technology-55165552

CSAIL. (2015). *Web inventor Tim Berners-Lee's next project: a platform that gives users control of their data.* Computer Science & Artificial Intelligence Lab. Massachusetts Institute of Technology. Retrieved from: https://www.csail.mit.edu/news/web-inventor-tim-berners-lees-next-project-platform-gives-users-control-their-data

Cummings, L., & Perrow, C. (1984). Normal accidents: Living with high-risk technologies. Book review. *Administrative Science Quarterly*, *29*(4), 630–632. doi:10.2307/2392945

Dain, S. (2002). Normal Accidents: Human Error and Medical Equipment Design. *The Heart Surgery Forum #2002-180891, 5*(3), 254–257. Retrieved from www.hsforum.com/vol5/issue3/2002-180891.html

De George, R. (2003). *The ethics of information technology and business*. Blackwell Publishing. doi:10.1002/9780470774144

De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. *Library Review*, *65*(3), 122–135. doi:10.1108/LR-06-2015-0061

Diamond, J. (1997). *Guns, Germs, and Steel: The Fate of Human Societies*. W. W. Norton & Company.

European Commission. (n.d.). *2018 Reform of EU data protection rules*. Retrieved from https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en#background

European Data Protection Board. (2020). *Statement by the EDPB Chair on the processing of personal data in the context of the COVID-19 outbreak*. Retrieved from https://edpb.europa.eu/news/news/2020/statement-edpb-chair-processing-personal-data-context-covid-19-outbreak_en

Farivar, C. (2014). *EU data protection reform could start 'trade war', US official says (Wired UK).* Retrieved from https://www.wired.co.uk/news/archive/2013-02/01/eu-data-protection-us-trade-war

Fink, M., Harms, R., & Hatak, I. (2012). Nanotechnology and ethics: The role of regulation versus self-commitment in shaping researchers' behavior. *Journal of Business Ethics*, *109*(4), 569–581. doi:10.1007/s10551-012-1431-2

Foundation, P. H. G. (2015). *Protecting the people behind biomedical Big Data*. Retrieved from https://www.phgfoundation.org/news/16565/

Fox, C., & Kelion, L. (2020). *Coronavirus: Russian spies target Covid-19 vaccine research*. Retrieved from https://www.bbc.com/news/technology-53429506

George, G., Haas, M., & Pentland, A. (2014). Big Data and management. *Academy of Management Journal*, *57*(2), 321–326. doi:10.5465/amj.2014.4002

Givetash, L. (2020). *U.N. warns of 'hunger pandemic' amid threats of coronavirus, economic downturn*. Retrieved from https://www.nbcnews.com/news/world/u-n-warns-hunger-pandemic-amid-threats-coronavirus-economic-downturn-n1189326

Goel, S. (2017). *Is part of Chelsea Manning's legacy increased surveillance? The conversation*. Retrieved from https://theconversation.com/is-part-of-chelsea-mannings-legacy-increased-surveillance-71607

Government, H. M. (2013). *UK data capability strategy: Seizing the data opportunity—Publications — GOV.UK*. Retrieved from https://www.gov.uk/government /publications/ukdata-capability-strategy

Groves, P., Kayyali, B., Knott, D., & Van Kuiken, S. (2013). *The "Big Data" revolution in healthcare. Accelerating value and innovation. Center for US Health System Reform*. McKinsey & Company.

GSMA. (2014). *GSMA guidelines on the protection of privacy in the use of mobile phone data for responding to the Ebola outbreak*. Retrieved from https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2014/11/GSMA-Guidelines-on-protecting-privacy-in-the-use-of-mobile-phone-data-for-responding-to-the-Ebola-outbreak-October-2014.pdf

Guillén, M., & Suárez, S. (2010). *The Global Crisis of 2007–2009: Markets, Politics, and Organizations.* .10.1108/S0733-558X(2010)000030A012

Heilbroner, R. (1967). Do machines make history? *Technology and Culture*, *8*(3), 335–345. doi:10.2307/3101719

Helmreich, R. (1997). Managing human error in aviation. *Scientific American*, *276*(5), 62–67. doi:10.1038/scientificamerican0597-62 PMID:11536800

Holm, S., Kristiansen, T., & Ploug, T. (2020). Control, trust, and the sharing of health information: The limits of trust. *Journal of Medical Ethics*. Advance online publication. doi:10.1136/medethics-2019-105887 PMID:32843438

Hong, W., & Thong, J. (2013). Internet privacy concerns: An integrated conceptualization and four empirical studies. *Management Information Systems Quarterly*, *37*(1), 275–298. doi:10.25300/MISQ/2013/37.1.12

Hurwitz, J. (2012). *The Making of a (Big Data) President*. Businessweek.com. Retrieved from https://www.businessweek.com/articles/2012-11-14/the-making-of-a-big-datapresident

IBM. (2015). *Bringing Big Data to the enterprise: what is Big Data?* Retrieved from https://www.ibm.com/software/data/bigdata/

International Medical Informatics Association. (2011). *Code of Ethics for Health Information Professionals*. Retrieved from: http://www.imia-medinfo.org/ new2/node/39

Jacobs, A. (2009). Pathologies of Big Data. *ACM Queue; Tomorrow's Computing Today*, *7*(6).

Jain, P., Gyanchandani, M., & Khare, N. (2016). Big Data privacy: A technological perspective and review. *Journal of Big Data*, *3*(1), 25. Advance online publication. doi:10.1186/s40537-016-0059-y

Jennex, M. E. (2017). Big Data, the internet of things, and the revised knowledge pyramid. *ACM SIGMIS Database: the Database for Advances in Information Systems, 48*(4), 69-79.

Johnson, M., Egelman, S., & Bellovin, S. (2012). Facebook and privacy: it's complicated. *Symposium on Usable Privacy and Security (SOUPS)*. doi:10.1145/2335356.2335369

Kuechler, W. (2007). Business applications of unstructured text. *Communications of the ACM*, *50*(10), 86–93. doi:10.1145/1290958.1290967

La Porte, T., & Consolini, P. (1991). Working in practice but not in theory: Theoretical challenges of high-reliability organizations. *Journal of Public Administration: Research and Theory*, *1*, 19–47.

Leigh, D. (2010). How 250,000 US embassy cables were leaked. *The Guardian*. Retrieved from https://www.theguardian.com/world/2010/nov/28/how-us-embassy-cables-leaked

Leveson, N. (2016). *Engineering a Safer World. Systems Thinking Applied to Safety*. MIT Press.

Leveson, N., Dulac, N., Marais, K., & Carroll, J. (2009). Moving beyond normal accidents and high-reliability organizations: A systems approach to safety in complex systems. *Organization Studies*, *30*(2–3), 227–249. doi:10.1177/0170840608101478

Likhterman, S., Lokshina, I., & Batugina, N. (1998). *The Use of Fuzzy Models to Estimate Stability of Mine Constructions.* In P. N. Martens (Ed.), *Roof-bolting in Mining* (Vol. 15, pp. 433–443). ABRW.

Lindenbaum, S. (2001). Kuru, prions, and human affairs: Thinking about epidemics. *Annual Review of Anthropology*, *30*(1), 363–385. doi:10.1146/annurev.anthro.30.1.363

Lokshina, I., Durkin, B., & Lanting, C. J. M. (2019a). IoT- and Big Data-Driven Data Analysis Services for Third Parties: Business Models, New Ventures, and Potential Horizons. In N. Meghanathan (Ed.), *Strategic Innovations and Interdisciplinary Perspectives in Telecommunications and Networking* (pp. 256–289). doi:10.4018/978-1-5225-8188-8.ch014

Lokshina, I., Gregus, M., & Thomas, W. (2019b). Application of Integrated Building Information Modeling, IoT, and Blockchain Technologies in System Design of a Smart Building. *Procedia Computer Science*, *160*, 497–502. doi:10.1016/j.procs.2019.11.058

Lokshina, I. V. (2001). Expert System Based on the Fuzzy Diagnostic Model to Support Coal Mine Ventilation Operator's Decisions. In A. Grmela & N. Mastorakis (Eds.), *Advances in Intelligent Systems, Fuzzy Systems, Evolutionary Computation* (pp. 118–122). WSEAS Press.

Lokshina, I. V. (2002). Intellectual Support of Investment Decisions Based on a Clustering of the Correlation Graph of Securities. *WSEAS Transactions on Systems*, *1*(2), 284–290.

Lokshina, I. V., Durkin, B. J., & Lanting, C. J. M. (2017). Data Analysis Services Related to the IoT and Big Data: Strategic Implications and Business Opportunities for Third Parties. *International Journal of Interdisciplinary Telecommunications and Networking*, *9*(2), 37–56. doi:10.4018/IJITN.2017040104

Lokshina, I. V., & Insinga, R. C. (2003). Decision support system for ventilation operators based on fuzzy methods applied to the identification and processing of gas-dynamic images. *Journal of Electrical Engineering*, *54*(9-10), 277–280.

Lokshina, I. V., & Lanting, C. J. M. (2018a). Addressing Ethical Concerns of Big Data as a Prerequisite for a Sustainable Big Data Industry. *International Journal of Interdisciplinary Telecommunications and Networking*, *10*(3), 34–52. doi:10.4018/IJITN.2018070104

Lokshina, I. V., & Lanting, C. J. M. (2018b). A Qualitative Evaluation of IoT-driven eHealth: Knowledge Management, Business Models and Opportunities, Deployment, and Evolution. *Proceedings of Hawaii International Conference on System Science (HICSS-51)*, 4123-4132. doi:10.24251/HICSS.2018.518

Lu, R., Zhu, H., Liu, X., Liu, J., & Shao, J. (2014). Toward efficient and privacy-preserving computing in the Big Data era. *IEEE Network*, *28*(1), 46–50. doi:10.1109/MNET.2014.6863131

MacGregor, D. (2003). 10 Public response to Y2K: social amplification and risk adaptation: or, ‘‘how I learned to stop worrying and love Y2K. In N. Pidgeon, R. Kasperson, & P. Slovic (Eds.), *The social amplification of risk* (p. 243). Cambridge University Press. doi:10.1017/CBO9780511550461.011

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. (2011). *Big Data: The next frontier for innovation, competition, and productivity*. Retrieved from http://www.mckinsey.com/Insights/MGI/Research/Technology_and _Innovation /Big_data_The_next_frontier_for_innovation

Marcus, G. (2012). The web gets smarter. *The New Yorker*. Retrieved from https://www.newyorker.com/culture/culturedesk/the-web-gets-smarter

Marr, B. (2015). *Big Data: Using SMART Big Data, analytics, and metrics to make better decisions and improve performance*. John Wiley & Sons.

Mayer-Schonberger, V., & Cukier, K. (2013). *Big Data: a revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

McAffee, A., & Brynjolfsson, E. (2012). Big Data: The management revolution. *Harvard Business Review*, *90*(10), 60–68. PMID:23074865

Miller, Z. (2013). *Former NSA Chief was worried about ‘‘Enemy of The State’’ Reputation | TIME.com.* Retrieved from https://swampland.time.com/2013/06/07/formernsa-chief-was-worried-about-enemy-of-the-state-reputation

Molok, N., Chang, S., & Ahmad, A. (2012). Information Leakage through Online Social Networking: Opening the Doorway for Advanced Persistence Threats. *Australian Information Security Management Conference*.

Mozur, P., Zhong, R., & Krolik, A. (2020). In Coronavirus Fight, China Gives Citizens a Color Code, With Red Flags. *The New York Times*. Retrieved from https://www.nytimes.com/2020/03/01/business/china-coronavirus-surveillance.html

Mundie, C. (2014). Privacy pragmatism. *Foreign Affairs*, *93*(2), 28–38.

Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review (Seattle, Wash.)*, *79*(1), 119.

Nuffield Council on Bioethics. (2015). *The collection, linking, and use of data in biomedical research and health care: ethical issues*. London: The Nuffield Council on Bioethics. Retrieved from: http://nuffieldbioethics.org/wp-content/uploads/Biological_ and_health_data_web.pdf

Nunan, D., & Di Domenico, M. (2017). Big Data: A Normal Accident Waiting to Happen? *Journal of Business Ethics*, *145*(3), 481–491. doi:10.1007/s10551-015-2904-x

OECD. (2019). *Government at a Glance 2019*. Retrieved from https://www.oecd.org/gov/government-at-a-glance-22214399.htm

Palmer, D., & Maher, M. (2010). Normal accident analysis of the mortgage meltdown. *Research in the Sociology of Organizations*, *30*, 219–256. doi:10.1108/S0733-558X(2010)000030A011

Perera, C., Ranjan, R., Wang, L., Khan, S., & Zomaya, A. (2015). Big Data Privacy in the Internet of Things Era. *IT Professional, 17*(3), 32-39. 10.1109/MITP.2015.34

Perrow, C. (1981). Normal accident at a three-mile island. *Society*, *18*(5), 17–26. doi:10.1007/BF02701322

Perrow, C. (1999). *Normal accidents: Living with high-risk technologies*. Princeton University Press.

Perrow, C. (2008). Disasters evermore? Reducing our vulnerabilities to natural, industrial, and terrorist disasters. *Social Research*, *75*(3), 733–752.

Perrow, C. (2009). What's needed is the application, not reconciliation: A response to Shrivastava, Sonpar, and Pazzaglia. *Human Relations*, *62*(9), 1391–1393. doi:10.1177/0018726709339120

Pidgeon, N. (2011). In retrospect: Normal accidents. *Nature*, *477*(7365), 404–405. doi:10.1038/477404a

Pillinger, M. (2020). Virus Travel Bans Are Inevitable but Ineffective. *Foreign Policy*. Retrieved from https://foreignpolicy.com/2020/02/23/virus-travel-bans-are-inevitable-but-ineffective/

Proposal for a Regulation of the European Parliament and the Council, European Commission. (2012). Retrieved from https://ec.europa.eu/digital-agenda/en/news/proposal-regulation-european-parliament-and-council-establishing-connecting-europe -facility

Rijpma, J. (1997). Complexity, tight–coupling, and reliability: Connecting normal accidents theory and high-reliability theory. *Journal of Contingencies and Crisis Management*, *5*(1), 15–23. doi:10.1111/1468-5973.00033

Roberts, K. (1990). Some characteristics of high-reliability organizations. *Organization Science*, *1*(2), 160–177. doi:10.1287/orsc.1.2.160

Rosenthal, C. (2014). *Big Data in the age of the telegraph*. Retrieved from https://www.mckinsey.com/insights/organization/big_data_in_the_age_of_the_telegraph

Sagan, S. (1993). *The limits of safety: Organizations, accidents, and nuclear weapons*. Princeton University Press. doi:10.1515/9780691213064

Scarpino, S., & Petri, G. (2019). On the predictability of infectious disease outbreaks. *Nature Communications*, *10*(1), 898. doi:10.1038/s41467-019-08616-0 PMID:30796206

Schlegelmilch, B., & Oberseder, M. (2010). Half a century of marketing ethics: Shifting perspectives and emerging trends. *Journal of Business Ethics*, *93*(1), 1–19. doi:10.1007/s10551-009-0182-1

Science. (2020). *The coronavirus seems unstoppable. What should the world do now?* https://www.sciencemag.org/news/2020/02/coronavirus-seems-un-stoppable-what-should-world-do-now

Sedayao, J., & Bhardwaj, R. (2014). Making Big Data, privacy, and anonymization work together in the enterprise: experiences and issues. *Big Data Congress*. doi:10.1109/BigData.Congress.2014.92

Shane, S. (2019). The Age of Big Leaks. *The New York Times*. Retrieved from https://www.nytimes.com/2019/02/02/sunday-review/data-leaks-journalism.html

Shane, S., Perlroth, N., & Sanger, D. (2017). Security Breach, and Spilled Secrets Have Shaken the N.S.A. *The New York Times*. Retrieved from https://www.nytimes.com/2017/11/12/us/nsa-shadow-brokers.html

Shrivastava, S., Sonpar, K., & Pazzaglia, F. (2009). Normal accident theory versus high-reliability theory: A resolution and call for an open systems view of accidents. *Human Relations*, *62*(9), 1357–1390. doi:10.1177/0018726709339117

Snook, S. (2000). *Friendly Fire: The Accidental Shootdown of U.S. Black Hawks over Northern Iraq*. Princeton University Press.

Szoldra, P. (2016). This is everything Edward Snowden revealed in one year of unprecedented top-secret leaks. *Business Insider*. Retrieved from https://www.businessinsider.com/snowden-leaks-timeline-2016-9

The Economist. (2020). *To curb covid-19, China is using its high-tech surveillance tools. Non-state firms are best-equipped to track people's movements and contacts*. https://www.economist.com/china/2020/02/29/to-curb-covid-19-china-is-using-its-high-tech-surveillance-tools

United Nations. (2020). *Senior Officials Sound Alarm over Food Insecurity, Warning of Potentially 'Biblical' Famine, in Briefings to Security Council*. Press Release. Retrieved from https://www.un.org/press/en/2020/sc14164.doc.htm

Wang, C., Ng, C., & Brook, R. (2020). Digital Smartphone Tracking for COVID-19—Balancing Public Health and Civil Liberties. *JAMA*. Retrieved from https://jamanetwork.com/journals/jama/fullarticle/2762689

Ward, P. (2017). Improving Access to, Use of, and Outcomes from Public Health Programs: The Importance of Building and Maintaining Trust with Patients/Clients. *Frontiers in Public Health*, *5*, 22. doi:10.3389/fpubh.2017.00022 PMID:28337430

Warren, S., & Brandeis, L. (1890). The right to privacy. *Harvard Law Review*, *IV*(5), 193–220. doi:10.2307/1321160

Waterman, S. (2013). NSA leaker Ed Snowden used a banned thumb drive, exceeded access. *The Washington Times*. Retrieved from https://www.washingtontimes.com/news/2013/jun/14/nsa-leaker-ed-snowden-used-banned-thumb-drive-exce/

Wheeler, N. (2019). Tracing outbreaks with machine learning. *Nature Reviews. Microbiology*, *17*(5), 269. doi:10.1038/s41579-019-0153-1 PMID:30742026

Witte, D. (2013). Privacy deleted: Is it too late to protect our privacy online? *Journal of Internet Law*, *18*(1), 1–28.

World Medical Association. (2016). *Declaration of Taipei on Ethical Considerations Regarding Health Databases and Biobanks.* Retrieved from: https://www.wma.net/policies-post/wma-declaration-of-taipei-on-ethical-considerations-regarding-health-databases-and-biobanks

Wu, J., Leung, K., & Leung, G. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modeling study. *Lancet*, *395*(10225), 689–697. doi:10.1016/S0140-6736(20)30260-9 PMID:32014114

Ylijoki, O., & Porras, J. (2016). Perspectives to Definition of Big Data: A Mapping Study and Discussion. *Journal of Innovation Management*, *4*(1), 69–91. doi:10.24840/2183-0606_004.001_0006

Zaslavsky, A., Perera, C., & Georgakopoulos, D. (2012). Sensing as a Service and Big Data. *Proceedings of International Conference on Advances in Cloud Computing (ACC)*, 21–29.

Zuboff, S. (1988). *In the age of the smart machine: Machine: The future of work and power*. Basic Books.

Izabella V. Lokshina, PhD, is Professor of MIS and chair of the Management, Marketing and Information Systems Department at SUNY Oneonta, USA. She is 2019 SUNY Chancellor Award for Excellence in Research and Creative Activities recipient. Her main research interests are state-of-the-art and emerging technologies, intelligent information systems and communication networks, as well as complex system modeling and simulation.

Cees J. M. Lanting, PhD, is Senior Consultant at DATSA Belgium, Leuven, Belgium, and is heading its IDA- and IDEA-lab for instrumentation and IoT development. His main research interests are system modeling, smart communications and result oriented IoT systems.