Shaping the ethical dimensions of smart information systems– a European perspective (SHERPA)

Deliverable No. 1.4

Report on Ethical Tensions and Social Impacts

Submission date

20th September 2019

This project has received funding from the
European Union's Horizon 2020 Research and Innovation Programme
Under Grant Agreement no. 786641

Document Control

Deliverable	Report on Ethical Tensions and Social Impacts
WP/Task Related	WP1 - Representation and Visualisation
Delivery Date	20th September 2019
Dissemination Level	Public
Lead Partner	The University of Twente
Contributors	Mark Ryan, Philip Brey, Kevin Macnish, Tally Hatzakis, Owen King, Jonne Maas, Ruben Haasjes, Ana Fernandez, Sebastiano Martorana, Isaac Oluoch, Selen Eren, and Roxanne Van Der Puil.
Reviewers	UCLan Cyprus
Abstract	
Key Words	Big Data; Artificial Intelligence; Smart Information Systems; Ethical Tensions; Social Impacts

Revision History

Version	Date	Author(s)	Reviewer(s)	Notes
0.1	June 6th 2019	Mark Ryan	UCLanCY	First Draft
0.2	June 28th 2019	Mark Ryan		Second Draft
1	September 18th 2019	Kevin Macnish		Final Draft

We would like acknowledge a special thanks to UCLan Cyprus, who were the quality assurance officers on this Deliverable.

Table of Contents

Executive Summary

The SHERPA consortium looked at the main ethical issues, tensions, and possible social impacts of smart information systems (SIS) in this document. The purpose for this is to provide a detailed analysis of the main ethical issues and tensions that have arisen throughout the previous four Deliverables in Work Package 1 (D1.1, D1.2, D1.3, and D1.5). This Deliverable brings all of the collective concerns and ethical problems into one document in a systematic and comprehensive manner.

The report approaches ethical analysis in terms of:

- general issues,
- aims of SIS,
- implications and risks of SIS,
- issues arising from specific techniques and technology,
- case studies and scenarios concerning application domains,
- research and innovation.

The Deliverable examines ethical tensions in the use of SIS in a pragmatic and comprehensive way, beginning with ethical issues related to the actual design of the technologies themselves. Whether or not there are inherent issues with their functioning, capacities, and programming (sections 2 and 5). The document then identifies the main ethical issues within the debate for the use of SIS in practice, outlining 24 of the key ethical concerns found within the literature (section 4). While the technologies themselves, and their use, raise important concerns that need to be addressed, it is important to not overlook specific domain applications and fields of practice, which is reviewed in section 6 of this report.

The Deliverable will also give a thorough analysis of the main ethical issues related to research & innovation aspects of SIS development (section 7). The report will subsequently finish with a detailed analysis of the main ethical issues and possible solutions within the report (section 8), in an attempt to identify, allocate, and group such a wide body of information into the most prevalent concerns for society today.

The ethical analysis is approached from both theoretical and empirical perspectives, leading to a thorough analysis of ethical issues in theory and practice. The report forms a solid groundwork for future deliverables, particularly Deliverable 3.2 (Proposals for Ethical Guidelines).

Revision Notes

An addendum has been added to the document summarising the work of Deliverables 1.1, 1.2, 1.3 and 1.5 (p118-42). As noted above, these deliverables have informed the current deliverable throughout. However, the appendix isolates and highlights the key findings, insights and recommendations of each of the other deliverables in Work Package 1.

List of figures

List of tables

List of acronyms/abbreviations

Abbreviation	Explanation

SIS	Smart Information Systems
AI	Artificial Intelligence
ICT	Information and Communications Technology
SDV	Self-driving vehicles
HE	Horizon Europe
GDPR	General Data Protection Regulation
ERAAI	European Regulatory Agency for AI
LEAs	Law enforcement authorities
ePR	ePrivacy Regulations
IoT	Internet of Things
R&I	Research and Innovation

Table 1: List of acronyms/abbreviations

1. Introduction

The SHERPA project aims to investigate, analyse and synthesise our understanding of the ways in which smart information systems (SIS; the combination of Artificial Intelligence (AI) and Big Data analytics) impact ethics and human rights issues.

Drawing on the results of Tasks 1.1-1.3, this deliverable will categorise the range of ethical tensions and social impacts raised by SIS, with an emphasis on privacy, discrimination, manipulation, inequality and security issues. Such tensions include the potential medical benefits of large-scale collection of personal health data weighed against privacy and discrimination concerns and the potential for re-identification of anonymised data. Wearable tracking devices, for example, provide safety for people with dementia but must be weighed against the implications of constant monitoring and the potential abuse of data generated from such devices. The Deliverable identifies the most pressing ethical benefits and concerns for each of the areas considered by SHERPA in the case studies (Deliverable 1.1) and scenarios (Deliverable 1.2), and pays attention to ethical tensions and social impacts relating to the use of SIS in research and innovation. The Deliverable studies how different uses of SIS in R&I could lead to a variety of ethical issues, which SHERPA will catalogue in a taxonomy, to evaluate the positive and negative implications of SIS and assess the ways these can be balanced against one other.

This Deliverable incorporates and builds on the results of Deliverables 1.1, 1.2 and 1.3, while also completing additional detailed analysis on issues, applications, and technologies not covered under the parameters of those Deliverables. This Deliverable will also make use of D4.4 from the SIENNA project (coordinated by Philip Brey), as there is a degree of overlap between the two projects' Deliverables, but with some differences in approach and focus (the SIENNA report focuses on AI and robotics in general, whereas this SHERPA report focuses more specifically on data-intensive SIS).

This Deliverable 1.4 offers a broad, overarching analysis of ethical concerns related to the development and use of SIS. Its aim is to provide the most comprehensive evaluation of ethical issues regarding SIS, for the SHERPA project. It examines many of the concerns found in tasks 1.1, 1.2 and 1.3, while developing further issues to capture the 'full picture' of SIS ethical concerns. It will provide a strong grounding and template for the SHERPA project in later Deliverables; in particular, task 3.2. Task 3.2 will integrate the findings from 1.4 to develop two sets of ethical guidelines, one for the ethical development of SIS, and a second on the ethical use of SIS. This Deliverable is intended to provide clear ethical guidelines to those developing and using SIS in the field.

This Deliverable contains seven sections (excluding the Introduction and Conclusion). We start by outlining what we mean by Big Data, Artificial Intelligence, and most importantly, Smart Information Systems, which leads into Section 3, outlining how these technologies are used in 16 social domains. Section 4 comprehensively details 24 pertinent ethical issues found within the literature regarding SIS. This section highlights the general ethical concerns found within the debate. Sections 5 and 6 focus on specific issues with the technologies themselves and within their application in particular social domains. Section 7 examines the main ethical issues found in the development and use of SIS in research and innovation (R&I). This is concluded by bringing together the main ethical tensions identified in sections 4 - 7. It considers which collisions between values and interests they involve, and how these conflicts could be resolved. It will offer a range of options that will be further developed in Task 3.2.

2. Smart Information Systems

This section aims to give an overview of the technical aspects of SIS, defined through the use of AI in Big Data analytics.

2.1 Defining Big Data

Due to the 'big' in 'Big Data', size is often the first characteristic that comes to mind when defining Big Data (Gandomi and Haider, 2015, p. 138). In practice, Big Data is often associated with datasets that have grown so large that their size is beyond the ability of commonly-used software tools and storage systems to capture, store, manage and process the data with reasonable performance (Elgendy and Elragal, 2014, p. 2; Ward and Barker, 2013, p. 1).

However, volume is only one of the three defining characteristics of Big Data, with "The Three V's" (3Vs) being commonly used to define it (Oussous et al., 2017, p. 3; Gandomi and Haider, 2015, p. 138). Gandomi and Haider describe the 3Vs as volume, variety, and velocity. However, an additional 3Vs were later added, as seen in Table 3.

Characteristic	Explanation
Volume	Refers to the magnitude of data. Important to note is that there is no specified threshold that defines Big Data volumes. Definitions of Big Data volumes are relative and depend on factors such as time and the type of data. What is considered Big Data could change in the future due to increases in storage capacities or processing power. Furthermore, two datasets of the same size could require different management and processing technologies (Gandomi and Haider, 2015, p. 138).
Variety	Gandomi and Haider describe variety as the structural heterogeneity in a dataset. Due to technological advances, structured, semi-structured and unstructured data can be used. Structured data refers to tabular data as found in spreadsheets and relational databases. Examples of unstructured data are text, images, audio and video, which often lack structural organization in data format. An example of semi-structured data is XML, which are documents that contain user-defined data tags without conforming to strict standards (Gandomi and Haider, 2015, p. 138).
Velocity	Refers to the rate at which data is generated and the speed at which this data should be analysed and acted upon (Gandomi and Haider, 2015, p. 138).
Veracity	Represents a degree of unreliability or uncertainty inherent in some sources of data.
Variability	Describes two additional dimensions of Big Data: 1) the velocity of Big Data is not consistent as there are peaks and downs in velocity' 2) Big Data is generated through various sources, which require connecting, matching, cleaning and transforming data from different sources.
Value	Describes how Big Data in its received form often has a low value relative to its volume. However, a high value can be obtained by analysing large volumes of such data (Gandomi and Haider, 2015, p. 139).

Table 3: The six Vs of Big Data

The most suggested keywords associated with Big Data show how Big Data is intertwined with Big Data analytics (Ward and Barker, 2013, p. 2). People associate Big Data not only with collecting large amounts of data. They also want to understand the meaning and importance of the data and use these insights as an aid in making decisions (Elgendy and Elragal, 2014, p. 219). The importance of analytics brings us to the next point; clarifying the differences between the key concepts in SHERPA; Big Data analytics, AI and machine learning.

2.2. SIS, Big Data Analytics, Artificial Intelligence and Machine Learning

Big Data analytics is often described as a science that aims to examine and draw insights from the data (Venkatram and Geetha, 2017, p. 16). Numerous techniques are available for Big Data analytics, such as statistical analysis and AI (Russom et al., 2011, p. 6; Venkatram and Geetha, 2017, p. 18). Big Data

analytics can be seen as data science, which employs various tools with the aim of drawing insights from Big Data, whereas AI is one of these tools, and thus*part of* rather than *equivalent to*, Big Data analytics.

Although it is not completely clear what falls under the label of artificial intelligence, the field of AI is commonly defined as a science with the goal of making machines do things that would require intelligence if done by humans (Negnevitsky, 2005, p. 18).

One of the most popular subfields within AI is machine learning. The key difference between machine learning and other approaches to AI is that instead of hand-coding software routines with specific rules and instructions, the machine is "trained", using large amount of data, to perform a certain task. One approach to machine learning currently gaining popularity is deep learning, which loosely models the biology of our brains, resulting in artificial neural networks with many layers, neurons and connections. Worldwide, data volume has also expanded, as a result of the Internet and all its applications, resulting in many Big Data sources (Upadhyaya Kynficlovfia, 2017, p. 7).

2.2.1. Enterprise Data

IBM has indicated that the internal data of enterprises are the main sources of Big Data (Chen et al., 2014, as cited on p. 179). This internal data of enterprises consists mainly of online trading and analysis data, which are historically static data and managed by RDBMs (see section 2.2), thus enterprise data is often structured data (Chen et al., 2014, p. 179). In addition to this data, an attempt is made to capture and record all data from data-driven activities in an enterprise, such as production data, inventory data, sales data and financial data. Lastly, web data is customer-level web behaviour data such as page views, searches and reviews, which can also be seen as enterprise data.

2.2.2. Text Data

Text data is one of the biggest and most widely applicable types of Big Data, as numerous websites, emails, forums, news sites, blogs and social media all present a lot of information in textual form. Su suggests that the focus of big text data is usually on extracting key facts from the text and using these as input for other analytical processes. Text data is considered to be unstructured data.

2.2.3. Audio, Video & Image Data

Another type of data that requires a massive amount of Big Data storage are audio, video and image data. These three data types are also considered to be unstructured data. Audio, video and image data have seen a massive increase in volume due to the rising popularity of media and social media platforms such as Spotify, Imgur and YouTube. Every day, billions of videos are viewed on YouTube (Sagiroglu and Sinanc, 2013, p. 1).

2.2.4. Social Media Data

Much of the content generated on social media falls under the category of text, audio, video and image data. However, social network data is more than the content posted. Within social network

sites such as Facebook, LinkedIn and Instagram, it is possible to perform a link analysis to uncover the network surrounding a particular user. The social networks keep track of connections between people and the content people like.

2.2.5. Biomedical Data

A lot of data used by frontier research in the biomedical field also deserves the label of Big Data, because:

• A series of high-throughput bio-measurement technologies are being developed which generate a lot of biomedical data (more detail in section 2.1.6)

• Massive amounts of data are generated by gene sequencing technology

• More and more data are being generated from clinical medical care (Chen et al., 2014, p. 180)

2.2.6. Internet of Things Data

Nowadays, an enormous number of devices and machines in the real world are connected to the internet and embedded with networking sensors. Due to this, various kinds of machines and devices can be sources of Big Data. Examples vary from sensors and devices in houses, which store information about heating, lightning, electricity etc., to sensors on cars, airplanes, oil pipes and windmill turbines, which could hold valuable information with respect to maintenance and performance (Chen et al., 2014, p. 177). Data generated by the Internet of Things is usually semi-structured or unstructured data (Chen et al., 2014, p. 177).

2.3. Big Data Storage

Since these various data sources generate more data than can be stored on a single computer's hard drive, there is an issue about how Big Data is stored in large scale distributed storage systems.

2.3.1. Distributed File Systems

File systems are the foundation of distributed storage systems. Distributed file systems have become quite mature after years of research and use in business and industry (Chen et al., 2014, p. 186). One of the most well-known distributed file systems is the Google File System (GFS). The GFS consists of a cluster of nodes (servers). There are two types of nodes: the master node and the chunk servers.

Each file that needs to be stored on the GFS is divided into fixed-size chunks and stored redundantly, to guarantee fault tolerance, on different chunk servers. By default, each chunk is stored three times, but this is configurable. On start-up the master node polls all the nodes to retrieve the information about which chunks are stored on which chunk server. All the read and write actions that a client wants to perform are done through the master node. Clients request the metadata (the mapping from files to chunks) and the position from the master node, and by using this data query the chunk servers directly for their information (Ghemawat et al., 2003). The GFS has cheap, high fault-tolerance and performance, as it uses cheap commodity servers. The downsides of the GFS are that it is not

optimised for small-sized files, and that it has a single point of failure (the master node). These limitations have been overcome by the successor of the GFS: Colossus (Chen et al., 2014, p. 186).

Besides the GFS, numerous well-known alternatives for storage systems have been developed by researchers and companies, such as the Apache Hadoop Distributed File System, which is derived from the GFS. Other examples include Microsoft Cosmos, which is used for their search and advertisements business, and Facebook Haystack, which is used to store large amounts of small-sized photos (Chen et al., 2014, p. 186).

2.3.2. NoSQL Databases

On top of this file system, a database technology is used. Due to the different types of structured, semi-structured and unstructured data, traditional relational databases alone are no longer sufficient to store Big Data (Chen, 2014, p. 186). With Big Data, noSQL (non-traditional relational databases) are often used. The main categories of noSQL can be seen in Table 4 (Saxena et al., 2014, pp. 4-5; Venkatram and Geetha, 2017, p. 13).

Database	Explanation
Key-value databases	This type of database is used when most of the access to the data is done through unique keys. It has a simple structure and is characterised by high expendability and shorter response times than traditional relational databases (Chen et al., 2014, p. 186). For example, Dynamo, which Amazon uses for most of its core services of the Amazon E-commerce platform, and Voldemort, which was developed and is still used by LinkedIn (Chen et al., 2014, pp. 186-187).
Column-oriented databases	This type of database is used when an application needs to access a few columns of many rows at once, and writes to the database are uncommon. Most of the column-oriented databases are based on the design of Google's BigTable (Chen et al., 2014, p. 187). A BigTable is a multidimensional sparse sorted map. Each row of the BigTable can store an arbitrary number of key-value pairs, making the BigTable suitable for data that scales to a large size (Singh and Reddy, 2015, p. 5). Google uses BigTable for many projects including web-indexing, Google Earth and Google Finance (Chang et al., 2008, p. 1). Well-known alternatives to BigTable are Cassandra, which was developed and made open-source by Facebook, and Apache Hbase (Chen et al., 2014, p. 187).
Document-oriented databases	This type of database stores data at a document level using a markup language such as JavaScript Object Notation (JSON) and eXtensible Markup Language (XMl). It is used when the structure of data is flexible and makes it easy to combine different data with different structures without losing access and indexing functionality. Popular document-oriented databases include: MongoDB, SimpleDB and CouchDB (Chen et al., 2014, pp. 187-188).

Graph databases	This type of database uses concepts of graphs (nodes, edges) to store data. In this type of database every data element is directly connected to adjacent elements (Singh and Reddy, 2015, p. 4). The most well-known example of a graph database is Neo4j.

Table 4: Types of Big Data databases

2.4. Big Data Analytics

Since Big Data is usually stored on clusters with numerous nodes, the traditional parallel programming models to process data are not sufficient. Therefore, different parallel programming models for Big Data have been proposed. These models provide a simplified programming model or API and, by doing this, hide the complexity of writing a distributed application.

2.4.1. Big Data Distributed Programming Models

The most well-known Big Data distributed programming model is probably MapReduce, which was proposed by Google in 2004. MapReduce is a simple yet powerful distributed programming model. As the name implies, it consists of only two functions: map and reduce. The map function takes as input a key-value pair. Users specify the map function which generates an intermediate set of key-value pairs. The reduce function merges all intermediate values associated with the same intermediate key. A typical MapReduce program processes many terabytes of data on thousands of machines. MapReduce programs can express many real-life tasks (Dean and Ghemawat, 2008, p. 1). Dean describes how MapReduce is used by Google for large-scale machine learning problems, clustering problems for Google News and the extraction of properties from web pages (Dean & Ghemawat, 2008, p. 10). MapReduce provides a simplified programming model that hides the complexity of writing a distributed application; the actual programs are scripted or programmed in a language such as Python, C, Java, R or Perl (Watson, 2014, p. 1259).

Although MapReduce is simple and fairly powerful compared to other programming models, a single MapReduce program has its limitations. If one were to use multiple MapReduce programs in succession, or iterative MapReduce, it overcomes these limitations. However, iterative MapReduce is slow due to latency and reuse of data across iterations. An alternative is YARN, which is more general than MapReduce, and provides better scaling, enhanced resource management and parallelism (Oussous et al., 2017, p. 8).

Another downside of MapReduce is that it is limited to batch processing, or in other words, not suitable for near real time applications (Watson, 2014, p. 1259). The most well-known distributed programming models for real time data processing are Storm (developed by Twitter) and Spark. Storm consists of a network of “bolts” and “sprouts”. A sprout is a source of streams, and a bolt processes input streams and output streams. By using these bolts and sprouts, Storm enables a user to perform transformation on real time data streams (Oussous et al., 2017, p. 10).

Spark is based on the Resilient Distributed Dataset abstraction. It provides Spark SQL, which enables users to perform queries on datasets, and Spark streaming, which enables users to stream tasks by performing a series of short batch jobs (Oussous et al., 2017, p. 10).

2.4.2. Machine Learning

As stated in the Introduction, Big Data analytics can be regarded as a data science, which uses various tools to analyse data. One of the most hyped techniques among these is machine learning. The distributed programming models discussed in the previous section provide the interface to implement machine learning algorithms in a parallel manner. Furthermore, many companies provide machine learning libraries that run on top of their Big Data storage and processing software stack. For example, Apache MLlib provides a scalable machine learning library that contains the most commonly used machine learning algorithms. The field of machine learning is often divided into three subdomains: supervised learning, unsupervised learning, and reinforcement learning - see Table 5 (Qiu et al., 2016, p. 2).

Type	Explanation
Supervised learning	This requires training with labelled data which has inputs and desired outputs. This technique is often used for classification, regression and estimation tasks. In supervised learning one can distinguish between computational classifiers, statistical classifiers and connectionist classifiers.
Unsupervised learning	This does not require labelled training data, the user only provides input data. This technique is most often used for clustering and making predictions. Unsupervised learning techniques can be categorised in parametric and nonparametric techniques.
Reinforcement learning	This enables learning from feedback with an external environment. For example, the machine could start by making random decisions, and based on the outcome of these decisions learn which actions yield success and thus should be appointed a greater weight. Reinforcement learning comes in the form of model-free and model-based techniques.

Table 5: Types of machine learning

Nowadays, the hottest research field in machine learning is deep learning (Oussous et al., 2017, p. 5; Qiu et al., 2016, p. 3). It is a widely used technique in analytics applications in the fields of computer vision, speech recognition and natural language processing (Oussous et al., 2017, p. 5). Deep learning uses mathematical models which are inspired by the human brain to automatically learn the underlying hierarchical representations, or data representations from large volumes of raw data (Qiu et al., 2016, p. 3). One of the reasons why deep learning is so popular is because of its increase in accuracy as the amount of data accumulates, whereas a traditional machine learning algorithm would require a change in coding.

Another subfield of machine learning which is worth mentioning here is Natural Language Processing (NLP). Using machine learning techniques, attempts are made to retrieve information from sources of textual data. Since so much information on the internet generated by social media and websites is in the form of text, NLP is a fundamental analysis technique. For instance, sentiment analysis can be applied to analyse consumer reviews on products. There are many machine learning algorithms not mentioned here, however, these are beyond the scope of this overview.

2.4.3. Other Analytical Tools and Technical Categories of Application Areas

Big Data analytics also use other tools and techniques that would not necessarily fall under the label of “machine learning” or “artificial intelligence”. For example, A/B testing, also called split testing, in which two different variants are used to see which variant yields better results. Other examples include raw statistical analysis. Applications of Big Data analysis tools and techniques can be divided into the following key technical fields: Structured data analysis, Text data analysis, Web data analysis, Audio data analytics, Multimedia data analysis, Network data analysis, Mobile data analysis.

2.4.4. Descriptive, Predictive and Prescriptive Analytics

There are three distinct types of analytics - descriptive, predictive and prescriptive (Watson, 2014, pp. 1250-1251):

• Descriptive: this type of analytics is like looking backwards; one aims to reveal what has occurred. This includes the reporting of data and visualisation of data. For example, what was the sales revenue in the first quarter of the year? What is our most profitable product? These types of questions sometimes require complex queries that need to be executed on a distributed computing platform. Machine learning may be used to answer questions about what people think about their product on social media (which could require NLP).

• Predictive: this is where machine learning algorithms become essential. Predictive analytics aim to predict what will occur in the future. Think of a question like: what is the next best offer for this customer? Another example is Microsoft analysis sensor data of aircrafts, to predict which aircraft needs maintenance.

• Prescriptive: Prescriptive analytics impose action, therefore becoming operational. Not only does the algorithm predict when an aircraft will need maintenance, it also automatically sends maintenance teams information based on the analytical predictions. While prescriptive analytics contains the most explicit ethical issues, the other two types of analytics are not devoid of ethically problematic issues, which will be the focus of this Deliverable.

3. Applications of Smart Information Systems

The University of Twente (UT) established that one of the main areas that needed to be identified, and which would make SHERPA unique amongst many other projects, is the ethical analysis of SIS in particular social domains. Between March and April 2018, the UT team carried out a broad literature analysis of SIS to uncover the most prevalent ethical issues being discussed, and to try to identify the different types of applications of SIS in practice and the types of social domains in whichthese SIS technologies would be used (see Fig 1).

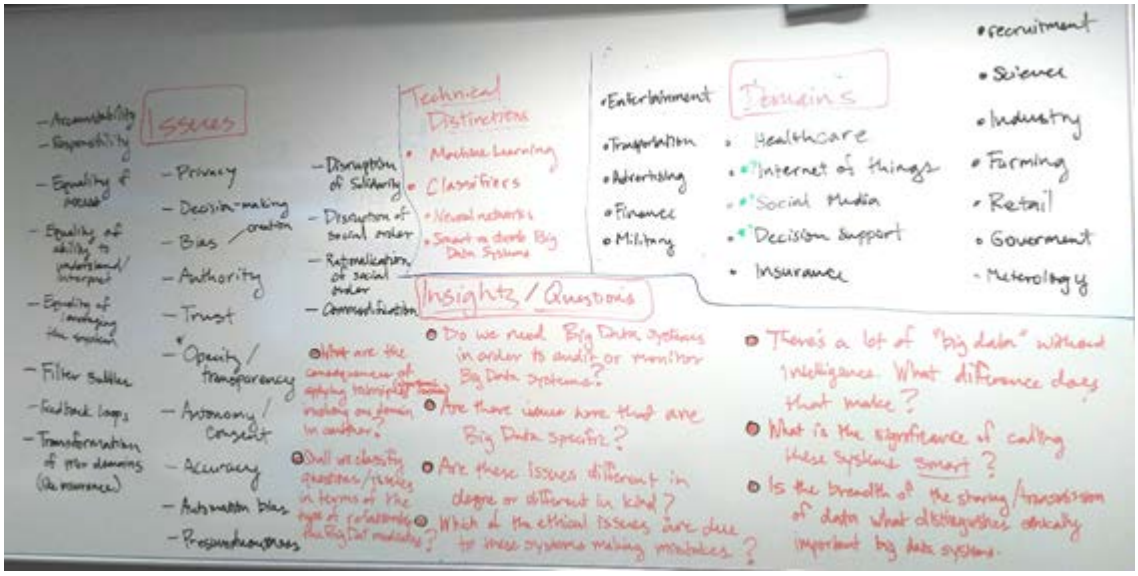


Figure 1: Image of Brainstorming at University of Twente

As a result, 16 specific social domains were established, most of which were thoroughly evaluated in the case studies and scenarios (D1.1 and D1.2). Each domain has its own particular use and application of SIS, so there may be specific ethical issues pertinent to that domain, which are not relevant for others. Similarly, there are specific ethical issues that appear in many, if not all, of the social domains. This section will briefly outline the 16 social domains, and how SIS are being implemented and used within those particular fields, prior to their ethical analysis in Section 6 of this Deliverable (see Table 6).

Social Domains
Banking and finance
Healthcare
Insurance
Retail and wholesale trade
Science
Education
Energy and utilities
Manufacturing and natural resources
Agriculture

Communications, media and entertainment
Transportation
Employee monitoring and administration
Government
Law enforcement and justice
Sustainable development
Defence and national security

Table 6: The 16 social domains

3.1. Smart Big Data in Banking and Securities

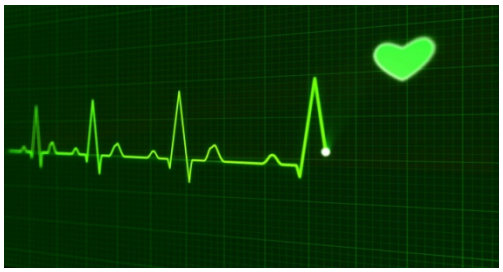
Retail traders, big banks, hedge funds and other large players in the financial markets use Big Data for trade analytics for high frequency trading, pre-trade decision-support analytics, sentiment measurement, Predictive Analytics, and risk analytics (the latter being used for purposes like anti-money laundering, demand enterprise risk management, "Know Your Customer", and fraud mitigation). In addition, oversight agencies like the US Securities and Exchange Commission use Big Data to monitor financial market activity and to catch illegal trading activity.

Financial services have been early adopters of AI, particularly in relation to high-frequency quantitative trading as a means to improve their trading decisions. Trading profits rely primarily on making the right decisions ahead of the competition. AI offers the potential to predict market dynamics, rather than simply respond to them. Hence, firms are increasingly relying on sophisticated mathematical models, Big Data analytics and AI to identify trading opportunities early, predict risks and even trigger timely trading decisions (Peng Zhang, Shi and Khan, 2017). AI has opened new trading horizons into cryptocurrency trading. As the value of cryptocurrencies is not regulated, the market is particularly whimsical and prone to fast-changing market dynamics. AI algorithms can override such short-term changes to identify trading opportunities (Tittel, 2018).

3.2. Smart Big Data in Healthcare

In the healthcare sector, four types of Big Data are used: (1) instrumentation data (sensors, monitors, RFID, barcode, video feeds); (2) diagnostic data (images, vital signs monitors, blood test results); (3) unstructured data (consultation recordings and notes, patient instructions, social media discussions, diaries); and (4) structured data (ERP, Transactional data, Hospital/Clinical Information Systems, prescriptions, payment records). They are used for understanding and serving patients, monitoring

and real time adjustments of operations, performance optimization and improvement. Beyond the realm of Big Data, smart data analytics may also be used in e-health applications and home medical equipment.



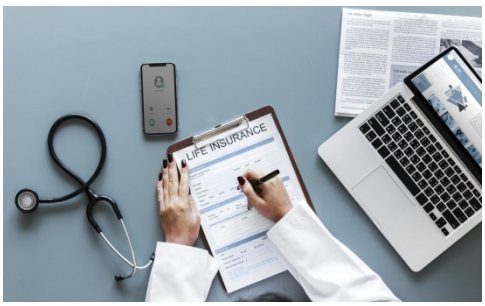
The healthcare sector has been adopting SIS in many different applications, ranging from early disease detection, identifying the spread of transmittable diseases, and improving the effectiveness of drugs and treatments. The use of Big Data is allowing healthcare scientists to advance their biomedical research, and AI is being integrated into disease analysis and other healthcare practices. The healthcare sector is also integrating embedded SIS in the forms of physical medical transportation robots, social robots, telepresence, and surgical assistants. The healthcare domain was one of the first, and most emphatic, adopters of SIS.

Big Data analytics is being used in hospitals to help medical staff work more efficiently, to provide a better service to patients and make a more accurate diagnosis. The development of different sensor technologies and wireless medical instruments (collectively termed as wearables) can monitor patients' health remotely by recording personal parameters such as blood pressure, heart rate or sugar levels. It also allows hospitals to save money, as fewer medical staff are needed for patients' daily care.

3.3. Smart Big Data in Insurance

Big Data has been used to provide customer insights for transparent and simpler insurance products, by analyzing and predicting customer behavior through data derived from social media, GPS-enabled devices and CCTV footage. Is is also being used for claims management and to offer faster service, since massive amounts of data can be analysed in the underwriting stage. Fraud detection has also been enhanced through Big Data. Through massive data from digital channels and social media, real time monitoring of claims throughout the claims cycle has been used to provide insights.

Research by Tata Consultancy Services revealed that the insurance sector invested in AI more than any other industry in 2015 ($124 million dollars) (Tata Consultancy Services Ltd (TCS), 2017). SIS are being used to process claims, detect fraud, risk management, marketing, and for insurance data analytics. Big Data is being analysed from a wide variety of sources, such as: social media data, registries, statistical data, personal data, sensors, and vehicle maintenance history (Bharadwaj, 2018; Deloitte Digital, 2017; Dutt, 2018; Foggan and Panagakos, 2018; Koh and Tan, 2018; Sennaar, 2018; Zagorin, 2018).



3.4. Smart Big Data in Retail and Wholesale Trade

Big Data systems and data analytics are being used for marketing and communications, optimization of staffing (in relation to predicted shopping patterns), story inventory and fraud reduction, amongst others. Big Data from customers and markets is being gathered amongst others from customer loyalty data, POS scanners, RFID, and local demographics data. SIS offers great potential for customer

identification, attracting new customers, customer retention, and customer development (Ngai et al., 2009, p. 2595). SIS offers businesses the opportunity to access customers easily online, and the ability to retrieve vast amounts of data about them to improve their marketing and sales. Customer relationship management SIS allows companies to develop their interactions with their clients (Chen and Popovich 2003). Companies have access to a wide array of data from their clients, with Cambridge Analytica having previously stated that they collect over 5,000 data points from over 230 million Americans (Cambridge Analytica, 2017; see also Cadwalladr and Graham-Harrison, 2018).

3.5. Smart Big Data in Science

In many fields, including natural sciences, engineering sciences, medical and life sciences, and social sciences, advances in research increasingly depend on the creation and mining of large data sets. The use of Big Data and AI is radically changing scientific investigation. Not only do they offer great potential to provide data on a scale never seen before, they are also being used to make predictive insights to progress biology, chemistry, physics, and earth science. They are being used for genome sequencing, cancer research, and to predict climate patterns.

With regard to biology, Big Data improves our capacity to sequence DNA. This is leading to benefits in the field of diagnostics, as it makes easier to detect genetic predisposition to diseases; moreover, DNA sequencing is positively affecting agriculture and livestock breeding. Concerning the environment and earth science, Big Data is helping scientists to monitor the planet to better understand and address climate change. In the field of chemistry, modern particle accelerators require Big Data analytics to detect relevant patterns arising from experiments. The same goes for astronomical observatories that collect a large amount of data and require Big Data analytics to move forward in the fields of astronomy and astrophysics.

Big Data is currently helping scientists to make progress in the field of medical and cognitive science. The causes and the best ways to address several diseases can be discovered with the contribution of Big Data analytics. Furthermore, the quantity and the complexity of data related to brain functions can be better handled by modern algorithms, which are helping neuroscientists to better understand how the human brain works. Big Data is also benefiting the field of artificial intelligence and robotics, as AI systems can exploit the enhanced processing capability provided by Big Data analytics to effectively interpret the surrounding environment and react to it.

3.6. Smart Big Data in Education

Big Data systems are used to monitor student performance at different educational levels, for example by logging online behavior and overall progress. They are also being used to provide customised educational programs and to improve the learning experience in real time. They are also used to measure teachers' performance and effectiveness, and fine-tune it against student numbers, subject

matter, student demographics, student aspirations, behavioral classification and several other variables. Big Data is also used by governments and educational organizations to develop analytics to monitor school performance and to reduce dropout numbers.

Large classrooms make it difficult to take care of every single student. Big Data offers a solution to help teachers monitor the educational path of everyone. Personalised learning outcomes can be assigned to each student and the same can be done with teachers - whose performance can be kept under surveillance and eventually improved by means of customised interventions. Big Data can also help students figure out what might be the best career for them, on the basis of their strengths, tastes and abilities.

Big Data can also be used to better organise classrooms on the basis of students' performance and learning goals. By combining behavioural data with information about students' general condition, administrators could find patterns in behaviour against data about family, location or socioeconomic background. In this way, classrooms and lectures can be better organised. Big Data could be used to make predictions about students' educational paths and future careers to help institutions (the state, schools, corporations) decide where, when and to whom resources should be directed.

3.7. Smart Big Data in Energy and Utilities

Energy companies use smart Big Data for energy management, energy optimization, energy distribution, and building automation in utility companies. Utility companies are using smart meter granular data to analyse consumption of utilities, which allows for better control of utilities use. The use of Big Data also allows for better asset and workforce management, which is useful for recognising errors and correcting them. SIS are being deployed in the energy sector to solve the Energy Trilemma: securing energy; producing affordable energy for all; in a sustainable manner. Smart grids provide the potential to improve the monitoring and control of energy consumption through the use of real time data. Smart meters provide SIS with the data needed to predict and optimise energy requirements.

3.8. Smart Big Data in Manufacturing and Natural Resources

In the natural resources industry, Big Data allows for predictive modeling to support decision-making that integrates large amounts of data from geospatial data, graphical data, text and temporal data. Areas of interest include seismic interpretation and reservoir characterisation. Big Data has been used in solving today's manufacturing challenges and to gain competitive advantage. It is being used to optimise production processes, for better forecast of product demand and production, for better understanding of plant performance, for providing service and support for customers faster, and for real time alerts based on manufacturing data. It is also being used to better control supply chains and manage supply chain risk, to perform predictive modeling of manufacturing data, to mine combinations of manufacturing and other enterprise data, to improve interactions with suppliers, better quality assurance and custom product design.

SIS is commonly used in the manufacturing industry, and is often integrated in what has now become known as 'Industry 4.0', which was first introduced in 2013 by the German government (Lee, 2014; Li, 2017; Wan, 2017). SIS in manufacturing promises responsive, or "agile" supply chains, through a better understanding of market trends and customer preferences (Tiwari, 2017, p. 15). There are many promising benefits from SIS in manufacturing, namely: insights about customers, improved services, and understanding customer behaviours and demands (Feki, 2016); identifying key customers (Sanders, 2016, p. 31); smarter pricing (Tiwari, 2017, p. 11; Zhong et al., 2016, p. 574); and new product development (Chae, 2015, p. 257). SIS also holds the potential to optimise supply chain and logistic operations by providing: ways to improve productivity (Auschitzky, 2014, p. 3); anticipate shipping times (Leveling, 2014, p. 4); SCM risk prediction (Chae, 2015, p. 257); and the reduction of hazardous material and carbon emissions (Zhao, 2017).

3.9. Smart Big Data in Agriculture

Smart Big Data is being used to provide predictive insights in farming operations, drive real time operational decisions, and redesign business processes in fundamental ways. Smart Big Data allows for the development of all kinds of precision farming tools, such as yield monitoring, field mapping, crop scouting and weather forecasting. It is being used for surveying crops (using drones and sensors), accurate crop predictions, automating planting and harvesting, improving seeds and other products, and reducing environmental impact. SIS are seen as the next step in the agricultural revolution to meet the world's growing food demands (Kumari, Bargavi and Subhashini, 2016; Morota et al., 2018; O'Grady and O'Hare, 2017).

SIS will take on a large role in developing innovative and effective ways to "improve water and air quality, improved soil health, food quality and security, protection of biodiversity, improvements to quality of life, increase output, cost reductions, crop forecasting, and improved decision-making and efficiency" (Macnish et al., 2019). This type of 'prescriptive farming' will revolutionise the agricultural industry, allowing farmers to maximise crop yields, identify plant disease, and manage their farms more effectively (Antle, Capalbo and Houston, 2015; Carolan, 2015; and Zhang et al., 2014). Most agribusinesses are now developing their SIS, such as Monsanto, Bayer, BASF, DuPont Pioneer and John Deere (Sykuta 2016).

3.10. Smart Big Data in Communications, Media and Entertainment

The use of SIS in communications, media and entertainment includes the following types of companies and organisations:

- Marketing, advertising and public relations companies
- Telecommunications companies
- Social media companies

- Publishing companies
- Information service companies and organizations (search engines, online databases, wikis)
- Entertainment companies (music, film, games)

As an increasing amount of data becomes available to media companies, algorithms are used to analyse large datasets to extract relevant facts, interesting stories and ultimately generate material of public interest. Big Data analytics also plays an important role in social media. Millions of tweets, images, status updates and visualisations are analysed in real time in order to create value-added services for users and to sell valuable information to other companies that are interested in people's habits.

In the field of communication, Big Data can be used to communicate a great variety of different content. Based on geolocation, smartphone applications can suggest to users places to visit or services to access. Advertisement companies can exploit Big Data analytics to access information about people and to create personalised ads, but also to see how people react to the ads. Data mining can be useful for all kinds of information sources on the internet, as it can be used to monitor users' behaviour, reactions to content, or preferences in general. Companies can optimise data to offer a better service to users, while all the information and services promoted by companies can be designed to better meet users' needs and tastes.

With regard to the entertainment industry, Big Data analytics is used to analyse catalogues of movies and TV series in order to draw conclusions about what people like the most and create new products of the same kind. In the music industry, algorithms can be used to understand what hits or kinds of music are more likely to be appreciated by listeners in the immediate future. Future hits can therefore be created on the basis of these predictions. Data mining can also be exploited by game makers to monitor gamers' behaviour so as to improve their gaming experience and create products based on their preferences. Organisations can analyse customer data along with behavioral data to create detailed customer profiles that can be used to:

- Create content for different target audiences
- Recommend content on demand
- Measure content performance

Smart Big Data is being used for digital advertising, for targeted, personalized marketing and for recommender systems. It is also being used to provide personalized and location-based services. In addition, an increasing number of consumer products collect data and send it to communications and media companies, retailers and manufacturers.

3.11. Smart Big Data in Transportation

Smart Big Data applications are being used by governments, private organizations and individuals:

- Governments use of Big Data: traffic control, route planning, intelligent transport systems, congestion management (by predicting traffic conditions)
- Private sector use of Big Data in transport: revenue management, technological enhancements, logistics and for competitive advantage (by consolidating shipments and optimising freight movement)

- Individual use of Big Data includes: route planning to save on fuel and time, travel arrangements in tourism etc.

In recent times, huge amounts of data from location-based social networks and high-speed data from telecoms have affected travel behaviour. Regrettably, research to understand travel behavior has not progressed as quickly. Smart and driverless cars rely heavily on data analytics, and Big Data and the car of the near future is essentially part of a gigantic data-collection engine. The cars have embedded computers, GPS receivers, short-range wireless network interfaces, and potentially access to in-car sensors and the Internet. Furthermore, they can interact with roadside wireless sensor networks on roads where these networks are deployed.

3.12. Smart Big Data for Employee Monitoring and Administration

Organisations in the private and public sector use Big Data and data analytics for employee administration and monitoring. They increasingly use it to enhance employee performance and work experience. Systems are being used to locate and profile potential employees, to enhance screening in the hiring process, for monitoring employee activity, for better task coordination between employees, for measuring and providing feedback on employee performance, for measuring employee well-being and satisfaction, for tracking employees, for predicting illness, and for predicting and identifying crime and fraud in the workplace.

The reason behind employee monitoring is to ensure that employees are not misusing their time in work, or doing illegal/harmful activities, and to generally improve productivity (Frayer 2002). More succinctly, using SIS for employee monitoring purposes is for the “prevention of related image damage, defence of corporate espionage, a general intended protection of corporate assets, detection of illegal software and missing data, increase of productivity, detection of reasons for a disciplinary warning letter or a termination, significantly reduced costs and increased availability of surveillance technologies, and others” (Macnish et al., 2019).

3.13. Smart Big Data in Government

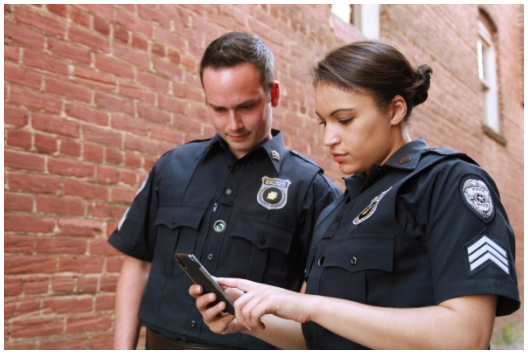
Joining up public sector data sources can make government more efficient, save money, identify fraud and help public bodies better serve their citizens. In *public services*, Big Data has a very wide range of applications, including energy exploration, financial market analysis, fraud detection, health-related

research and environmental protection. Amongst others, Big Data is being used in the analysis of large amounts of social disability claims, to inform social workers about their clients, to detect and study patterns of food-related illnesses and diseases, and to manage climate change.

SIS offers the public sector the ability to improve local and national services and amenities, while reducing costs, environmental impact, and management constraints (Zanella et al., p. 23). It is estimated that a variety of governmental bodies will require the use of SIS in the future in areas such as: housing, offices, transportation, security, decision-making, e-services and healthcare (Bibri 2018; and Rjab and Mellouli 2018). Data mining can reveal delicate situations that are usually hidden from the public eye, such as human rights violations or natural disasters. Big Data can provide information about populations, groups' and individuals' situations, and governments can use it to become more responsive towards the people represented.

3.14. Smart Big Data in Law Enforcement and Justice

Big Data analytics is also being used for intelligence gathering, surveillance, and prosecution. Smart Predictive analytics allow for predictive policing through hotspot mapping and predictive risk assessment of individuals. Tablets, smartphones and mobile biometric devices are integrated into smart Big Data systems, which allow for mobile policing with real time analysis. Drones and cameras connected to smart Big Data systems allow for smart visual surveillance. Smart case databases allow legal professionals to draw insights and connections using advanced analytical algorithms. 'Legal analytics' is the application of data analysis methods and technologies within the field of law to improve efficiency, gain insight and realise greater value from available data. Legal analytics can, amongst others, help lawyers predict the behavior of judges and juries. Blockchain and smart contracts are changing the way in which agreements are made and documented in law.



3.15. Smart Big Data in Sustainable Development

In 2015 the United Nations (UN) developed 17 Sustainable Development Goals (SDGs) to meet, or strive towards, by the year 2030. These range from reducing inequalities, ensuring environmental sustainability, eliminating hunger/starvation, and ensuring justice and fairness around the globe (United Nations 2018). SIS are being heralded to achieve many of these objectives by providing policymakers with insights and predictive analysis that they would not have had previously. Smart Big Data systems use as data sources data that is relevant to understanding human well-being, development of capabilities and infrastructure, climate change and environmental degradation. They are being used to make smarter and better decisions and provide better monitoring and evaluation, for example to see where funds are going, whether change has occurred and what caused it, and to ensure better collaboration between different agencies.

3.16. Smart Big Data in Defence and National Security

Smart Big Data is increasingly being used in defense (including cyber-defense) and security. Key national security missions include conventional military defense, counter nuclear proliferation, counter chemical/biological WMD, counter terrorism, cybersecurity, and counter-intelligence, and may also include counter narcotics, counter money laundering, and actions against organised crime.

Intelligence gathering includes human intelligence, gathered from a person on the ground; geospatial intelligence, gathered from satellite, aerial photography, and mapping/terrain data; measurement and signature intelligence, gathered from sensing instruments for the purpose of identifying distinctive features associated with the source, emitter or sender, to facilitate the latter's measurement and identification; cyber-intelligence, gathered from cyberspace; financial intelligence, gathered through analysis of monetary transactions; and several others. Similarly, smart Big Data can provide strong support for planning and operations, including predictive analytics and real time decision support.

4. Ethical Analysis: General Ethical Issues

This section focuses on the most predominant ethical issues which will be faced by individuals and society as a result of the implementation and use of SIS. The first part examines a number of concerns related to the aims of SIS and the epistemological challenges, followed by 24 ethical issues arising from the development and use of SIS. The aim of this section is to provide a comprehensive analysis of the most pressing ethical concerns related to SIS in general, before we examine potential issues related to these technologies specifically (section 5 of Task 1.4), and issues related to their application in 16 social domains (section 6 of Task 1.4). This analysis will provide the backbone for the SHERPA project's ethical analysis of SIS and will provide us with insights as to how to incorporate these issues into our Ethical Guidelines in Deliverable 3.2.

4.1. Concerns Regarding the Aims of Smart Information Systems

4.1.1. Epistemological Concerns Regarding the Aims of Big Data

Big Data, which embodies high volume, velocity, exhaustiveness in scope and huge variety of data, is widely discussed today as it offers tremendous opportunities (Kitchin, 2014). Conceptually, the term Big Data leads people to believe that this phenomenon is about the amount and 'bigness' of data, however, Big Data is really about the capacities that it offers, namely to "search, aggregate, and cross-reference large data sets" (Boyd & Crawford, 2012, p. 663). The real problem and epistemological challenge relates to finding the small patterns (Floridi, 2012). It is the small patterns that hold value

and competitive edge. Alongside new techniques and technologies, we need an epistemology to help us find these small patterns. Small data will remain hugely valuable in the era of Big Data, as it enables greater control and allows researchers to pose and answer specific questions (Kitchin & Lauriault, 2014).

Some commentators argue that Big Data knowledge will fundamentally change knowledge production, creating a new epistemology, but Kitchin (2014) refutes these claims. Moreover, creating separate discussions and policies will make it difficult to integrate Big Data practices into existing frameworks, and can lead people to use the novelty as a means to "undermine hard-won ethical, legal, and other norms" (Lipworth, Mason & Kerridge, 2017). There is also the claim that Big Data may encourage the 'end of theory' by replacing the uncovering of causal relationships with correlations. However, others propose that data-intensive science "aims at identifying causal structure" and is situated within a hierarchical structure that is not too far removed from conventional scientific modelling or general epistemological frameworks (Pietsch, 201, p. 2).

Data-intensive science requires representing configurations of phenomena that are relevant to a specific research question, whereby configuration refers to a specific combination of values for different variables, captured by eliminative induction in a specific research context (Pietsch, 2016, p. 4). Another characteristic of data-intensive science is the automation of scientific discovery. In data capture, processing and modelling allows for overcoming some of the limitations found in depending on human cognition to uncover patterns of significance, but this also has the drawback of reducing our understanding of the results. Thus, while data-intensive science may aim to make sense of complex phenomena through various algorithmic techniques (such as using classification trees) and speed up the process of discovery, this may be at the risk of incurring issues of interpretability by reducing human participation in the discovery process.

Furthermore, algorithms and data-intensive modelling can handle various elements of causal complexity (Pietsch 2016). The utility of Big Data is in its capacity to store, manage and interpret volumes of data, and the ability to find and deduce information in a manner that exceeds human capability (Ekbia et al., 2015, p. 1528). Big Data analytics methods, for some commentators, ought to be considered as a supplementary tool rather than as a replacement for the scientific method (Calude & Longo, 2017). Algorithms may appear to be 'scientific' and value-neutral, i.e. the belief of 'scientism', but algorithms are scientifically flawed instruments (Johnson 2014). There is also the belief that data can 'speak for itself', which reveals a number of ideas which underpin the rise of empiricism and pseudo-positivism in Big Data-driven science (Kitchin 2014). This belief is based on:

- the idea that big data can display a whole domain in full resolution;
- that there is no need for *a priori* theories or models;
- that the data analytics techniques/software used are agnostic (i.e. free from human bias); and,
- that the meaning found in the data "transcends context or domain-specific knowledge" (Kitchin, 2014, p. 265).

There is an assumption that "big data will lead to much better forecasts" in a diverse range of fields and disciplines, including scientific discovery, medical diagnosis, along with financial, commercial and political applications (Hosni & Vulpiani, 2017, p. 2). But this assumption may come as a result of accepting the extreme inductivism at work in the use of Big Data for predictive analytics and forecasting. This inductivism relies on two assumptions: "Similar premisses lead to similar conclusions (Analogy)" and "Systems which exhibit a certain behaviour, will continue doing so (Determinism)" (Hosni & Vulpiani, p. 7). However, reliance on analogies and determinism are prone to mistake correlation with causation (Hosni & Vulpiani, 8). The inductive character of predictive algorithms may for example, lead to racial profiling, because they require "one to think of the disposition to commit

crimes as a persistent feature of certain people, who in turn, tend to conform to certain specific features" (Hosni & Vulpiani, p. 7).

Similarly, the relationships between data analytics and implementing machine learning techniques often involve "accounting realism" (Rieder, 2016). Algorithms do not test or apply a hypothesis, but instead assess truth or validity in relation to a specified objective such as profit maximisation (Rieder, 2016, p. 44). Thus, when algorithms make decisions (such as hiring new employees based on a statistical model related to number of sales to infer performance level), then the decision made by the algorithm reflects the criteria that they are trained to look for.

Alongside accounting realism, data analytic methods also depend on a datafication process which is based on the belief in the objectivity of quantification, and especially the potential of tracking all kinds of human behaviour and what can be termed sociality in online data for predicting future human behaviour (Van Dijck, 2014, p. 201). But the data generated by online platforms is not purely objective, because of the role of intervention by the owners of these platforms in the process of algorithmic optimisation. In the case of social media platforms, "trending topics" may for instance be commonly perceived as representations of spontaneous online sociality, but the algorithms underlying what is listed are systematically fine-tuned to channel user responses (Van Dijck, 2014, p. 201).

To understand 'sociality' requires specific analytical methods that require critical interrogation, such that researchers in different disciplines may observe and ask questions differently given their discipline, even if looking at the same data. This has problems for how Big Data (specifically in terms of visualisation techniques) can have impacts on how individuals gain knowledge (Lewis & Westlund, 2014). In the domain of journalism especially, infographics and interactive data visualisation tools can encourage audiences to "play" with the data to comprehend a particular and personalized version of the news narrative (Lewis & Westlund, 2014, p. 7). Such interactivity presents three epistemological concerns:

- the form of the knowledge matters, given the news medium utilized,
- the production of knowledge (based on journalistic norms) is tied to the visualizations used, and
- public acceptance of knowledge claims (based on what conditions legitimize these claims) may be based on the visualization rather than concerns of truth (7-8).

Transparency, press councils, clear codes of conduct and healthy media criticism are necessary to verify these techniques, and the content and interpretations made in the curation of news produced via algorithmic decision-making (Diakopoulos, 2017, p. 27).

4.1.2. Epistemological Concerns with Regards to the Aims of AI

Increasingly, employing algorithms may imply that we are altering knowledge production, where knowledge is conformed to the logic of the algorithms (Gillespie, 2016). Kitchin (2017) makes pertinent four concerns on the increasing dependency and use of Big Data analytics:

- their increasing influence in shaping human life necessitates critical investigation;
- algorithms "are best understood as being contingent, ontogenetic and performative in nature, and embedded in wider socio-technical assemblages";
- access to how they are formulated, their heterogeneous character, contextual and contingent unfolding complicate research; and,
- there are various ways in which the constitution and work of algorithms can be studied but that employing a combination of such methods is best to overcome an array of challenges

caused by algorithms. There is not a one-size fits all form of Big Data (Kitchin & McArdle, 2016, p. 9).

Conversely, research in AI is not aimed at the construction of super-intelligent machines with traits like modesty or honesty, but instead focuses on goal fulfillment and what can be called "optimization power" (Muehlhauser & Helm, 2012, p. 3). But with the increased computational power of AI there is also the issue of "singularity": If humans are not the most intelligent beings on earth raises questions as to how do we stay in control of a complex intelligent system, or if AI will have some advantage over us (Bossmann, 2016).

AI may create an imbalance of power between individuals and societies. For example, Bostrom (2013) and Bostrom and Yudkowsky (2014) argue that super-intelligent systems may be capable of making their own plans. The question arises about superintelligent and therefore independent AI: Where does the data of the AI come from? How much data can an AI have? Are there any limitations to the AI? The moral status of AI is also questioned: "the prospect of AIs with superhuman intelligence and superhuman abilities presents us with the extraordinary challenge of stating an algorithm that outputs superethical behavior" (Bostrom & Yudkowsky, 2014, p. 18).

In a similar line of analysis, an important way to make sense of the structure of algorithms is to look at agency, and specifically what algorithms are capable of (van Otterlo, 2017). van Otterlo (2017) distinguishes five broad classes of algorithms:

- algorithms that can reason, search and infer based on the training data they are supplied (e.g. for translation or image recognition);
- algorithms that learn and find generalized patterns from within the data;
- algorithms that optimise for the best possible action and rank items (e.g. best food or matches on dating apps);
- physical manifestations such as robots; and
- superintelligence (van Otterlo, 2017, p. 4).

Algorithms can rank and classify individuals from their identity (given the data they are fed), and can have adverse effects on opportunities (such as employment or credit scoring) as well as exposing vulnerabilities (such as increased surveillance, manipulation, exclusion and discrimination) (Balkin, 2017, p. 1235). These algorithms work in a manner very different from human intelligence, and "achieve the results that we see today [because programmers] abandoned the ambition to reproduce in digital form the processes of the human mind" (Esposito, 2017, p. 4). Algorithms are impacting more human lives the more they are deployed, and the fact that they function without the parameters of human intelligence makes it more problematic, because they only process data and make decisions, and ethical consequences are not properly framed in their functioning.

4.1.3. Ethical Concerns Directed at Smart Information Systems

We are increasingly using Big Data and algorithmic decision-making, thus, algorithms are increasingly shaping human life. Algorithms are best described as 'mathematical constructs' that have specific purposes with "given provisions" and help translate large amounts of data into meaning (Mittelstadt et al., 2016, p. 2). The more complex and 'intelligent' an algorithm is, the more autonomously it can operate. The increased investment and deployment of algorithms in decision-making processes has led to issues concerning how exactly algorithms make their decisions and what kinds of ethical issues arise as a result.

Algorithms reaching conclusions from statistics or machine learning "produce probable yet inevitably uncertain knowledge", which means that while they may usefully find correlations and patterns, such findings "are rarely considered to be sufficient to posit the existence of a causal connection" (Mittelstadt et al., 2016, p. 4). Algorithms are not infallible, as the output (i.e. decision) may be based on inconclusive knowledge. While it may be assumed that the connection between the data being processed and the output reached by the algorithm is accessible, this is not always the case. A lack of interpretability, and verifiability of the data being used (especially the scope, provenance and quality of data), may cause epistemic as well as practical problems in assessing the decisions made by algorithms, making them inscrutable.

The predictions and decisions reached by algorithms are only as reliable as the data that is input, which means that biases in the inputted data will affect the neutrality or lack thereof of the output. There is also the potential for algorithms to produce unfair and discriminatory actions. There is an additional difficulty with regards to algorithms that has to do with not being able to confer responsibility and accountability when algorithms cause negative effects. In cases where harm is caused by an algorithm's decision, it is difficult to find the cause and to identify accountability, due to algorithms' 'traceability problem'. More so, the effects of algorithms and Big Data analytics are not uniform or homogenous. They may affect a wide range of stakeholders in a variety of ways, from individuals to organizational and societal groups.

- For individuals, the ethical concerns they raise are: data ownership, data control, awareness of data procurement and use, trust in the agencies concerned, privacy, self-determination and fear from the pervasiveness of algorithmic decision-making (Someh et al., 2016, p. 6-7).

- For organisations, they are: competitive pressure concerning algorithmic performance, data quality, data sourcing, data sharing, algorithmic decision-making, presentation of data, ethical capability, ethical culture, ethical governance, ethical performance and reputation (Someh et al., 2016, p. 7).

- And for society, the issues are: power asymmetries, dependency, social awareness of the public, surveillance, the need for guidelines and authority (Someh et al., 2016, p. 7).

The following section will evaluate many of these ethical issues, while also adding additional ethical concerns.

4.2. Ethical Issues Regarding the Implications and Risks of SIS

4.2.1. Access to SIS

An area of concern, especially for researchers, is the diverging levels of access to Big Data. Some companies restrict access to their data entirely, and others sell the ability to access the data for a fee, while others offer small datasets to university-based researchers (Boyd & Crawford, 2012, p. 674). This uneven access to SIS may produce a power asymmetry, whereby only students and researchers from top universities have access to data sets, while everyone else is left without (674). Another worry is that researchers who do get access, may not have full freedom to investigate the datasets as they wish, as any contentious questioning may lead to their access being revoked (675). A point of concern is the level of access, as well as exclusion from access, which makes for proper investigation of analytics techniques and the methods that are used difficult.

4.2.2. Accuracy of Data

SIS technology holds the potential to supersede the scientific method in importance because of the belief that algorithms sufficiently trained on large databases can discover patterns and regularities that lead to predictions and decisions independent of meaning or context (Calude & Longo, 2017, p. 3). This data-oriented methodology relies on the size of the databases used for the algorithms to find correlations. Accurate data is important because these correlations may allow us to predict future outcomes (8). One of the main criticisms against this understanding of SIS is that for "any coding of an arbitrary database of a large enough size into a string of digits, there will be correlations of a pre-determined arbitrary length" (12). The predictive power of correlations is not given by an algorithm (or criteria/relevance that the algorithm is meant to explore); it is given by the size of its database. If numerous correlations become observable in an immense database, the correlations may be arbitrary, and not necessarily because of any relevance/criteria such as proximity or separateness of observable phenomena.

The old adage 'garbage in, garbage out' is powerfully relevant to SIS development and use. Big databases that generate poorly curated, gamed or biased data will likely produce predictions that have weakened validity that lower the utility of the analytics methods used (Kitchin and Lauriault, 2015, p. 466). The curation and interpretation of data is an important aspect in garnering the value of Big Data and the accuracy of any correlations or patterns found in datasets. While large amounts of data are being collected and analysed about individuals, this data is only meaningful after aggregation, correlation or calculation (Couldry & Powell, 2014, p. 3). Thus, the accuracy of Big Data is determined from the processing of the data that can improve or decrease the validity of predictions made from analytics methods. Thus, while Big Data algorithms may appear reliable and value-neutral, they require the active interpretation of researchers, who may bring their own biases and interpretations (Crawford et al., 2014, pp. 1669-9).

While advocates of Big Data analytics may think that the data is able to speak for itself, "meaning emerges from the interaction of data and an analyst", and so the interpretation may contain "biases or misreadings of big data which are consequent on the method of its analysis" (Fuller, 2015, p. 578). There is a need to focus on how Big Data is 'read' and interpreted (Van Dijck, 2014), for data scientists to be skilled in understanding, interpreting, and presenting data. Furthermore, data scientists should be self-aware of their interests when claiming that the data they are using is objective or free from bias. Making statistical claims about datasets relies on knowing where the data is coming from, accounting for weaknesses in the datasets (i.e. inaccuracies or missing labels), and looking out for biases - not just in the data, but also in the interpretation of the data (Boyd & Crawford, 2012, p. 668). If this approach is lacking, there is the possibility of misinterpreting data, implementing biases, and diminishing the accuracy of SIS recommendations.

4.2.3. Accuracy of Recommendations

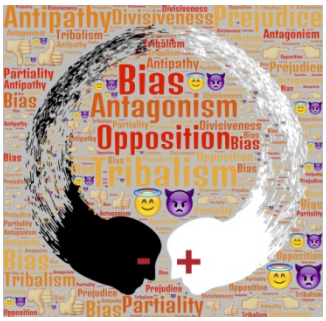
The growing demand for and use of predictive algorithms in varying sectors (e.g. healthcare, insurance, education, and banking) has led to a scoring trend in these sectors. The development of robust learning algorithms has meant incremental removal of humans from predictive algorithm processes, whereby new forms of learning are projected by data mining programs once they have found a range of correlations and inferences (Citron & Pasquale, 2014, p. 5). Reduced human scrutiny will mean decisions based on scores (e.g. who will receive a loan, and who will not, based on specified indicators), which can lead to a chain of programs that not only make decisions but also decide which indicators to look for.

Furthermore, the use of statistics in algorithms may produce probable, but sometimes uncertain outcomes (James et al., 2013). Patterns discovered by SIS do not always justify a causal connection,

so there is a risk of inaccurate or wrong outcomes and conclusions, as well as a simplification of models that can in turn be inaccurate, and/or discriminatory (Ananny, 2016; Barocas, 2014; Hildebrandt, 2011; Illari & Russo, 2014; and Miller & Record, 2013). It is also difficult to reproduce falsified algorithmic results (Ioannidis, 2005; and Lazer et al., 2014). However, inscrutable evidence resulting from SIS should be accessible in order to expose how the data used by ML and AI contributed to the conclusion (Miller and Record, 2013).

4.2.4. Algorithmic Bias

If algorithms draw conclusions using inferential statistics and/or machine learning techniques, they may produce probable, but essentially uncertain, information (Mittelstadt et al., 2016, p. 4). For instance, individuals might be mistakenly denied some public services based not on their own actions but on the actions of others with whom they have some commonalities (Lepri et al., 2017). The outcomes of such algorithms are called *inconclusive evidence* (Mittelstadt et al., 2016, p. 4). It is inconclusive because the above-mentioned techniques can only help identify significant correlations, not causal connections, between phenomena. Therefore, it is often not sufficient to motivate public actions on the basis of insights of such a connection (Mittelstadt et al., 2016). If there is not enough evidence to justify an action, then there is both a problem of legitimacy and also a problem of potentially biased algorithms (Kraemer et al., 2011; Newell and Marabelli 2015; and Macnish 2012). Bias can appear in social values, or bias included in the data (Diakopoulos 2015; Friedman and Nissenbaum, 1996). For instance, the example of Amazon's[1] AI to hire people was shown to be gender-biased because it concluded that male candidates were almost always better suited for the job.



Biases in the design and implementation of algorithms can take three forms: pre-existing bias, technical bias and emergent bias (Friedman & Nissenbaum, 1996). Pre-existing bias is a prejudice already existing in society or in particular individuals, which is transmitted by the algorithm's programmers during the design process. Since data about human beings represents the ultimate training source for the algorithm, all human biases and prejudices are inevitably absorbed and repeated by the algorithm (Barocas & Selbst, 2016; Kim, 2018). Biases embedded in hiring algorithms may exclude some vulnerable groups and therefore lead to discrimination in the decisions of who is hired and who is excluded (Barocas & Selbst, 2016; Kim, 2018; Lepri et al., 2018). For example, by means of heterogeneous data collected from social networks (e.g. concerning someone's preferences, the kind of pages visited or their network of friends), algorithms are able to make predictions about people's ethnicity, sexual orientation, political views, or even calculate their happiness and intelligence. An employer using this data for personnel recruitment may inadvertently, or purposefully, do so in a discriminatory manner (Raub, 2018).

An instance of discrimination created by biases embedded in an algorithm can be seen in Amazon's "prime-lining", where low-income minority neighbourhoods were excluded from their service. In this case, the "low income" and the "minority" labels were actually proxies for race (Jackson, 2018). In another case, Google was showing men advertisements for higher-paying jobs, while women were shown more generic advertisements (Datta, Tschantz & Datta, 2015). Predictive policing is being used to better identify and catch criminals through the use of algorithms to create profiles of people who are deemed to be indicative of criminal behaviour (Jackson, 2018).

[1] Retrieved from: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Technical bias can arise from flaws in computer tools, difficulties in ascribing social meaning when the algorithm is developed out of context, or other innate technical imperfections (Friedman & Nissenbaum, 1996). For instance, if an algorithm is supposed to implement a system of random choice, possible flaws may lead to discriminatory consequences. Some individuals or groups may be excluded or over-represented by the criteria that algorithms look for, which inform the decisions they make. For example, Kim (2018) imagines that a hiring algorithm could come to the conclusion that "liking curly fries on Facebook predicts intelligence", due to the detection of a casual recurrent pattern. If the algorithm in question starts labelling candidates as qualified or unqualified on the basis of this wrong correlation, its decisions will lead to bias.

Emergent biases materialise in the context of use after the design is complete, for instance as a result of the changing of societal norms or shared values (Friedman & Nissenbaum, 1996), as well as the applications that algorithms are used in (Mittelstadt et al., 2016). Even though the bias may not be embedded in algorithms, it may well be rooted in the user's mindset. For instance, people are likely to spend more time on social media and see more ads that are related to their interests and opinions. Social media creators and advertisers are therefore likely to use algorithms to exploit this in a biased way to ensure their own interests (Sleeman & Rademan, 2017).

In these contexts, the algorithm will be biased towards the advertisers and away from the needs of the customer (Brin & Page, 2000). Consumers end up being negatively affected by the presence of bias in this use of algorithms, since it encourages "the formation of 'echo chambers' or 'social bubbles' that could significantly entrench the ideologies of users without providing an opportunity for these views to be challenged" (Sleeman & Rademan, 2017, p. 3). In other words, users' ideas risk stagnating, as the possibility to question one's own opinions is undermined since their newsfeeds are curated to show what they have already 'liked'. This becomes an important concern, especially in the spreading of political information. Potential voters can easily be identified and targeted with personalised ads to mould their political ideas (Kim, 2016) while alternative views are not made visible to them.

4.2.5. Discrimination

Discrimination occurs when individuals are profiled based on their online choices and behaviour, but also their gender, ethnicity and belonging to specific groups, affecting the type of information they are provided with, and/or how they become treated (Calders et al., 2009; Cohen et al., 2014; and Danna and Gandy 2002). Data-driven algorithmic decision-making may lead to discrimination that is then adopted by decision-makers and those in power (Lepri et al., 2017, p. 4). SIS may become powerful tools to stigmatize and discriminate, so regulators should have the ability to test the fairness and accuracy of algorithmic scoring systems, and citizens should be able to challenge when these algorithms cause them harm (Citron and Pasquale 2014). While ensuring non-discrimination in classification models is a challenging task, and the desire to fully eliminate discriminatory attributes may be naïve, action still needs to be taken to reduce discriminatory outcomes from SIS (Pedreschi, Ruggieri and Turini 2018). In the use of Big Data analytics, discrimination can arise from four different sources:

- how the input data is weighted can lead to disparate impact;
- categorization (e.g. classifier variables) may be considered a form of direct discrimination that leads to disparate treatment of those categorized under certain labels;
- the misuse of certain models in different contexts; and
- if biased training data is used then biases will be perpetuated leading to discoveries appearing as evidence of proof (Lepri et al., 2017, p. 4).

Discrimination in algorithms may be conscious or unconscious acts by those employing the SIS, or a result of algorithms mirroring society by reflecting pre-existing biases (Baroccas and Selbst 2016). SIS have the potential to affect the level of inequality and discrimination, and if institutional biases are not highlighted or corrected, these systems can reproduce existing patterns of discrimination and inherit the prejudices of prior decision-makers (Barocas & Selbst, 2003, p. 674). While in some cases discrimination and bias may be intentionally embedded in algorithms, discrimination may be an emergent by-product of the data mining process itself when arbitrary correlations or weights (in the sense of relevance) are given to certain variables (674).

Consequently, what a model learns depends on the training data it learns on. Decisions based on incorrect or misleading data hold the potential to be used to discriminate against individuals and groups of people. The predictive power and efficacy of SIS is tied to the training data it learns from, such that the quality of the data, and type of representation (either overrepresentation or underrepresentation) can lead to discriminatory decision-making. The outputs of algorithms that are fed with biased data is in turn called *misguided evidence* (Mittelstadt et al., 2016, p. 4-5). Existing patterns of discrimination, based on the prejudices and/or misinformation embedded in society, can be easily reinforced by data-driven algorithmic decision-making processes. Inequalities resulting from such patterns of discrimination (especially against gender, race and class) can exacerbate historically disadvantaged groups that "deserve" less favorable treatments based on their current situation, without considering why they are part of such groups (Lepri et al., 2017).

4.2.6. Economic

Big Data has the potential to create massive economic benefits to those developing and employing them: "harnessing the promise of big data through the widespread collection of disparate online transactions and interactions coincides with its cost efficiency in targeting niche markets and providing oversight of populations" (Crawford et al., 2014, p. 1666). Besides its computational structure and dynamics, the rise of Big Data is also tied to its role as a spur for innovation. However, there is a rhetoric that providing Big Data "is to contribute to the advancement of science, innovation and learning" (Crawford et al., 2014, p. 1666). However, there are also dangers of Big Data gathering, for example "repositories of data are characteristically unstable; data is leaky, and it escapes in unexpected ways, be it through errors, hacks or whistleblowing" (Crawford et al., 2014, p. 1666).

Big Data analytics has the potential to boost the economy and improve the efficiency and productivity of corporations. More specifically, Big Data can help optimise the utilisation of resources, eliminate wastage and increase performance (Badri, Boudreau-Trudel & Souissi, 2018). Moreover, Big Data analytics can provide accurate and detailed information about the market, as it can measure even small changes in wages, employment and commercial operations (Einav & Levin, 2014). While the information about the socio-economic situation can be used by governments to address economic issues, productivity and innovation in the free market can lead to a general increase in consumers' welfare.

However, the flow of information generated by data mining does not always lead to positive outcomes. In fact, the whole process of data mining, elaboration and generation of valuable information introduces substantial new asymmetries of power and knowledge. Corporations can gain accurate knowledge about people's tastes and behaviour through their data, often unknowingly to the end user (Zuboff, 2015). Big Data corporations can facilitate new forms of price discrimination aimed at extracting the highest price for goods from each customer. This form of "predatory

marketing" has the effect of enriching Big Data companies at the expense of consumers' welfare and privacy. As a result, economic inequalities are likely to be consolidated and exacerbated (Newman, 2014).

According to "surveillance capitalism", the relationship between companies and the population is not equal anymore: while traditionally people and companies needed one another for employment and consumption, nowadays this mutual relationship is increasingly weakening. In this new model, the tools and services made available by Big Data companies are not exchanged for something of equal value. Instead, "they are the 'hooks' that lure users into extractive operations" (Zuboff, 2015, p. 83). In other words, even if consumers are usually requested to give formal consent to the collection of their personal data, oftentimes they are probably not fully aware of the implications that follow the exchange. For this reason, consumers may be regarded as the passive targets of data extraction. In this context, individual behaviour emerges as a new kind of commodity exploited by private companies. That is to say, Big Data corporations monitor people's lives in order to nudge their behaviour and eventually make a profit (Zuboff, 2015).

Another economic-related issue arises from the "filter bubble" that affects people when they have their minds and behaviour nudged by advertisements. This "filter-bubble" could be imagined as a soundproof environment in which personal ideas keep echoing, resulting in their thickening and radicalisation. A consequence of filter bubbles is that as more people live in worlds of personalised information, the less likely they are to be confronted with information that does not fit their beliefs and tastes (Helbing, 2015, p. 59). In other words, as more and more personalised advertisements aim at meeting their tastes and desires, people are less and less confronted with content that does not fit their ideas. In this way, our minds end up being increasingly programmed and standardised by these manipulative technologies, while the wealth of ideas generated by the circulation of different points of view is inevitably reduced.

When such a variety of perspectives is lacking in society, social, cultural and economic diversity risks being undermined (Helbing, 2015). If we think of an ecosystem, the reduction of diversity corresponds to the loss of biological species. Analogously, we may argue that diversity and innovative solutions are necessary to keep societies healthy. Without the creativity necessary to give birth to unprecedented ideas and creations, the whole socio-economic system runs the risk of withering and eventually collapsing. As a result, mass unemployment and economic depression might occur (Helbing, 2015).

4.2.7. Employment

The accelerated growth in AI and ML technologies means that it is inevitable that AI will replace many jobs, such as doctors and bankers, and even parts of government will be automated. For example, Tesla has promised to introduce self-driving trucks within a decade, which is likely to lead to the loss of millions of jobs (Bossmann, 2016). The Financial Times in 2016 showed the possibility of AI leading to mass unemployment (Cookson, 2016), and a recent University of Oxford Study estimates that with computerization and ML in the next 20 years "47% of total US unemployment is at risk" (Frey & Osborne, 2013, p. 44). It is therefore necessary to address this, as well as questions regarding the fair distribution of wealth created by machines in a "post-labour" economy, where the people who own AI-driven companies will be the ones to obtain all the benefits (Bossmann, 2016). Indeed, there is a

growing consensus that the increasing introduction of AI in the labour market will negatively affect wages and job creation, as more and more jobs will be automated (Wallach, 2018).

SIS in recruitment could lead to a more impartial selection of personnel (Wilson, 2017). However, algorithms are not always free from human biases on account of biased training data (see Section 4.2.4), such that machines can incorporate and replicate human biases when it comes to selecting personnel (Ajunwa et al., 2016). Even though employers cannot overtly select their audience on the basis of racial distinctions, there exist categories that can be closely identified with race. For instance, groups of people living in disadvantaged neighbourhoods and having low-paying jobs may easily overlap with some disadvantaged ethnic categories. On the basis of such proxies, algorithms may be inadvertently or voluntarily used to send ads on the basis of racial considerations (Kim, 2018; Raub, 2018). Also, if academic credentials are treated as having priority relevance for the algorithm's decision-making, the reputations of some universities could be assigned enormous weight, even if applicants' competencies may be unrelated to the name and rank of the university they are associated with (Barocas & Selbst, 2016). In this case, a bias that is structurally embedded in society may be exacerbated by the use of SIS in employment.

Hiring algorithms may give birth to discrimination even when the data used for training is not biased. Correlations detected by algorithms do not always correspond to actual causal relationships, which may lead to these correlations being inconsistent or completely wrong. For example, a hiring algorithm might observe that visitors of manga sites are often good coders. Even if this correlation turns out to be reliable, it is unlikely to be based on a causal relationship. For this reason, the correlation in question might easily vary and eventually dissolve. An algorithm strongly relying on this assumption could therefore be misled and unjustifiably used to hire people unsuited for particular jobs (King & Mrkonich, 2016). Another possible way for algorithmic bias to undermine employment does not concern the quality of the analysis, but the quantity or the representativeness of the sample addressed. People who are less involved in the digital economy or have unequal access to SIS are likely to be excluded from the new processes of job recruitment (Madden et al., 2017).

But besides the use of algorithms in instances of employment (or exclusion from employment as a result of algorithmic discrimination), there is also the potential for Big Data systems to be used to collect information about employees in the workplace. Emails, phone calls and web searches can be monitored and analysed by trained algorithms, in order to promote efficiency and productivity in the workplace. Data mining can be used to monitor employees' activities in order to eliminate the subjective nature of performance assessment (Wilson 2017). This may lead to a situation of constant surveillance that is likely to undermine employees' privacy and other important human rights (Ajunwa, Crawford & Schultz, 2017). Moreover, data mining may jeopardise people's employment and careers, as it may encourage demotion and replacement (Edwards, Martin & Henderson, 2018).

4.2.8. Freedom

As Big Data systems can be used to gather information on individuals' online choices and behaviour, some distance between the individual and organisations using their data is necessary to guarantee people's freedom (Broeders et al., 2017). Individuals can be protected from particular institutions that might want to monitor their activities, as Big Data constitutes a useful source of personal information that can be exploited to gain information about people's lives. Furthermore, as less human oversight

is present, and algorithms constrain the possibility of understanding the decision-making process, increasing reliance on algorithms can bring the threat of algocracy (Danaher, 2016, p. 246).

We can immediately see how the issue of freedom is strictly connected with those of surveillance and privacy. The more the individual's activities are put under observed control, the more their privacy is undermined and their freedom jeopardised. We can imagine a situation in which data concerning someone's health is collected in order to better address an existing disease. In this case, the collection and the accessibility of this data could seriously hamper the search for appropriate health insurance or future job opportunities, as some sensitive information may be disclosed. As such, the use of Big Data analytics could lead to information about individuals (such as their general interests and behaviour) through surveillance measures and profiling which delimit their ability to secure a job (Wolf, 2014, p. 14).

SIS has the potential to threaten freedom of choice and democracy (Helbing et al., 2018). While today the manipulative power of algorithms results in nudges towards some preferred behaviors, free will and the self-determination of people, which are the preconditions for democratic constitutions, run the risk of being compromised. Increasingly, overwhelming and personalised forms of digital control can methodically regulate and restrict discourse (Balkin 2018). Finally, Big Data affects our consumption freedom - by exploiting the personalised information collected algorithms can be used to instill unnecessary desires and needs in people's lives (Helbing et al., 2018). In this way, the individual's capacity to freely control their choices are compromised.

The rising role of algorithms in societal decision-making, can also be considered a form of technocratic governance (Janssen and Kuk, 2016). This type of governance attempts to deconstruct complex societal problems into neatly defined and well-scoped problems that can be solved through algorithms (Janssen & Kuk, 2016, p. 371-72). The notion of political realities having a diminished role arises because both political decision-making and those who face the decision-making, are under the determining effects of algorithms. Algorithms may diminish actors' ability to voice their concerns (Couldry & Powell, 2014, p. 4). The belief in technocratic or algorithmic governance relies on the assumption that algorithmic automation occurs without human bias (Janssen & Kuk, 372), which we have already seen to be incorrect.

4.2.9. Human Rights

The human rights discourse around SIS comprises two kinds of rights: existing rights that need to be extended into the digital sphere, and new digital rights (Kuriakose & Iyer, 2018). The right to equality and the right to work are part of the first group. The right to equality concerns equality of data flow and access. For instance, according to their social position, people may be subjected to a different degree of surveillance. The right to equality aims at preventing the implementation of privileged internet services and at providing everyone with the same benefit of these services. Since the digital revolution, and in particular the advent of Big Data, have not created the same amount of jobs generated by the industrial revolution, the right to work is increasingly becoming a sensitive issue, as discussed in Section 4.2.7.

Digital rights focus on the right to privacy and the right against propensity-based discrimination. The right to privacy emerges as Big Data aims at collecting a large and varied amount of information about individuals (Kuriakose and Iyer, 2018). The issue of privacy "encompasses (i) the right to erasure such as the right to remain anonymous or be forgotten and (ii) the right to be excluded from surveillance, targeting and censorship" (Kuriakose & Iyer, 2018, p. 16). Privacy is of high concern as protecting individual's data (especially concerning their choices and what information is presented to them) also protects their right to freedom of expression, association and related rights (Latonero, 2018, p. 7). SIS

may be used to collect information about individuals' lives in order to steer people's behaviour, thus jeopardising their freedom and decision-making.

However, human rights are not always negatively affected by SIS. SIS can also promote the right to health, as it can be used to make better predictions with regard to the progress of a disease or to prevent particular groups of people from contracting a disease (Peterson, 2017). Nonetheless, the right to privacy may conflict with the right to health. Sensitive information concerning people's health may be disclosed and result in issues with health insurance and employment. This is especially relevant when private actors such as employers, financial institutions and insurance companies have a strong incentive to discriminate against persons who are deemed to not only have existing impairments but also the potential to develop impairments in the future (Peterson, 2017, p. 3). In other words, the disclosure of health issues may encourage private actors to exclude people from fundamental services, jeopardising the right to equality in conjunction with the right to privacy.

The right to privacy may also conflict with the right to science (Vayena & Tasioulas, 2016). The right to science concerns people's opportunity to share, access and benefit from the knowledge deriving from the collection of Big Data. Evidently, this is likely to come at the cost of infringing the individual right to privacy, together with all the possible consequences mentioned before (such as infringement of the right to equality, the right to work, the right to autonomy or the right to freedom). Informed consent and the possibility to freely set privacy preferences could be helpful tools to address this conflict.

SIS can also be utilised to provide early warning signs through real time detection of human rights violations, emerging humanitarian crises and other vulnerabilities (Sarfaty, 2017, p. 13). For example, they were used to reveal recent human rights abuses in Syria, or may be used to prevent human trafficking or slavery. However, it is not always easy to determine whether human rights are actually at stake in places we are not really familiar with (Aronson, 2016). Due to the different cultural context, an apparently threatening situation may be ordinary for a different country. Moreover, the collection of Big Data with the purpose of safeguarding human rights may threaten to infringe upon other human rights, track individuals, or be misused by states surveilling the population.

4.2.10. Individual Autonomy

Algorithms that steer citizens' behaviour in the public space have a *transformative effect* (Mittelstadt et al., 2016, p. 9), because they influence how we perceive the world (Floridi, 2014). SIS can reontologise the world by categorizing and conceptualizing it in new and unexpected ways (McQuillan, 2017; Lake 2017). Such transformative effects on citizens' behaviors can lead to the violation of citizens' autonomy, especially if an individual's decision-making is compromised when their choices are curated by third-parties that are not working in the individual's interest (Applin and Fischer, 2015; Stark and Fins, 2013). Algorithms are not exclusively used to detect customers' desires - rather, they have become increasingly capable of conceiving ads and content customised for each individual (Grafanaki, 2017). What users see on their screen is often decided by an algorithm and not necessarily based on personal choice (Newell and Marabelli, 2015).

But the problem with this "personalization" is that since the content presented is meant to be consistent with users' tastes and beliefs, the diversity of the information received is inevitably reduced (Barnet, 2009). By consequence, users' online freedom in exploring alternative content is likely to become more and more difficult and their range of choices is expected to decrease. The more they 'Like' the suggested content, the narrower their range of content becomes until they are caught in a fortified self-fueling prophecy of 'personalised' information (Grafanaki, 2017, p. 803). If the authenticity of one's behavior is undermined, autonomy will also be affected.

Moreover, predictive algorithms that present curated content lack the capacity of allowing for spontaneous discoveries, which are often part of our human condition (Raymond, 2015). The idea of always being watched can be perceived as a threat to one's self-expression and self-determination (Grafanaki, 2017). Moreover, not knowing what personal information is recorded and stored by institutions can lead to a sense of helplessness and vulnerability. Third-parties may curate content to exploit individuals' desires and cognitive irrationalities and compromise their decision-making abilities (Pan, 2016; Yeung, 2016, p. 124). Since the individual capacity for making informed and rational choices is distorted, users' autonomy ends up being irremediably undermined.

Algorithms affect how people analyse the world and modify their perception of the social and political environment (Ananny, 2016; Floridi, 2014). Personalisation algorithms may influence individuals' decisions based on vulnerabilities (Bozdag 2013; Goldman, 2006; and Newell & Marabelli, 2015). Deciding what is the relevant information for an individual is inherently subjective (Johnson, 2013). Personalisation algorithms limit the diversity of information that the users receive (Pariser, 2011; Raymond, 2014), and thus a condition for autonomous decision-making (van der Hoven & Rooksby, 2008). The right to information is the right of identity, as it manages the information about the self that constitutes one's identity (Floridi, 2010; and van Wel & Royakkers, 2004). SIS black-boxes prevent us from constructing informed decision-making (Kim et al., 2014).

4.2.11. Inequality

Human judgements are often affected by various biases that can be unveiled and subsequently avoided by means of more "objective" tools, such as SIS. However, if SIS training data is biased, then the algorithms are likely to reproduce these biases in their processing, as well as in decision-making based on this training data, which may lead to inequality, either in terms of exclusion and over-representation, or in terms of different treatment between social groups. Data used to train algorithms may exclude some minorities who do not have access to the internet, or social groups excluded from society. In this way, the analyses carried out by the use of algorithms may not be representative of the whole population under examination (Schradie, 2017). Some groups that are already disadvantaged may face worse inequalities, especially if those belonging to historically marginalised groups have less access and representation (Barocas & Selbst, 2016, p. 685).

Additionally, Big Data activities can give rise to inequalities through the quality of the analysis itself, in the difference in treatment and consideration received by specific social groups. For instance, if a company has always tended to exclude women candidates in the past, the algorithm trained with the

corresponding dataset will keep reproducing that bias (Kim, 2018). Another example can be the practice of predictive policing, whereby some areas identified as high-risk become overly monitored by the police, leading to over-policing in these areas. In this way, people living in different neighbourhoods end up being treated unequally by law enforcement (O'Neil, 2016) especially if the areas marked as high-risk are areas with ethnic minorities or historically marginalised groups.

While in the majority of cases inequalities come as a result of unconscious biases and unintentional acts of discriminations, sometimes algorithmic biases are used to mask intentional discriminatory acts. For instance, on the basis of the combination of some information about people's location, personal tastes or network of friends, their possible membership of disadvantaged groups can easily be inferred. As a result, someone might use these traits revealed by algorithms to set up future models and pretend to be unaware of such discriminatory patterns (Barocas & Selbst, 2016). Similarly, SIS may be used to do business at the expense of the worse-off. For example, in banking, algorithms can easily be used to trace people who urgently need money, and their difficult situation can be exploited (O'Neil, 2016).

4.2.12. Informed Consent

Informed consent is required in the use of SIS to ensure human dignity (Ioannidis, 2013). Some of the issues relating to informed consent in SIS use are: consent forms not only listing physical harms but also pointing out the possibility of individuals' information being distributed on the internet; samples being transported to different jurisdictions, meaning different judicial taxonomy concerning data sharing; regulatory bodies having different standards concerning the definitions and laws to cover how to deal with anonymizing data for research (Chow-White et al., 2015). The sheer size of information collected and curated by researchers, makes it difficult to consider users as participants in research (Fairfield & Shtein, 2014). Individuals "were not asked, have not consented, and do not know most of the time" when and by whom their data is being used (Fairfield & Shtein, 2014, p. 44). Given the size and depth (i.e. of personal preferences, race, gender and ethnicity) of the data accumulated, the responsibility for ensuring individuals are properly informed falls on the researchers.

An additional issue is the aggregation of information about entire communities. Aggregation has an adverse effect because when the datasets are from multiple individuals, while researchers may gain consent from a number of these individuals, studying the data can lead to revealing information about other individuals in the community. And these individuals may not have provided or been asked to provide consent to be included in the formation and study of aggregated datasets (45). In order to mitigate any negative impacts on community members, researchers use methods such as "participant observation to ensure that there is sufficient connection between researcher and research subjects to enable the minimization of harm" (48).

Informed consent may be difficult to uphold in SIS when the value and consequences of the information that is collected is not immediately known by researchers, thus lowering the possibility of upfront notice (Politou et al., 2018, p. 5). However, consent may be the last line of defence for individuals to avoid loss of control of their personal data. An additional option is the revocation of consent, something introduced by the GDPR, whereby individuals can declare to have their data either removed or deleted from where the data is stored and used after consent had been originally given. This revocation can take the form of refusing the data to be held, through the deletion of records and their backups, stopping the live tracking of individual information, as well as physically grinding hard

disks where the gathered information is stored (5). The future role of informed consent will be dependent on protections that can be placed on information and intellectual property, along with whether individuals and groups will put their trust in researchers (Clayton, 2003, p. 20).

4.2.13. Justice

Big Data poses a threat to justice in three ways (Johnson 2014, 2018). Firstly, social privileges can be already embedded in the data collected. Data may over-represent some people or social groups who are likely to be already privileged by other existing institutions. This gap in the representation of certain groups, in contrast to the under-representation of others in data collection may exacerbate existing social patterns and power relations. Big Data may be comprehensive but nonetheless biased, which means that it may reflect racial and class privileges and negatively affect disadvantaged groups. Secondly, the differential capabilities of data users may lead to unjust situations. People who are better positioned to gain access to data and have the expertise to interpret them may have an unfair advantage over people devoid of such competencies.

Thirdly, Big Data can work as a tool of disciplinary power, as it can be used to evaluate people's conformity to the norms representing the standards of disciplinary systems. So individuals that deviate from these norms end up being either marginalised or disciplined. The norms reflected by Big Data are often built on the power relations that constitute society, and "with the norms reflecting the power structure of the society in which they developed, they reiterate the patterns of justice and injustice that open data set out to ameliorate" (Johnson, 2014, p. 270).

One of the reasons why the rise of datafication and algorithmic decision-making has an effect on issues of justice is its burden on predominantly poorer members of society (Taylor, 2017). For example, data-driven law enforcement may concentrate on poor neighbourhoods that have historic criminality. Furthermore, characteristics such as race, ethnicity, religion, gender, location, nationality, and socio-economic status, determine "how individuals become administrative and legal subjects through their data" and how their data can be used to draw up policies as well as personalised commercial strategies (Taylor 2017, p. 3). Dataveillance is also being increasingly taken over by the private-sector, which leaves the responsibility for delivering accountable and transparent systems to them (3).

Big Data has the potential to be used for revealing and addressing issues of environmental justice, for instance by monitoring, mapping and elaborating upon strategies against toxic pollution. However, the data collected may be non-representative, as many hidden or invisible people may be excluded from data collection (Mah, 2017). Secondly, the high speed of Big Data may lead people to overlook the historical causes of environmental problems and therefore fail to address them. Thirdly, the difficulty of Big Data analytics could potentially introduce even more uncertainty into an already contentious field. If taken together, all these issues have the potential to further exacerbate the environmental justice issues that they aim to prevent.

4.2.14. Ownership of Data

In order to be a useful source of information, Big Data has to constantly flow. If it did not leave its point of origin, it would not be able to communicate anything to anyone. However, at the precise moment Big Data is extracted and collected by an external source, some issues emerge: who is the owner of this data? Who should have control of the data? Sax (2016) draws on Kirzner's theory of "finders-keepers" to show how data miners could claim rights to the data extracted from people.

According to this theory, as any resource does not exist as such before the extraction, the person who takes possession of it automatically becomes its owner. Analogously, as long as an entrepreneur extracts valuable data in a just way, they become the owner of a radically new resource whose value is given by the original act of collection. Likewise, Big Data companies are the finders-creators of the insights derived by the algorithms they employ, and so the fruits of these insights are legitimately theirs to own and control (Sax, 2016, p. 29). According to this view, the property and control of personal data would legitimately be owned by Big Data companies.

Even though Kirzner's idea seems plausible when it refers to inanimate objects, in the case of Big Data people are often involved. Kirzner's idea seems to introduce some sort of separation between goods, as the resource extracted has to part from the initial raw material (Sax, 2016). Is this separation still valid when it comes to individuals? One could argue that someone owns his or her information in the same way as he or she owns his or her body. In other words, the information extracted could still remain an indivisible part of his or her self (Floridi, 2005). When it comes to Big Data collection, the identity of people is therefore at stake; anyone who gains control over personal information, is also dealing with individuals' identities. Even if Big Data companies have rights over personal data, the way data is acquired remains problematic. And even though the transaction of data is supposed to be based on an informed consent procedure, this process remains quite controversial. As the people concerned are often either not competent to take a decision or not willing to spend time reading informed consent documents, the legitimacy or the acquisition of personal data could be undermined (Sax, 2016). A third source of concern relates to the difficulty in making predictions on the implications of the data flow. Even though Big Data companies could claim rights to personal data and succeeded in acquiring them in a just way, the impact on the people from whom data were collected should also be taken into account. As the consequences would be hard to predict and could be unpleasant for the people involved, it is questionable whether Big Data companies should maintain control over personal data (Sax, 2016).

There are multiple ways in which people from whom data are collected may be negatively affected. When people lose control over their personal data, they risk having their privacy violated; conversely, the idea of a right to privacy implies that individuals maintain control over their personal data (Someh, Breidbach & Davern, 2016). Individuals who have their privacy violated may suffer issues that are mainly related to their identity and autonomy. With regards to identity, the collection of information about individuals may affect the way they conceive themselves or are seen by others; for instance, algorithms might profile people according to their race, gender or social status and lead to discriminatory situations (Someh, Breidbach & Davern, 2016). Concerning autonomy, Big Data may turn against the interests of the individual when organisations use it to customise offers and steer consumer behaviour for their own benefit (Zuboff, 2015).

Despite the fact that data is often extracted from individual activities, individuals are not the only entities affected by data circulation. Algorithms are often used to make predictions about whole groups of people in order to monitor their activities and/or steer their behaviour. As such groups of people are often the final target of Big Data companies, it is difficult to maintain that individuals should be the only owners of the information extracted from their activities (Purtova, 2017). Given that individuals and groups seem to have little control over their personal data, it is plausible that companies themselves should take care of their privacy. Unless people who have their data stored were free to opt out whenever they feared their privacy was in danger, institutions holding personal

information should protect data from malicious attacks and prevent their data being misused (Wallach, 2018).

Despite the presence of risks for individuals, the collection and the free circulation of data might be desirable when they achieve important societal goals (Wallach, 2018). What if universities and other research institutes need to collect a large amount of data to improve agricultural techniques or address some urgent healthcare-related issues? If the spread of information about particular individuals or groups turned out to affect their reputation or reduce their opportunities, would it still be a desirable solution? Since software giants like Google or Facebook are the main collectors and administrators of personal information, they maintain control over a big slice of the data flow and play the part of the gatekeepers of information. This creation of large monopolies of information and knowledge has the potential to generate new forms of inequalities or exacerbate existing ones (Purtova, 2017). However, personal data is not easy to conceptualise as an object in the same sense as an individual's private property, and their right over it (Purtova, 2017). More so, personal data can point to the group (or multiple groups) that the individual belongs to (Purtova, 2017, pp. 17-18).

4.2.15. Potential for Military, Criminal, Malicious Use

Military personnel can collect data about local populations from social network or other internet sources used by civilians. Big Data may be used to make predictions about future possible scenarios and to elaborate advantageous strategies accordingly. Such forecasting capabilities along with the strategic advantages they present, show that Big Data analytics has the potential to facilitate the accomplishment of military missions, especially if used to save lives, both of civilians and soldiers (Haridas, 2015). Big Data may also be used to improve the real time decision-making process, for which the capacity to instantaneously process a large and diversified amount of data is essential. Finally, Big Data analytics may improve anti-terrorism operations or assist military intelligence and cyber-defence (Çintiriz, Buhur, & Şensoy, 2015).

When it comes to managing and selling personal information, the boundary between legitimate and malicious use is not always clear-cut. There exist companies that gather and resell personal information (such as personal internet browsing history, email address or state records) to other corporations interested in using it to make a profit. These companies are called "data brokers" (Asta, 2017). In addition, they can use the data collected to create "people search" websites, which allow people to find information about specific individuals in the world. In the majority of cases, the information spread by data brokers is used by corporations to show people personalised advertisements or to directly contact individuals for commercial purposes. In other cases, these services have been used to facilitate criminal acts like tax fraud, but also legal though unethical acts such as predatory targeting of rape victims, individuals with AIDs and the elderly with dementia (Asta, 2017, pp. 271).

Big Data is also used in intelligence and national security systems, and potential hackers may manage to open a breach and steal or alter precious information. If that happens, national security and the functioning of the state mechanism may be undermined (Johnson, 2013). However, it may be the case that cyber-attacks are morally justifiable, while cyber-defence should not be implemented. For instance, if a state violates human rights and a cyber-attack is directed against it by another state, its intervention might be ethically justifiable (Smith, 2018). However, most cyber-attacks and cyber espionage seem to be directed against private companies using Big Data, which often represents a

precious source of information and valuable knowledge. In this sense, a further "V" (other than the three usual "Vs" of volume, variety and velocity), standing for "voracity", can suggest how the hunger for information has increased in the development of Big Data systems (La Torre, Dumay & Rea, 2018). Data may be stolen, altered or even destroyed in order for companies to gain an advantage over other companies. As a consequence, corporations may suffer financial loss and have their structural, reputational and human capital negatively affected. Moreover, the whole economic system may suffer a lack of competitiveness and innovation. Finally, as data breaches could alter or damage datasets, the reliability of the information and knowledge obtainable from them may be irremediably undermined (La Torre, Dumay & Rea, 2018).

Big Data not only constitutes the target of hackers; it can also be used to enhance cyber-security. By means of Big Data analytics, investigations, predictions, and even prevention of cybercriminal activities can be carried out more efficiently (Brewster et al., 2015). In addition, Big Data can constitute a valuable decision-making tool in information security management (Fan, 2016). For example, in large-scale events with potential for public order issues, a large amount of information can be collected from social media or other internet sources so as to identify possible criminals. In addition, potential witnesses to crime incidents or terrorist attacks can be traced by means of geo-tagged data from smart devices (Brewster et al., 2015). In relation to cyber-security, Big Data analytics offers the opportunity to improve situational awareness, as it may help to recognise anomalous or suspicious behavioural patterns indicating an attempt of fraud or other security threats. For instance, financial companies can use Big Data and behavioural analytics to identify potentially fraudulent transactions, which can be detected thanks to the large number of regular transactions previously processed (Eastman, Versace & Webber, 2015).

4.2.16. Power Asymmetries

Big Data has the possibility to create power asymmetries by causing higher energy intensity (via energy demands to sustain data centres), data vulnerability, security requirements, the global (digital) divide, and potential for misuse by the powerful (Portmess & Tower, 2014). Big Data can be understood as a phenomenon reflecting not simply computational machines and their infrastructures but also the human intentionality behind these machines and infrastructures, deploying particular patterns of power and authority (Portmess & Tower, 2014, p. 1). The knowledge offered by Big Data and its practices, and how to regulate this knowledge is in the hands of a few powerful corporations (Wheeler, 2016). For example, the real outrage which followed the 'emotional contagion' study of Facebook is not limited to concerns over informed consent or the lack of ethical review boards in corporations (Boyd, 2015). Individuals and groups feel uncomfortable with the power imbalance held by SIS companies. Such power imbalances are heightened given that companies and governments can deploy powerful means for surveillance, and privacy invasion, as well as manipulation through personalised marketing efforts and social control strategies (Lepri et al., 2017, p. 11).

The private and public sectors play a role in the ascent of datafication, especially when specific groups (such as corporate, academic and state institutions) have greater unrestrained access to Big Data

datasets, along with the public perceiving datafication as a leading paradigm (van Dijck, 2014, p. 203). The ideology of datafication is therefore backed by institutional demands as well as public interests in the datafication process. As data firms advocate the objectivity and effectiveness of their computational tools, by adopting online platforms for measuring social traffic, government agencies and academics interchange the control of data collection and analysis from the public sector to private industries (203). This leads to a tripartite alliance between governments, academia and data firms that are increasingly interconnected by their exchanges in personnel as well as innovative technologies (203). Altogether, there is the potential that both the public and private sectors can create power dynamics from using SIS: "Companies can bid on certain combinations of words to gain more favourable results. Governments are probably able to influence the outcomes too. During elections, they might nudge undecided voters towards supporting them—a manipulation that would be hard to detect. Therefore, whoever controls this technology can win elections—by nudging themselves to power" (Helbing, 2019, p. 7).

4.2.17. Privacy

Big Data can generate personally identifiable information (PII), but sometimes this is done in a way that does not violate legislation but may also pose a threat to privacy (Crawford and Schultz, 2014). Privacy has come to embody a power struggle between those seeking information through surveillance, and as a means through which privacy can enable a democratic and free society for consumers (Coll, 2014). Privacy self-management has been constructed in such a way that it has become a currency with which other goods can be purchased (Hull, 2015). This market-oriented understanding of privacy takes away the larger meaning and understanding of privacy.

There is a need to strike a balance between privacy and allowing data analytics to generate value for our economy (Tene & Polonetsky, 2018). To do so, policy makers must address the legal concerns of personally identifiable information, control over what is done with this information, and define the purposes for which data may be used.

Other than the issue of individual privacy there is also a need to acknowledge group privacy concerns (Floridi, 2014). There is often a tension between security and privacy, and the beneficiaries are framed to be the individual or society, thereby ignoring the impact on groups. Oftentimes it is not the individual that is of interest, but the group to which that individual belongs. Just as fishers try to catch the whole shoal and not just the one sardine, so those using data often try to capture information about the groups which individuals may belong to. Group privacy must be recognised alongside individual privacy and societal benefits when assessing analytics from an ethical viewpoint (Mittelstadt, 2017). However, data analytics often allows for the creation of groups in such a way that avoids violating legislation on privacy.

There are numerous online sources and platforms that make use of re-identification techniques endangering the privacy of users, such as geotagging and content uploaded to social media extracted from user choices on websites (Marabelli and Markus, 2017, p. 2). The increasing presence of ubiquitous and affective computing is linked to the continuous collection of large volumes of user data from smartphones, wearables, and sensors to pick up the emotions, traits and behaviour of users (Politou et al., 2018, p. 3). The coupling of Big Data infrastructures and novel sources of behavioural data (such as smartphones and social media data), allows inferences about individuals' sexual orientation, ethnic origin and recreational habits to be identified (Lepri et al., 2017, p. 11). These

monitoring techniques allow research to be conducted on social media and e-commerce platforms, such as Facebook's 'contagion experiment' on how altering users' news feeds could affect their emotional states (Marabelli & Markus, 2017, p. 2). In situations such as this, protection does not extend to non-subjects, such as those who are identified (without consent) by research done on subjects who give consent, which is made more difficult to legislate by variations of rules and laws in different countries or jurisdiction (Marabelli & Markus, 2017, p. 3).

4.2.18. Responsibility/Accountability

Due to the sometimes intrinsic opaque nature of algorithms, it is difficult to trace a particular problem to its sources (i.e. the traceability problem) (Mittelstadt et al., 2016). The problem can be caused by a bug in the system, a systemic failure or bias in the data - but determining the real cause may be impossible with non-interpretable machine learning algorithms. When the source of the problem is difficult to find, it also makes it challenging to identify who is to be held responsible for harm caused by algorithmic decision-making (Mittelstadt et al., 2016, p. 5). The accountability issue for algorithms is under-researched and insufficient attention has been paid to distributed responsibility between humans, algorithms and organisations (Mittelstadt et al., 2016, pp. 12). Accidents concerning Big Data often raise three concerns: firstly, data isn't a physical artifact, so identifying accidents occurs after it has happened; secondly, accidents are not geographically specific due to the global infrastructure of Big Data systems, making it difficult to identify who and where responsibility lies; thirdly, it is difficult to predict the timeframe of impacts and thus identify responsibility (Nunan & Domenico 2017, pp. 497-498).

When problems arise, traditionally, the designers and the users of algorithms, who are public managers in this case, would be the ones to blame (Kraemer et al., 2011). However, it is only justifiable to attribute blame when the actor possesses some degree of control and intentionality in carrying out the action that leads to harm being caused (Matthias, 2004). Accordingly, having control over the algorithmic process while designing and/or using the algorithms are the two conventional criteria by which to be considered responsible/accountable. However, sometimes the logic of deep neural networks cannot be interpretable by the engineers who design them, let alone the policy makers attempting to legislate how they should be deployed. Thus, in the context where such algorithms are deployed in public decision-making processes, it is not easy to hold public managers responsible for the public actions motivated by insights derived from algorithms. Only when "decision-making rules are 'hand-written', their author retain responsibility" (Bozdag, 2013).

Even when the steps taken by the algorithm are known, the rationale as to why certain variables have influenced the decision reached by the algorithm is unclear. This is due either to the use of multiple overlapping models and classifiers, or specific techniques such as boosting and bagging, which may increase accuracy of decisions through optimization, but lead to reducing the ability to interpret how the decisions are reached (de Laat 2017, p. 14). One way of recovering interpretability would be the implementation of "Quantitative Input Influence" which focuses on "how much individual variables (as well as combinations of them) have contributed to the final algorithmic outcome" (de Laat 2017, p. 15). There may be a tradeoff between accuracy and accountability of algorithms, because sometimes the more transparent algorithms are demanded to become, the more they lose out on accuracy and richness.

The difficulty in establishing control of the algorithms due to the complexity and volume of code and techniques used (e.g. deep neural nets) means that the traditional concept of responsibility cannot deal with algorithms that perform without or beyond the capacity of human oversight. And this also means that the standard notion of accountability, framing an actor as accountable only when they have control and intention when carrying out an action, is also difficult to make use of in the case of highly complex algorithms (Matthias 2004). This therefore creates an "accountability gap", in which accountability can be assigned to several moral agents given the composite nature of algorithmic design (i.e. with there being multiple programmers, lines of code, computing devices and infrastructures, and public managers who decide when an algorithm can and can't be used) (Burrell 2016; Cardona 2008; Matthias 2004; and Zarsky 2016). Some have stated that certain machine-learning algorithms should be considered moral agents with moral responsibility, while others argue that moral responsibility requires intentionality (Floridi and Sanders 2004; Sullins 2006). Regardless, there should be a collaborative discussion and development of ethical requirements to start an operational ethical protocol (Turilli 2007; and Turilli and Floridi 2009), in order to make responsibility and accountability concepts that can be applied when accidents and harms are caused by algorithmic decisions.

4.2.19. Security

Another area of concern in the deployment of AI and ML is the area of security. AI and ML are becoming powerful tools, but are also sensitive given the amounts and kind of data that they hold. As a consequence, data security will become one of the most important sectors, to protect these systems against hackers, ensure a positive impact and reduce risks. Security will be the starting point to both protect and decide who has access to these technologies, and who designs the algorithm, including the training in security awareness, will be important to consider for both users and technical staff (Macnish and van der Ham 2019). Bostrom and Yudkowsky affirm that "Verifying the safety of the system becomes a greater challenge because we must verify what the system is trying to do, rather than being able to verify the system's safe behavior in all operating contexts" (Bostrom and Yudkowsky, 2014, p. 6).

With regards to the relation between security and Big Data, ethical issues can be split into two categories: issues protected against by cybersecurity and issues arising from cybersecurity. The former refers to all the ethical concerns that are supposed to emerge due to a lack of security and that cybersecurity aims at preventing. The latter relates to all the ethical issues that may emerge with the application of cybersecurity systems, which may themselves give rise to ethical questions. The following paragraphs will focus on the first kind of concerns with security and Big Data.

4.2.19.1 Issues protected against by cybersecurity

The multiplicity of data sources, the diversity of data formats and the different types of data storage make it hard to create a coherent security system, due to the different access restrictions and security policies that every source may implement (Jha, 2016; Wang, Jiang & Kambourakis, 2015). Security issues may emerge within different domains and involve a multiplicity of parties. For instance, security breaches may concern single individuals. In this case, personal data or facts about people are stolen or revealed to third parties who are interested in spying and profiting from individual behaviour. If personal data are illegally disseminated among third parties, the people about whom data are collected risk having their identity revealed and their everyday life negatively affected. The spread of sensitive data may eventually lead to discriminatory situations (Matturdi et al., 2014). For instance, on the basis of data about people's location, tastes or network of friends, their possible membership in disadvantaged groups can easily be inferred. An employer who uses this information for personnel

recruitment may give birth to discriminatory situations. Personal data can be collected from an increasing number of sources, such as healthcare data centres, social media or internet of things devices. For instance, sensitive information about patients' health status may be stolen and sold to insurance companies, which may be used to discriminate among their clients. In short, as the data flow is increasingly expanding and involving various domains, security and privacy concerns are also likely to increase (Agrawal & Tripathi, 2015).

As the data market is becoming one of the main sources of revenue, hackers may extract and sell valuable information from social platforms, e-commerce or even private emails to make a profit (Zou, 2016). However, data theft is not the only source of risk for security. Once data are collected, the individuals concerned have little or no control over the flow of personal information. As Big Data companies might have the means to trace the identities of the persons from whom the data originate, the simple trading of personal data may represent a threat for personal data security. Since personal data are related to individual identity, behaviour and habits, people in control of personal information are also in control of some important aspects of their lives (Zou, 2016).

Security breaches may also concern larger groups of individuals, such as private corporations or society as a whole. For instance, public institutions and infrastructures that make use of Big Data may be targeted by hacker attacks. If some administrative or institutional processes rely on Big Data collection and analytics, the functioning of our democratic system may be jeopardised. Big Data are often used in contexts such as intelligence and security; if there is a security breach in one of these systems, national security itself may be undermined (Johnson, 2013).

Alternatively, datasets can be hacked so as to steal precious company secrets for financial gain. When the stolen data contain valuable knowledge, the cyberattack may cause information to lose its socio-economic value (La Torre, Dumay & Rea, 2018). For instance, as reported by Mengke et al. (2016), a few DNS servers in China broke down as a result of a hacker attack in 2014. As a consequence, thousands of people could not access the internet, resulting in damage to the tourist industry, the aircraft industry and e-commerce. Furthermore, security threats do not only come from confidentiality theft or leaks, which occur when data are stolen or disclosed by unauthorised parties. They can also come from damage to data integrity or loss in availability. When data are altered or destroyed, the knowledge and the operational functions they should provide are at risk of becoming unreliable. For instance, in a context in which IoT devices are increasingly interconnecting people and smart environments, the sabotage of datasets might result in catastrophic consequences for the whole infrastructure network (La Torre, Dumay & Rea, 2018).

4.2.19.2 Issues protected against from cybersecurity

While the previous paragraphs focused on the ethical issues arising from security flaws in the management of Big Data, the following will address the problems emerging together with the use of security systems as such. First, the target of security interventions should be defined, as in some cases the use of cybersecurity may be ethically problematic. For instance, one may argue that a cyber-attack would be justifiable when conducted by one state against another that violates human rights (Smith, 2018). If that is the case, an attempt to defend the malicious state by means of cybersecurity systems would be morally wrong. Depending on whom or what security is directed to, completely different ethical considerations could be made.

If we stick to the ethical issues arising with cybersecurity and Big Data, there is a concern about the effectiveness with which security is implemented: insufficient funding, the way in which data are stored and accessed, poor training of staff and professional negligence may result in serious deficiencies in cybersecurity systems. As a consequence, the data protected are endangered (Macnish and van der Ham, 2019). The second concerns privacy: as several practitioners in the field of

cybersecurity are supposed to handle personal data on a regular basis, privacy issues are likely to arise (Macnish and van der Ham, 2019). Third, both human and machine biases have turned out to affect the algorithms used to process Big Data. As some cybersecurity systems need to profile potential suspects, it is crucial to prevent them from operating with discriminatory patterns similar to those emerging in other fields (Macnish and van der Ham, 2019).

4.2.20. Surveillance

Big Data has intensified surveillance as a result of the interconnection between datasets, the greater analytical tools available, the increased quality, and persistent traditional privacy notions (Lyon 2014). In the ascent of Big Data, ICT services and the challenges to privacy, "metadata appear to have become a regular currency for citizens to pay for their communication services and security" and this trade-off may be partly explained by the "gradual normalization of datafication as a new paradigm in science and society" (Dijck, 2014, p. 198). Datafication is seen as a "legitimate means to access, understand and monitor people's behaviour" by researchers and scholars as well, who find in this avenue "a revolutionary research opportunity to investigate human conduct" (198).

With the rise of Big Data, the way in which surveillance takes place in society has changed. When it comes to understanding Big Data and surveillance, the Big Brother metaphor is probably no longer effective: the idea of a central governmental institution that is constantly watching people from a centralised perspective has become outdated. Big Data is collected by a multitude of corporations that are not necessarily related to the state, and the information extracted by private companies are often voluntarily provided by individuals themselves (for instance, when they make use of social networks). In short, as private entities monitor individuals, surveillance becomes fragmented and decentralised. Bentham's panopticon structure, which required that everyone was permanently monitored by an invisible actor situated at the centre of the system, is similarly no longer as relevant (Doughty, 2014). Instead, the ascent of Web 2.0 infrastructures and social media are producing the emergence of "participatory panopticism", a situation in which "the many" (i.e. governments, corporations, researchers and even users) watch "the many" (Mitrou, 2014, p. 12). Bentham's panopticon is replaced by the new picture of the "omniopticon", where everyone is watching everyone, with surveillance mainly produced and reproduced by large user groups providing user-generated content. For this reason, modern dataveillance (i.e. surveillance of and through data sources) can be defined as a mass self-surveillance (Fuchs, 2011).

The state is one of several actors involved in modern surveillance activities, whose rationale is determined by the power relations and the interests that constitute society. Conversely, dataveillance has become one of the main means of exercising power, especially through monitoring techniques of personal data (which points not just to individual users, but also groups of users and their communication networks) (Fuchs, 2011, p. 240). In short, the flow of information and the way surveillance is conducted go together with the power relations and the particular interests that constitute society. The decentralisation of surveillance is emphasised by Zuboff (2015), who replaces the "Big Brother" character with the "Big Other": the new institutional regime, according to which dataveillance is nowadays exercised by those who are in possession of knowledge on people's lives and have control over people's behaviour. In this way, people's everyday lives end up being monitored

and eventually commodified by means of data mining. This phenomenon, which could be called "surveillance capitalism", aims to use Big Data to predict and steer human behaviour as a means to produce revenue and market control (Zuboff, 2015).

Individuals' identities may be affected by the awareness of being under surveillance, as individuals could be discouraged from exercising their freedom of expression. In other words, when people think they are being spied on, they also fear being judged and become less willing to behave in accordance with their own selves (Mitrou, 2014). In addition, people's identity is threatened by the fact that the data collected by companies may not be representative; rather, the persona under surveillance is produced and reproduced as a result of the data extracted from different sources about them. For instance, Big Data analytics uses techniques that can recombine, draw relations, and move data across different contexts, all of which can lead to the creation of a new digital subject constructed through dataveillance (Matzner, 2016, p. 206). Since the authorities' intentions are to assess whether a particular person may be dangerous for society, the subject created out of data-based surveillance ends up looking like a suspect. This happens because the profile made by authorities does not correspond to any real identities existing in reality; rather, it is constituted by an agglomeration of personal data collected by disparate sources and merged together in a new "persona" or "identity" (Matzner, 2016).

In other words, the persona created out of the combination of disparate personal data does not reflect the existing person from which data are collected. Different combinations of personal data are meant to give birth to different personas. However, as individuals are diverted from their personal selves, they also end up being exiled from their own behaviour. New lifestyles are likely to correspond to the introduction of new identities. And new possibilities of control and subjugation are made available by corporations with the new power of digital surveillance. People are encouraged to shape their lives on the basis of predetermined identities conceived by Big Data companies; for instance, as corporations are interested in producing new consumers, people are encouraged to consume more. Consequently, since corporations need people to consume, people's identity is moulded so as to give birth to new consumers. Even though this may be true for all kinds of advertising, the personalised advertisements enabled by Big Data makes this power of nudges even more tangible. In conclusion, Big Data allows companies to process the data collected from surveilling people's activities. By means of data analytics, companies may eventually nudge people's behaviour in order to make a profit (Zuboff, 2015).

With regard to its particular fields of application, dataveillance either has the potential to lead to desirable outcomes (as it can be used to improve the efficiency of some social practices and the well-being of people), or it may result in bad consequences. For instance, Big Data is currently used in the fields of health care, policing and employment. With regard to health care, Big Data can be used to monitor and predict the course of disease outbreaks or to prescribe suitable treatments for specific patients. On the other hand, since this kind of dataveillance is supposed to provide more and better calibrated prescriptions to the population so as to improve its health status, it also has the potential to significantly influence its behaviour. In this way, individuals end up with a smaller set of options from which to choose and therefore with a smaller degree of freedom (Garattini, 2017). With respect to criminal justice, Big Data can be used for predictive policing. Even if this new technology may reduce human bias, increase efficiency and improve prediction accuracy, the use of predictive analytics also has the potential to generate multiple forms of algorithmic bias and exacerbate existing patterns of inequality (Brayne, 2017). Concerning employment, even if dataveillance in the workplace aims at increasing efficiency and productivity, it also has the potential to negatively impact employees' motivations and their trust in their employers. Moreover, employees' awareness of having their privacy violated may lead to a decrease in their well-being (Connolly, 2015).

When it comes to assessing whether surveillance activities should be carried out, a useful hermeneutical tool has been proposed by Macnish (2014). The principle of proportionality used in the field of just war can be applied to the ethics of surveillance. According to this concept, harms and benefits should be weighed against one another in order to decide whether a military action should be carried out or not. In just war theory, this notion can be applied both to the context of *jus ad bellum* and *jus in bello*. With regards to the first, the ethical reflection should focus on whether it is morally justifiable to start a war; with regards to the second, ethical reflection is meant to provide an ethical evaluation of the means used in battle. In both cases, the proportionate or disproportionate character of the actions to be undertaken should be evaluated. If this rationale is moved to the field of surveillance, first of all one should focus on the ends and the consequences of surveillance activities.

There are several reasons why people could be surveilled; for instance, the government may need to watch over a potential criminal, or employers may monitor employees in order to increase the productivity of a company. The principle of proportionality aims at determining whether the benefits that follow the implementation of surveillance activities outweigh the harms. Second, when surveillance is carried out, proportionality must be applied to the means used to monitor people's activities. Again, benefits and harms have to be weighed against one another. Depending on the context, several pros can be listed to justify surveillance activities, such as security, efficiency, productivity, welfare or accountability. On the other hand, potential harms affecting individuals or society can also be found, such as privacy violations, discrimination, paternalism, behavioural uniformity, abuse of power or diminution of trust. Even if It is not always easy to weigh up costs and benefits, the principle of proportionality remains a useful tool to approach the matter of surveillance in an ethical way (Macnish, 2014).

4.2.21. Sustainability/Environmental Impact

Big Data's application in environmental and sustainability applications aim to leverage the infrastructure that Big Data systems offer, with projects that use these systems for the monitoring of energy consumption, supply chain management, biodiversity, deforestation and carbon emissions (Keeso, 2014, pp. 13-16; Song et al., 2016, pp. 492; Wu et al., 2016, pp. 875; Hampton et al., 2013, pp. 156; Dubey, 2017, pp. 1-3). These projects, conducted through collaboration between private organisations, non-profit organizations and government agencies, allow for greater speed in analysing the data collected, near real time insights and maps to be generated, which can be used for more effective decision-making. As such the use of Big Data can present a number of opportunities. These opportunities include: i) fostering partnerships (e.g. between private technology companies and NGOs); ii) merging of sustainability and strategy (by integrating sustainability into marketing, finance and R&D); iii) personalising the environment (tools that aid human health can also point towards environmental issues, such as asthma inhalers measuring air quality); iv) emerging and accessible technology (e.g. environmental science technology providing new tools for more effective monitoring, and public governmental datasets keeping individuals informed); and v) emerging sources of funding (such as financial awards for innovation in environmental and sustainable impact) (Keeso, 22-23).

But there are a number of barriers to the full utilization of Big Data systems in environmental and sustainability applications. These constraints are: i) financial (not all NGOs have the necessary financial resources to fully invest in Big Data systems);

ii) skills-based (most data scientists are working in other sectors, or are too expensive to retain); iii) access-based (NGOs and developing countries are at an information disadvantage, lacking finances to invest in these systems); iii) conservatism (those in the conservation community may not readily shift to depending on Big Data systems, having their own way of doing things); and iv) confidentiality (governments as well as private companies may be against full transparency and open data sharing, with concerns over privacy as well as keeping species' locations undisclosed for their safety) (18-19). There is also a concern about the data centres needed to run SIS, as the demand for huge computing power (for the increasingly interconnected ubiquitous devices), along with greater resources and energy required for data collection, storage and analytics, all contribute to the production of greenhouse emissions as well as impacting the environments they are built on (Wu et al., 2016, pp. 875-876).

4.2.22. Transparency

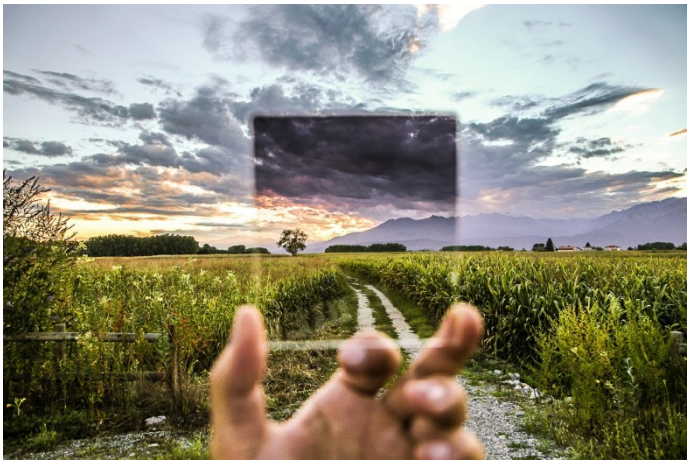
When humans cannot see or understand the connection between the data that are processed to produce evidence and the conclusion derived from such evidence, in other words, if humans cannot interpret how the data used by algorithms contribute to the conclusion generated by them, the legitimacy of those conclusions/insights would be considered problematic. These types of evidence are called inscrutable evidence (Mittelstadt et al., 2016, p.4), and cause the problem of intrinsic opacity, which is due to the nature of certain machine learning methods that are too complex for humans to interpret (Lepri et al., 2017, p.12). Such opacity creates a legitimacy problem because "many theories of political legitimacy insist that decision-making procedures must be rationally acceptable to those who are affected by them" (Danaher, 2016, p. 252).

There are three forms of algorithm opacity: opacity that stems from corporate self-protection, opacity as a result of technical illiteracy, and opacity from the divergence between the mathematical idealization of machine learning contra the demands of human reasoning and interpretation (Burrel, 2016, pp. 1-2). While the term 'algorithm' may not be something new, what *is* new is the capability of algorithms to be used for data collection and storage, as well as the types of information that are tracked, which includes purchasing activities, the links individuals click on, and geospatial movement from mobile devices, services and applications (2). In situations where companies may purposefully conceal how their algorithms work, a way to deal with this kind of opacity is to make their code available for regulatory scrutiny. From a regulatory standpoint, corporations making the design of the algorithms they use open access, lowers the difficulty of tracing instances of consumer manipulation by reading the code that is used (4).

But leaving their algorithmic design open can also lead to what is called "gaming the system" (Kitchin 2017; and Voinea and Uszkai 2018). Too much transparency may lead to the ineffectiveness of algorithms, whereby the classification variables (e.g. knowing red flags in profiles of tax evasion, what proxies there are for predicting potential terrorist behaviour) can become known by the individuals/groups that the algorithms are meant to track and deter. To fight against this, "adversarial prediction games" that make models and algorithms more robust against the manipulation of classifiers are part of an emerging field combining classification systems with game theory (de Laat 2017, p. 12). Transparency can also lead to companies losing their competitive edge, and is another reason for opacity to remain an issue to properly understanding how algorithms work, especially if it

is seen as necessary to keep them secret and protect them as intellectual property (Voinea and Uszaki, 2018, p. 936).

Besides intentional opacity, writing algorithms "demands a special exactness, formality, and completeness that communication via human languages does not" (Burrel 2016, p. 4), which leaves the structure and mechanics of algorithms opaque for those who lack the necessary computational literacy. A concern for algorithmic governance, especially with regards to the transparency of how decisions are made, is problematised by the increasing complexity of more automated systems, and the more data automated systems have to handle. A tradeoff exists between the ability to deal with more complex phenomena and the lack of explicit human understanding of how the complexity is handled by algorithms. "[W]ithout some knowledge of computer science and mathematics, this makes participation in the co-creation or open innovation of algorithms challenging" (Janssen & Kuk 2016, p. 373) which means that only a select few can properly question and investigate algorithmic decision-making processes, at the exclusion of the majority of the public who lack the literacy to both understand and question these processes. This may lead to an imbalance of knowledge by users on how the algorithm is processing their data (Granka 2010; and Zarsky 2016).



Even while extremely large databases are "possible to comprehend and code may be written with clarity, the interplay between the two in the mechanism of the algorithm is what yields the complexity (and thus opacity)" (Burrel 2016, p. 5). Also, algorithms are not always static and fixed in form or practice, since when they are in operation they unfold in multiple ways (some foreseen and some unexpected). In some cases a company may run different versions of an algorithm to assess their respective merits, or "randomness might be built into an algorithm's design meanings its outcomes can never be perfectly predicted", making the algorithm's effects highly context- and use-dependent (Kitchin 2017, p. 21). This highlights how algorithms are contextual as well as contingent in how they develop/evolve, as well as in how they reach their decisions. These aspects complicate the transparency of algorithms, since why and how the code may work in one case or for one user but not in another situation or for a different user remain unclear.

Transparency "involves encountering non-obvious information that is difficult for an individual to learn or experience directly, about how and why a system works the way it does and what this means for the system's outputs" (Rader et al., 2018: 1). Thus, mechanisms that aim to increase the transparency of algorithms allow individual's greater ability to question and critique not only the implementation of algorithmic systems, but also the design of these systems, since the issues that these systems have (such as bias and discrimination) may often be traced to the inputs and data sets that they are trained on. Mechanisms that increase transparency can take the form of explanations of how the system reached its output. "How" explanations are referred to as "white box" descriptions that elaborate on how, in the case of recommendation systems, the recommendation that was given was reached, given the reasoning and data sources the system used. Whereas "Why" explanations treat algorithmic systems as "black boxes", and aim to "fill an intention gap between a user's needs and interests and the system's goals, but do not provide any visibility into how the system works" (Rader et al., 2018, p. 2). "Why" explanations are therefore meant to correlate the user's goals with

the system's goals, and the more the two match, the more willing users are to accept, for instance, recommendations from an algorithm.

There is the possibility that some may "favour algocratic governance systems for appropriate instrumental reasons, impressed by their greater speed, accuracy and insight (when compared to human systems)", but "in favouring them we may end up with systems that are increasingly opaque" (Danaher 2016, p. 255). However, algorithms should be accessible (Turilli and Floridi 2009), and if the algorithm is difficult to interpret, it makes it difficult to carry out accurate risk assessments (Schemer 2011). However, it is not enough for the code to be transparent; ethical behaviour should include requirements for algorithms to be explainable or interpretable (Tutt 2016). More work needs to be done in auditing to design better applicable ethical mechanisms for algorithms (Adler et al., 2016; and Sandvig et al., 2015).

4.2.23. Trust

Woolley (2017) discusses the notion of trust in the context of big data analytics for policy development purposes, and identifies the relationship between consent, trust and justice. Rieder and Simon (2016) try to explain the push for numerical evidence within governance from a wider socio-political context, diving into historical explanations and how the epistemological claims on Big Data can be explained by understanding the political culture that has been shaped by public distrust and uncertainty. For van Dijck, there is a presumption of trust in the quantitative methods of datafication, and the integrity of the institutions that utilize these methods, when government, academia and data firms use this data (2014, p. 204). But such trust and integrity is difficult to easily agree as verifiable, given that the "custody over data flows appears to be mired in a fuzzy delineation of territories; access and restrictions to data are fought over both before the public's eye and outside people's realm of knowing" (204). This is further complicated by the move from mere surveillance to dataveillance, the former being the monitoring for specific purposes, while "dataveillance entails the continuous tracking of (meta)data for unstated preset purposes" (205).

4.2.24. Use of Personal Data

Bentley et al. (2014) illustrate that in the rising volumes of data, it is becoming increasingly clear that the digital shadow that individuals using the internet generate from their online choices, to become subjects of big data research through "a form of mass ethnography - a record of what people actually say and decide in their daily lives" (63), also becomes useful in studying behaviour in non-Western countries, with increasing internet users in the developing world (63). Metcalf and Crawford (2016) point out that not only does data science create a distance between researchers and subjects of research, but that research data has become "infinitely connectable, indefinitely repurposable, continuously updatable and easily removed from the context of collection" (Metcalf & Crawford, 2016: 2). This change in the structural composition of data complicates the proper definition of ethical guidelines, or ensuring that individuals are aware of the extent to which their data is used by researchers. For instance, regulatory organisations are ordinarily concerned with the impact of research mostly when there is direct 'intervention' in an individual's life or body. But such direct impact is not the case with data analytics techniques, which may "create a composite picture of a person from disparate datasets that may be innocuous on their own but produce deeply personal insights when combined" (3). Thus despite the data research happening away from the body of the individual whose data is being used, this does not mean no harm may be suffered by the individual.

Landau (2015) highlights the somewhat co-dependent relationship between users' offering their data to companies such as Google, and the responsibility of such companies to use this data to provide better personalised experiences for users (Landau, 2015: 2). There is therefore a relationship between

privacy, consent, control of user data and responsibility, as data "about the user also enables more targeting advertising or services" (2) and for researchers, "massive data illuminates connections that might not have been apparent" (2). Such connections include companies being able to infer physiological states (e.g. pregnancy) from a user's buying patterns, as was the case when the company Target concluded a teenager would be giving birth in five months from what she had bought (3). In a similar line of analysis, new data collection, storage and curation techniques and technologies are transforming how individuals become seen and defined. Individuals are moving from mere data consumers to co-producers, as behaviour in online interactions are recorded, which include internet searches, purchases and participation in learning activities, with this data increasingly becoming a valuable asset for private companies (Souto-Otero & Benito-Montagut, 2016: 19; Hutton & Henderson, 2017: 149). As there is such a variety of techniques from mathematics, statistics and computer science are used at each step of big data analytics implementation, which make it possible for researchers and companies to not only gauge what individuals are likely to buy, or how to better design products, but can also be used to predict how people will respond to epidemics, or emotional content (Wang et al., 2016: 758). This kind of access to individual data is also found in what Lepri et al. (2017) refer to as social good algorithms, i.e. algorithms which influence decision-making and resource optimisation of public goods using behavioural data. The decision optimisation of these social good algorithms is "facilitated by both the design of the algorithms and the data from which they are based" (Lepri et al. 2017: 7).

Oostveen and Irion (2018) bring to our attention that there is a fundamental relationship between laws concerning privacy protection and protection of human dignity and autonomy in European countries (Oostveen & Irion, 2018: 6). These laws "offer normative underpinnings of the fundamental rights to privacy and the protection of personal data" in order to safeguard individual choice and freedom (6). As an example of this relationship between privacy protection and protection of individual rights (e.g. freedom of expression), they point out that tracking of individuals' online behaviour clashes with the freedom of users to inform themselves. This is because "users are no longer free to inform themselves without being tracked" (7).

Additionally, the ability to track individual choices affords companies the ability to tailor personalised communications (e.g. advertisements for products or political campaigns). There is the fear that these communications, by targeting individuals (and groups), could "isolate them in a world that consists of limited information [which can create so-called filter bubbles] that always confirm their beliefs and opinions" (10), as they are not exposed to information from contrasting opinions or perspectives. More so, greater personalisation of individual (and group) data in big data utilisation in algorithmic decision-making can also lead to intentional or unintentional discrimination. The example of redlining, whereby areas such as neighbourhoods "are denied services, comes down to denial of services to people of a certain ethnic background" (10), simply because the individuals are living in that area. In this example, the neighbourhood becomes a proxy for ethnicity, and leads to discriminatory decision-making by the algorithm on ethnic grounds. Such discrimination may appear unharmful if the variables are pet breed or dietary requirements, but becomes harmful when based on an individual's ethnicity or religious beliefs (10).

Golder and Macy (2014) attend to the fact that with greater use of digital technologies, it is becoming easier for researchers to make analyses of social life from the data people generate. The "digital

footprints" left by emails, phone calls and social media posts "affords unprecedented opportunities for the collection of both experimental and observational data on a scale that is at once massive and microscopic" (Golder & Macy, 2014, p. 131). Massive in the sense that the volume of data covers millions of individuals, and microscopic in that these digital footprints can be traced to individual microinteractions as well, which allows for "a detailed record of daily activities and the frequency and intensity of social relationships" (131).

One of the areas of research into online behaviour that big data focuses on is the tracking of words in big data analytics (Hilbert, 2016). The analysis of social media, as a source of tracking words, allows researchers to predict book sales, social predictions, pandemics (such as swine flu), as well as travel patterns (Hilbert, 2016: 10). One limitation of using social media as a source of behaviour analysis, "is the potential differences between digital and real world behaviour" (11). Another area of research is the potential to draw insights on human mobility from the capacity of mobile phones to be used to track locations. Geographic mobile records can be used to show migration patterns of low-income settings, population movements following an earthquake or disease (11), as well as being used for gaining insights into real time consumer behaviour or traffic information (12). There are three themes related to human-data data mining process: i) legibility of the data mining process, ii) agency for individuals to be promoted by reducing the opacity of the data mining process, iii) improving negotiability between individuals whose data is mined and those who make use of their data (Hutton & Henderson, 2017, p. 151).

5. Ethical analysis: Ethical Issues with Specific Types of SIS and SIS Techniques

5.1 The Ethics of Algorithms

An algorithm is a sequence of instructions that specifies in an unambiguous manner how to solve a class of problems or perform a certain task. Algorithms do not only exist in computing; they exist also, amongst others, in mathematics, and are implemented in biological neural networks and electrical circuits. Computer algorithms are algorithms that are implemented in a formal programming language and are part of a computer program. A computer program centrally consists of algorithms and can even itself be considered to be a complex algorithm. Algorithms are effective methods for producing a result. They start from an initial state with (optional) initial input, and then describe a computation that involves a finite number of well-defined successive states that results in eventual "output" and a final ending state. The instructions from going from state to state can be described as rules. For example, an algorithm can contain a rule specifying that if the input consists of the letter "y", then display the text "Are you sure?" on the screen and wait for further input.

At first glance, it might be believed that although algorithms may be used in programs that raise ethical issues (for example, programs designed to collect personal information without consent, or programs that can copy themselves and infect a computer), the algorithms themselves are morally neutral. Take, for example, an algorithm that calculates the sum of two numbers: what could possibly be morally controversial about it? Similarly, an algorithm within a car navigation system that

calculates the shortest route between two points does not seem to raise any moral issues. So can there be an ethics of algorithms?

There is an emerging consensus that many algorithms are not ethically neutral because they are value-laden: they have orientations in favor of or against certain values (Kraemer, van Overveld and Peterson, 2011; Mittelstadt et al., 2016). As Kraemer, van Overveld and Peterson argue, they can be conceived of as an instance of a broader phenomenon, which is that technological artifacts can be value-laden (see also Van den Hoven, Vermaas and van de Poel, 2015; Brey, 2010). These authors are not making the claim that all algorithms are value-laden. Presumably, an algorithm that merely adds up two numbers is not value-laden in any interesting sense. However, as Kraemer et al. claim, many algorithms are value-laden in that they cannot be designed without implicitly or explicitly taking a stand on ethical issues. There are multiple ways of designing them to perform the specified task, and different designs involve different value choices.

It is often possible to design different algorithms to perform the same task. For example, a program can employ different algorithms to play chess, for example ones that do exhaustive searches of several moves ahead, or ones that instead focus on investigating a limited set of moves. Different algorithms can exist at the algorithmic (logical) level for the same task, and they can then also be implemented differently in programming. Moreover, specified tasks that algorithms need to carry out are often not defined in a formal manner, but are defined using terms and concepts from ordinary language that include vaguenesses, ambiguities, and unstated background assumptions. For example, an algorithm that is to identify running behavior in a video feed must translate a vague concept, "running", into an exact specification, and there are multiple ways to do that. In addition, there are often additional requirements, explicitly stated or implicit, that algorithms must satisfy which could affect its design. For example, a navigation algorithm may be designed to calculate the shortest distance between two points, but requirements may be added that waterways and unpaved roads are excluded, or that the vehicle does not cross international borders.

So algorithm design often involves choices. The next argument to make is that some of these choices are morally charged. That this is sometimes so can be seen by considering two central functions of algorithms. Some algorithms have an informational function: the outcome they generate is a piece of information (a number, a string, a record, a picture) that can then be used by either humans or machines. (They can also be input for other algorithms.) Other algorithms rather have the function to recommend or cause action: they issue a particular recommendation to human users (or machines), as for example, when a navigation system tells the driver to make a left turn, or they may cause certain events to occur, as when an algorithm embedded in a robot causes it to raise its arm.

It is easiest to see moral charge for those algorithms that recommend or cause actions. Actions, in general, may be moral or immoral, so it follows that if an algorithm recommends or causes an action, it takes a moral position. Not all actions involve significant moral choices, of course, but a good many of them do. So, for example, algorithms that recommend or cause actions that violate people's rights or are discriminatory are clearly not morally neutral.

It can also be shown that moral choice is often involved in algorithms that do not recommend or cause actions, but merely produce information. The production of information is a process that involves the selection and interpretation of data, and the use of standards of evidence for drawing conclusions from data, and the use of categories to interpret and categorize data. All of these processes can be construed as actions that involve choice, and in some cases these choices can be seen to be morally charged.

To begin, the use of certain categories to represent reality involves moral choices. Some categories, for example, are morally controversial by grouping or depicting entities in a way that some say they should not be grouped or depicted. It would, for example, be morally controversial (and

possibly illegal in some jurisdictions) to have an algorithm that classifies people as "racially pure" and "racially impure". Similarly, it involves an (often implicit) moral choice to employ only two categories for categorizing gender ("male" and "female"), thereby excluding the existence of non-binary genders. In general, the choice of categories used in algorithms and in the representation, interpretation, categorization and organization of data, involves implicit or explicit choices to highlight or "construct" certain aspects of reality, while downplaying or leaving out other aspects, and to invoke certain attitudes in users and prime them in a certain way (Lakoff, 1987). Some of these choices are moral in nature.

The inferences drawn by algorithms can also be morally charged. Except for logically valid inferences, inferences tend to be underdetermined by the evidence. Algorithms may, for example, make generalizations based on a limited number of positive instances, or assume causal relations where there are only statistical correlations. Such inferences are not always morally charged. For example, the inferences drawn by an algorithm from data from a quantum physics experiment are not likely to involve implicit moral choices. In other cases, however, inferences may be based on moral biases or prejudices. For example, algorithms may be structured to make prejudicial inferences to associate low socioeconomic status with crime. When no prejudices are involved, algorithms may also involve implicit moral choices. Felicita et al. give the example of MR-scans of the heart, in which algorithms that produce the image contain threshold values for categorizing parts of an image as light or dark grey. This threshold value influences whether an MRI-scan is classified as indicating possible pathology or not, and can create a bias towards false positives or false negatives. But whether there are more false positives or false negatives is an implicit moral choice: it is a choice between avoiding inconvenience to a lot of people and unnecessary tests, and avoiding undetected pathologies.

We have seen that algorithms can be morally charged for two broad reasons: either because the actions that they take or recommend involve moral choices, or because the inferences they draw and categories they use involve moral choices. Orthogonal to these two types of value-ladenness is the notion of *algorithmic bias*. Algorithmic bias is a type of value-ladenness of algorithms that results in unfair outcomes, either disadvantaging social groups (by gender, race, ethnicity, age, etc.), people with certain characteristics (e.g., people whose surname is more than ten integers long, people with dual citizenship), or randomly selected individuals or groups. It can be found in the categories used, inferences drawn, decisions made and actions taken. It may also result from a bias in data used.

A third general way in which algorithms can be value-laden is by the degree to which they can be understood by their users and stakeholders. This specifically relates to algorithms that make decisions or recommend choices. *Algorithmic transparency* is the principle that the purpose, inputs and operations of algorithms must be knowable to its stakeholders. Advocates, such as the High-Level Expert Group on AI of the European Commission, hold this to be a moral principle: those affected by an algorithm should have the ability to understand why the algorithm makes the decisions that it does. This is considered especially imperative in cases in which the rights of people are affected by the algorithm's decisions, for example in cases in which computer programs provide sentencing guidelines or decide on the creditworthiness of loan applicants. Algorithmic transparency is also considered to be a requirement for algorithmic accountability, which is the principle that organizations that use algorithms should assume responsibility for the decisions made by those algorithms (Binns, 2017; Mittelstadt et al., 2016).

5.2 Data Ethics: Ethical Issues with Data Types and Sources

Many ethical analyses of Big Data and AI do not distinguish between different data sources and types of data. However, some types and sources can be associated with specific ethical issues. For example,

biomedical data can involve issues of medical privacy, and mobile data can raise concerns of location privacy. In this section, we review key types and sources of data and the ethical issues they raise.

5.2.1 Enterprise Data

Data analytics (specifically data mining techniques) have been an important part of the infrastructure of business organization for a number of reasons. These include: i) the volume of data that businesses have access to has significantly increased, meaning traditional means of identifying useful information is no longer feasible; ii) changes in the structure of business organisation, placing knowledge workers to be responsible for optimizing business processes; iii) companies seeking to expand and broaden their product cycles quickly by identifying new markets; and, iv) as companies' computing infrastructures become more globally spread, new techniques are needed to manage and take advantage of these distributed information networks (Kleissner, 1998: 1; LaValle, 2011: 21-22). Data mining has been applied to customer resource management (CRM) by companies looking to gain better insights on their customer base and improve supply chain organisation, as data mining is useful for extracting interesting, non-trivial information that can exploit specific patterns on customers, suppliers and inventory items (Symeonidis et al., 2003: 590). These techniques allow companies to gain a better understanding of customer-base segmentation, sales and market opportunities, business changes, planning forecasting, quantification of risks, detection of fraud, and identification of root causes of cost (Russom, 2011: 11). Data mining can be thus described as having a number of key characteristics: it is a process (rather than a one-time activity) that is complementary to decision support tools, by finding insights that may be buried in volumes of data discovered through algorithmic means, that allow business professionals (not just ICT professionals) to gain new insights that can be valued for the performance of companies (Kleissner, 2). Data mining can offer enterprises solutions in varying fields including retail, healthcare, banking and securities, insurance and transportation logistics. Companies that aim to utilize data mining techniques need access to heterogeneous data sources (e.g. relational database systems, object oriented database systems, Web pages and mainframe hierarchical database systems), sampling (i.e. using a subset of data sources to build models, evaluate models and use models for prediction), and improving scalability by merging and incrementally updating models (Kleissner, 7).

But this entails that companies will also be able to have improved ability to track online consumers. When it comes to managing and selling personal information, the boundary between legitimate and malicious use is not always clear-cut. There exist companies that gather and resell personal information (such as personal internet browsing history, email addresses or state records) to other corporations interested in using it to make a profit. These companies are called "data brokers" (Asta, 2017). In addition, they can use the data collected to create "people search" websites. This kind of business can result in a wide range of effects, some of which may be of great ethical concern. In the majority of cases, the information spread by data brokers is used by corporations to show people personalised advertisements or to directly contact the individuals for commercial purposes. The whole process of data mining, elaboration and generation of valuable information introduces substantial new asymmetries of power and knowledge. On the one side data about people are extracted by some corporations, which can therefore gain accurate knowledge about people's tastes and behaviour. On the other side, people themselves do not precisely know the nature of the data collected and what they will be used for (Zuboff, 2015). This asymmetry in information is also likely to give Big Data companies an advantage over private individuals on an economic level. Online profiling and personalised advertisement may negatively affect people. For instance, Big Data corporations can facilitate new forms of price discrimination aimed at extracting the highest price for goods from each customer. This form of "predatory marketing" has the effect of enriching Big Data companies at the expense of consumers' welfare and privacy. As a result, economic inequalities are likely to be consolidated and exacerbated (Newman, 2014).

5.2.2 Text Data

Text-based data has become a site of analysis as it conveys voluminous information from the Internet about arguments, worldviews and the values of individuals and groups, for exploration by both computer scientists and social scientists. Text analysis or mining is an interdisciplinary endeavour that involves information retrieval, machine learning, statistics, computing linguistics and data mining (Chen et al., 2014: 195). Although the use of algorithms for analysis text-based data makes it difficult for social scientists to engage with the data at the same rate as computer scientists, conversely computer scientists may lack the theoretical concepts to interpret the meaning of this data (Bali, 2014: 2). Machine learning methods (including supervised and unsupervised techniques) have been used by social scientists for text analysis for a number of applications including identifying the sentiments of users of social media (Bifet, 2013: 18), patents, assessing the political leanings of publications, and historical trends in activists' movements (DiMaggio, 2015: 2). But an important point is that the aim and goal of machine learning is viewed differently by computer scientists, who are more concerned with designing, testing and describing models, while social scientists are more concerned with statistical significance and causal validation (DiMaggio, 2). Additionally, the promise and hope of algorithms being robust enough to work independently, is based on attempting to overcome issues of human judgement including errors of reasoning, ideological leanings, vulnerability to priming, stress, pride and prejudice which would alter interpretations of observations reached by machine learning (DiMaggio, 3). But similarly to the concerns outlined regarding data mining of data from social media and smartphones, text data mining means that individuals and groups in varying sectors are monitored and classified, and discriminatory actions may follow from the mining of their data. Consequently the better the mining techniques used and the insights revealed, the more invasive such analytics is likely to be.

5.2.3 Social Media Data

Social media or social life data includes data from social media services (such as Facebook, YouTube and Twitter), as well as online forums, online games and blogs (Tsou, 2015: S70). Data mining techniques used on social media data aim to make sense of opinions, identifying groups amongst the masses of a population (as well as which online groups users of social media platforms such as Facebook are likely to join), which individuals have influence, as well as recommending products and activities (Barbier and Liu, 2016: 327-328). There are a number of motivating reasons for why data mining techniques are used for social media data, which include: i) social media data sets are quite large, meaning that without automatic information retrieval and processing, gaining insights in a reasonable amount of time would be difficult; ii) social media datasets can be full of trivial or noisy data; iii) data from social media is often dynamic, with frequent changes and updates, which mean analytics methods to keep up with these changes are required (Barbier and Liu, 332-333). Some of the most common data mining applications for social media data include group detection (where discovery of group dynamics lead to insights about activities, goods, and services that individual users in the group engage in), group profiling (such as tracking topic taxonomies to discover how group values change, by looking at the tags used by members in the group over time), and recommendation systems (which can recommend products to individuals, but also recommend new friends and groups that individuals would be interested in joining, based on the individual's profile) (Barbier and Liu, 337-340). The capability of retrieving, classifying and making decisions based on what social media users engage in, means that companies are able to track and monitor users' online behaviour in a way that benefits these companies. For instance, people who spend a lot of time on social media are going to see more ads when they are exposed to content that are related to their interests and opinions (mined from what they 'Like', or the comments they make). Social media creators and advertisers are therefore likely to use algorithms to exploit this in a biased way to ensure their own interests (Sleeman & Rademan, 2017). In these contexts the algorithm will always be beneficial towards the advertisers

and diverted from the needs of the customer (Brin & Page, 2000). Moreover, there are numerous online sources and platforms that make use of re-identification techniques, further endangering the privacy of users. These techniques include geotagging from Facebook and Instagram, as well as extracting user information from cookies on websites where content is uploaded to social media, meaning that information is increasingly monitored by governments, private companies and academic researchers (Marabelli and Markus, 2017, p. 2). Web 2.0 infrastructures and social media networks are therefore producing the emergence of "participatory panopticism", a situation in which "the many" (i.e. governments, corporations, researchers and even users) watch "the many" (Mitrou, 2014: 12).

5.2.4 Biomedical Data

The implementation of Big Data in biomedicine is based on the aim of shifting from population-based healthcare to personalised medicine programs based on each patient's history, ancestry and genetic profile (Costa, 2014: 433; Mittelstadt and Floridi, 2015: 3-4; Luo et al., 2016). Companies and institutions that make use of Big Data for generating, interpreting and combining genomics and health data include Appistry, Beijing Genome Institute, CLC Bio, Context Matters, DNAnexus, Genome International Corporation, GNS Healthcare, NextBio and Pathfinder, offering cloud computing services, web-based applications, big data analytics and business intelligence from research to clinical settings (Costa, 435). Despite the slow progress of healthcare professionals in incorporating these connections between patient and disease information with Big Data infrastructures, the collection and analysis of health and disease data is projected to enhance the quality and longevity of life by giving healthcare professionals data-centric and predictive models for personalising treatment plans (Costa, 436). Medical records contain a range of data including demographics, laboratory values, prescribed medications, imaging and other diagnostics, clinical interventions, clinical notes and free-form text (Peek et al., 2014: 43). In a similar way to how corporations like Amazon, Google and Facebook leverage consumer data to offer individuals specific products based on their actions, healthcare provides can leverage patient data with analytics tools for providing personalised healthcare (Costa, 436). Companies that offer storage, analysis and processing of biomedical information include Amazon Web Services, Cisco Healthcare Solutions, DELL Healthcare Solutions, GE Healthcare Life Sciences, IBM Healthcare and Life Sciences, Intel Healthcare, Microsoft Life Sciences and Oracle Life Sciences (Costa, 437). At the same time, however, an important concern is the security of patient information as it is transferred across different storage sites and processed across different data infrastructures (Costa, 438), as well as the privacy that may be infringed given the different parties involved in analysing and processing patient information (Bellazzi, 2014: 10). This could mean that the more interconnected these systems become patient consent may also be endangered(as they no longer know who exactly is making use of their data) (Costa, 438-439).

5.2.5 Mobile Data

While the term data mining may not be new, what *is* new is the capability of mining algorithms and developments in computing infrastructures to be used for data collection and storage, as well as the types of information that are tracked, which includes purchasing activities, the links individuals click on, and geospatial movement from mobile devices, services and applications (Burrell, 2016: 2). Geotagging functions offer the potential to draw insights on human mobility from the capacity of mobile phones to be used to track locations. Geographic mobile records can be used to show migration patterns of low-income settings, population movements following an earthquake or disease (Hilbert, 2016: 11), as well as being used for gaining insights into real time consumer behaviour, as well as traffic information (Hilbert, 12). The coupling of Big Data infrastructures and novel sources of behavioural data (such as smartphones and social media data that individuals engage with on their phones and other devices), allows inferences about individuals' sexual orientation, ethnic origin and recreational habits to become disclosed (Lepri et al., 2017: 11).

5.2.6 Web Data

Web content mining and web usage mining is the application of data mining for discovering patterns and useful information from user data on websites consisting of text, image, audio or video data, while web structure mining uses graph theory to discover the authorities and hubs of any web document (such as the appropriate web links for a web page) (Zhang and Segall, 2008: 684; Olson, 2008: 192). Web mining processes can be divided into five subtasks: i) resource finding and retrieving; ii) information selection and preprocessing; iii) patterns analysis and recognition; iv) validation and interpretation; v) visualization (Zhang and Segall, 685). Web content mining uses methods including: relevance feedback algorithms (for content-based image retrieval), keywords search (for homepage analysis), correlation mining/machine learning/partial tree alignment (for web query interface integration and opinion mining), and transforming multiform data into a unified format (for warehousing web data) (Zhang and Segall, 686). Web usage mining techniques include: association rule hypergraph partitioning (for automatic personalization), association rules/classification/sequential patterns (for system improvement, site modification and business intelligence), fuzzy clustering (for analysing and responding better to user behaviour for generating web promotions), pattern analysis (for identifying subjectively interesting web usage patterns), and clustering/association/classification (for large-scale web log and customer data mining) (Zhang and Segall, 689). Web structure mining techniques include: clustering/sequence alignment methods (for mapping user navigation patterns), frequent access path identification algorithms (for mining web browsing patterns for e-commerce), and heuristic approach (for hyperlink selection) (Zhang and Segall, 692-3).

These data mining techniques can thus allow companies to better understand (through monitoring) user behaviour on their websites, and improve the experiences of users by evaluating and improving site features (Jones and Gupta, 2006: 63). The data sources for these methods include individuals' homepages, server and client logs, weblog data, cookies, explicit user input, data from university websites, URLs from search engines and proprietary data sets (Zhang and Segall, 686-693). The methods used for mining data from the web and data sources involved, mean that those mining the data have access to information about individuals' online behaviours in multiple forms (from what individuals are clicking on, search terms, website preferences and online purchases). As such, an initial ethical concern is the invasion of user privacy along with the lack of informed consent, as users may not always be aware of who is obtaining, using or disseminating the data that is acquired when they are online (Wel and Royakkers, 2004: 130-131). Despite the fact that individuals should be informed about what the data being collected from them is used for, this is problematic to uphold with automatic data retrieval and classification by algorithms, because it is not clear beforehand what kind of patterns will be revealed in the data and therefore it becomes difficult to specify the exact purpose of the data in advance (this is further complicated if data is mined from historical datasets rather than in real time (Wel and Royakkers, 131). An additional issue arises when data is mined from an individual's home page or profile, and used outside the context in which it was originally posted, thus even if the data is public it does not mean it can be collected and used freely by data miners (Wel and Royakkers, 131). And in the case of web usage mining, how users navigate through a website can be tracked by the website owners, and though the log data may not identify the person's characteristics, it does identify their IP-address, time of entering and leaving the site, as well as hyperlinks followed, with cookies re-identifying the user upon return (Wel and Royakkers, 131).

5.3 Ethics of Big Data Analytics and Learning Techniques

Big Data analytics comprises the capabilities, techniques, and processes for gaining insights about patterns in very large data sets. We can evaluate techniques and applications of Big Data analytics from both epistemic and ethical standpoints. Relevant epistemic standards have been developed in the fields of probability and statistics, and their foundations have been investigated by philosophers of science (Hacking, 2001). Ethical issues emerge when we consider the application of analytics to moral subjects, especially human beings. This section provides an outline of ethical issues in Big Data analytics.

It is common to categorize data analytics capabilities as *descriptive* analytics, *predictive* analytics, or *prescriptive* analytics (Lustig et al., 2010). It is useful also to distinguish *diagnostic* analytics as a fourth category (Chandler et al., 2011). Ethical issues can be identified with respect to each of these four categories.

5.3.1 Descriptive Analytics

Descriptive analytics provides insights about the past and present states of objects represented in data sets. A primary ethical problem is the risk of advancing distorted representations of human situations. Such distortion may be due to inaccurate, spurious, or missing data. For example, a dataset might under-represent or over-represent particular segments of a population. This is especially problematic when patterns of under- or over-representation reflect patterns of social disadvantage.

Additionally, distortion may be due to emphasizing the factors that are easiest to quantify. Reductionism like this may oversimplify complex processes and cause distrust (Beresford, 2010). Since money is more easily quantified than other sorts of value, it is natural for descriptive analytics to present value in monetary terms. This may produce distorted representations of situations where the value is not monetary.

Finally, some descriptive analytics may jeopardize individual rights or dignity. With sophisticated descriptive analytics, an individual's privacy may be violated, even if the original data was collected and accessed only in ways that respect subjects' privacy (Barocas & Nissenbaum, 2014). In addition, some inferences may fail to respect the individuality of persons (Vedder, 1999).

5.3.2 Diagnostic Analytics

Diagnostic analytics, like descriptive analytics, focuses on the past and future, but adds inferences about causation and other aspects of explanation. Thus, it adds *why* questions to the *what* questions of descriptive analytics. A general worry about diagnostic analytics is the temptation to draw causal or explanatory conclusions, even when the data justify claims only about mere correlation. This is ethically significant when drawing conclusions about moral responsibility for actions and outcomes, since responsibility is not reducible to mere correlation.

Beyond mere correlation, even conclusions about causation do not straightforwardly entail moral responsibility. A person or group may be the immediate cause of some effect without being responsible for it (Wolf 1987). Moreover, persons or groups in complex situations may exhibit behavior that is not due to core character traits (Harman 1999). In general, overzealous application of diagnostic analytics may result in attributions of full or major responsibility when attributions of partial or distributed responsibility would be more appropriate. When diagnostic analytics is used to determine praise, blame, reward, or punishment, failure to properly attribute responsibility may result in morally objectionable outcomes.

5.3.3 Predictive Analytics

Predictive analytics relies on descriptive and diagnostic analytics to construct models that yield predictions about future events or other unknowns. In addition to epistemic challenges, predictive analytics raises several ethical issues. The first involves barriers to accountability: If a system makes predictions about a person, but the way the prediction was reached cannot be inspected and explained, then the system cannot be held accountable. However, especially with predictions by machine learning systems, it is difficult to specify and achieve the kind of interpretability and explainability required to engender trust in the system (Lipton, 2018).

A second issue also relates to accountability: Predictions issuing from data analytics may be self-fulfilling prophecies (Salganik & Watts, 2008). Furthermore, feedback loops of prediction may produce or exacerbate patterns of unfair treatment (Ensign et al., 2018). With self-fulfilling prophecies and feedback loops, predictions affect the outcomes they predict. For this reason, it is difficult to judge whether the prediction was warranted in the first place, but the predictions may impact people nonetheless.

Third, predictions about individuals' future behavior, especially their likelihood of success or failure, may discriminate against particular populations, especially historically disadvantaged groups (Barocas & Selbst, 2016). Along similar lines, the concern mentioned earlier about descriptive analytics failing to treat people as individuals is intensified when predicting individuals' future thoughts and behavior.

5.3.4 Prescriptive Analytics

Prescriptive analytics extends the other analytics capabilities already discussed by identifying options and recommending choices among alternative future courses of action. Prescriptive analytics is inherently normative, since it does not stop with conclusions about how things have been or will be, but also advocates particular future courses of action. Prescriptive analytics is typically concerned with optimization of future outcomes, and the selection of criteria to be optimized may depend on ethical considerations. Notably, optimizing for a moral value is difficult and requires ethical reflection, particularly when the value in question is fairness. There are several different conceptions of fairness that do not always ground the same prescriptions (Friedler et al., 2016).

Further ethical issues for prescriptive analytics arise in the selection or construction of the alternative possible courses of action from which a prescription is to be selected. Insufficient imagination regarding alternatives may result in the pursuit of a morally inferior course of action (Werhane 1998).

Since the primary purpose of data analytics is the production of new knowledge, it is unsurprising that many evaluations of analytics focus on its epistemic dimensions. However, this section's examination of different categories of analytics capabilities has shown that data analytics has numerous ethical aspects that extend beyond epistemic concerns.

5.4 Machine Learning

In this section ethical issues entailed by machine learning are addressed. Machine learning (ML) is an important technique in artificial intelligence (AI) that has drastically changed everyday life. ML may be categorized into three sections: supervised, unsupervised and reinforcement learning. All categories use 'data sets'; a training and a test data set. For supervised learning, both input and output is given in the training data, thereby 'supervising' the algorithm towards the correct answer. In unsupervised learning, only the input data is given. The algorithm then needs to develop a model to discover underlying patterns in the data. In this sense, it creates its own output. Reinforcement learning 'learns' by trial and error, thereby including the idea of 'making mistakes' into its working. The algorithm is given a begin state and an end state (the goal). Based on the trial-and-error procedure it learns which

steps or decisions are 'good' (i.e. lead to the goal) and which are 'bad' (i.e. impede reaching the goal). Either the algorithm is given which steps are considered good or bad (supervised), or it is left in the dark and must figure it out by itself (unsupervised).

Machine learning algorithms are popular due to the fact that the algorithm is able to update its abilities by itself. Thus, not everything needs to be preprogrammed (contrary to rule-based systems). Since the increase of data availability, machine learning algorithms have been able to significantly develop, surpassing many human achievements.[2] This increased use of ML algorithms, however, does not come without worries. This section highlights some of the most important concerns.

5.4.1 Bias and Discrimination

Currently, one of the major concerns revolves around discrimination. The common thought that algorithms are objective has been rejected by many specialists, arguing that the algorithms' dependency on data allows for a bias to be exposed or perhaps even augmented. The algorithm does not think for itself, but merely does what it is told. Therefore, if data shows a bias (for example, women never being recruited for a specific job), the algorithm will include this bias in its model, and therefore score women lower on the employee list for this particular job. This bias then creates an *unfairness*, by limiting possibilities for a specific group over another. A second type of bias concerns an unequal distribution in the input data. This allows for an "uncertainty bias" to occur (Goodman & Flaxman, 2017, 54, see also Kamishima, Akaho, & Sakuma, 2011). The uncertainty bias entails that an algorithm prefers certainty over uncertainty, therefore more often choosing options with more certainty (i.e. more data in the training set) and dismissing those with more uncertainty (i.e. little training data available).

In addition, Burrell (2015, 3) warns of biased impact the design of an algorithm may have, besides the input data. She argues that developers may (inadvertently) design their own bias into the algorithm. She therefore states that since algorithms always carry some human aspects, algorithms should not be regarded as objective.

5.4.2 Explainability

ML algorithms, specifically neural networks, have an opaque character. This implies that their computations are not transparent, not interpretable to humans. It is therefore difficult to explain why an algorithm reached its conclusion. Neural networks are based on the human brain. Like the brain, the algorithm utilizes different 'layers'. What happens in these layers is unclear to humans. All that is known is that certain features are mixed and matched, eventually resulting in an *output*. Parameters are used for this mix and matching, all with a specific weight that accounts for the value of a feature. Algorithms increasingly take decisions (or at least advise decision-makers by their output) that affect people's lives. For this reason, it is not surprising that some people feel the need for an explanation of why their mortgage request was rejected, or why they were not hired for a certain job. The problem then is that these outputs cannot be explained. Although there is some debate on this problem, there is no clear consensus on definitions (Doshi-Velez & Kim, 2017; Lipton 2018).

Neural networks' accuracy improves with an increase of data. This increase of data, however, may contribute to the opaqueness of the algorithm. More data implies more features (Burrell, 2015), which complicates the analysis of the algorithm. Furthermore, when an algorithm uses a great number of features, it becomes necessary to apply 'dimensionality reduction' to keep the algorithm's computational capabilities. Dimensionality reduction, however, implies neglecting some features and

[2] Activities that could be labelled as 'easy' (such as throwing and catching a ball) are incredibly hard to program, thereby illustrating significant limitations of machine learning algorithms.

merging several features with similar correlations. It remains unclear exactly which features are used and to what extent, therefore adding to the opacity problem.

5.4.3 Reliability

The reliability of an algorithm depends on the input data and the algorithm's ability to *generalize.* If the accuracy of an algorithm approaches 100%,[3] it is likely that the model is too fitted ('overfitted') for the input data, resulting in an overall worse performance on external data. In this sense, there is an on-going trade-off between accuracy and robustness in algorithms.

5.4.4 Privacy

An important concern regarding machine learning is the ability to connect hidden relations between features. If some features are left out due to fear of discrimination, the algorithm may still be able to link available features to this ignored feature, the so-called "red-lining effect" (Kamishima, Akaho, & Sakuma, 2011, 644; see also Dwork & Roth, 2014, 7). This also creates the possibility of turning anonymous data into personalized data.

5.4.5 Responsibility

Responsibility and accountability are a pressing issue considering ML algorithms. Cerna Collectif (2018) has addressed the problem of assigning responsibility by arguing that when the system is flawed the designer is responsible, but when it is used in the wrong way, the user is responsible. However, this neglects the impact of the training data, and assumes that if it is developed correctly then it will work correctly. While input data is assigned by the designer, the system updates itself outside the designer's control, for instance using the user's data. This problem has been addressed by Matthias (2004), who has formulated the concept of 'responsibility gap', arguing that there is a moral distance between AI machines and their developers. AI machines now sometimes act according to their plans, resulting in a loss of "control over the device" (Matthias, 2004, 176) by the developers. The device creates and revises its own plans, leaving the programmer as a mere 'creator' rather than the 'coder' of the system (Matthias, 2004, 176).

5.5 Natural Language Processing (NLP)

This section addresses ethical issues arising in the field of Natural Language Processing (NLP), a relevant subfield of AI. NLP deals with analyzing and synethisizing human language. It analyzes and derives meaning from texts (Natural Language Understanding; NLU), and synthesizes text such as responses to queries or translations (Natural Language Generating; NLG). NLP is used for tasks such as machine translation, speech recognition and text analysis or summarization.

NLU converts human language into "internal computer representations of information" (Reiter & Dale, 3), and NLG converts such internal representations into human language. Although they have similar end points, their internal workings and the problems arising are different. A problem for NLU is to understand incorrect grammar and paraphrasations equally, while for NLG a main concern is to develop language that is understandable for humans.

Currently, an increased use in machine learning algorithms has been observed in NLP, due to the highly accurate results. NLU involves text analytics (TA), which deals with comprehending the meaning of a

[3] An accuracy of 100% is practically impossible, see Brynjolfsson & Mitchell, 2017, 4.

text by extracting information. NLG is mainly concerned with translating languages using machine translation. Speech recognition aims to understand spoken queries by humans. This is different from voice recognition, which is focused on identifying a person based on their own individual speech.

The debate about ethical issues present in NLP has increased in recent years. One reason for this, given by Hovy and Spruit (2016), concerns the increased use of social media for developing NLP tools. The relation between the text and the author has become much more noticeable, allowing authors to be more easily identified, for example. In addition, the use of social media has revealed hidden biases in previous methods for NLP implementation. The most important issues that arise with general NLP that are relevant to NLU, NLG and speech recognition are elaborated in the following paragraphs.

5.5.1 Privacy

Personal privacy is especially at risk where NLP methods are used for medical reasons. Here, data sensitivity plays an even greater role. However, while anonymization of data (partly) maintains a level of privacy, it has a negative influence on the progress of NLP. The more data is used, the better the methods work. Privacy concerns in this sense then limit the progress of NLP, as data sets are restricted (Suster, Tulkens & Daelemans, 2017).

The use of social media in NLP techniques facilitates the identification of people. NLP tools are privacy sensitive, as they can categorize individuals into specific groups. This makes it possible to (at least partly) identify a person, based on his or her communication style. Communication style may hint at personal information such as potential living area, gender, etc. This, then, increases the possibility for identification (Hovy & Spruit, 2016). Interestingly, anonymization of data is not necessarily the standard means in the NLP community (Mieskes, 2017). Thus, especially if the data is used without someone's permission, the possibility for identification may raise privacy concerns.

There are privacy issues in speech recognition systems as well. In order for such systems (e.g. 'Amazon Echo', 'Google Home') to be able to respond, they need to recognize when a human is speaking. Therefore, many of these devices remain in a so-called "always-on" mode (Carlini et al., 2016, 513). In this mode, they are not actively recording the conversation, but are constantly listening for their 'trigger word'. A trigger word is a word that activates the device, such as 'OK Google'. The hearing of these devices is not perfect, leading them to sometimes interpret a spoken word as their trigger word. They start to record a conversation, even if this was not planned by the user. This is a privacy concern, as it may record people without their intention. Furthermore, a security issue is involved as well. The systems can pick up voice commands that are unrecognizable - and therefore unnoticed - by humans (Carlini et al., 2016, 525). The always-on mode creates the opportunity for others to hack the system and give commands without the permission and knowledge of the owner (Carlini et al., 2016).

5.5.2 Bias and Discrimination

While the use of demographic factors increases privacy concerns, it allows for more accurate results concerning minorities, younger people, etc. In the past, factors such as age and ethnicity have been neglected by NLP tasks, treating language as a uniform phenomenon (Hovy 2015, 752). However, the data sets used were specifically related to a group "older, richer, and more well-educated than the average population" (Hovy & Søgaard, 2015, 483). This then created a bias, disadvantaging people not part of this group. Social media, however, shows a clear distinction between different groups. To avoid exclusion, it is necessary to have more diverse input data. Hovy and Spruit (2016, 593) argue that a misrepresentation in data "in itself already represents an ethical problem for research purposes,

threatening the universality and objectivity of scientific knowledge" (Hovy & Spruit 2016, 593). This bias resides in the data (Hovy & Søgaard, 2015, 487; Hovy & Spruit, 2016, 593). Hovy and Spruit point out a bias existing on modeling and design levels as well. A model may produce false positives, risking "bias confirmation and overgeneralization" (Hovy & Spruit, 2016, 593). A design may lead to both bias confirmation and overexposure, which in turn may maintain or develop stereotypes (Hovy & Spruit, 2016, 594).

In addition, speech recognition for women and ethnic minorities are less accurate than for men, which shows a racial and gender disparity (Blodgett et al., 2016). Tatman (2017) has shown that automatic captioning on YouTube produces worse results for women than for men (Blodgett et al., 2016, 1). The impact is two-folded: the viewers (especially those completely relying on captions) have less information at their disposal, and the speakers have a smaller audience (Blodgett et al., 2016).

5.5.3 Transparency & Explainability

As NLP tools are increasingly developed with neural networks, transparency is reduced (Lei et al., 2016, 1). Currently, the best method for NLP is sequence to sequence (Seq2Seq) learning, which builds on deep language modeling (Wiseman & Rush, 2016). Due to the hidden layers in these networks, transparency and explainability is severely affected.

5.5.4 Specific to NLU: Text Analysis

NLU may be used for nudging people into a certain behaviour (Pryzant et al., 2018, 1), which might be considered as an interference with their autonomy. For example, a course description may contain specific words to nudge students into choosing that particular course, or what words in a consumer complaint will cause the management to act respond.

Secondly, if the input data is too narrowly focused on one specific type of phrasing, it may appoint different interpretations to different sentence structures that have the same semantic meaning.

5.5.5 Specific to NLG: Machine Translation

Machine translation may raise ethical concerns when a sentence has a certain ambiguity. The system then needs to either keep the ambiguity or choose a specific way to translate it. The way it is eventually translated might be due to a bias in the input data (Knight & Langkilde, 2000). Additionally, a sentence may overtranslate or undertranslate, implying that an NLG tool may translate a certain word or sentence more often than mentioned in the original language, or it may neglect certain words in the translation, resulting in a different meaning of the sentence and/or a miscomprehension by the reader (Zheng et al., 2019, 3).

Translation is done with the help of 'word embeddings'. A word is placed on a particular vector and compared. For example; Man is to X as Woman is to Y. A 'correct' relation would be Man is to King as Woman is to Queen. However, research shows that there are harmful relations included in these embeddings (e.g. Man is to Computer Programmer as Woman is to Homemaker) (Bolukbasi, 2016).

5.6 Embedded AI and Ambient Intelligence

Ambient Intelligence (AmI) is an emerging field related to AI that aims to assist people in their life using technologies embedded in the environment. For example, a fridge connected with kitchen cabinets could be able to create a grocery list based on what is missing inside the fridge and the cabinets (so-called 'smart technologies') (See Raisinghani et al. 2004). Some important characteristics of an AmI device include that it is embedded (the device is 'invisible' to the user), context-aware (the

device recognizes users), personalized (the device is tailored to the user's needs), adaptive (the device is able to change according to its environment and/or user), anticipatory (the device can anticipate a user's desires), unobtrusive (the device is discrete) and non-invasive (the device can act independently without interfering with the user) (Gams et al., 2019, 76).

AmI is a combination of ubiquitous computing, ubiquitous communication and user adaptive interface (Raisinghani et al., 2004). Ubiquitous computing refers to "omnipresent computers that serve people in their everyday lives at home and at work, functioning invisibly and unobtrusively in the background and freeing people to a large extent from tedious routine tasks" (Raisinghani et al., 2004). Ubiquitous communication implies that computers have the ability to interact with one another. This can also be seen as a part of ubiquitous computing. User adaptive interface, or intelligent social user interface (ISUI) has profiling as its main characteristic ("ability to personalize and automatically adapt to particular user behaviour patterns"), and context-awareness ("ability to adapt to different situations") (Soraker & Brey, 2007, 8). AmI devices can "infer how your behaviour relates to your desires" (Soraker & Brey, 2007, 9). ISUI include the ability to recognize visual, sound, scent and tactile outputs (Raisinghani et al., 2004).

Ambient Intelligence has the potential to save humans costs and time, provide a more convenient life and increase the level of safety, security and entertainment (Raisinghani et al., 2004). This, then, may lead to "an overall higher quality of life" (Raisinghani et al., 2004). While AmI surely realizes some - if not all - of these benefits, several ethical concerns arise with its use, relating to privacy, identity, trust, security, freedom and autonomy (Brey, 2005; Wright, 2005, 4). Furthermore, AmI may influence humans' individual behavior (Soraker & Brey, 2007) as well as their relation to the world (Araya, 1995).

5.6.1 Privacy

Privacy concerns are considered of utmost importance by both critics and proponents of AmI (Brey, 2005). Four properties of ubiquitous computing that make it especially privacy sensitive compared to other computer science domains have been highlighted (Langheinrich, 2001, 6). These properties are ubiquity, invisibility, sensing, and memory amplification. Thus, ubiquitous computing is everywhere, unnoticed by humans, with the ability to sense aspects of the environment (e.g. temperature, audio) as well as of humans (e.g. emotions), potentially creating "a complete record of someone's past" (Brey, 2005, 9). Regarding the Social Interface, one may add the properties of profiling (i.e. constructing unique profiles of users), and connectedness (wireless connection between devices) (Brey, 2005, 9). Privacy risks of AmI are considerable due to the interaction between devices. It is the combination of the sensitivity of the recorded information, the scale of this recording, and the possibility that interaction of devices facilitates distribution of personal information to other parties that makes AmI so vulnerable to privacy violation (Brey, 2005).

5.6.2 Freedom and Autonomy

While AmI may be regarded as augmenting freedom due to time and money savings, it may also be regarded as diminishing human autonomy and freedom (Brey, 2005, 4). Autonomy is commonly regarded as dependent on an individual's ability to make their own decisions and is seen as important due to the opportunity for "self-realization" (Brey, 2005, 4). Brey has analyzed AmI in relation to our freedoms and concludes that AmI has a chance to enhance our freedom in two ways: it may "enhance control over the environment by making it more responsive to one's needs and intentions" as well as improve "our self-understanding and thereby helping us become more autonomous" (Brey, 2005, 8). However, it simultaneously limits both freedoms by confronting "humans with smart objects that perform autonomous actions against their wishes" and "by pretending to know what our needs are and telling us what to believe and decide" (Brey, 2005, 8).

In addition, Soraker and Brey (2007) state that the use of AmI may influence a person's behavior. They argue that for AmI to understand what we want, the behaviour humans need to show to a device is similar to the behaviour they need to show to a pet; it must be "discrete, predictable and overt" (Soraker and Brey, 2007, 10). They claim that this may change our natural behaviour. Thus, AmI may force us into changing who we are and how we act; we will then be forced to fit ourselves within this technology. Moreover, some AmI devices may promote their use in solitude, risking isolation of individuals and a degeneration of society. Also, as some devices may replace tasks such as grocery shopping, the "face-to-face interaction between people" might diminish (Raisinghani et al., 2004), potentially adding to a feeling of isolation. Furthermore, as AmI devices are fabricated globally, there is a risk of cultural bias, resulting in discrimination of some cultures and encourage "homogenization of cultural expressions" (Soraker and Brey, 2007, 11). Finally, AmI systems impede manual resets. Soraker and Brey warn of a potential widening between users who simply go along with the requirements of the device and people that try to 'game' the system.[4] Not only is there an influence on an individual level, Araya (1995, 235) has argued that the whole relation between people and the world may be altered, as the entire world is transformed into a surveillable object.

6. Ethical Analysis: Ethical Issues in Different Application Domains

The use and implementation of SIS holds great potential to positively transform the lives of people around the world in a wide variety of ways. However, there is also the possibility that the use of these applications may have adverse ethical implications This section looks at 16 social domains, and the ethical issues identified within each, during the implementation and use of SIS. The purpose of doing so is to identify contrasting, diverging, and unique ethical issues pertaining to a range of social applications..
Many of the ethical issues in this section are derived from the SHERPA project's ten case studies and five scenarios, WP 1.1 and 1.2. The case studies focused on ten specific social domains, compiling detailed literature reviews and carrying out empirical research into organisations integrating SIS. The scenarios focused on five specific social domains using SIS, presenting future-focused accounts of ethical issues in these domains by the year 2025. Both of these Deliverables provided insights into the ethical analysis of the 16 social domains established by the University of Twente (UT) during an intensive brainstorming session prior to the commencement of the SHERPA project, as discussed in Section 3 of this Deliverable.

The purpose of this section is to analyse each social domain, and its use of SIS, to determine what ethical issues are relevant to that area, to create a broader, and all-encompassing, picture of the ethical issues faced in application domains. Table 1 classifies the ethical issues found within the 16 social domains that UT identified.

Social Domains	Ethical Issues

[4] Gaming the systems entails that someone may understand how a device responds to a user's behaviour, and therefore intentionally behaves in a specific way to conform the device to his/her own desires. This is problematic if a device is not merely for individual use, but rather for an AmI meant for multiple users. See Soraker & Brey, 2007, p. 11.

Banking and finance	Wellbeing; Digital Divide; Power Asymmetries; Market Manipulation; Accountability; Malicious Use
Healthcare	Privacy; Accountability; Informed Consent; Accuracy of Algorithms; Algorithmic Bias; Surveillance; Use of Personal Data; Data Ownership
Insurance	Accessibility of Data; Privacy; Bias; Employment; Discrimination; Transparency; Responsibility; Ownership of Data; Informed Consent; Security
Retail and wholesale trade	Manipulation; Privacy; Informed Consent; Bias; Competitive Disadvantage; Transparency and Vulnerability
Science	Privacy; Data Ownership; Accountability; Discrimination and Bias; and the Digital Divide
Education	Privacy and data protection; Bias; Public good or not; Inequality and asymmetries; Freedom of thought
Energy and utilities	Health and Safety; Privacy and Informed Consent; Cybersecurity Risks; Energy Equity; Sustainability
Manufacturing and natural resources	Digital Divide; Privacy Issues; Security; Contextual Integrity; Data Quality; Deskilling; Transparency; Integrity
Agriculture	Accuracy of Data and Recommendations; Data Ownership Intellectual Property and Power Asymmetries; Inaccessibility Economic and Digital Divide; Privacy; Animal Welfare and Environmental Protection; Employment
Communications, media and entertainment	Over-Representation and Bias; Research Ethics; Informed Consent; Re-identification; Profiling Individuals; Surveillance; Privacy; Filter Bubbles; Fake news
Transportation	Safety and prevention of harm; Autonomy; Rights; Insurance and discrimination; Privacy
Employee monitoring and administration	Harm to Employees; Privacy; Dignity; National Legal Differences; Informed Consent; Inequalities; Malicious Use of SIS; Transparency

Government	Accuracy of Data; Accuracy of Algorithms; Technological Lock-in/Power Asymmetries; Security; Manipulation; Access to SIS Availability of Data; Data Ownership
Law enforcement and justice	Discrimination; Human Rights Issues; Policing Biases; State and Corporate AI Collaborations; AI Explainability and Social Responsibility
Sustainable development	Conflict of Interests; Economic Pressure; Inequalities and the Digital Divide; Privacy; Accuracy of SIS and Bias; Availability and Accuracy of Data; Transparency and Trust
Defence and national security	Collateral Damage; Ethical Principles; Autonomous Decision-making; Counterattack; Informed Consent; Protection from Harm; Privacy; Control of Data; Vulnerabilities and Disclosure; Competence of RECs; Security Issues; Transparency; Trust; Risk; Responsibility; Business Interests and Codes of Conduct; Anomalies

Table 7: Social Domains

6.1. Banking and Finance

Economic and Social Well-being: High-frequency trading that utilises AI can lead to "flash crashes" that can trigger economic imbalances. Today, the majority of pension funds, insurance funds and government bonds are invested in stock market products traded via SIS trading. A flash crash has societal and well-being effects and implications as a result of financial market turbulence (Stankovic, et al., 2017).

Digital Divide and Power Asymmetries: Currently, only large investment houses can afford to, and have the necessary experience to run AI infrastructure. Inextricably, access to such technology is associated with access to information and the power to exploit market information unavailable to smaller firms or private investors. Similar information asymmetries that give an unfair advantage to some traders (i.e. insider trading) have been regulated against, but not for AI. In addition to uneven access to technology, there is the issue of access to quality data to train algorithms and validate models in real time (Harris, 2017). This has implications for furthering economic inequalities between those who can afford to engage in high-frequency trading over long-term investors and those who cannot.

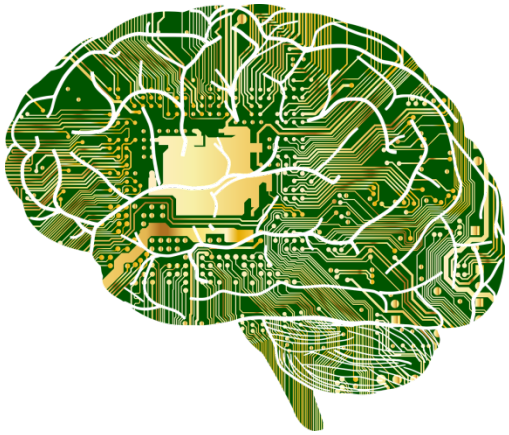
Cybernetic Market Manipulation: Commonly used methods of cybernetic market manipulation are pinging, spoofing, electronic front running, and mass misinformation. Pinging refers to a large number of small trades submitted and immediately cancelled with the intention to elicit a response that reveals the trading intentions of another party. Spoofing refers to trading stocks below their normal market value to trigger the offloading of their stock leading to a drop in market prices. Electronic front running is enabled by special feeds privy to privileged actors, often for a fee, that allows them to see

submitted, but not yet executed, trades. This may allow their systems to react ahead of the competition. Finally, mass misinformation is the contamination of publicly available data (say for example via fake news) to manipulate stock or commodity prices. This can have knock-on effects on the value of retirement funds, destroy companies, and facilitate rogue trading (Lin, 2017).

Accountability: While trading algorithms are in principle programmed by people, the lines of accountability are often unclear about who bears the legal and moral responsibility for the consequences of their actions. Because of the shared responsibility between those officially responsible (i.e. technology players, programmers, and data sources), this can lead to unaccountability (Cave, 2017). The view of inscribing ethical reasoning in autonomous agents is salient, but not particular to the field of trading. It is part of the more general move toward explainable AI, which enables AI to reflect and explain their decisions and actions.

Artificial agents pose a philosophical question about agency more generally that relates to intent behind trading actions. This has implications for distinguishing lawful and unlawful actions (Cave, 2017). This is exacerbated by the gap between the technologies available to traders and those available to regulators to detect and counter such actions (Busch, 2016). Accountability for algorithmic decision-making means being able to justify why a particular output/decision was reached by an algorithm. In the context of automatic credit scoring systems, "the bank might justify their decision by reference to the prior successes of the machine learning techniques they used to train their system; or the scientific rigour involved in the development of their psychometric test used to derive the credit score" (Binns, 2017, p. 2).

Malicious and Illegal Use: AI effectively creates an alternative currency system outside the realms of the monetary policy of governments and central banks. By facilitating the trade of cryptocurrencies, AI potentially may create shadow banking, i.e. financial activity outside the remit of the law. This has implications for tax evasion, money laundering and trading in the dark web. The latter ranges from trading illegal goods (e.g. drugs trade) to rogue financing or illegal activities (terrorism funding) (Dierksmeier and Seele, 2016).



6.2. Healthcare

Privacy: In SIS projects, there is a risk of privacy violations if the identity of the data provider is uncovered. If SIS health data repositories can be de-anonymised, there is the potential for privacy harms against the data subjects (Rommelfanger et al., 2018). When SIS is used in clinical settings to e.g. implement genomic discoveries, there is a risk of privacy infringements (Chow-White et al., 2015). If genomic data is widely incorporated in online networks, the management of privacy and consent will be greatly challenged.

Accountability: When using SIS in healthcare, there is the possibility of breaches of privacy and varying harms, so there should be a level of transparency and accountability.

Informed Consent: There is a difficulty asking for *specific* consent in healthcare when using SIS, given the possibilities that Big Data can re-identify individuals without consent (Rumbold and Pierscionek, 2017). Being able to have a clear way of dealing with informed consent is also complicated by the fact that different countries may have different regulations with regards to the use of public health data.

In the application of SIS in healthcare, informed consent is problematic due to the unknown extent to which the data gathered from individuals may end up being used. The kinds of questions that may emerge include: "how to obtain consent for future unspecified and/or "secondary" research; how to protect donors' confidentiality; whether, when, and how to return research results and incidental findings; how to conceptualize the ownership and property status of donated data and tissue and of research results; and, in the case of biobanks, how to manage the return of materials to communities and disposal of unused material" (Lipworth et al., 2017, p. 486).

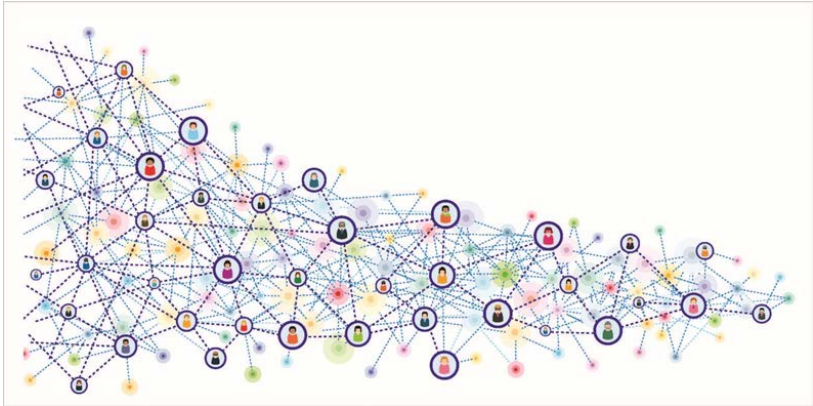
Accuracy of Algorithms: The accuracy of recommendations given by infectious disease outbreak algorithms depends on high volume of data based on individuals' health records as well as surveillance of their behaviour concerning their healthcare choices. But at the same time, there are also risks including "wrongful conclusions, potential misuse of personal information, well-publicised privacy breaches and ongoing profiling of individuals for commercial purposes" (Garattini et al., 2017, p. 2).

Algorithmic Bias: In order to make use of aggregated population data in the domain of infectious diseases, profiling is used that stratifies individuals into smaller groups based on ethnic group, gender and socio-economic status. A reliance on algorithmic analysis at the cost of reduced general understanding can "provide the basis for (intentional or unintentional) discrimination among individuals or groups by downstream policy makers and implementers" (Garattini et al., 2017, p. 8). There is a need for individuals to be aware of how algorithms reach their decisions so that "decision-making through profiling will not impact future healthcare provision and that an alternative human intervention can be provided" (Garattini et al., 2017, p. 8).

Surveillance: The utility of Big Data analytics for devising effective infectious disease control policies relies on surveillance strategies. Healthcare organizations have the ability to "continuously observe and monitor behaviours through mobile phone apps or wearable devices, offering personalized services and advice" (Garattini et al., 2017: 9).

Use of Personal Data: The growth of databank research has also prompted ethical concerns and questions surrounding it, as databanks contain "both laboratory-generated data (e.g. genomic, proteomic, or metabolomic data) and various forms of "real world data" from sources such as electronic medical records, clinical audits, administrative databases, registry reviews, mobile health applications, social media, census data" (Lipworth et al., 2017: 486).

Data Ownership: There is the concern that public health data holds the possibility of being used for monetary purposes, so it is important to identify who owns this data. In terms of brain research data, for example, there are enormous datasets that combine copyright data, publicly accessible data, patents, and trademarks, making it difficult to identify data ownership (Alexiou et al., 2013; Anagnostopoulos et al., 2016).



6.3. Insurance

Accessibility of Data: Insurance companies using SIS are heavily reliant on data (Deloitte Digital, 2017; Dutt, 2018). However, it is often difficult to access data because they are contained in many different systems, such as laboratories, clinics, registers, and private companies (Koh and Tan, 2018).

Privacy: If SIS can determine if a client has a disease or disorder through public information, i.e. through their social media posts, insurance SIS may infringe on the privacy of that individual (Dutt, 2018).

Ownership of Data: In the insurance industry, it is important to clearly establish who is the owner of the data provided to the insurance company. Sometimes, it is unclear how and why an insured person's data is being used, and if they are aware of its use.

Transparency: As a result of some machine-learning being black-boxed, it increases the difficulty of having transparent insurance SIS (Bharadwaj, 2018). However, in the insurance industry, customers have a legal right to be informed about how their personal data is used, so there is a tension between the useful implementation of SIS and how transparent their outcomes are.

Responsibility: There is an onus of responsibility on insurance companies implementing SIS, as they are directly working with data from the insured persons. European organisations need to abide by European data protection regulations, ensure strict in-house quality control procedures and that customers' data is protected.

Bias: If data is corrupted, lacking, or inaccurate, it may cause SIS to provide false insurance policy recommendations, which may cause prejudice, discriminatory, or harmful decisions against the policy-holder (Bharadwaj, 2018).

Discrimination: If SIS correlate publicly available data on individuals to make decisions about their insurance policies, this may lead to discrimination against groups of people and individuals (Dutt, 2018). SIS may further exacerbate discrimination based on individuals' income, home address, job, education level, marital status, and ethnicity (Foggan and Panagakos, 2018).

Security: Because of the level of personal information used by insurance companies, they need to ensure strong security measures are set in place to protect insured individuals' data.

Employment: There is a concern that SIS will replace jobs in the insurance industry. For example, Japanese insurer Fukoku Mutual Life replaced over 30 employees with AI systems (Newton Media,

2018). However, there is also a lack of available skilled AI professionals in the insurance industry, making it a challenge to implement SIS effectively without further human input (Bharadwaj, 2018).

6.4. Retail and Wholesale

Privacy: There are many ways that privacy can be infringed upon through the use of SIS in retail and marketing, such as Target identifying a pregnant teen based upon her online activities (Duhigg, 2012); social networks use facial recognition on dating websites to identify people; and data analytics has the potential to uncover people's anonymised identity (Braun A., Garriga G., 2018, p.666). The use of SIS in marketing creates opacity because it is unclear what customers' data will be used for (Ghosh and Moorthy, 2015). Data may be stored for long periods and may be used in a number of different ways that the consumer is unaware of (Braun A., Garriga G., 2018, p.671).

Informed Consent: Data derived from SIS may be used by companies for purposes that the customer has not consented to (Foster and Young, 2011). Even when terms & conditions are supplied, most people do not read these as they use legalistic jargon and are cumbersome to read, which raises concerns around the validity of informed consent in these situations.

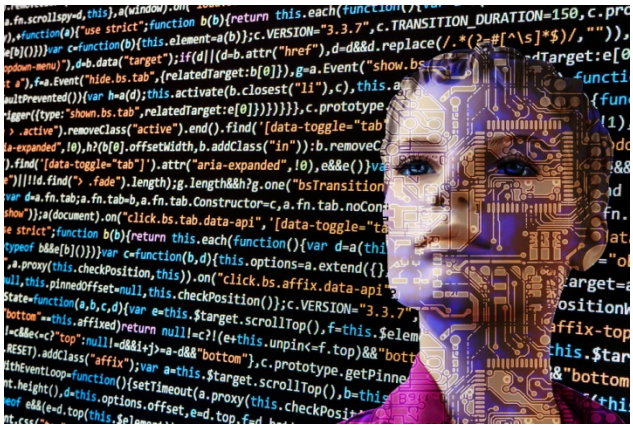
Bias: There is the possibility that retail and marketing SIS will enforce or create prejudice and bias towards certain groups within society (Macnish, 2012; Mittelstadt et al., 2016; O'Neil, 2016). SIS models may use ethnicity, an individual's background, their economic status, home address, or even technology used, to create biased marketing campaigns for that person. For example, the hotel search engine Orbitz split hotel searches among users depending on the computer they were searching from. Apple users were given details about more expensive hotels than Windows users (Mattioli, 2012).

Manipulation: In some areas of retail, companies are using SIS to create models to identify financially vulnerable customers who would be more likely to take out a loan (Harrison and Grey, 2010, p. 438). There is the possibility that vulnerable individuals will be preyed upon, manipulated, and exploited by the use of SIS techniques in the retail sector.

Competitive Disadvantage: Particularly in marketing and retail, competitive advantage is key to a business' success or failure.

Companies working under EU legislation need to abide by the GDPR, which may impact EU businesses developing and using SIS. SIS is heavily dependent on available and accurate data, and it is proposed that companies working outside of the EU will be able to improve their SIS because of greater access to data that is inaccessible in the EU.

Transparency and Vulnerability: Sometimes, the more transparent SIS companies are, the greater likelihood that it will be used against them, particularly in retail and marketing sectors. For example, with traditional rule-based systems, there is the possibility that customers can game the system, gaining access and control over the company's different offers.



6.5. Science

Privacy: SIS have the potential to create privacy violations when used in scientific research by uncovering the data provider's identity or other sensitive information. There is the possibility that data used in SIS could be used by third-parties to re-identify research subjects (Rommelfanger et al., 2018).

Data Ownership: Data from scientific research may be used and distributed to third parties for commercial benefits. Because SIS involve a myriad of intellectual properties (such as copyrights, trademarks, and patents), it is difficult to pinpoint who owns what data (Alexiou et al., 2013; Anagnostopoulos et al., 2016).

Accountability: The use of SIS may lead to issues of algorithmic bias, injustices, and harms to users, so there needs to be accountability for using SIS for scientific research (Anagnostopoulos et al., 2016).

Bias: There is the possibility that certain segments of the population will be over-, or under-, represented in scientific research using SIS, leading to bias or discriminatory recommendations.

Digital Divide: Scientists will have varying levels of access to SIS, which may progress or hinder their research as a result. This may lead to a digital divide, with some organisations, research bodies, cities, countries, or continents, having greater access to SIS than others.

6.6. Education

SIS tools are deployed in education with a view to enhance the education process via personalised support. The use of AI as learning support requires repeated interactions between AI tools and students. Unavoidably, minors' personal data must be shared to allow AI tools to adapt to students' habits and learning styles and make better decisions on how to support students and classrooms. Therein comes the challenge that all data has to be kept safe and anonymous.

Privacy and data protection: Classroom robots and learning buddies are constantly monitoring students and their environment, via video and audio surveillance tools. Hence a key ethical issue identified is that of privacy.

Bias: Bias is a risk because of AI's learning capabilities. It is hard to fix because of unknowns, imperfect processes of data collection and annotation, lack of social context and most importantly, different definitions of concepts such as fairness. Learning is very much a matter of motivation. As AI cannot account for changes in motivation, it can trap a student in a self-fulfilling cycle based on historical behavioural data.

Inequality and asymmetries: Ownership and access to AI learning tools may only be accessible to affluent students, which would create power asymmetries and inequality in terms of opportunities. This can be the result of retrieved AI skills or even different capabilities of human-to-robot interaction.

6.7. Energy and Utilities

Privacy and Informed Consent: Data retrieved about householder energy use may contain information that could infringe upon the household members' privacy. For example, smart meters could be used to detect when someone is home, taking a shower, or watching TV, based on the appliances' usage (Gray, 2018, Knapman, 2018). Some families may be forced to provide their smart meter data because that energy provider is the only viable one in their area, thus limiting their ability to control the use of their data.

Energy Equity: There is the possibility that affluent consumers will be prioritised in the future, if the smart grid manages energy distribution unevenly. Some have claimed that there is the possibility that algorithmic bias may prioritise providing wealthy individuals with their energy needs over poorer families. There is also the possibility that dynamic energy pricing may benefit the more affluent in society, while forcing those who cannot reduce their energy levels to become worse off (Faruqui, 2010).

Health and Safety: There are health concerns relating to the use and implementation of smart grids. Some believe that radio frequency radiation has carcinogenic effects, despite the Electric Power Research Institute indicating that these levels are acceptable. So the retrieval of data for SIS may cause health and safety risks to users.

Cybersecurity Risks: There is the potential that cyberattacks could cause serious damage to power control equipment, which may severely impact the energy grid (Eder-Neuhauser, et. al., 2017). For example, in West Ukraine, in 2016, cyber-attackers hacked the local power grid, cutting electricity in 250,000 homes for several hours (Cherepanov, 2016; and Cherepanov and Lipovsky, 2017)

Sustainability: Using SIS to model energy production and consumption may allow us to use it more sustainably. However, the creation and use of SIS and smart meters comes at an energy cost to a country's energy consumption levels.

6.8. Manufacturing and Natural Resources

Privacy Issues: Within the manufacturing industry, there may be privacy infringements if companies analyse the performance of their employees (Tiwari, 2017, p. 17). Supply-chain management (SCM) and predictive analysis uses multiple different sources to identify patterns and trends for effective prediction and implementation. However, using a wide variety of different data sources may create privacy concerns (Bates et al., 2014; Petersen, 2018). If sensitive data is used that identifies individuals,

or has the potential to identify individuals, there is a concern that this data may be used illegitimately or maliciously.

Digital Divide: There is the possibility that using SIS within the manufacturing industry will lead to a greater digital divide amongst people, companies, and countries. There is the possibility that SIS will reinforce power asymmetries and inequalities between large and small companies because of the availability of data (larger companies having vastly more available information); access to closed external data; determining how valuable a data asset is; and who will control these datasets and be able to access them.

Security: If there are poor cybersecurity measures in SCM SIS, there is the possibility of attacks on that system, leaving users, owners, and stakeholders of that organisation in jeopardy. Security is worsened by poor threat-detection procedures, misclassifications, and misuse of data (Gupta, 2018; Horvitz, 2017).

Data Quality: Zhong mentions how manual-based data collection approaches are still widely used in SCM (Zhong et al., 2016, p. 581). However, data obtained through these approaches are prone to be incomplete and inaccurate, which could lead to unreasonable or unrepresentative decisions. Therefore, determining how one can verify the quality of social media data, whether news is fake, and if the manufacturing industry is vulnerable to AI influences, are all difficult challenges.

Transparency: Auschitzky et al. (2014) report how a chemical company was able to reduce waste of raw materials by 20%, and energy costs by around 15%, by using data analytics and deep neural networks. While these optimisations are impressive, there is often an opaqueness around the types of data used, which makes it difficult to identify potential mistakes resulting from SIS use (Hacker, 2018; Horvitz and Mulligan, 2015; Meira, 2017; and Wachter et al., 2017b).

6.9. Agriculture

Data Ownership: Farmers are worried that their farm data will be used against them by regulatory bodies, governmental officials, and commodity traders (Coble et al., 2018, p. 84; Ferris, 2017; Rosenheim and Gratton, 2017, p. 403; and Sykuta, 2016). This may result in harm to the farmer and their livelihood, from increased fines, restrictions, unfair market pricing, selling unnecessary products

to the farmer, or threats/blackmail against them (Ferris, 2017; Kamilaris, Kartakoullis, and Prenafeta-Boldú, 2017, p. 29; and Ksetri, 2014, p. 13). In the agricultural community, there is apprehension about giving farm data to third-parties because of concerns about data ownership and what can be done with it (Coble et al., 2018, p. 84; Kosier, 2017; and Schönfeld, Heil and Bittner, 2016).

Privacy: Some claim that privacy is less of an issue in the agricultural industry because fewer personal data is retrieved from farms (Ferris, 2017; and Zhang et al., 2014). However, farmers' personal information and farm data still creates several privacy concerns and farmers want their data to be stored and used in a safe and secure manner (Ferris, 2017; Lokers et al., 2016; and Tzounis et al., 2017). In addition, SIS may also retrieve third-party individuals' data without their knowledge or consent, thus infringing upon their privacy as a result (Schönfeld, Heil and Bittner, 2016).

Accuracy of Data and Recommendations: SIS are used in agriculture to improve farm decision-making (Talavera et al., 2017), but some claim that they may provide misleading recommendations as a result of inaccurate data (Taylor and Broeder, 2015, p. 13; and Zhang et al., 2014). Data may be jeopardised as a result of animal interference and false readings as a result of varying climatic conditions (O'Grady and O'Hare, 2017; Tzounis et al., 2017). Data may also be affected by local idiosyncrasies, farmers' limited knowledge of and thus provision of inaccurate data (Byarugaba Agaba et al., 2014, p. 21; Lokers et al., 2016; and Taylor et al., 2014). Inaccurate data and recommendations may subsequently result in lost harvests, sick livestock, and general harm to the farmers' business.

Inaccessibility: There is a concern that farmers will not have the necessary skills or knowledge to understand the use of SIS on their farm or the ability to understand the data, thus placing a greater dependency on agricultural technology provider (ATPs) (Schönfeld, Heil and Bittner, 2016; and Sykuta, 2016, p. 60).

Intellectual Property and Power Asymmetries: While agribusinesses want to retain intellectual property of their data, there is a tension between data they retrieve from farmers, the analytics involved to produce valuable insights, and the intellectual property used to do so. Many agribusinesses have been creating strict contracts that bind farmers to contractual arrangements with them, preventing them from changing agricultural technology provider (ATP), and often facing penalties and court cases as a result (Darr, 2014; Sykuta, 2016): "ATPs may have concerns about receiving data from farmers that the farmer herself does not own, giving rise to potential violations of intellectual property or licensing restrictions" (Sykuta, 2016, p. 66). John Deere has implemented policies prohibiting farmers from adjusting their machines because of potential intellectual property infringements – the companies' hardware is contained on/within the vehicle – reducing the farmer's control over their farm (Carolan, 2015; and Wolfert et al., 2017).

Economic and Digital Divide: Most agricultural data analytics is done on large monocultural farms, is often expensive to implement, and requires good local technological infrastructure to be adopted (Carbonell, 2016; Kosier, 2017, p. 11; Micheni, 2015; and Schönfeld, Heil and Bittner, 2016). Agricultural SIS may create a digital divide between those that can afford to implement it and farmers, regions, and countries who cannot. This may worsen inequalities and wealth disparities (Kamilaris, Kartakoullis, and Prenafeta-Boldú, 2017, p. 29; and Poppe, Wolfert and Verdouw, 2014).

Animal Welfare and Environmental Protection: Implementing SIS on farms has the potential to injure, stress, and harm farm animals and surrounding wildlife. SIS may become damaged and leak toxic material, electrical voltages, polluted water run-off, and generally become a hindrance on the farm (Kosier, 2017). Algorithms may also prescribe harm to surrounding farmlands and the environment by not factoring those in as externalities in their recommendations (Antle, Capalbo and Houston, 2015).

Employment: There is the possibility that SIS will replace the need for many human jobs in the agricultural sector. For example, if agricultural SIS can provide recommendations that would have traditionally been done by an agronomist, then it may eventually replace them. At present, most agricultural SIS are not advanced enough to do this, but this is a limitation on the state of the technology, rather than any particular social, economic, or political barrier within the field.

6.10. Communications, Media and Entertainment

Research Ethics: There are questions about how much information can be used about individuals for research purposes when that is obtained through social media. Whether data that is publicly and voluntarily posted by users can legitimately be used for research purposes or if it infringes upon users' privacy is widely contested (Townsend and Wallace, 2016, p. 5).

Informed Consent: It is often difficult to determine the level of consent that users are giving on social media and whether adequate informed consent procedures are followed by companies collecting and processing this information for research purposes (Townsend and Wallace, 2016, p. 6). For example, in June 2014, a computer scientist and two academics at Facebook conducted an "emotional contagion" test, which "altered the content presented in the news feed of 689,003 people during one week to assess whether or not exposure to emotional content by one's contacts would alter what a person posted" (Boyd, 2016, pp. 4-5). This was defended because the "practice of A/B testing is commonplace in and essential to the production of algorithmically produced recommendations, which are the cornerstone of Facebook's news feed" (5). One of the main issues that added to the backlash against Facebook's experiment was the lack of informed consent.

Re-identification: While there are great efforts to anonymise individuals for research purposes, there is an ethical concern that their identify will be established as a result of the re-identification method.

Profiling Individuals: Social media companies and researchers may use social media content to profile users by correlating their trends and behaviours online. There is now concern that they can predict personalities (Golbeck et al., 2011, p. 261) and detect depression (Shen et al., 2017, p. 3838). However, whether the use of SIS to determine these patterns is accurate or not is debatable. Furthermore, there are a wide range of potential ethical issues arising as a result of increased profiling, such as privacy, surveillance, and algorithmic bias (Bekkers et al., 2013, p. 341).

Privacy: Facebook micro targets ads system has vulnerabilities that could breach a user's privacy: "We experimentally investigate the workings of the Facebook's advertising system and establish that … which we show leads to an attacker being able to design and successfully run advertising campaigns that: A) Infer information that people post on the site in "only me", "friends only", and "hide from these people" visibility mode B) Infer private information not posted on Facebook through ad content and user clicks C) Display intrusive and creepy ads to individuals" (Korolova, 2010, p. 3). The Cambridge Analytica scandal also raised a number of privacy concerns related to the social media users' data: "Cambridge Analytica used Big Data and advanced ML techniques to provide a full suite of services to enable highly targeted marketing and political campaigning, which raised concerns with regards to the privacy of those whose data had been accessed (Gupta, 2018; Isaak and Hanna, 2018)."

Filter Bubbles: Social media platforms have to present content in a certain order to the end user. Since it is in the best interest of social media platforms to display content that is relevant to the user, content is often personalised. This could introduce risks related to how to determine which content is presented to users, given the vast amount of available content: "Search engines, news aggregators, and social media networks are increasingly personalizing content through machine learning models, potentially creating "filter bubbles" in which algorithms inadvertently amplify ideological segregation by automatically recommending content an individual is likely to agree with" (Flaxman et al., 2016, p. 299).

Over-Representation and Bias: Researchers may conflate Big Data, for instance proposing that data retrieved from social media platform is representative of an entire population. Boyd and Crawford highlight how researchers using data scraped from Twitter may make it appear as representational of all individual online activity. As they point out, however, "Twitter does not represent 'all people', and it is an error to assume 'people' and 'Twitter users' are synonymous: they are a very particular sub-set" (Boyd & Crawford, 2012, p. 669). Such over-representation rests on assumptions of what 'users' are, what 'participation' is and what 'active' means (Boyd & Crawford, 2012, 669).

Fake news: Fake news can be amplified and exacerbated through the use of social media platforms (Guess et al., 2018, p. 1). Social media plays a major role in the emergence of "fake-news" (Peters, 2017, p. 564): "Facebook was among the three previous sites visited by respondents in the prior thirty seconds for 22.1% of the articles from fake news websites we observe in our web data" (Guess et al., 2018, pp. 8-9). However, it is very difficult to create particular algorithms to combat and censor the spread of fake news, because of the classification of what constitutes misinformation (Parry, 2018, p. 1).

Surveillance: Social media data may be used by public organisations and governments as a means of surveillance and control: "social media monitoring is gaining a fully-fledged position alongside the more traditional ways of gauging sentiments and views among target groups and clients" (Bekkers et al., 2013, p. 341).

6.11. Transportation

Self-driving vehicles (SDVs) offer great benefits for society, in terms of reducing parking space requirements, traffic jams and congestion, identifying better routes to take, driving more sustainably, and a reduction of crashes holding up traffic flow. They can also turn needless 'driving time' to extra productive, social or relaxation time for passengers. However, they raise a wide number of ethical issues.

Safety and prevention of harm*:* The most important discussion around SDVs is that of safety for passengers and bystanders. A key question of moral agency that remains open is how should the SDV be programmed and who should determine the course of action in the case of an unavoidable collision (Contissa et. al., 2017; Johnsen et. al., 2017). If algorithms aim to only protect the driver, they may crash into children or light vehicles, instead of other cars, walls, or lampposts, to protect the driver (Contissa et. al., 2017; Nyholm, et. al., 2016). If algorithms target those less at risk, then people may take more unsafe activities in order to become 'safe', i.e. cycling without a helmet so that SDVs view you cautiously, thus avoiding collision.

Autonomy*:* There has been a concern that SDVs are threatening our free will and responsibility, because of the removal of accountability from the individual as a result of overreliance on algorithms and AI (CNIL, 2017). Pre-programmed responses remove control from the human being in specific circumstances (FMTDI, 2017).

Rights: While SDVs hold the promise of safe personalised mobility for the elderly, blind or otherwise disabled people are still disadvantaged, as there is a question about who could deny people the right to use SDVs.

Privacy*:* As a result of the large amounts of data retrieved from SDVs, policymakers need to identify methods to ensure privacy and data security; determine who should have access to this data; how it should be securely stored; and if law enforcement should be allowed to hack an SDV if it is breaking the law (CNIL, 2018).

6.12. Employee Monitoring and Administration

Harm to Employees: Some claim that employee monitoring is used "to keep workers under pressure, to threaten, to appeal, and to make them feel the power over" (Karahisar, 2014). The use of SIS employee monitoring tactics creates a sense of pressure and control over the employee, making them feel humiliated, stressed, demoralised, and anxious (Karahisar, 2014). These psychological harms may thus materialise into physical harm as a result of the stress caused by SIS monitoring (Alder, 1998).

Privacy: One of the outcomes of employee monitoring is the infringements upon employees' privacy, with the feeling that privacy in the workplace does not exist (Mujtaba, 2004; Mishra and Crampton, 1998). If employees' privacy is being harmed, they may have less trust in their management, less commitment to the company, and feel less motivated (Chory, Vela and Avtgis, 2016).

Dignity: While employers have a legitimate reason for being concerned about what their employees are doing during working hours, some claim that this does not justify constant monitoring because of its infringement on their dignity as human beings (Fairweather 1999). Employees should not be forced to uncover all aspects of themselves, even in the workplace (Fairweather 1999).

Informed Consent: Informed consent is an important issue when using employee monitoring SIS. It is important to obtain informed consent from users before implementing SIS in the workplace: "In our system, we give the ability to our customers to take consent from their customers. We give them the ability to configure how the system will work depending on the state of consent. For example, if the customer has not consented, it is not possible to allow the customer to use the system in a full functionality or even delete the customer from the system" (Macnish et al., 2019, p. 47).

Inequalities: There is the possibility that employee monitoring SIS may create inequalities in the workplace if they are controlled by one or only a few individuals. This may lead to the monitoring of certain individuals within the company, placing a greater focus on some, or ignoring others that have preferential treatment.

National Legal Differences: There is a wide variation between the implementation of privacy laws and workplace monitoring, "in this respect the EU, United States, and Canadian approaches are similar; all give some due to the business reasons for electronic monitoring" (Lasprogata, King and Pillay, 2004). However, their approaches are quite different: "The lack of a common legal paradigm for the EU, Canada, and the United States is due to inconsistencies in the privacy laws and underlying value systems of the different countries, and the variety of factors that alter the lawfulness of employee electronic monitoring from country to country" (Lasprogata, King and Pillay, 2004).

Malicious Use of SIS: There is the possibility that employee monitoring SIS may be used for malicious or illegitimate purposes, for instance, threatening, extorting, or terrorising employees with details found out about them using SIS. Employers may install SIS monitoring without the knowledge or consent of their employees and use this information for marketing or financial purposes.

Transparency: If employee monitoring SIS is transparent, in how it works and is used, it may reduce many of the harms and ethical concerns raised using this technology.

6.13. Government & Law

Accuracy of Data: If there is insufficient or inaccurate data, there is the possibility that many important details will be missed from analysis (Batty et al., 2012; Kitchin, 2016a). The data used to inform policymakers and guide decisions may be contextually loaded, biased, inconsistent, unreliable, misclassified or insufficient (Bibri, 2018, p. 197; Glaeser et al., 2018; Kitchin, 2015a p. 15, Kitchin, Lauriault, and McArdle, 2015, p. 28).

Accuracy of Algorithms: One of the resultant issues of inaccurate data is that it will cause algorithms to be misleading or biased. However, algorithms may be incorrect, regardless of appropriate data. There is the possibility that algorithms will reduce the complexity of districts, cities, and nations by formulating that they are something rational, knowable, mechanical, highly routinized, predictable and limited (Batty, 2018; Creemers, 2018; Kitchin, 2015a; Kitchin, 2013; Kitchin, 2016a; and Kitchin, 2016b). If human bias is inputted into SIS algorithms, there is the possibility that state services will be provided unfairly and unequally (Capgemini Consulting, 2017; and Sholla, Naaz, and Chishti, 2017). An additional concern is that citizens will be 'nudged' to do certain activities, manipulated, and socially-controlled (Cardullo and Kitchin 2017; and Kitchin, 2018, p. 25)

Technological Lock-in/Power Asymmetries: The costs of Big Data analytics, storage systems, and AI research costs a lot of money for governments to implement and there is not always a guarantee of return on investments (Glasmeier and Christopherson, 2015, p. 7; and Hashem et al., 2016, p. 749). The public sector often forms partnerships with private companies on SIS projects, but there is the possibility that they may become dependent on these companies, leading to: the jeopardization of public decision-making; privatisation of public goods and state facilities; private spaces becoming used for private organisations' data retrieval and marketing purposes; and public bodies and places being open to cyber-attacks (Batty et al., 2012; Hollands, 2015; Kitchin, 2015b; and Kitchin, 2016a).

Privacy: There is a great concern that with the amount of data being retrieved by governments that the public's privacy will be infringed upon by geo-targeting, profiling, social sorting, surveillance, marketing purposes, and general feelings of being monitored (Elmaghraby and Losavio, 2014; Kitchin 2015a, p. 9; Kitchin, 2016c, p. 8; and Li et al., 2016). Citizens are worried that the use of SIS in the public place will lead to dataveillance, with calls for citizens to be informed about what is done with their data, who will have access to it and how it will be stored (Galdon-Clavell, 2013, p. 720).

Security: Data retrieved from SIS may harm citizens, but it may also jeopardise different governmental departments' functionality, and thus the safety and security of citizens. While greater digitalisation may help improve many aspects within different governmental bodies, it also increases their vulnerability to digital hacking, data abuses, increased security costs, and security threats (Kitchin, Lauriault, and McArdle, 2015, p. 20; Li, Cao, and Yao, 2015; and Sen et al., 2013). Adequate security measures are one aspect required for public acceptance of SIS (Zhang et al., 2017).

Manipulation: As a result of SIS governmental use, it may lead to active and/or passive manipulation. Active manipulation is when algorithms are deployed to manipulate/nudge citizens for illegitimate and direct political purposes. Passive manipulation is nudging through algorithms employed for legitimate purposes, such as to instil healthy behaviours in the public to decrease the burden of the public health on the state budget: "nudges that deliberately seek to exploit cognitive weaknesses to provoke desired behaviours entail a form of deception" (Bovens, 2008; Yeung, 2012, as cited in Yeung, 2017). Nudges circumvent the individuals' rational decision-making processes, thus disrespecting individuals as autonomous, rational beings capable of making decisions concerning their own affairs (Yeung, 2012, p. 137).

Access to SIS: There is a difficulty with implementing SIS within governmental projects, particularly if citizens use and benefit from them. The ethical concern of who has access to SIS and who should benefit from them raises concerns around equality, justice, and discrimination. The successful adoption and use of SIS by citizens strongly depend on how accessible, user-friendly, and integrative they are (Ryan, 2019).

Availability of Data: If there is a lack of available data repositories, it may impact the effectiveness of governmental SIS. If there are poor, limited, or inaccurate datasets in SIS' algorithms, it "may lead to discriminatory recommendations, inaccurate predictions, and harmful consequences" (Ryan, 2019).

Data Ownership: When different governmental bodies are using SIS, and often in partnerships with private organisations, the issue of data ownership is raised. If data is collected in a public space, then there is a tension if this should be the property of that district or municipality, or does the private organisations have a claim to it? One way around this problem is the clear articulation of data ownership in data partnerships with the private sector. It is also important to clearly articulate what responsibility the private organisation has within the SIS governmental project to ensure public data sovereignty and control over the running of their SIS (Ryan, 2019).

6.14. Law Enforcement & Justice

Smart policing systems and algorithmic-based predictive policing are expected to predict the location and sometimes the perpetrators of crimes to prevent crime and minimise the cost of policing. Facial recognition CCTV systems, now standard in China, proliferate, along with other means of surveillance, such as biometrics on public transport and Social Credit Systems that track bank records and buying patterns and other lifestyle choices. These may be used to surveil the public, raising a number of ethical issues.

Discriminatory use: This is not directly related to the AI technology per se, but to the prioritisation of AI's deployment in law enforcement. For example, its deployment against petty theft and crime is oriented to protecting society from crimes instigated by people in low socio-economic classes, while its deployment against tax evasion, fraud or corporate crime would be oriented to protecting society from crimes instigated by people in high socio-economic classes.

Human rights issues: Civil society organisations protest that predictive policing technologies are an affront to Europeans' fundamental rights. There is much debate within police ranks and others about whether when a police officer responds to an algorithm that has 80% predictive capabilities, they are infringing on a person's civil rights by treating them as a suspect on the basis of a statistical calculation, rather than observing anything to warrant suspicion.

Policing biases: AI data relies on crime and arrest data, which correlate suspicious behaviours to crimes that people have committed, and create some heuristic biases. In addition, those with an existing profile in a police database are much more likely to be identified as a future threat. This means predictive techniques are not able to detect first-time offenders, and also makes people with no record easy targets for exploitation by criminals. It is also ineffective to protect domestic abuse homicide victims, whose perpetrators often have no record. On the other hand, AI can be used to reduce discriminatory and embedded police practices,, such as discriminatory stop-and-search practices that led to arrests for spurious reasons.

State and corporate AI collaborations: Social media and ICT companies collect reams of personal data. This information, however, is rarely turned over to the police. There are increasing demands to share it with the police and/or intelligence agencies in order to tackle e.g. terrorism, paedophilia, and domestic abuse. Such questions are still open in national legislatures and a matter of current social debate.

AI explainability and social responsibility: Ethical issues have risen on policy agendas within law enforecement authorities (LEAs) themselves as well as in their oversight bodies. LEAs recognise that to improve trust with the public, they need to be more transparent about their priorities and how they operate. Similarly, progressive LEAs expect the AI systems they use to be explainable and not simply

black boxes. In other words, the AI systems used by LEAs should be capable of interrogation, should explain their purposes and whom to contact for more information.

6.15. Sustainable Development - Smart Cities

Conflict of Interests: Some claim that smart cities are top-down neoliberal ideologies that use SIS and algorithmic governance to prioritise vested interests within cities, instead of being citizen-focused (Cardullo and Kitchin, 2017; Kitchin, 2014; Kitchin, 2015b; Kitchin, 2016a; and Owen et al., 2013). Smart city infrastructure is devised by large SIS companies and civic partnerships but place a much greater focus on the 'smart', while neglecting the needs of the city (Foth, 2017; Galdon-Clavell, 2013, p. 718; Grey, Dyer, and Gleeson, 2017, p. 48; and Hollands, 2015).

SIS technical fixes are implemented, instead of tackling underlying political and social issues, with SIS corporations using cities as test-beds for their products (Kitchin, 2015a, p. 9). There is the possibility that SIS companies: place the interests of the city secondary, overlook the variability of cities, view cities as homogenous, see their SIS as being mutually compatible with all of them, while eradicating their diversity (Foth, 2017; Kitchin, 2015a, p. 9; Kitchin, 2016a; Kitchin, et al., 2017; O'Grady and O'Hare, 2012, p. 1581; Shelton, Zook, and Wiig, 2015).

Economic Pressure: The use of smart city SIS is still in early stages of development, so it is difficult to say if cities will economically benefit or become successful from investment in SIS (Kitchin, 2016b). Despite this, smart city SIS is being signalled as an innovative way to encourage investment in a city, as it is seen as smart, creative and forward-thinking (Kitchin, Lauriault, and McArdle, 2015, p. 25). Cities are feeling pressured to adapt, innovate, and integrate SIS, or face being left behind (Batty et al., 2012; Voda and Radu, 2018, p. 110)

Inequalities and the Digital Divide: There is a concern that SIS will replace many jobs within cities, such as customer service, taxis, bus drivers, factory workers, and delivery drivers (Capgemini Consulting, 2017; and Munoz and Naqvi, 2017, p. 7). Also, cities that use SIS require innovative and educated people, which may lead to a 'brain-drain' on rural areas. There is the possibility that SIS in smart cities will cause a digital divide between urban and rural, between neighbouring cities, and even between countries (Kitchin, 2015a, p. 9; and Kohli, 2014). SIS may create inequalities between those who can afford to use and implement them, and those who cannot (Chourabi et al., 2012, p. 2291; and Glasmeier and Christopherson, 2015, p. 10).

Privacy: There is a concern that SIS will infringe upon citizens' privacy by tracking their movements, scanning their bodies, and recording their conversations (Bartoli et al., 2011). SIS are being integrated in cars, homes, public spaces, workplaces, and even nighttime lighting in our cities. A major tension is ensuring that data is retrieved for the effective use of SIS, but that citizens' privacy is not breached at the same time. SIS holds the potential to single out voices among a crowd in a public place (Tung, 2018), lip-read what individuals are saying (Condliffe, 2016), or determine the movements, purchases, or activities of citizens (Kitchin, 2016c, p. 5), which may all create huge privacy concerns within the smart city of the future.

Accuracy of SIS and Bias: Within the development and use of smart city SIS, there is the possibility that there will be bias in SIS recommendations. There needs to be a distinctive process to identify, and account for, issues within SIS applications. Public organisations are often limited in their training datasets, which also limits the effectiveness of their smart city SIS. Problems could range from minor inconveniences such as chat-bots with glitches on municipality websites, to crash fatalities resulting from incorrect SIS traffic signalling. Public organisations need to identify the levels of potential harm caused by SIS and how to minimise these impacts, through internal remedial procedures, but also by conducting regular third-party audits of their SIS.

Availability and Accuracy of Data: In smart city partnerships, it may be difficult for public organisations to access certain datasets because private organisations do not want their intellectual property and competitive edge impacted. Furthermore, many organisations are fearful of the repercussions from GDPR legislation of exchanging data. However, data availability is fundamentally important for the success of smart city projects, so there is a tension between what is achievable and what is desirable. Some public bodies, such as Amsterdam municipality, are taking proactive measures to this concern by creating and developing their own datasets to ensure data sovereignty from private companies. Other public bodies create data partnerships with private companies that are mutually beneficial for both.

Transparency and Trust: In smart cities projects there needs to be a strong degree of trust between private organisations and public bodies and a clear division of responsibility and controls in place so that power asymmetries do not materialise. Both partners need to have strong degrees of transparency to ensure that neither is being misled in their development and use of SIS in cities. There also needs to be a degree of transparency for the citizen and how SIS will affect and impact their daily lives.

6.16. Defence & National Security

Many actors are engaged in cyberwarfare: governments, organised crime, terrorists and big companies engage in cyberattacks for a variety of reasons (Singer et. al., 2018). Foreign powers use cyberwarfare to disrupt critical infrastructure or other key state functions in other countries, as well as to discredit leaders and manipulate public opinion, with a view to strengthening their own power and the political regimes of their choice. This can be done covertly to mask their intentions, or explicitly as part of a military operation to save costs, or in the mix of their general warfare capabilities. In response, countries invest in cyberwarfare and digital forensics to better identify who is behind an attack. The fear of being overwhelmed by foreign powers, fear of defeat and fear of subjugation drive governments to invest in modern warfare. Fear of the unknown is a factor, reminiscent of the cold war era, as it is hard to estimate how many cyberwarriors the enemy has. Intelligence agencies,

especially, recognise the need for faster decision-making by the military and its new cyberrecruits. Information warfare, however, raises a number of moral issues.

Ethical principles: For most civilians, the investment in and use of AI in warfare is an anathema. Some civil society organisations and leftist politicians call for a strategic and moral re-allocation of national priorities from combating other countries to refocusing on the collective challenge facing civilisation from the ravages of climate change. Employees of the big five have pressured senior executives to disengage from selling AI technology to the military and install codes of ethics and codes of acceptable corporate practice. Such codes enable the companies to object to and even deny military contracts to guarantee their commercial success.

Autonomous decision-making: For many researchers, giving machines the decision over who lives and dies crosses a moral line (Sample, 2018). Despite scientists' opposition to the development or manufacturing of autonomous killer robots, we expect countries, including the US and those in Europe, to continue such developments under the guise that bad guys will be doing it and hence should be countered (ibid). Informed opinion is divided: some say information warfare requires instant decision-making that obviates the possibility of human intervention. Others say that some untoward events involving AI show that human intervention must always be possible. In any event, there is widespread agreement among stakeholders and the public that algorithms must be able to explain what they are doing and why, and be able to defer to a human for advice in case of a dilemma.

Collateral damage: The US and Israel developed Stuxnet specifically to target Iran's centrifuges, but an unintended consequence was the eventual release of the software into 'the wild', where it infected "thousands of computers across the world that had nothing to do with Iran or nuclear research" (Harwell, 2018). Hence, critics in the US and Europe have questioned the development of cyber weapons, especially those that could cause collateral damage or have unintended consequences.

Counterattack*:* For several years, there has been debate about when to retaliate against cyberattacks and who should do so. The US and European governments have warned companies and citizens not to take the law into their own hands. They should share any information about attacks they've suffered with others in their sector and, especially, with national cybersecurity centres, but this policy has not been an adequate response, in part because there are so many cyberattacks and because national cybersecurity centres are unable to defend companies and citizens against all such attacks. Warnings from national politicians about retaliation have been rebuffed. Hence, companies and governments have adopted a different policy, i.e., it is acceptable to retaliate in certain circumstances. Government officials and companies have set up working groups to debate under which circumstances, and how measured, retaliatory responses should be made against different types of attacks.

Cold war dynamics and loss of trust: With so many countries potentially engaged in cyberwarfare, trust between countries is the number one casualty. While foreign powers may deny any involvement, evidence to the contrary shatters their credibility. This can spur a new era of cold war dynamics between countries, with wider socio-economic implications. We anticipate an escalation of information warfare in everyday life which can potentially involve anyone using the Internet, either as a victim or a warrior. Decision-makers, from parents to parliamentarians, are confronted with ethical dilemmas. Should children and vulnerable people be advised to limit their use of the Internet to the absolute essentials? Should they be trained to recognise aggression and how to respond? How do we spot manipulation? Should we embed algorithms with morality – i.e., to do good and to shun evil – when questions inevitably arise about what is good and evil.

Uncertainty in determining responsibility: How should we act when we have limited certainty of who is likely responsible for a cyberattack? Uncertainty means we don't know who is responsible, who should be held accountable (if it is possible to hold someone responsible). Uncertainty provides space

for double standards and the invocation of different value systems. Uncertainty arises from the processing of flawed data.

6.16.1. Cybersecurity

Informed Consent: Informed consent is an important topic of concern within the cybersecurity industry (Johnson et al., 2012; Miller and Wertheimer, 2009); however, one of the major obstacles is the complexity of information when informing individuals (Burnett and Feamster, 2015): "One cannot speak about informed consent if one gives too little information, but one cannot speak about informed consent either if one gives too much. Indeed, giving too much information might lead to uninformed dissent, as distrust is invited by superfluous information" (Pieters, 2011, p. 61).

Privacy: Privacy is a core concern in cybersecurity in order to prevent attacks on users' safety; however, as a result it may also infringe on users' privacy itself. Cybersecurity systems need to understand and detect 'typical' from 'atypical' behaviour of the user to identify when abnormal behaviour, and potential attacks, are occurring. Understanding typical behaviour requires a level of monitoring the user and their personal profile, which may cause infringements on their privacy.

Protection from Harm: Cybersecurity may cause harm to individuals, through the disclosure of vulnerabilities, which would potentially allow hackers to breach flaws in users' security systems. If these vulnerabilities are not disclosed, then it may also put users in jeopardy, for example, a fault in medical devices being used by the patient (Nichols, 2016; Spring, 2016).

Control of Data: If a cyberattack is successful, then the control of that data is lost: "data dumping, in which research is carried out in countries with lower barriers for use of personal data, rather than jump through bureaucratic hurdles in Europe. The result is that the data of non-European citizens is placed at higher risk than that of Europeans" (Macnish and van der Ham, 2019, p. 8).

Competence of Research Ethics Committees: There have been demonstrable cases where RECs have allowed cybersecurity tests on non-consenting individuals because of their lack of expertise in understanding the risks posed by these technologies (Burnett and Feamster, 2015). For example, there was a case where two RECs allowed the study of testing firewalls in totalitarian states, because they did not view it as ethically relevant. However, the IP addresses used could have easily been linked to the individuals using them, putting those people at risk (Macnish and van der Ham, 2019).

Vulnerabilities and Disclosure: Identifying and addressing vulnerabilities in cybersecurity is an important task, but if these vulnerabilities are leaked to hackers, it may have harmful effects on those systems (Macnish and van der Ham, 2019, p. 9). For example, if there are vulnerabilities in e-voting systems, they will not be trusted during the election, and afterwards the results may be questioned (Pieters, 2011). But if the vulnerability is not revealed, the vulnerability may actually cause the election to be compromised.

Transparency: There is a concern about transparency in the cybersecurity industry, because if they are too transparent, and to the wrong people, it may open the possibility of attack. Therefore, "how far to push transparency: should it extend to government agencies or even other companies? On one hand sharing information increases vulnerability as one's defences are known, and one's experience of attacks shared, but on the other it is arguably only by pooling experience that an effective defence can be mounted" (Macnish and van der Ham, 2019, p. 14).

Trust: Trust often stems from how transparent that system is, showing how secure a system is and being able to describe why it is secure (Glass et al., 2008). This needs to be explainable by the designer, where black boxes provide unease among SIS users (Bederson et al., 2003; and Pieters, 2011). "Explanation-for-trust is explanation of how a system works, by revealing details of its internal operations. Explanation-for-confidence is explanation that makes the user feel comfortable in using

the system, by providing information on its external communications. In explanation-for-trust, the black box of the system is opened; in explanation-for-confidence, it is not" (Pieters, 2011, p. 57).

Risk: In cybersecurity, it is very important to effectively determine possible risks, acceptable risks, and ways to calculate them (Hansson, 2013; see also Wolff, 2010). There may be varying risks to users from cybersecurity threats, depending on where they live, the group they belong to in society, and the culture they belong to (Byers, 2015; Macnish and van der Ham, 2019).

Responsibility: It is unclear who should be held responsible for protecting against cyberattacks, if it should solely be cybersecurity companies or if governments should assist, as well (Guiora, 2017, pp. 89–111). Cyberattacks may create national insecurities, so there is the question of a state's "responsibility for protecting its own economy on the internet as it does in physical space, by providing safe places to trade" (Macnish, van der Ham, 2018, p.14). Also, cyberattacks typically refer to attacks from outside entities attacking one's systems. However, as the boundaries of actors expands, it becomes unclear who is inside these boundaries: "mobile devices [that] can access data from anywhere, and smart buildings [which] are being equipped with microchips that constantly communicate with each other" (Cleeff et al., 2009, p. 50).

Security Issues: Security is obviously an important ethical concern for the cybersecurity industry: "Insufficient funding, poor oversight of systems, late or no installation of "patches" (fixes to security flaws), how and where data are stored, how those data are accessed, and poor training of staff in security awareness" (Macnish, van der Ham, 2018, p.11-12).

Business Interests and Codes of Conduct: Sometimes there is a perceived conflict between business interests and alerting users of a security threat or flaw in their system. Marissa Meier, then CEO of Yahoo, did not inform users of the cyberattacks in 2013 and 2014 because it would have led to a massive loss of profit (Stone, 2017). Therefore, "public-spirited motivations should be protected from predatory practices by companies seeking to paper over cracks in their own security through legal action. However, current conventions as to how to proceed with disclosure of vulnerabilities seem to be skewed in favour of corporations and against the interests of the public" (Macnish and van der Ham, 2019, p. 9).

Anomalies: The issue of anomalies may have a striking impact on the cybersecurity of SIS, such as fake base stations. There is a discrepancy among nations in how they deal with these, as well. The U.S. has been trying to prevent these fake base stations because of the damage of trust that people place in networks, while in China they are prolific, and in France they are actually used by the police.

7. Ethical issues of SIS in Research & Innovation

This section provides an overview of ethical issues and principles arising from the use of AI in various types of scientific research and technological innovation relating to SIS. This is an important element in the call and in the DoA. It will consider ethical issues that researchers and innovators in different fields that use SIS may encounter in their work. It also examines ethical issues in the development of SIS in both the public and private sectors, which is a distinct issue.

7.1 Uses of SIS in Responsible Research and Innovation

To understand the implications of using SIS in scientific research and technological innovation, we need to examine the use of advanced analytics on Big Data in understanding and theorizing on various types of phenomena. SIS utilise Big Data analytics techniques to cluster as descriptive, diagnostic, predictive, prescriptive analytics, each of which plays a different role in observation of reality and hypothesis development, data collection and data analysis about a phenomenon, systematisation of knowledge by establishing logical relations among previously disconnected facts and hypotheses, and explanation of events and hypotheses through systematic theory development. The key difference between doing analytics with other IT systems and AI is not it's computational power or even its autonomy to undertake confirmatory research driven by researcher hypotheses. AI can play a more transformative role by introducing a new "method of invention" (Cockburn, et. al, 2017). DAPRA for example has invested in "Automating Scientific Knowledge Extraction (ASKE)" a program that aims to develop approaches and tools that automate scientific knowledge discovery, expose assumptions in pre-existing models and documentation, identify new data and infer useful information, integrate information into machine-derived models, and run these models for the purpose of deriving scientific results (Elion, 2019). The role of AI in responsible research and innovation (RRI) can be conceptualised in three ways: as an automation tool, as a generalised technology, and as a substitute for human subjects. More autonomous processes of examining data to discover deeper insights, make predictions or generate recommendations require the use of AI to cope with its volume and complexity. In 2003 researchers in Ross King at Aberystwyth University in Wales, created Adam, a robot able to independently undertake genomic research - from hypothesis, to experiment, to reformulated hypothesis - on the behaviour of yeast; "armed a model of yeast metabolism and a database of genes and proteins involved in metabolism in other species."(Mosaic, 2018).

7.2 Current Uses of SIS in Research and Innovation

To organise our insights we discuss the contribution of SIS to the scientific process, which can be broadly described as follows:

1. Make an observation.
2. Ask a research question or make a hypothesis.
3. Research prior literature/research/stats
4. Form a hypothesis or testable explanation.
5. Make a prediction based on the hypothesis.
6. Test the prediction.
7. Analyse results
8. Interpret the results
9. Synthesise findings/reflections
10. Write up of publication/report/etc.
11. Publication of publication/report/etc.

7.2.1 Making Observations

With respect to making observations, descriptive analytics on Big Data focus on classification and clustering of data and their visualisation, which can allow humans to visualise a holistic picture based on relationships predetermined by humans. This reduces complexity, allowing humans to comprehend the big picture and develop hypotheses. Deep learning can undertake exploratory research by analysing Big Data (i.e., multiple, large datasets of static and real time data) to generate sense-making about a phenomenon. Recognising and extracting patterns from observation and analysis of Big Data datasets can lead to the generation of new hypotheses, leading to alternative theories about how constructs and phenomena relate and interdisciplinarity. Algorithms for Big Data analytics have been used to extensively in life science informatics, in particular with respect to genomic research. By using algorithms, for example, scientists analyse large samples of genomes and then use an algorithm to compare how frequently a certain DNA variant appears in people with a certain trait or condition, and people without it, to generate hypothesis about a possible cause. Often tens or hundreds of parameters of such variants are flagged, requiring the comparison at any single time, making the use of AI and bioinformatics necessary. This has implications for exploring degenerative diseases such as cancer and autoimmune conditions, or even the impact of environmental factors on epigenetics (Mosaic, 2018).

7.2.2 Ask a Research Question

While formulating broad research questions remains the prerogative of researchers, diagnostic analytics techniques can be used for data mining to discover data and correlations that can identify relationships between constructs and flag unexpected outliers that can raise researchers' insight into how to further specify a research question, or even a research hypothesis. Depending on the complexity of data, algorithms can generate many and often inappropriate hypotheses that need to be scrutinized to reduce not only the number of considered hypothesis but also their value in producing useful results. Nowadays, that value is effectively driven by the motivation of the researcher or the information stakeholder, i.e. those whose needs for the derived knowledge or insight the research aims to satisfy. For example, health data can be analysed to satisfy the information needs of doctors and/or patients (Shillabeer and Roddick, 2006). Alternatively, we may allow AI to test all probable hypotheses as a means of engineering machine-learning creativity and "computational serendipity". Once, and if, such hypotheses are verified we can look further into post-rationalising research outcomes (Kitano, 2016). While not yet a reality, a future can be envisioned where such hypotheses will be derived by a machine in the quest for problem solving.

7.2.3 Research Prior Literature/Research/Statistics

SIS can increase the efficiency of R&I by automating the more routinised, yet labour-intensive research processes, such as literature search, data collections and clustering. Reading, understanding, developing insights and synthesising prior literature, as well as flagging gaps, inconsistencies and omissions remains still within the sphere of human interpretation. Companies, such as Iris.ai, work towards providing AI-powered literature search and clustering tools. Papers however can contain data errors, missing information, and even exaggerations. Yet, humans are still able to surpass such challenges to make scientific progress; an ability machines will need to replicate to synthesise prior research.

7.2.4 Form a Hypothesis or Testable Explanation

Diagnostic analytics can help us develop alternative conceptual models that may provide potential explanations or alternative theoretical models of a phenomenon. While in science a hypothesis is "a provisional explanation for observations that is falsifiable." or an "explanation about the relationship

between data populations that is interpreted probabilistically", in AI, "hypothesis is a candidate model that approximates a target function for mapping inputs to outputs." These models can contain not one but several complementary hypotheses that can be concurrently tested to explain a phenomenon, depending how the phenomenon was framed and how the model was configured for testing (Brownee, 2019).

7.2.5 Make a Prediction Based on the Hypothesis

Use of predictive analytics can have applications in many research domains by research models that test a hypothesis. In astrophysics, for example, Generative Adversarial Networks (GAN) are used to predict what changes galaxies go through when they shift from low- to high-density regions, and what parameters are involved in the process (Falk, 2019). Generative models can even have applications in social sciences. As early as 2007, X advocated the use of Agent-Based Computational Modeling in Social Science, to make predictions about social behaviour by emulating social conditions (Bainbredgem, 2007, Epstein, 2007). Yet, real life applications in this domain are still in their infancy.

7.2.6 Test the Prediction and Analyse the Results

Predictive analytics are future-looking and can undertake the collection and analysis of data related instantaneously. To test predictions, one should develop and execute a research protocol for data collection and data analysis. This can be intimidating for most people without technical expertise, although IBM's suite of AI solutions, Watson OpenScale, Watson Studio, and Watchson Machine seek to bridge this gap by creating interfacing solutions that enable interaction between non-technical domain experts with AI technologies. Other solutions such as the Wings workflow system help researchers translate their conceptual model into executable AI workflows, that can be executed by execution Machine learning engines such as such as OODT and Pegasus. Execution engines perform rapid analysis of data against previously developed models to reduce complexity and produce insights about future scenarios in relation to a phenomenon. Researchers can even instruct algorithms to manipulate the model parameters to identify the best fit, leading to a revision of the initial model, if desirable. Because of this, predictive analytics can be used as decision support systems for decision-makers. Machine learning is often used to tweak and optimize the initial model based on learning that takes place as more and more data is analysed. In theory, using AI could resolve some global disagreements on issues of paramount importance, by minimising professional and disciplinary disagreements.

7.2.6 Interpret the Results

AI in the sense of an autonomous system which can make its own decisions can be used not only to learn about, develop value judgments about, and develop expectations about a phenomenon, but also interpret and make decisions based on such knowledge. IBM's Watson OpenSearch suite of tools claims to be able to help researchers to do so, though the availability of large quality datasets is currently a bottleneck. or hypothesis testing, thus enabling researchers to concentrate on more creative parts of research, such as hypothesis generation, model development, interpretation and synthesis.

The process of reflection on the findings and the resulting synthesis of findings remains a human task, as is the write up of publications, reports and their publication. For reflection to take place, sentient machines would need to be developed, which requires the development of general artificial intelligence; an innovation on the far distant horizon.

7.3 Ethical Issues Arising from the Use of SIS for Responsible Research and Innovation

The primary source of ethical problems with SIS is their opacity, or at least humans' current inability to understand their 'way of thinking', their choices and the decision criteria that led to them. Related to this is the fact that, as designers, we have not made progress towards developing the technology's moral code for it to distinguish between moral and non-moral decisions and understand the social consequences of their actions. In addition, we have not developed emotional intelligence for SIS to empathise with humans and establish a sense of the psychological harm their actions can cause at the individual and social level, which can have deleterious effects (Hiroaki, 2016).

Another ethical issue comes from the removal of plurality from the scientific process. AI tools can be used directly by decision-makers for eliciting scientific hypotheses and knowledge, without the need to consult a scientific team.

Roberts et al. (1991) made a critical distinction between purpose or agenda-driven decision support, and scientific knowledge creation. In science generation, created knowledge must be scrutinised for consistency and completeness, and through this additional problems and knowledge gaps are generated, leading to a holistic understanding. Hence, while this may lead to quick, evidence-based decisions, it also removes voices of dissent or even objection. This can be useful but also dangerous, as decision-makers are often under pressure to act, that often leads to compromises and errors of judgment.

In her keynote speech during The Digital Ethics Summit held in December 2017, Elizabeth Denham, Information Commissioner, the UK's Information Commissioner's Office, suggested that "there's no dichotomy between ethics and innovation. But ethical considerations should dictate the direction of travel.

What does this mean in practice? What particular issues does the use of SIS in Research and Innovation raise for consideration?

To better understand the ethical implications of SIS in Research and Innovation, we need first to understand the ethical principles for research and innovation and their sources (European Committee For Standardization, 2017).

- **Professional principles and codes of conduct.** These are ethical principles that specifically concern the behaviour and practices of individual researchers and innovators and the way they treat others. Assessment of this behaviour is not normally the responsibility of ethics committees, but rather is the responsibility of research integrity boards, professional ethics boards or disciplinary committees.

- **Ethical guidelines for institutional responsibility and integrity.** These are ethical principles that concern the way in which the institutional setting for research and innovation ought to be constructed so as to support ethically sound research and innovation practices. These principles are not normally applied by ethics committees, although ethics committees sometimes address them in their work.

- **Ethical guidelines for the conduct of research and innovation.** These are ethical principles for the assessment of plans, procedures, and practices in research and innovation. This latter category of principles is normally considered by ethics committees and is therefore central to their functioning as ethics committees.

To organize our findings, we use the framework of ethical principles and issues developed by the SATORI project (Jansen, et al., 2017). The framework was based on several general sets of principles for the ethical conduct of research and innovation, such as the Singapore Statement on Research Integrity (2010), and the European Code of Conduct for Research Integrity (2011).

General Ethical principles: Overall, the resulting framework consists of 8 main ethical principles that are applicable to all (or most) fields of research and innovation (Shampoo and Resnik, 2015).

(1) Research integrity;

(2) Social responsibility;

(3) Protection of and respect for human research participants;

(4) Protection of and respect for animals used in research;

(5) Protection and management of data;

(6) Dissemination of research results;

(7) Protection of researchers and research environment;

(8) Avoidance of and openness about potential conflicts of interest.

In this analysis, we have excluded from consideration principle (4) due to its similarity to principle (3) in relation to the impact of AI; and principles (7) and (8) as these relate explicitly to the wellbeing and motivational interests of people.

7.4 The Impact of SIS in RRI Ethical Principles

7.4.1 SIS and Research Integrity

The principle of research integrity suggests that researchers should carefully select and declare their sources of knowledge, research methods and biases and make them available to the public and fellow researchers. To this end, it is the responsibility of research institutions to ensure that research and innovation takes place in a fair and accountable way.

- Ensure careful and honest presentation of data and research findings.
- Practice universalism (hold research to equal standards, regardless of where and by whom it was performed) and disinterestedness.
- Ensure that institutions act according to their purpose, in a transparent and accountable way.

Like most cybernetic systems, AI relies on sound models of the relationship between information inputs to produce information outputs. Hence, when a machine learning model is designed or trained poorly, or used incorrectly, flaws may arise. Common flaws can be broken into three categories - incorrect design decisions, deficiencies in training data, and incorrect utilization choices. Flaws arising from design decisions are related to designing the system on an inaccurate theoretical and research model. Hence, biases can stem from choosing constructs (features) that have little or no effect on a phenomenon, and linking them in does not produce sound generalizable results.

It is worth highlighting that these are not biases produced by the technology per se, but perpetrated by it. While in service delivery, AI models and systems can be trained and/or tested on a test set, a sample of data where expected outputs are known, in the case of research where outputs are to be

explored and determined, such cross-checks and balances can be impossible to apply. Hence, latent algorithmic biases that researchers are not aware of can go unnoticed. In addition, as humans often cannot understand the final connections and choices made by AI, it is difficult to judge its academic rigour. Hence, AI transparency and explainability that can 'translate' for humans the research assumptions, and the inference processes for making decisions and producing insights, is paramount to ensure the careful and honest production and presentation of research findings, and for ensuring that research institutions act in a transparent and accountable way. Iris.ai, using AI to systematize and semi-automate literature search and clustering, can have a positive effect on the principle of universalism, by bringing to the fore relevant research from less considered publications or authors, which is often excluded by other researchers for the sake of efficiency rather than merit.

7.4.2 SIS and Social Responsibility

The principle of social responsibility highlights researchers' obligations towards society and their role in guarding social justice in accordance to what is morally right and proper. It points to the need for researchers to be aware, raise awareness and mitigate against any negative societal impacts from their work on the rights and welfare of the individuals and communities involved. They also need to ensure research is responsive to the needs and desires of those involved or to be impacted by the outcomes of their research.

Social responsibility issues arising from the poor use of AI in research relate to epistemic concerns around inconclusive evidence; inscrutable evidence; misguided evidence; and normative effects around unfair outcomes; transformative effects; and traceability. (Brent Mittelstadt et al., 2016). Much like outputs of other statistical models, AI-driven insights are rarely meant to explain a phenomenon, but to indicate correlations between constructs (variables) that comprise it. Even their comprehensiveness is limited to the constructs a designer has included in the machine learning model, while other variables not currently entertained may potentially play a significant role. Mis-interpretation and misuse of quantitatively driven research is not exclusive to AI and is not a problem with the appropriation of the technology as a research tool per se. The air of objectivity that surrounds AI-driven research and lack of reporting about the limitations of AI-driven studies and their suitability to inform recommendations for action is a systemic issue that needs to be addressed, especially when AI research informs public policy, or social and economic practices with implications for social justice. For example, DNA profiling and identified correlations of a gene to a type of disease may lead to lifestyle recommendations, adherence to which may determine access to free healthcare.

Issues of prejudice in AI research can impact on the rights and welfare of the individuals and communities involved. AI systems rely on easy-to-measure proxies based on assumptions about their appropriateness to represent a phenomenon. For example, a face recognition system is not able to distinguish waiting from checking out an area with the intent to steal a car, as it is only capable of tracking movement and duration of time a person remains within a restricted radius. Such biases are possible within any kind of research and it is standard practice amongst researchers to flag potential biases when reporting on their studies. AI may make it difficult for researchers to reflect on latent algorithmic biases and the extent of their impact, as AI can perpetuate and accentuate mis-profiling of people. There are, of course, issues of generalizability. For example, the SUBITO (Surveillance of Unattended Baggage and Identification and Tracking of its Owner) project, generalized findings about people walking together at an international scale by training its algorithm on university students at one British university (Macnish, 2012), who have distinct patterns of behaviour. This may partly stem from the fact that access and ability to utilize AI for research is not equal. Social scientists do not tend to have AI skills, hence AI is used by tech people who make inroads into Life and Social Sciences, setting agendas and carrying their own biases and viewpoints with them. For example, technological feasibility tends to take precedence over domain rigor, be it in humanities (sociology, anthropology, psychology), life sciences or other. Hence a multidisciplinary approach to AI model developments and

testing may be required. Some companies (Alphabet, Microsoft, etc.), technological universities (see, for example, MIT, Oxford) have established ethics departments, and along with dedicated institutes have develop as guardians of the ethical use of AI in society. There are also independent organisations, such as the Responsible AI institute, that advocate responsible research and innovation.

7.4.3 SIS and the Protection of and Respect for Human Research Participants

Researchers respect the autonomy and dignity of research participants and those impacted by research and innovation, irrespective of gender, cultural, ethnic, and geographic identities. To this end, researchers should:

- Ensure that research participants are provided with adequate information about the research, including its purpose, its funder(s), who will use its results, the consequences for them of participation in it, and policies regarding privacy and confidentiality;
- Obtain consent from research participants that is informed, given freely, and provided in an explicit form (informed consent);
- Treat human participants with due consideration for their dignity, autonomy and personal integrity;
- Ensure that research participants are not exposed to serious physical or psychological harm or strain as a result of the research;
- Ensure that any risks or burdens to research participants are balanced by benefits to the participants or to society; Ensure that the privacy of research participants is protected and that identifiable information about them is kept confidential;
- Respect cultural diversity and pluralism, meaning that the cultural background, values and viewpoints of research participants are respected, as well as the cultural values and norms that apply in research settings;
- Ensure that one's pool of human research participants adequately represents society or the social group being investigated, with respect to categories such as gender, age, race, ethnicity, social class, religion, culture and disability; or discuss and, where possible, compensate for limitations in one's selection.

AI issues arise from the fact that most Big Data research is based on secondary data, i.e. on the meta-analysis of already collected data or on web crawling of public personal data (e.g. social media data). When publishing their data, most individuals were under the assumption and understanding that such analytical capabilities were not in existence, or even possible. Nobody was ever asked, or will ever be asked, for permission to use such data; how any resulting knowledge will or can be used is not explicit, and there is no means to find out, particularly when it comes to generating/augmenting commercial products/services. Hence, no researcher has attempted to ensure that research participants are provided with adequate information about the research, including its purpose, its funder(s), who will use its results, the consequences for them of participation in it, and policies regarding privacy and confidentiality. In addition, no processes or standards of obtaining informed consent for such research exists. Hence, research participants are often not informed adequately and explicitly in order to opt-in to such research. In addition, IoT has made it possible for covert monitoring to take place. For example, real time satellite data of logistics routes of major FTSE organisations has been used to draw inferences about their near future financial performances and predict the FTSE index trends. This can be done covertly without permission from participants, as open satellite data can be used to do so. The inferencing patterns of business people with access to AI analysis and the societal implications for discriminatory pricing or exclusion from services is not well understood, or even considered to be within the remit of most researchers' responsibility.

Treating human participants with due consideration for their dignity, autonomy and personal integrity is precarious in AI research due to potential violations of privacy, especially when personal opinions

and data are monitored, as in the case of natural language processing (NLP) (semantic and sentiment analysis) in social media research.

Algorithmic biases with regards to cultural diversity and pluralism have been discussed above. These, however, are not endemic to the technology per se, but rather to inappropriate sampling of training algorithms and overgeneralization of findings without regard for deployment of AI. Another related issue is the misattribution of AI correlations to causality, hence drawing inaccurate inferences about the cultural practices, values or viewpoints of research participants.

7.4.4 SIS and the Protection and Management of Data and Dissemination of Research Results

Researchers should:

- Store all research data securely, and render them difficult to access or hard to use for unwanted third parties;
- Be aware of all actual and potential data flows;
- Ensure that all personal data that researchers plan to collect are necessary for the research;
- Obtain informed consent from research participants for the collection and use of their personal data, or verify that such consent has been given;
- Ensure that data related to identifiable participants are stored securely, and that such data are not stored any longer than is necessary to achieve the objective for which they were collected;
- Ensure that, for any secondary use of data, the data in question are openly and publicly accessible or that consent for secondary use has been obtained;
- Consider and anticipate the effects that gaining access to personal information could have on third parties (e.g., persons related to the data subject).
- Consider whether publicly available information should actually be considered sensitive personal information and treated as such;
- Take precautions when merging multiple data sources to ensure that anonymity and or pseudonymity are maintained;
- Inform participants in open online forums about systematic registration or reporting of information when possible;
- Only conduct research with appropriate ethical approvals in place.

To date, Big Data analytics is basically a meta-analytics methodology performed on secondary data, i.e. data originally collected for a different purpose that is now re-analysed to test new hypotheses. Researchers who use AI for Big Data analytics, in particular, usually work on existing large datasets (often anonymized or pseudo anonymised), or publicly available material. As such, they do not have to reveal their identities to participants. Stringent informed consent procedures are more prevalent across academia but have always been common practice to specify the lawful basis for data processing, on the basis of it being a 'task in the public interest' and 'necessary for scientific research in accordance with safeguards'. Often Big Data analytics do not rely on downloading and storing big databases but accessing and querying big databases that reside elsewhere.

7.4.5 SIS and Dissemination of Research Results

Researchers are expected to openly disseminate research findings in order to benefit society and ensure constructive dialogue with and scrutiny by fellow researchers, stakeholders and the public, unless there are compelling reasons for not doing so. Researchers should:

- Wherever possible, strive towards open access publications, which provide free online access to any user;

- Where possible, make research results available to different audiences that may have an interest in them, using different formats and media. Aim to include the general public, if results may be of interest to them, and aim to include regions that are otherwise excluded for reasons of economic disadvantage.

A key issue of using AI for scientific discovery is the need to make research results discoverable by AI. Hence, AI or other forms of Big Data analytics become a new category of 'user' or 'audience', and perhaps new types of 'formats and media' should be used to make them more accessible to the technology.

8. Main Ethical Issues and Possible Solutions

This section has three objectives:

(1) To summarize key moral values and issues that have been presented in this study
(2) To identify possible ethical tensions in the application of the ethical principles identified under the first objective
(3) To identify possible mitigating actions that can be taken to reduce ethical tensions

These objectives will be taken up in the following three sections that correspond with them.

8.1 Key Moral Values and Issues

In this section, we will identify the main moral values for smart information systems. A moral value is an idea that is expressive about what is right and wrong. It abstracts from specific things, situations or events, to express a general quality of goodness or rightness. Examples of moral values are "justice", "freedom", "privacy", "integrity" and "dignity". They are to be distinguished from moral norms or principles, which are prescriptive statemens, often derived from moral values, that identifying standards to be adhered to or actions to be carried out, for example "Personal information should not be collected or distributed without the bearer's informed consent".. In this study, we do no endeavor to formulate moral norms or principles, as this is something that will be taken up in later works. Formulating moral norms and principles is also more difficult and more controversial than identifying relevant moral values or moral issues, since it requires a greater degree of agreement. For example, people can agree more easily that privacy is important, and that there are privacy issues with internet use, than that they agree on specific privacy norms.

Let us now turn to the question of how we can identify the main moral values and issues for smart information systems. The major input for this comes from this study, which is based on a review of the academic ethics literature on artificial intelligence and big data, and on case studies and scenario studies of ethical issues in AI and big data. However, this study references dozens of moral values, and even more ethical issues. We want to have an idea of those moral values and issues that are most important or fundamental. How do we get this?

There are several methods by which this may be established. First, ethical analysis can reveal certain moral values to be more fundamental than others by identifying hierarchical relationships. For

example, accountability is a more fundamental moral value than transparency of decisions, because transparency is normally analyzed as a condition for accountability. Second, when two moral values are not hierarchically related, there may be standards or agreements, in ethical theory, in policy, or in stakeholder consultation, that certain moral values have greater moral weight than others. For example, people may decide that for them, security normally outweighs privacy. Third, we may appeal to the recurrence or broad applicability of moral values and moral issues to establish their importance. If a moral issue only plays out in law enforcement and defense, for example, then it is presumably less important or fundamental than one that also plays out in healthcare, education, government and other domains.

Building on these considerations, we will proceed as follows. First, we will establish which are the most important values identified in this study. We will do so using the well-established method of reflective equilibrium (Rawls, 1971), which is a method frequently used in ethics for finding coherence among a set of beliefs through deliberative mutual adjustment among general moral principles, moral intuitions about particular cases, and theoretical considerations about these principles and intuitions. In this case, we will use the method to mutually adjust judgments about the values that are relevant for the ethical assessment and guidance of smart Information Systems. We will use the following sources for this:

1. *Key moral values enshrined in the Charter of Fundamental Rights of the European Union* (ref.). The values enumerated in the Charter also have a legal status, but many are recognizable as moral values as well.
2. *Key moral values recognized in philosophical ethics* (as evidenced in philosophical ethics publications, including ethics handbooks and introductions, e.g., Singer, 1993).[5]
3. *Key moral values established in earlier guideline documents for Smart Information Systems*, specifically the following three documents:
    a. the *Ethics Guidelines for Trustworthy AI* of the High-Level Expert Group on Artificial Intelligence (HLEG) (2019);
    b. the *Recommendation of the Council on Artificial Intelligence* of the OECD (2019) and
    c. *Ethically Aligned Design*, a publication of the Institute of Electrical and Electronics Engineers (IEEE) on the ethical development of intelligent and autonomous systems (The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 2019).

These three documents were selected because they are documents that result from extensive deliberations amongst a diversity of experts and stakeholders and that are adopted internationally across a large number of countries.[6]

4. *Key moral values referenced in this study.* We will map and analyze references to moral values in the ethical analyses that were performed in sections 4 through 7. For section 4 (general ethical issues), we map the 24 moral values and related moral issues discussed in it. For section 5, we will map significantly referenced moral values in each part (i.e., moral values that are explicitly proposed as relevant). For section 6, we will map the

[5] When we refer to philosophical ethics, we mean Western philosophical ethics. The values and principles in nonwestern traditions are often quite different. This being a report for the European Commission, our focus is on Western ethics.

[6] These three documents have a focus on AI, and are less concerned with big data. We did not find there to be ethical guidelines documents for big data that with the same level of sophistication and international support as exist for AI. However, we will make reference to some guideline documents on big data later on as supporting evidence.

moral values and associated moral issues as summarized in the table at the beginning of this section. For section 7, finally, we will also map the significantly referenced moral values in the section.

In what follows, we will present and discuss the result of this analysis. The values listed below are (a) recurring across many or all of the sources mentioned; (b) held to be key values in these sources, that moreover deserve to stand on their own rather than be subsumed under other moral values. In some cases, the values are not explicitly mentioned in the sources but can be shown to be implicit. Some values are proposed as important even though they are only referenced in a few of the sources. Explicit justifications are given when this is the case.

We will also discuss proposals for values and principles that do not belong to the realm of ethics but that have been proposed to be relevant. In ethical guidelines, reference is sometimes made to values and principles that are arguably not ethical, but that relate to social desirability (e.g., "social cohesion", "economic growth"), professional desirability (e.g., "objectivity", "efficiency"), or that are believed to be instrumental to the realization of moral ends (e.g., "transparency"). The reason that we will do so is that such values are sometimes referred to in ethics guidelines, and we want to examine the justifications for doing so.

8.1.1 Individual rights

Individual rights correspond with values such as "freedom", "dignity" and "privacy". They have major importance in philosophical ethics and are also central in the Charter of Fundamental Rights of the EU. They also have an important place in the three SIS guideline documents. The HLEG makes extensive reference to the rights enumerated in the EU Charter, and identifies many of them as relevant for providing a foundation for its ethical guidelines. The OECD guidelines assign an important role for "human-centred values". The values mentioned in the OECD document correspond strongly with the fundamental rights found in the EU Charter. The IEEE guidelines list "human rights" as amongst their key principles. The document does not enumerate these rights, but rather it requires adherence to international conventions, including the Universal Declaration of Human Rights, the Geneva Convention, and others. In our study, finally, there is frequent and extensive reference to different kinds of individual rights, across sections 4-7. It can be concluded that a strong case exists for inclusion of individual rights in a listing of moral values that should guide the development and use of Smart Information Systems. There are, however, many individual rights, some well-established and some more controversial. We need to consider more explicitly which individual rights are most relevant for SIS. This is to which we will now turn.

(1.1) **Autonomy**. Autonomy is self-governance or self-determination. It is the ability to have one's own thoughts and to construct one's own goals and values, and the freedom to make one's own decisions and to perform actions based on them (cf. Dworkin, 1988). Autonomy is one of the fundamental moral values recognized in philosophical ethics. It is a key value in the HLEG guidelines and is also amongst the "human-centered values" enumerated in the OECD document. The IEEE document also makes extensive reference to "human autonomy" as a guiding principle. The EU Charter does not explicitly refer to autonomy, but as the HLEG points out, it refers to strongly related values and principles. In particular, it refers to the right to freedom of thought, conscience and religion (art. 10) and the right to integrity of the person, including the right to respect for mental integrity (art. 3), and it also references a general right to liberty that could be interpreted to include freedom to make one's own decisions (art. 6). Our study, finally,

has identified autonomy as an important moral value that is referenced in all each of the four sections 4-7.

Autonomy is especially relevant in relation to AI because AI can undermine it by processing information and making decisions in place of humans and by choosing goals and values for them. Also, the non-transparency of many AI systems means that humans are often not in a situation to reassert control. We can conclude that the different sources provide a strong argument for including autonomy as one of the core moral values for SIS.

(1.2) **Privacy.** Privacy is a key value in philosophical ethics, and is paramount in ethical studies of information technology. The EU Charter recognizes a right to privacy in two principles: respect for private and family life, home and communications (art. 7) and protection of personal data (art. 8). The HLEG lists "privacy and data governance" as one of its seven requirements for trustworthy (i.e., ethical) AI. The OECD guidelines recognize "privacy and data protection" as among the "human centered values" that must be upheld and make frequent references to privacy elsewhere in the document. The IEEE guidelines do not have privacy as one of its key principles, and instead refer to a principle of "data agency", which refers to control of individuals over their personal data. This is similar to a right to privacy, and the right to privacy is extensively referenced in other parts of the document. In our study, we also frequently found reference to privacy, and in fact, it is more frequently referenced than other values in our study. In section 4, it comes to play in our sections on privacy, surveillance and use of personal data, it is frequently referenced in sections 5 and 7, and it is referenced in almost all studies of application domains in section 6. Altogether, these sources provide a strong argument for including privacy as one of the core values for SIS.

(1.3) **Freedom.** Freedom (or liberty) is a key value in philosophical ethics (Mill, 1859, Berlin, 1969), and is frequently referenced in ethical studies of information technology, particularly in reference to freedom of expression and information. The EU Charter devotes a chapter to "freedoms" and lists 14 articles enumerating various kinds of freedoms, such as freedom of expression and information, freedom of assembly and association, and the right to liberty and security of person.[7] The HLEG lists "freedom of the individual" as one of the fundamental rights at the basis for trustworthy AI. It also references freedom in its principle of autonomy for AI, which appears broader in its description than just autonomy, to also include other freedom rights. The OECD guidelines recognize "freedom" as among the "human centered values" that must be upheld. The IEEE guidelines do not have freedom or liberty as one of its key principles, but make frequent reference to "freedom" and "freedoms" in its document, and also adheres to freedom rights by adhering to international conventions on human rights. In our study, we frequently reference freedom in sections 4 and 5, but only once in section 6 and not at all in section 7.

Isaiah Berlin (1969) has argued in an influential paper that freedom or liberty has two dimensions. Negative liberty is the ability to act without obstruction or interference by others. Positive liberty is to be self-determined: the ability to be one's own master, having one's own thoughts and making one's own decisions. This type of freedom is often associated with autonomy. Positive freedom involves control over the environment, while negative freedom involves self-control, including control over one's own thoughts

[7] It also identifies privacy as a freedom (art. 7 and 8) as well as the right to property (art. 17), and as said before, several freedom articles jointly also imply a right to autonomy.

and decisions. Since we already established autonomy as a separate value for SIS, a separate value of freedom must refer to negative liberty. This refers to matters like freedom of speech, freedom of assembly, freedom of movement, and other negative freedoms that could be either enhanced or limited by SIS. Freedom of information (the ability to access information) is also a negative liberty, although it could also adversely affect positive liberty (autonomy). In conclusion, the different sources provide a strong argument for including freedom (especially negative liberty) next to autonomy as a core value for SIS.

(1.4) **Dignity:** Dignity (Düwell et al. 2014) is an important value in philosophical ethics, although perhaps less referenced than the other values listed so far. Dignity is the right of persons to be treated with respect, as beings that are respected and valued for their own sake. Four violations of human dignity are typically associated with it: humiliation, instrumentalization or objectification, degradation and dehumanization (Kaufman, Kuch, Neuhäuser and Webster, 2011). Examples of specific violations are torture, rape, labor exploitation, bonded labor, slavery and social exclusion. Dignity is a key principle in the EU Charter, and the first chapter and article is devoted to it. The HLEG lists "respect for human dignity" as one of the fundamental rights as the basis for trustworthy AI, but in formulating its principles for trustworthy AI it does not mention it separately but subsumes it under the principle of prevention of harm. In its ultimate seven requirements for AI, it is no longer referenced, although perhaps implicit in its requirement of human agency and oversight. The OECD guidelines recognize "dignity" as among the "human centered values" that must be upheld. The IEEE guidelines make frequent reference to dignity, although they do not list it as a separate principle. Presumably, however, dignity is included in the IEEE's generic reference to "human rights", which is a key principle in its guidelines. In our study, "dignity" is referenced eight times, which is significantly less than the previously mentioned three moral values.

Some of the worries in the ethics literature on SIS concern the possible occurrence of instrumentalization, objectification and dehumanization of human beings, for example by disrespect for individuality in profiling and data processing in SIS, and objectification of humans in decision-making by AI. Also given the importance of dignity in the EU Charter in particular, as well as in the UN Universal Declaration of Human Rights and in the OECD guidelines, it would therefore be justified to adopt dignity either as a separate value or to include it under some broader value if possible.

(1.5) **Property:** The right to own property is often conceived of as a fundamental right in philosophical ethics. Likewise, the EU Charter includes a right to property (art. 17), which includes a provision that intellectual property must be protected. The HLEG guidelines do not contain a principle of (intellectual or other) property. The document does contain some references to the intellectual properties of companies with respect to AI system, which it is claimed must be taken into account when auditing AI systems and must be balanced against user rights. The OECD guidelines also do not contain a principle that refers to a right to property. However, its preamble does mention intellectual property rights as amongst those that must be recognized. It also points out that national and international legal, regulatory and policy frameworks for intellectual property rights with relevance to AI have already been developed. The IEEE guidelines do not have a principle for property rights, but do make several references to the importance of intellectual property rights. In our study, intellectual property and data ownership are mentioned frequently. We have a segment on ownership of data in

section 4, and issues of intellectual property and data ownership are mentioned in section 5, and in five of the application areas discussed in section 6.

The mixed image that emerges from the different sources makes it unclear whether property is important enough to include as one of the central principles for smart information systems. However, several considerations can be brought to the table to the effect that the three AI guideline documents underestimate the importance of intellectual property as an ethical issue. First, they may underestimate the importance of data ownership. Their references to intellectual property rights are not to data but predominantly to AI systems, which are software-based constructs. They concern whether or not third parties have a right to access to their inner workings by argument of transparency, user rights or accountability or whether intellectual property rights trump these rights.

What is not referenced in the three documents are controversies over *data* ownership. Many AI systems are data-intensive and make use of large databases. In contemporary society, data has become a valued resource, and has been called "the new oil". Questions of who has a right to ownership, access and control are therefore becoming paramount. This applies both to personal data and other types of data. In the EU, the GDPR gives persons strong ownership claims over their personal data, including strong rights of access and informed consent. However, in practice, there are still contested issues, such as whether and when informed consent is sufficiently given for other parties to assume ownership and control, when information produced by mixing personal data with other data sources still qualifies as personal data, what rights still apply to anonymized personal data, and how ownership rights of database owners relate to ownership rights of persons whose data has been collected.

Furthermore, questions of ownership and access also apply to other types of data, amongst others when multiple parties are involved in the collection and integration of data and when data is collected from private property (e.g., surveys of farmland), public property or natural resources not owned by the collecting agent. Some have advocated policies of open data, according to which data should be, as much as possible, a public good that is freely available to all without access and use restrictions (Kalin, 2014). This would especially apply to government-held data. The United Nations has advocated a principle of data philanthropy, according to which companies should share data for public benefit. All considered, there appears to be good reasons to include property (with special reference to intellectual property rights and data ownership) as a value for AI, either as a separate value or included under some other value. It is different from other moral values discussed so far, however, in that it appears to be a more contested concept.

(1.6) Safety and Security

The EU charter recognizes a right to life (art. 2), a right to the integrity of the person (including bodily and mental integrity, art. 3) and a right to security (art. 6). Similar rights are also recognized in philosophical ethics. In relation to SIS, these rights could arguably be subsumed under a value of "safety" or "safety and security".[8] This is what the HLEG does by including a requirement of "Technical robustness and safety" for AI. Similarly, the OECD includes a principle of "Robustness, security and safety". The IEEE guidelines do not seem to have any such principle, but they include principles of effectiveness and awareness of misuse that cover safety and security aspects, and they have a principle of human rights, that includes the rights included by the EU charter. Safety and security are

[8] It should be noted, however, that computer security is a very different concept than the ethical and legal concept of security of person. Computer security may contribute to security of person (and safety), but should not be conflated with it.

also frequently referenced principles in our study. There appear to be reasons, therefore, to include a value (and corresponding principles) of safety (or safety and security) for SIS.

It is also possible to subsume safety and security under a more general value of well-being or prevention of harm. For further discussion, see the later entry on "well-being and prevention of harm".

(1.7) Other rights

Rights to equality and nondiscrimination are prominent in both philosophical ethics and the EU Charter, and could be significantly harmed by SIS. However, we choose to discuss them below under the heading of "Justice and fairness". This is because these rights are often discussed in the context of justice and fairness. Justice (or fairness) is often considered to be not a moral right but a value that expresses a moral condition or procedure, and is therefore not included here under the heading of "individual rights".

The EU Charter also recognizes labour rights, citizen's rights, justice rights (including rights to a fair trial), rights of children and the elderly, and includes an article for consumer protection. These rights could all be affected by SIS, and therefore deserve to be taken into account in ethical guidelines for SIS. However, they are perhaps not as fundamental as the rights enumerated so far, and could perhaps better be discussed under more general values. This is what existing sets of guidelines for SIS tend to do.

The principle of informed consent, finally, is frequently mentioned in relation to SIS, and deserves a place in ethical guidelines. It is, however, a principle that is often seen as derived from autonomy, since the ability to make one's own choices and decisions, which is central in autonomy, involve the ability to make informed decisions about procedures that involve oneself. It is therefore best subsumed under that value.

8.1.2 Justice and fairness

Justice and fairness are two closely related concepts, and following scholarship in theories of justice since the 1970s (Rawls, 1971), we will treat them as (near-) synonyms. Ethicists distinguish several types of justice, the most important one being distributive justice, which is the socially just allocation of goods in society – pertaining to goods such as opportunities, rights, income, and resources. The question of distributive justice is to determine which allocations of such goods are fair and which are unfair. Most theories of distributive justice agree that a principle of equality between persons must play a role in such distributions. But most of them do not hold a strict egalitarianism, according to which each person should have the same level of social and material goods and services. Rather, equality is usually sought in rights and opportunities, rather than in material goods and services. This results in a notion that it is unfair to deny rights to persons, and deny equality of opportunity, but that it is not necessarily unfair for there to be differences in wealth, property and income. Disagreements exist in society about the rights and opportunities that people should be entitled to and what level of basic services should be provided to all in a just society.

There can sometimes be good reasons to assign different rights in society to different groups of individuals. For example, prisoners have temporarily restricted freedom rights, and recent immigrants may have restricted civic rights. Principles of nondiscrimination state that in regard to human rights, there should not be any differentiation that is based on inalienable parts of one's identity, including gender, race, age, sexual orientation, national origin, religion, income, property, health, disability and opinions. There is debate in society about which identity markers should be included in nondiscrimination principles and what specific kinds of actions are discriminatory. Another principle related to fairness and equality is diversity (or

"respect for diversity"), which goes beyond nondiscrimination to include positive valuation of individual differences, recognition of differences in individual need and support for the diverse composition of organisations and communities. Another related principle is that of inclusiveness, which is the inclusion of marginalized people within social practices and treating them fairly and equally.

The EU Charter includes principles of equality before the law (art. 20), nondiscrimination (art. 21) and cultural, religious and linguistic diversity (art. 22), integration of persons with disabilities (art. 26) and fair and just working conditions (art. 31), amongst others. The HLEG includes a principle of diversity, nondiscrimination and fairness that includes the considerations of the previous paragraph. The OECD includes fairness within a principle of "human-centered values and fairness", with specific reference to non-discrimination and equality, diversity, fairness, and social justice. The IEEE guidelines surprisingly do not include a principle of fairness or justice. However, equality and nondiscrimination are implicitly contained in their principle of human rights, and in their detailed discussion of the principle of well-being, they make clear that it should be understood to not only include individual well-being, but also societal and environmental well-being, including psychological, social and economic fairness. The document also makes reference to cultural diversity and inclusiveness. In our own study, fairness is frequently referenced, as are the associated concepts of justice, diversity, algorithmic bias, (non)discrimination, power asymmetries and digital divide.

We conclude that there is considerable agreement between the sources to include a broad value of fairness for AI, that includes principles of justice, equality, nondiscrimination, diversity and inclusion. This moral value would moreover cover several types of unfairness, which were distinguished in the various documents, including unequal access to AI systems and services (countered by a principle of universal access), algorithmic bias (biases and inaccuracies in algorithms that lead to the unequal or unfair treatments of individuals or groups - countered by algorithmic fairness), and other unfair aspects of SIS design and impacts of SIS.

8.1.3 Responsibility and accountability

Moral responsibility relates to agents who are expected to perform certain actions towards the entity they bear responsibility for. Performing these actions is considered praiseworthy and not performing them is considered blameworthy. Moral responsibility thus comes with social expectations and corresponding social responses concerning the actions of agents. While responsibility and accountability are sometimes used synonymously, accountability can be usefully distinguished from responsibility by defining it as the obligation or willingness to accept responsibility for one's actions. Differences between the two are that in accountability there is a the (prior) commitment of an agent to accept responsibility, and that accountability cannot be shared, whereas responsibility can be.

It is not surprising that the EU Charter of fundamental rights does not include principles of responsibility or accountability, since it is concerned with those who are recipients of moral actions (bearers of rights) rather than the agents who engage in moral action. However, both the HLEG, OECD and IEEE documents include accountability as a key moral principle. The HLEG guidelines reference it as one of the seven requirements of trustworthy AI, the OECD guidelines references it as one of its five principles, and the IEEE guidelines include it as one of its eight principles. In the SHERPA study, accountability and responsibility are referenced frequently in sections 4-6. There therefore appears to be strong arguments to include accountability as a key moral value for SIS.

As the HLEG report proposes, accountability in the context of AI means that mechanisms are in place in society, and within organisations, "to ensure responsibility and accountability for AI systems and their outcomes, both before and after their development, deployment and use"(HLEG 2019, p. 19).

Included in accountability is algorithmic accountability, which is the principle that organizations that use algorithms should assume responsibility for the decisions made by those algorithms. However, accountability, as described by the HLEG, is a broader concept than algorithmic accountability, involving additional actors and additional responsibilities for the functioning and impacts of SIS.

8.1.4 Individual well-being

Well-being is recognized as a key value in ethics. It concerns a person's welfare or quality of life. Well-being is not generally recognized as a right, but rather as a (morally) desirable condition. For this reason, it should come as no surprise that well-being is not referenced in the EU Charter on fundamental rights. However, the EU Charter references many aspects of life that are conditions for well-being, including life, human dignity, bodily and mental integrity, privacy, and freedoms. The HLEG guidelines do not reference individual well-being as a separate moral value, but reference it as part of the fundamental value of societal and environmental well-being. It moreover proposes a key principle of prevention of harm, which protects two key components of human well-being: human dignity and mental and physical integrity. t is noteworthy that in its first draft, the HLEG also employed a principle of beneficence, "do good", which advocated that AI systems should be designed and developed to improve individual and collective wellbeing. This was removed after objections by stakeholders, especially from industry.

The OECD guidelines include well-being as one of its key principles, as part of a broader principle of "inclusive growth, sustainable development and well-being", and appears to reference augmenting human capabilities and enhancing creativity as examples of fostering well-being. The IEEE guidelines have well-being as one of its key principles. They define this notion in terms of the OECD Guidelines on Measuring Subjective Well-being, which measure well-being through people's self-reported judgments on their well-being, and also reference Amartya Sen's and Martha Nussbaum's capability approach as a second way of measuring well-being. (The OECD guidelines also reference the enhancement of capabilities as a desirable goal.) In our study, well-being (and quality of life and welfare) are only mentioned a few times, but this is possibly an artifact of the way we conducted our study, as we have not included a significant focus on this value.

A choice that presents itself is whether to adopt a well-being principle that requires the active promotion of well-being, or merely the prevention of harm to well-being. That is, should Smart Information Systems be built and used to promote and protect human well-being, or should they operate according to a more limited no-harm principle, according to which they should not harm human well-being. According to this more limited principle, SIS should protects life, health, physical and mental integrity - which could largely be captured by the previously mentioned "safety and security" principle – but they do not have to actively promote well-being. The case for the more limited option is that most ethical theories do not claim that there is a moral duty to enhance human well-being, but only to not harm it. The case for the enhanced option is that it need not be presented as a duty but rather as an aspirational goal. Its inclusion in the OECD, IEEE and (less explicitly) the HLEG guidelines suggests the existence of enough support to choose the latter option.

8.1.5 Societal and environmental well-being

An exclusive focus on the rights and welfare of individuals in ethics of SIS could lead to the neglect of responsibilities for other objects of moral concern, including social and political arrangements and institutions, and the environment. This is not to say that these other entities are separate from individuals: their flourishing will also contribute to the flourishing of individuals. Conceptions of moral and social responsibility have traditionally included not only responsibilities towards persons,

but also responsibilities towards social institutions and arrangements, society as a whole, and the natural environment. Societal well-being, the flourishing of society and its institutions, and environmental well-being, are therefore worthy moral goals, either in the minimal sense of not harming societal and environmental well-being, or in the more ambitious sense of actively promoting them.

Because of the EU Charter's focus on individual rights, it makes little reference to societal and environmental well-being, although it does contain a principle of environmental protection (art. 37) and refers to fundamental principles of democracy and the rule of law in its preamble. The EU's vision of societal and environmental well-being (including principles of sustainable development, democracy, rule of law and others) is however obvious in other policy documents such as the Lisbon Treaty. The HLEG proposes a requirement of "societal and environmental well-being" which includes consideration of the impacts of AI on individuals in all spheres of life, including consequences for their physical and psychological well-being, as well as broader social and environmental impacts, including impacts on institutions and society at large. Explicitly mentioned are impacts on democratic processes as a topic of concern.

In the OECD guidelines, societal and environmental well-being is covered in its principle on "inclusive growth, sustainable development and well-being", which involves a pursuit of "beneficial outcomes for people and planet". The IEEE guidelines include a principle of well-being that is explicitly stated to include societal and environmental well-being. In our report, finally, we reference issues of employment, economic development, sustainability, and democracy, as well as other broader social issues, although we should add that in our research design, we did not assign a large place to such issues. Considering the various sources, we can conclude that there is strong support for the inclusion of societal and environmental well-being as moral values for SIS.

8.1.6 Transparency

Transparency of AI systems refers to the idea that the purpose, inputs and operations of AI programs and algorithms should be knowable to its stakeholders so that they can understand how and for what purpose these systems function and how their decisions are arrived at. It is associated with other principles that concern the understandability of AI systems, including explainability, traceability and interpretability.

All three guideline documents include transparency as a key principle. The OECD guidelines reference a principle of "transparency and explainability" for AI. The HLEG and IEEE guidelines merely refer to "transparency", but both recognize explainability as a component of transparency. Transparency is not recognized as one of the fundamental moral values in philosophical ethics, although it is sometimes recognized as a personal or professional virtue. It is not referenced in the EU Charter. So should it be included in a listing of moral values or principles for SIS? In light of this, it should be noted that the HLEG and IEEE recognize that transparency is not a *moral* value. The HLEG document identifies it as a requirement for trustworthy and ethical AI, but not as one of the fundamental moral principles for AI.

The IEEE document notes that its general principles for ethically aligned design includes principles that do not correspond with what it calls universal human values. It however holds that these principles are nevertheless necessary for ethical development and deployment of AI. In our study, we also reference transparency as an important concept for SIS. The justification for inclusion of transparency as an important value (albeit a nonmoral one) is (according to our study and to the HLEG and IEEE documents) that it is instrumentally important for the proper realization of moral values, particularly accountability, fairness and individual rights. It can be concluded that there are good grounds to

include transparency as a guiding value for AI, though not a moral value, but an instrumentally necessary value for the realization of moral values.

8.1.7 Other (moral and non-moral) values

Very few other moral values have been found in our study other than those listed above. The HLEG, OESO and IEEE documents also highlight no other moral values other than those covered above. However, these documents, as well as our study, do reference several nonmoral values and principles that are considered to be important for ethical AI. The HLEG refers to a nonmoral value of "technical robustness" in its principle of "technical robustness and safety", and relates it to other nonmoral notions such as accuracy, reliability and reproducibility of AI systems. It considers these as necessary conditions for safe AI that minimizes harm. So like transparency, these nonmoral values are considered instrumentally necessary for the realization of moral values. Similarly, the HLEG makes reference to a principle of "data governance", which is included in its principle of "privacy and data governance", and relates it to requirements of quality and integrity of data and the inclusion of data protocols that govern data access. These are considered necessary for the realization of moral values like privacy, fairness and harm prevention.

The IEEE guidelines, which have a focus on professional developers and deployers of AI, contain nonmoral principles of effectiveness, competence and awareness of misuse that are similarly believed to be necessary for ethical AI. The OECD's principle of "robustness, safety and security", finally, includes the nonmoral value of robustness, and makes reference to principles like systems robustness and traceability, which are believed to be needed for the realization of safety and other moral values.

What these documents suggest is that there are several nonmoral requirements for SIS, other than transparency, that are similarly important in realizing ethical SIS. These include, most centrally, the reliability, accuracy and security of smart information systems and the data included in them, and the awareness of, and prevention of, misuse and dual use.

8.2 Ethical Tensions

Ethical tensions are potential conflicts between either (a) two moral values or principles, or (b) a moral value or principle and a nonmoral value, principle, or interest, or an existing tendency or practice. Ethical tensions occur when it appears that it is difficult or impossible to realize both desired outcomes simultaneously, so that a choice will have to be made. Ethical tensions can be coincidental (occurring in a special case or instance) or systematic (recurrent across a large number of cases).

An example of an ethical tension between two moral values is the tension between privacy and security in video surveillance. It appears that the choice to engage in video surveillance increases security but reduces privacy, and the choice not to engage in it enhances privacy but limits security. This then raises the question of what is the right decision to make: to engage in video surveillance or not to do so. An example of an ethical tension between a moral value and a nonmoral quantity is the tension between privacy and commercial interests in social media. Social media companies have a commercial interest in the marketing and exploitation of personal information of its users, which reduces privacy. It seems that one cannot easily have both: strong privacy protection for users and strong advancement of the commercial interests of social media companies.

In what follows, we will consider recurrent, systematic ethical tensions involving smart information systems that involve moral values that were discussed in section 8.1. I will distinguish three types of tensions: tensions between moral values, tensions between moral values and (nonmoral) interests,

and tensions between moral values and features or properties of SIS.[9] Our discussion is based on tensions that we have identified in the academic literature in ethics of SIS and in our own case studies

Tensions between moral values

- *Privacy vs. security.* SIS have many applications for security purposes, especially as used by law enforcement, but also by private companies to ensure the safety and security of personnel or customers. In many of these applications, personal information is collected and processed, in ways that often do not involve full informed consent. There is hence a frequent tension between security and privacy in SIS, as already noted above.

- *All moral values vs. intellectual property rights.* Companies have a legitimate interest not to give third parties access to the inner workings of its SIS, as this could lead to a loss of intellectual property. However, such access is often needed in order for third parties to be able to establish that their rights and interests are respected by the system, and in order for companies and individuals involved in the development and use of SIS to be accountable. In addition, SIS may contain confidential data that companies may (legitimately) want to be kept secret, but doing this could also violate other moral values.

Tensions between moral values and interests

- *All moral values vs. commercial interests.* The commercial interests of private companies may sometimes be aligned with the rights of consumers, employees, or other stakeholders, or at least neutral with respect to them, but they may also be in tension, as when private companies consider it in their interest to have good access to personal information about customers or employees, or to perform certain actions that restrict their freedom and autonomy. Similarly, commercial interests could lead to consumers or employees (e.g., those with more purchasing power) to be treated differently or discriminated against. Commercial interests could also induce companies not to invest in ethical practices if these are considered too costly and time-consuming.[10]

- *All moral values vs. misuse of SIS for selfish or malignant reasons.* Misuse of SIS can occur by actors within private or governmental organisations that use systems for personal ends, against organizational policy. It can also occur by third parties that have been licensed to use or access systems of data, but do not use it according to expectations or policy. Systems and data can also be hacked, stolen or otherwise accessed without permission and be used for unauthorized selfish or malignant ends. These forms of misuse can harm any of the moral values we considered in 8.1. Malignant uses include different forms of theft, revenge, cyberterrorism, cyberwarfare, and pursuits of political and ideological agendas.

Tensions between moral values and technical properties and organizational conditions relating to SIS

[9] In addition, there can be tensions between moral values and the organisational or institutional embedding of SIS. These will not be considered here.

[10] This is not to say that not-for-profit orgganisations, such as governmental organisations and NGOs, cannot have interests or policy priorities that conflict woth moral values. They can, and therefore potential value conflicts should be considered for these organisations as well.

- *Accountability, individual rights and fairness vs. opacity of SIS*. SIS that are opaque and whose workings cannot be properly understood or explained risk violations of several moral values, including accountability, fairness, and various individual rights.

- *Fairness, freedom and well-being vs. inaccuracy and unreliability of SIS.* SIS that contain a high error rate and therefore make poor decisions risk violation of many moral values considered in 8.1. Systems that represent people and groups and that make erroneous judgments that concern them (false positives and false negatives) may cause harm to the interests of some individuals or groups, thereby violating moral values of fairness, freedom (e.g., when persons are falsely stopped and searched as a result) and well-being.

- *Autonomy vs. decision-making capabilities of SIS.* SIS that make autonomous decisions may affect the autonomy of stakeholders affected by the decision-making process by limiting their choices and determining goals and choices for them.

- *Privacy, well-being, and other moral values vs. lack of security in and for SIS*. If SIS are not secure, then unauthorized access to systems or data, and possible theft or vandalism can harm various moral values, particularly the privacy of individuals whose personal data is contained in the system, and the well-being of individuals and society.

8.3 Mitigation of Ethical Tensions

In our discussion of mitigation actions that could be taken to reduce ethical tensions regarding SIS, we will make use of the overviews of methods for the implementation of ethical guidelines provided in the HLEG and IEEE reports. Both reports suggest methods for the implementation of ethical guidelines that can be associated with different actors (developers, policy makers, organizational users and others). The HLEG makes a distinction between what they call technical and non-technical methods, both of which apply to all stages of the lifecycle of SIS. Technical methods include ethics by design methods, explanation methods for transparency, methods of building system architectures for trustworthiness, extensive testing and validation, and the definition of quality of service indicators. Non-technical methods include regulation, codes of conduct, standardization, certification, accountability via governance frameworks, education and awareness to foster an ethical mindset, stakeholder participation and social dialogue, and diverse and inclusive design teams.

The IEEE report has a chapter on "methods to guide ethical research and design" for researchers, technologists, product developers and companies (pages 124-139), and a chapter on policies and regulations by governing institutions and professional organizations (pages 198-210).[11] In its methods for ethical R&D chapter, it considers both individual and structural approaches, and distinguishes between three overall approaches: interdisciplinary education and research, corporate practices on SIS, and responsibility and assessment. Interdisciplinary research involves the integration of applied ethics into education and research to address issues concerning SIS, and includes educational programs, interdisciplinary collaboration that brings engineers and scientists into contact with social science and humanity scholars, attention for intercultural information ethics, and institutional ethics committees in AI fields.

[11] It also has a separate chapter on law, but legislation as a method for ensuring ethical SIS seems to be covered already in its chapter on policy.

The section on corporate practices on SIS proposes structures to be put in place for creating and supporting ethical systems and practices around the funding, development and use of SIS. This involves instituting an ethical corporate culture for SIS that facilitates values-based design, instituting value-based leadership roles, incentivizing and empowering technical staff to raise ethical concerns, training staff to consider broader societal and ethical issues, the inclusion of ethics review boards, stakeholder inclusion, and values-based design. The section on responsibility and assessment includes oversight procedures for algorithm development, an independent review organization and certification agency for ethical SIS, ethical assessment and cautious use of black box software, and better technical documentation.

In its policy chapter, the IEEE advocates the founding of national policies and business regulations for SIS on human rights approaches, the introduction of support structures for the building of governmental expertise in SIS, the fostering of SIS and ethics training in educational programs, governmental support for ethical research, development, acquisition and use of SIS through standards, national ethics guidelines, funding programs, and research groups for SIS for the public good, the development of policies for SIS to ensure public safety and responsible SIS design, and educating the public on the ethical and societal aspects of SIS.

The methods proposed by the HLEG and IEEE are partially overlapping and in part complementary. We believe that they jointly provide a strong set of methods for fostering the ethical development and use of SIS. At its core, ethical SIS requires, in our view, five key ingredients:

(1) methods of incorporating ethics into the design of SIS;

(2) corporate social responsibility cultures that support ethical development and use of SIS;

(3) national and international standards and certification for ethical SIS;

(4) education, training and awareness raising in the ethical and social aspects of SIS, and

(5) governmental policy and regulation to support and require ethical practices in SIS.

These ingredients will help to mitigate many of the identified tensions as well. Specific methods may however be needed to address specific tensions. For example, mitigation of privacy vs. security tensions requires the development of privacy-enhancing techniques for security systems and the development of alternative security systems that do not involve the massive processing of personal information.

It can be concluded that many strategies exist for mitigating ethical tensions associated with SIS. A large number of general strategies for ethical SIS have been proposed, which will often have a positive effect on alleviating ethical tensions. In addition, specific strategies are needed for alleviating specific ethical tensions, that will often draw from the general methods that have been proposed for ethical SIS.

9. Conclusion

This report has presented a thorough and systematic analysis of the ethical issues arising from smart information systems. It opened with an overview of the technology involved in the development and use of SIS, paying close attention to particular tools and techniques currently in use.

The report then considered applications of smart information systems in particular domains before examining general ethical issues. These issues were broken down into concerns relating to the aims of SIS and concerns regarding the implications and risks of SIS. Further attention was paid to ethical issues raising from specific types of SIS and techniques used in SIS technology. This focused particularly on ethical issues arising from the development and use of algorithms and concerns arising from data ethics (including types and sources of data). The cross-cutting approach taken to ethical analysis ensured that general ethical issues were therefore approached from multiple angles to ensure that the field was examined thoroughly.

Following the general overview of ethical issues arising from aims, implications and types of SIS, the report examined ethical issues arising in each of the different application domains which informed deliverables D1.1 (case studies) and D1.2 (scenarios). This ensured that ethical issues arising from real-world, as opposed to largely theoretical, concerns were brought under the spotlight. It was noted here that while there were no ethical issues arising from the case studies or scenarios that were not already covered in the general ethical issues (and thus also supporting the thorough approach taken to the general ethical issues), reporting in individual application domains often did not cover ethical problems which were of concern to people working in those fields.

The report then looked at ethical issues arising in research and innovation before summarising the main ethical issues and presenting possible solutions to some of the more pressing concerns. In this, the report drew on methods proposed by the HLEG on AI and the IEEE to suggest mitigation strategies for key ethical concerns. These methods will form a central element in the development of ethical guidelines (Task 3.2).

In summary, the report has presented a broad, overarching analysis of ethical concerns related to the development and use of SIS. In so doing, it provides a strong grounding and template for the SHERPA project in later deliverables; in particular, task 3.2, which will integrate the findings from th report to develop two sets of ethical guidelines, one for the ethical development of, and a second on the ethical use of, SIS. This deliverable is intended to provide clear ethical guidelines to those developing and using SIS in the field.

10. Addendum: Overview of Deliverables in Work Package 1

Introduction

This appendix to D1.4 brings together a brief overview and key findings from the other four deliverables in SHERPA Work Package 1. These deliverables were:

- D1.1 – Case Studies
- D1.2 – Scenarios
- D1.3 – Cybersecurity
- D1.5 – Human Rights Issues

By presenting these findings here, it should be clear how each of these deliverables has contributed to the overview presented in the main body of D1.4. The findings of each of these prior deliverables has been used in the creation of D1.4, but it is not always obvious where the input has been made. Rather than cross-reference the earlier deliverables throughout D1.4, and thus impeding the reading of an already lengthy document, an overview and key insights and recommendations of each of the earlier deliverables was felt to be of more use.

In each of the overviews, the report starts with a description of the deliverable task in terms of defining the problem, followed by an overview of the key issues which were identified in the course of the research of that deliverable. The most significant insights are then summarised, followed by recommendations and a conclusion. In each case, the conclusion looks forward to future research in the SHERPA project, identifying the Work Packages towards which the findings of the deliverables will contribute.

As such, this appendix provides an overview of the key work and findings of Work Package 1, as well as indicating how these contribute to further work which will be undertaken in the remainder of the SHERPA project.

Ethical Issues and Responses to Smart Information Systems

Definition and Problem

This briefing document provides an overview of 10 case studies into the ethics of Smart Information Systems (SIS) involving the combination of artificial intelligence (AI) and big data analytics. The case studies were conducted to establish ethical issues arising from the implementation and use of SIS, and responses to these issues.

There is a considerable literature on ethical issues arising from SIS. However, there is little empirical research, and fewer case studies, exploring these issues across different applications of SIS. Through conducting case studies, the issues raised in the literature could be applied more specifically, and responses to ethical issues identified. Each case study involved background research on academic and trade literature regarding ethical issues related to different applications of SIS (Table 1). Practitioners in each area of application were then interviewed for a fuller understanding of ethical issues experienced in the workplace and the responses taken. Results of the case studies were then combined and contrasted to identify gaps in knowledge and provide a comprehensive analysis.

No.	Case Study Domain	Case Study Focus
CS01	Employee Monitoring and Administration	A company using IoT for Employee Monitoring and Administration
CS02	Government	A division within government, a municipality, using SIS
CS03	Agriculture	Large agribusiness using SIS
CS04	Sustainable Development	1. Large Municipality; 2.Public Organisation; 3. Telecommunications Company; 4. Large Municipality
CS05	Science	A large scientific research project
CS06	Insurance	Health insurance companies
CS07	Energy and Utilities	Energy and utilities company
CS08	Communications, Media and Entertainment	Cybersecurity department within a multinational telecommunications company
CS09	Retail and Wholesale Trade	A national telecommunications company developing SIS for retail customer-relation management
CS10	Manufacturing and natural resources	A company developing SIS for risk prediction in supply-chain management

Table 1: Case Study Domains

Issues

There were 26 ethical issues identified in the case studies (Table 2). Privacy, which has received a great deal of attention as a result of GDPR, was the only ethical issue addressed in all 10 case studies. Security, transparency, and algorithmic bias are also regularly discussed in the literature, so were expected to be significant. However, there were many issues that received less attention in the literature (e.g. access to SIS, trust, and power asymmetries) but were discussed frequently in the interviews. There were also ethical issues that were heavily discussed in the literature and which received less attention in the interviews than expected (e.g. employment, autonomy, and criminal or malicious use of SIS).

Ethical Issues	CS01	CS02	CS03	CS04	CS05	CS06	CS07	CS08	CS09	CS10
Access to SIS	●	●	●	●		●	●			●
Accuracy of Data		●	●	●				●	●	●
Accuracy of Recommendations			●	●		●		●	●	
Algorithmic Bias					●	●		●	●	●
Discrimination	●				●	●		●		●
Economic		●	●	●				●		
Employment			●	●		●		●		
Fairness			●	●	●					
Freedom							●			
Human Contact			●							
Human Rights					●			●		●
Individual Autonomy								●	●	
Inequality	●		●	●						
Informed Consent	●		●	●		●	●	●		●
Integrity					●					●
Justice		●	●	●				●		
Ownership of Data		●	●	●		●				●
Military, Criminal, Malicious Use	●			●				●	●	
Power Asymmetries	●	●		●	●		●	●		

Privacy	●	●	●	●	●	●	●	●	●	●
Responsibility	●		●	●		●		●		
Security	●	●	●	●	●	●		●	●	●
Sustainability			●	●						
Transparency	●		●	●	●	●	●	●	●	●
Trust	●	●	●	●		●	●	●		
Use of Personal Data	●	●	●	●	●	●		●	●	

Table 2: Ethical Issues Identified in Each Case Study

Key Insights

The individual ethical issues are extremely important in themselves. However, our findings extended beyond the issues to the responses taken by the organisations interviewed. These were broken down into the following six areas.

Organisational Methods

The organisations interviewed were aware of the issues which may affect relations with users of SIS. They are trying to establish mitigating approaches to deal with these issues, especially by ensuring that responsibility is upheld when developing, deploying and using SIS. These approaches include commitments to responsible data science, stakeholder engagement, ethics review boards, following codes of ethics and good standards of practice. Organisations found they were often conflicted by legal, economic, technical or practical abilities to follow through with many of their goals in this area. For example, one company attempted view their use of SIS as ways to protect ethical standards but noted conflicts arising between integrity and the most profitable ways to use SIS. Furthermore, in many cases SIS are made up of components across organisations, e.g. when the data is owned by one company, the algorithm by another, the processing is happening on the hardware of a third for the purposes of a client that is a fourth. This complicates locating responsibility for ethical issues.

Technical Methods

In order to ensure privacy, many technical procedures were noted in the case studies. These included: encryption, government-supported secure storage, and the anonymisation or pseudonymisation of data (through automated or manual means). Some companies employed third-party penetration testers to examine their systems for weaknesses, others held regular hackathons and sent fake phishing emails to test staff. However, those engaging in such tests were large multinationals with significant funds. It would be harder for an SME or most municipalities to offer the same level of protection, or engagement with the hacker community. While some companies were happy to rely on mostly technical solutions to privacy concerns, those with greater technical expertise in computer security were more cautious. This may suggest that those who have the greater technical competence are more aware of the limitations of technology, and so less prepared to put their faith in technology to resolve complex social concerns.

Human Oversight

Trust in SIS is often affected by the lack or loss of human involvement and expertise. Despite the promises often made about SIS, these systems retain some inadequacies which demand human oversight and intervention. For instance, in the agricultural sector it was noted that '*SIS cannot replace agronomists but can support them and there is still a need for a knowledgeable person to provide further support'*. The key issue here is that there is greater trust in people and their expertise than in SIS. It was noted that the mistrust in many cases could be a result of SIS or human designers making unfair decisions, or of a lack of transparency in how decisions are made (see below).

Ethics Training for Developers

The weight of transforming technology often falls on technology experts, who are typically the decision makers for issues surrounding data collection, data manipulation, and computational aspects of SIS applications. Considering ethics within decision making is not an aspect with which computer scientists, statisticians and data analysts are familiar, as ethics is rarely taught on computer science and related degree courses. It was noteworthy that members of technology teams generally find it difficult to identify and discuss ethical issues. Ideally, the technical experts should be able to identify and discuss the ethical and policy implications of SIS, since they have significant impact on the successful use of the software product itself. Evidently, this is not always the case.

There is also a need for positive and imaginative responses to the introduction of policies that safeguard the ethical use of data. For example, GDPR has been adopted successfully by experts who have translated the regulation into the design and use of SIS.

Data Control and Transparency

Some interviewees aimed to place more control in the hands of citizens and/or those to whom the data pertain. In one case an explicit link was made between citizens having control over their data and ensuring privacy, although privacy breaches may still occur when control is given to citizens. A related concern is whether citizens would know what happens to their data and why. This again raises the desirability of ethics education and transparency in decision making. However, private companies may avoid full transparent about their processes for reasons of intellectual property or fears that some users might learn how to cheat their system. Despite this, it is feasible that while the details of specific processes might not be made transparent, codes of conduct and general principles could be made publicly available.

Computer Science Training for End-Users

Educating those using data is important for security. However, many users lack a basic knowledge of computer science and mathematics, limiting the potential for the informed co-creation of SIS. There is a related concern that this lack of education may lead to an imbalance of knowledge among users as to how algorithms process their data. Once the public has sufficient understanding of the methods and purposes of data collection and processing there will be the scope to gain genuinely informed (rather than uninformed) consent. Long recognised as central to research ethics, as well as GDPR, informed consent helps to guarantee the dignity of the subject and limit harm to that subject. However, a lack of understanding on the part of the subject prevents informed consent from occurring.

Conclusion

Firstly, between general and specific literature (both academic and trade) the ethical issues arising from SIS in practice have been described. However, the experience of at some practitioners suggests that some issues are more prevalent than the literature would suggest, and others less so. There may be a discrepancy between what academics and journalists think *should be* the main ethical issues, and what actors experience *as* the main issues. This would benefit from further empirical research.

In response, SHERPA WP2 will carry out this empirical research. Identification of stakeholders has begun and continues (Task 2.1). Large-scale surveys (Task 2.3), Delphi studies (Task 2.4) and interviews with stakeholders (Task 2.2) are planned. This includes ongoing engagement with interviewees from the case studies (Task 2.2). We will hence develop a more complete picture of which ethical issues are primary concern to the practitioner community.

Secondly, the need to establish ethical understanding and behaviour in organisations is clear, although the means by which this should happen is not. Developers are typically under-educated in ethics and users under-educated in computer science. In some cases, this means unethical systems may be developed but not recognized as such. Various responses have been tried, focusing on the developer community (organisational, technical, and human oversight methods) and on the user community (increasing stakeholder control of data and transparency of data use). The focus on informed consent alone as a means of guaranteeing ethical SIS is clearly insufficient. Guidelines are needed for developer and user communities to clarify duties and boundaries of responsibility. Also needed are methods to ensure that these guidelines are successfully implemented.

In response to this, SHERPA WP3 is developing a series of options for these next steps. Guidelines are being developed separately for the user and developer communities, along with implementation recommendations to see these incorporated into standard practice (Tasks 3.2 and 3.4). Regulatory options are similarly being explored include considerations regarding the creation of new regulatory bodies (Tasks 3.3 and 3.6). Technical options and interventions are also being explored (Task 3.5)

Taking the work of the case studies forward, the SHERPA project will therefore gain a deeper understanding of developer and user needs and practice. This will in turn inform the development of practical suggestions to shape these communities so that SIS will become more ethical in shape and use in the coming years.

More information

- SHERPA Workbook on Case Studies (https://www.project-sherpa.eu/category/case-studies/)
- Orbit Journal - Special Issue on SHERPA Case Studies (https://www.orbit-rri.org/ojs/index.php/orbit/issue/view/7)

Future Scenarios relating to Smart Information Systems

Definition and Problem

This briefing document reports on policy scenarios addressing the use of Smart Information Systems (SIS) in five different application domains. The scenarios considered how new and emerging technologies may raise various social, ethical and human rights issues in the year 2025. Policy scenarios provide a useful methodology to engage key stakeholders in exploring the ethical, legal, social and economic issues influencing the development and take-up of emerging technologies that are relevant for policymakers. Stakeholders identified actions required of policymakers and other stakeholders, ethical guidelines, data protection policies and other measures needed now to address the issues five or six years hence when there may be fewer policy options. The legitimacy of our scenarios stems from our inviting stakeholders to participate in the scenario development process from the outset and thereafter inviting increasing numbers of stakeholders to comment on each iteration of the scenario.

No.	Scenario domain	Scenario focus
Sc1	Social services	Deepfake technologies powered by AI
Sc2	Energy sector	Information warfare
Sc3	Policing	Predictive policing
Sc4	Transport	Self-driving vehicles (SDV)
Sc4	Education	Learning buddies

Table 1: Scenario domains and focus

Issues

During the scenario development process, stakeholders raised similar concerns with those raised in the case studies. Of those, the following echoed across the scenario domains.

It is already obvious that SIS are having far-reaching impacts and that those impacts will only amplify as the technology evolves, as algorithms are used in ever more applications. Some applications, e.g., Google Translate or the Duck Duck Go search engine, are useful, while others (such as targeted advertising) offend many consumers, invade their privacy and use people's personal data without their explicit informed consent. Hence participants saw SIS as bringing great benefits, but also great threats.

While some of the recommendations from the five scenarios are specific to the specific technology area, there are some common themes that appear, one being data protection. Workshop participants were concerned about the use of personal data without the consumer citizen's consent. They were concerned about AI being used to improve advertising targeted at individuals and other invasions of privacy.

Another common theme was the need for greater, more coherent regulatory oversight in the application of the technologies. Participants all agreed that SIS, while offering great benefits, also create great risks. Sometimes malefactors deliberately develop SIS systems and algorithms to achieve their gains at the expense and harm of society as when the WannaCry malware attacked the UK National Health Service (NHS). This seems to have been a clumsy but moderately successful plot by North Korea to gain foreign revenue at the expense of the NHS and various other organisations.

There was a shared sense that that the big five technology companies – Amazon, Apple, Facebook, Google and Microsoft – wield too much power with little effective oversight. The big five were seen to effectively control the SIS market, taking a disproportionate share of the available human talent. Their resources and the amount of data at their disposal dwarfs anything by any other organisation. Hence the big five are driving the future of SIS and putting algorithms to work in a vast array of different applications to understand people better.

There was some discussion about the need to bring explainability into algorithms, by which was meant the need for algorithms to inform users (or those affected by those algorithms) the purpose of the algorithm, who was funding the development of the algorithm, and whom to contact for more information. This currently rarely happens but participants hoped it would more likely be the case by 2025.

Another issue that arose was that of inequality, arising from the fact that some people were more likely to benefit from SIS (e.g. robotic "learning buddies" or holographic companions) than others, particularly those in lower socio-economic strata. The related issue of fairness also arose, e.g. predictive policing algorithms were more likely to target street crime than corporate crime.

As SIS penetrate further into our economies and societies, they speed decision-making such that SIS-powered decision-making becomes more needed. Human decision-makers cannot respond fast enough, especially in the instance of attacks on cities and critical infrastructure. SIS-powered decision-making raises apprehensions about decisions gone wrong or without an appreciation of the consequences.

SIS often raise complex ethical issues regarding legal and moral liability. Some SIS scientists have already signed a petition against working on killer robots; some employees have rebelled against working on SIS use in military technologies. Questions of liability proliferate. Who is liable for an algorithm on which many data scientists have worked? Is it the organisation who is funding

development of the algorithm? Is it the programmer who feeds the data to train the algorithms? Is it the client who is using the algorithm? Do the middlemen, the suppliers, have some liability? Or the insurance companies? Other issues worth debating are those relating to autonomy. Are SIS creating dependencies, and thereby reducing our autonomy? Some of these issues are also being explored in the SIENNA project.

Key Insights

The individual ethical issues are extremely important in themselves. However, our findings extended beyond the issues to the responses experts would like to see by policy makers. These were broken down into the following seven areas (Table 2) and described in more detail below.

	Mimicking technology	*Information warfare*	*Predictive policing*	*Driverless cars*	*Learning buddy*
1. Create codes of conduct for technology and domain professionals	√				√
2. Promote ethics by design technology development approaches			√	√	
3. Engage the public/domain experts in AI policy development and governance	√				√
4. Develop educational and training material about the ethical use of SIS technology		√		√	
5. Stir technology investment policy towards embedding ethical components of SIS		√	√	√	
6. Create a super-regulator for SIS to create and enforce regulation across different jurisdictions	√	√			
7. Update legal definitions in laws and policies regulating technologies to encompass novel issues raised by SIS		√	√		

Table 2: Desired responses from policy makers.

Create codes of conduct for technology and domain professionals

Organisations offering SIS-powered holograms or robots for social care should make public their codes of conduct, which spell out, *inter alia*, how information and data will be collected, what will happen to such data, how they will be used, and who will have access to them. Similar codes of conduct should apply to "learning buddy" robots.

Promote ethics by design technology development approaches

Given the many issues raised by SIS, proponents and developers should take ethical principles into account during the different stages of the development process.

Engage the public/domain experts in SIS policy development and governance

Engaging stakeholders in the SIS policy development process is essential in view of how far-ranging the impacts of SIS are. Constructing scenarios is one way of engaging stakeholders in a deliberative exercise aimed at producing recommendations for policymakers.

Develop educational and training material about the ethical use of AI technology

Education and training regarding the ethical use of technology should obligatory companions to any SIS-related engineering courses.

Steer technology investment policy towards embedding ethical components of SIS

While companies developing SIS should embed ethical practice in the good and services, we recognise that in some circumstances such practice may be questioned, e.g., in information warfare.

Create a super-regulator for SIS to create and enforce regulation across different jurisdictions.

As SIS are used in many domains already, there is an apparent need for a "super-regulator", one that can operate across those different domains and can interact with existing regulators in those domains that are already regulated. The super-regulator should have adequate enforcement powers.

Update legal definitions in laws and policies regulating technologies to encompass novel issues raised by SIS.

The rules of information warfare need to be written. Existing rules, such as the Geneva Convention, are no longer fit for purpose. The nature of warfare has changed completely. The attackers do not wear uniforms; they may be proxies so that states can deny initiating aggressive behaviour; armies are no longer needed.

Conclusion

It is obvious from the scenario-construction process and from the scenarios themselves that SIS offer many benefits but also raise many ethical issues, especially as regards third-party unauthorised use of personal data, intrusions upon privacy, manipulation of social media, consumers and citizens, and ready-made opinions. AI pervades our societies and economies and will increasingly do so. SIS will affect individuals, communities and societies. The transformational power of SIS far exceeds other regulated products and services, such as cigarettes, motor vehicles, medicines or industrial waste, yet SIS go largely unregulated or only partly regulated in some narrow areas.

None of the scenarios discussed regulatory models or went into any depth on the nature of appropriate regulatory models, but all reflected the need for some form of regulation. The diversity of issues and applications illustrated by the scenarios suggest that regulation needs to be

multidisciplinary in scope. One of the recommendations in the first scenario stated: “Existing regulators should adopt a co-ordinated (co-regulatory) approach to SIS mimicry to ensure harmonised, consistent rules for industry. As holograms like Lucy raise various issues beyond the remit of a single regulator, some mechanism is needed to ensure regulatory harmonisation.”

Most regulators are sector specific, but SIS crosses all sectors. To be effective, a regulator needs enforcement powers. A new regulator with a remit to challenge SIS practices in whatever domain may lead to conflict with sector-specific regulators. So, when policymakers and legislators are thinking about regulatory options, they will need to take into account the sensitivities and the mandates of other regulators (where they exist).

Regulatory options are the subject of future SHERPA deliverables, but suffice to say, based on the scenarios and as an input to those later deliverables, that any new regulator or regulatory scheme will need to consider the inclusion of a wide range of competencies – technical, legal, ethical, organisational, economic, political, cultural – with enforcement powers across sectors and jurisdictions and with the sensitivities and diplomatic skills required to interact with other regulators, some of whom will already have formidable powers of their own.

Furthermore, SIS-powered technologies cross borders. The scenarios do not suggest situations confined to specific countries. Hence, any regulatory scheme will need to have a trans-border, international dimension. Finally, as the scenarios depict, SIS touch the lives of many (even most) consumers and citizens, hence any new regulatory scheme will need to raise public awareness about the dangers of SIS. The benefits speak for themselves, but the dangers hide in black boxes.

More information

For full information on SHERPA scenarios, see our workbook webpage.

Organisations working on future AI scenarios: https://www.eurai.org, https://futureoflife.org/ai-policy/,

Policy papers: Mannino, A., Althaus, D., Erhardt, J., Gloor, L., Hutter, A. and Metzinger, T. (2015). Artificial Intelligence: Opportunities and Risks. Policy paper by the Effective Altruism Foundation (2): 1-16.

Leslie, D. (2018) Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector, Alan Turing Institute

Security Issues, Dangers and Implications of Smart Information Systems

Definition and Focus

This briefing document provides an overview of the study and resulting report on security-related challenges and implications of the growing role of Smart Information Systems (SIS) in our society, with a particular focus on machine learning-based systems. While many SIS risks, weaknesses, and dangers are demonstrated today only in academic experiments, we are already observing a growing *practical* interest in understanding methods and techniques for developing attacks against SIS as well as ways of using those systems for malicious purposes. In the study, we explored how flaws and biases might be introduced into machine learning models powering SIS; how machine learning techniques might, in the future, be used for offensive or malicious purposes; and how SIS can be attacked, and how those attacks can presently be mitigated. Ethical consequences of the flaws, attacks, and defences were considered with a focus on new issues and challenges brought by specific characteristics and properties of SIS differentiating those from traditional ICT systems. The cybersecurity element of SIS, including its weaknesses and benefits in terms of ethical values, was incorporated into D1.4 §4.2.19 (p48-49).

Structure and Scope

The study covers three major topics:

- Bad SIS
- Malicious use of SIS
- Adversarial attacks against SIS and defence approaches

In the first topic, attention focused on flaws arising from incorrect assumptions and poor understanding of machine learning methods applicability to a specific problem, bad design decisions, problems with training data, and mistakes in utilization of SIS.

In the second topic, on applying SIS for malicious purpose, the report focused on methods of intelligent automation used for effectively and efficiently preparing and carrying out attacks and crime; use of SIS for generating and propagating fake news and disinformation; increasing effectiveness of phishing and spam attacks; generation of fake or maliciously modified audio and visual content for impersonation; scams, and various types of social engineering; and finally obfuscation techniques used by malware writers.

The third topic, adversarial attacks against SIS and defence approaches, focused attention on the main types – and notable examples – of attacks against machine learning models: confidentiality, integrity, availability, and replication. Attacker motives were also analysed in the examples, as motive

understanding is crucial for selecting defence strategies. Also within this topic, recent work on detecting and mitigating attacks against SIS was presented, with notes on additional serious challenges for the defenders brought by the nature of machine learning-based systems.

Key Insights

Artificial intelligence has already become powerful to the point that trained models have been withheld from the public over concerns of potential malicious use. This situation parallels vulnerability disclosure, where researchers often need to make a trade-off between disclosing a vulnerability publicly (opening it up for potential abuse) and not disclosing it (risking that attackers will find it before it is fixed).

Malicious Use of and Attacks against SIS

As artificial-intelligence-powered systems become more prevalent, it is natural to assume that adversaries will learn how to use them maliciously and attack them. With this in mind, machine learning will likely be equally effective for both offensive and defensive purposes (in both cyber and kinetic theatres), and hence one may envision an "AI arms race" eventually arising between competing powers. Already we can see that in conventional cyber security complex attack methodologies and tools initially developed by highly resourced threat actors, such as nation states, eventually fall into the hands of criminal organizations and then common cyber criminals. This same trend can be expected for attacks developed against machine learning models. This is particularly concerning given that adversarial attacks against machine learning models are hard to defend against, owing to the fact that there are many ways for attackers to force models into producing incorrect outputs.

Flaws and Bias

The capabilities of machine learning systems are often difficult for the lay person to grasp. Some humans naively equate machine intelligence with human intelligence. As such, people sometimes attempt to solve problems that simply cannot (or should not) be solved with machine learning. Where problems are in principle soluble through machine learning, though, many challenges can be observed. Even knowledgeable practitioners inadvertently build systems that exhibit social bias due to the nature of the training data used. It is very difficult to verify if a machine learning model contains any flaws or biases. Public services exist that are powered by flawed machine learning models. People use these systems without understanding that they are flawed. This problem exists due to the inherent complexity of the field.

The understanding of flaws and vulnerabilities inherent in the design and implementation of systems built on machine learning and the means to validate those systems and to mitigate attacks against them are still in their infancy, complicated – in comparison with traditional systems – by the lack of explainability to the user, heavy dependence on training data, and oftentimes frequent model updating. This field is attracting the attention of researchers, and is likely to grow in the coming years. As understanding in this area improves, so too will the availability and ease-of-use of tools and services designed for attacking these systems.

Ethical Challenges of Defence and Mitigation

Methods of defending machine-learning-based systems against attacks and mitigating malicious use of machine learning may lead to serious ethical issues. For instance, tight security monitoring may negatively affect users' privacy and certain security response activities may weaken their autonomy.

Monopolisation

Companies that devote substantial resources to artificial intelligence research (such as Google, Facebook, Apple, Amazon, etc.) already have a distinct advantage over companies that don't. As those advantages pay off, the gap will continue to widen, perhaps to the point where it is no longer possible to compete in the marketplace. In an effort to remain competitive, companies or organizations may forgo ethical principles, ignore safety concerns, or abandon robustness guidelines in order to push the boundaries of their work, or to ship a product ahead of a competitor.

Disinformation

Text synthesis, image synthesis, and video manipulation techniques have been strongly bolstered by machine learning in recent years. Our ability to generate fake content is far ahead of our ability to detect whether content is real or faked. As such, we expect that machine-learning-powered techniques will be used for social engineering and disinformation in the near future. Disinformation created using these methods will be sophisticated, believable, and extremely difficult to refute.

Conclusion

SIS are playing an increasing role in the cyber-landscape, in terms of both attacking and defending networked systems. This role can be understood in three ways: vulnerabilities introduced by the poor use of SIS, the malicious use of SIS, and the use of SIS in defending against malicious attacks. Each of these raises serious ethical issues for developers and users of SIS, some of which are well-rehearsed in the public imaginary, such as privacy and fake news, but others are not. These under-reported (and, arguably, under-researched) areas would benefit greatly from further research.

Within SHERPA, the ethical issues arising from the highly technical and complex landscape of cybersecurity have provided insights into concerns not raised in the Case Studies (D1.1), the Scenarios (D1.2), or the Human Rights concerns (D1.5). As such, the research carried out in D1.3 has provided unique insight into ethical issues arising from specifically technical issues relating to SIS. This has added a depth to the Ethical Overview report (D1.4).

In response to this, SHERPA WP3 is developing a series of options for dealing with the identifying ethical issues. Guidelines are being developed separately for the user and developer communities, along with implementation recommendations to see these incorporated into standard practice (Tasks 3.2 and 3.4). Regulatory options are similarly being explored to include considerations regarding the creation of new regulatory bodies (Tasks 3.3 and 3.6). Technical options and interventions are also being explored (Task 3.5).

More information

- SHERPA Workbook on Case Studies (https://www.project-sherpa.eu/category/case-studies/)
- SHERPA Deliverable 1.3 (https://doi.org/10.21253/DMU.7951292)

Current Human Rights Frameworks relating to Smart Information Systems

Definition, Problem and Issues

This briefing document provides an overview of 11 concrete challenges for SIS from a human rights' perspective. Setting out the legal and ethical framework relevant to tackling each challenge, both current and future, this research examined potentially negative impacts on human rights in an effort to raise awareness about the nature of the problems.

The 11 challenges were identified through deskwork (academic and grey literature) and by analysing scenario and case study output from SHERPA. The challenges were as follows:

1. Dignity and Care for the Elderly
2. Digital Divide
3. Unemployment
4. Privacy and Data Protection
5. Accountability and Liability
6. Bias and Discrimination
7. Democracy, Freedom of Thought, et al
8. Security, Dual Use and Misuse
9. Health
10. Environment
11. Rights, including Robot Rights

This approach allowed for a conceptualisation of each thematic area/challenge and the determination of potentially optimal solutions for the implementation of the Charter of Fundamental Rights of the European Union and other hard and soft legal instruments which aim at providing a fair, just and equitable society founded on principles of human rights. Each section looked at the meaning of the identified challenges in the framework of law and ethics and discussed how to ensure adequate protection and promotion of principles, which are paramount for the functioning of a democratic society.

Key Insights

The key insights are reproduced below individually from every challenge chapter, including recommendations.

Dignity and Care for the Elderly

Europe currently has an ageing population. By 2050, over 36% of Europe's population is expected to be older than 65 years. Care for the elderly in Europe relies heavily on informal care provided by women, which thereby provide an unacknowledged and unpaid workforce within society. In the UK

alone, current estimates are that the formal care sector will be under-staffed by 400,000 by 2028, putting further pressure on the informal sector.

Social robots might contribute to care for the elderly by reducing the overall (and therefore also women's) informal care load, including formal and informal staffing, as well as helping with loneliness and isolation. However, social robots could also involve a loss of dignity and a loss of privacy. They may also be open to manipulation, reduce important human contact and create harm through malfunctioning.

Recommendation

The Canadian Supreme Court ruled that dignity is too contentious to be applied in court (R. v Kapp (2008) 41 §22*)*. At least five different meanings of dignity can be identified, some contradicting each other. If we take this to be the case, then the twinning of human rights with dignity is not helpful in meeting the challenges of social robots and elderly care. Rather, human rights should be seen as distinct from issues regarding dignity, autonomy, etc. as they stand on their own in international law.

Digital Divide

52% of the world's population do not have access to the Internet. However, within knowledge economies Internet access is necessary for access those economies, including to jobs. The social digital divide between those with access to the Internet and those without increases as citizens without access become increasingly marginalized from political participation.

The technical requirements and costs for SIS are considerable. Hence, new digital divides are predicted as only a limited number of companies will be able to make full use of SIS. Furthermore, the global digital divide will increase as SIS use favours already high-income regions which have the capacity to invest further.

Human rights solutions to the digital divide are not an obvious choice, as there is no human right to Internet access, although such access promotes other human rights (e.g. education). Only the non-legally binding Sustainable Development Goal 9 promotes a significant increase in universal and affordable access to the Internet in least developed countries by 2020.

Recommendation

The world economy is seriously and life-threateningly biased towards the rich. Lack of access to computers, the Internet and SIS is only one symptom of this divide and its solution requires broader governance frameworks based on solidarity.

Unemployment

Unemployment, a major challenge in the world of work, can lead to serious poverty-related problems. Possible displacement through robots could lead to unemployment-related poverty and health risks. Such displacements are predicted to occur across the work spectrum, including in jobs that appeared safe to date. However, there is considerable uncertainty on the impact of automation and robot use on unemployment.

Human rights legislation does not guarantee jobs, only equal access to jobs in the EU and access to social security in the case of unemployment. However, the non-legally binding Sustainable Development Goals specifically aim towards "full and productive employment and decent work for all" (United Nations 2015).

Recommendation

Unemployment is worst where no access to social security can be provided. Human rights solutions to the displacement through robots are therefore most urgently needed in resource-poorer regions.

Privacy and Data Protection

The right to privacy can be broadly defined as a person's right to control access to his or her personal information. SIS fosters the ability to gather, analyse, and combine vast quantities of data from different sources, thus increasing the information-gathering capabilities of actors that use this technology. The potential impact of SIS on privacy is immense. It is increasingly used in information gathering and processing, especially for huge quantities of data, because of (i) the speed of analysing data, (ii) the scale of data that can be processed in a reasonable amount of time, (iii) and the automation of AI processes. The most important challenge faced from the perspective of privacy and data protection is internet users' data breaches, including the exploitation and misuse of data of individuals as consumers but also as voters or simple citizens.

The right to privacy is expressly and broadly protected in Europe in several human rights instruments (ECHR and EU Charter), in the so-called GDPR (General Data Protection Regulation) as well as by the European and national Courts. The movement is led by the Court of Justice of the EU with new rights granted to users, such as the 'right to be forgotten'. The GDPR has fundamentally reshaped the way in which data is handled across every sector of the economy and social life, in Europe but also globally. For example, companies trading in Europe, including foreign multinationals, must have users' permission, given via a clear affirmative action, before they can receive their personal data or override their privacy preferences. The number of data received/generated/controlled must be minimised, processes transparent and supervised.

Recommendation

Privacy is a fundamental human right, although not an absolute one. This means that it can be limited in the public interest. Global economic interests involved in the development of SIS also play a crucial role. A single set of rules addressing all stakeholders is therefore necessary, particularly data processors, controllers, economic actors and/or relevant authorities. The GDPR is a step in the right direction. Further work is however warranted in this area, so that the provisions of the GDPR may be reflected universally. This is likely to be problematic given different emphases placed on the value of privacy viz. the value of business development in non-European countries such as China and the US. A universal approach based on multilateralism or pluralism could provide an adequate regulatory framework.

Accountability and Liability

Advanced SIS systems are able to perform activities which used to be typically and exclusively reserved for humans. The development of certain autonomous and cognitive features (e.g. the ability to learn from experience and take quasi-independent decisions) has made SIS more similar to agents that interact with their environment and they are now able to alter it significantly. The more autonomous such systems are, the less they can be considered mere tools in the hands of other actors such as the

manufacturer, the operator, the owner or the users. Advanced SIS systems can therefore be considered as actors and/or subjects in their own right, raising issues of accountability and responsibility, where the damage caused by a machine cannot be clearly linked to a defect or a human wrongdoing.

EU legislation deals primarily with product and machine safety (Product Liability Directive, Machinery Directive, General Product Directive) and must be adapted to emerging technologies. The European Commission (2017) recommends that the types of damage which users could recover should not be limited. The European Parliament (2017) has expressly stressed the paramount importance of legal certainty on liability for innovators, investors and consumers. The question of civil and criminal liability of advanced SIS systems has been paused, though, as a result of several accidents caused by robots.

Recommendation

If it is assumed that SIS systems/entities can be held liable for a criminal offence through their action, sentencing must follow. Sentences could take the form of (i) the deletion of the SIS software controlling the entity; (ii) the suspension of the SIS-enabled entity for a set period of time; or (iii) the conduct of community service by the SIS system (Gabriel Hallevy, 2010, 199). The rapid development of SIS systems and the dangers deriving therefrom require legal changes to safeguard the welfare of society especially from criminal conduct by those systems, which can lead to serious threats on the social order if not properly regulated.

Bias and Discrimination

Discriminatory AI decisions may be affecting a growing number of cases in finance, health care, and education. Examples include discrimination based on language/accent, racial and gender. Bias and discrimination can also lead to security risks.

Recommendation

People involved in the design, development and use of SIS need to be vigilant about how they design and train machine-learning systems, or one will see ingrained forms of bias built into the artificial intelligence of the future. Making algorithms fair and non-discriminatory is a daunting exercise, but there are steps which could help move society in the right direction.

Democracy, Freedom of Thought et al

The recent scandal over Cambridge Analytica's participation in electoral manipulation and gross breaches of privacy is a very good example which demonstrated that AI and Big Data can pose threats to democracy, as they intervene with freedom of thought. Thus, the danger of manipulation, which is one of many types of attack on elections, is quite clear.

Recommendation

The human rights analysis within Sherpa (D1.5), in the 'Democracy, Freedom of Thought et al' chapter, paid particular attention to the right to freedom of expression as it is strongly related with democracy and freedom of thought. Following the analysis of the challenges that AI poses for freedom of expression, some recommendations in dealing with the issues have to do with Respect for the rule of law, transparency, accountability and others.

Security, Dual Use and Misuse

The rapid development of SIS can pose security threats, requiring the detection and prevention of the misuse of data or the framing of their dual use, for both civil and military purposes. The technological challenges presented by SIS are global phenomena, present in multiple fields of the economy, governance structure and national defence. The sustainability of human rights is under pressure when developing and using SIS, including autonomous weapons and machines with a will of their own, which may come into conflict with humanity and human rights values. A balancing exercise between national security and the protection of human rights must therefore take place.

EU legislation caters for such issues in EU Regulation 428/2009 which sets up the EU 'regime for the control of exports, transfer, brokering and transit of dual-use items'. The rationale is that dual-use items, including software and technology, should be "subject to effective control when they are exported from the European [Union]", for security reasons and for the purpose of ensuring non-proliferation and/or misuse of such technologies, software and data while transferred. A number of soft law instruments are also currently being developed at the EU and national level.

Recommendation

It is important to carry out regular human rights risk assessments in SIS-related fields, mitigating risks of the misuse of human rights through a mapping and policy commitment exercise. The EU needs a good blueprint of European digital integration, leading to enforceable common standards and encompassing multiple aspects of SIS, including but not limited to the Digital Single Market and fundamental human rights.

Health

Medical big data is a particularly rich but sensitive type of big data as it consists of patients' electronic health records (EHR), insurance claims, prescriptions etc. The most notable example of a serious breach to human rights is the Google Deepmind and the Royal Free Trust case.

Recommendation

Experts suggest that is time to better evaluate apps and debate the consequences of substituting big data for accurate data in health research and this is because there are concerns regarding the accuracy of some apps and their technical limitations. Other recommendations in academic literature include data audits, workforce and technical solutions, and federal regulations as umbrella solutions. Under the second (workforce and technical solutions) some specific propositions are: Scribes, Automation, Natural Language Processing, and Best Practices Standards and Training Programs. Under the third (federal regulations) some specific propositions are: Meaningful Use Regulations, The HIPAA[93] Privacy and Security Rules, and The Common Rule.

Environment

Technology experts warn that AI advances could harm our environment. One example is the manufacture of digital devices and other electronics — which go hand-in-hand with the development of AI. The introduction of new technologies necessary for development brings with it irreversible ecological (and other) consequences. Some important facts worth mentioning are that electronic waste is expected to reach 52.2 million tons in 2021, that the UN Environment Program (UNEP) reported in 2015 that 60 to 90 percent of the world's electronic waste is illegally dumped and that in 2014, an estimated 42 million tons of e-waste were generated.

Recommendation

The World Economic Forum, in its publication 'Fourth Industrial Revolution for the Earth Series, Harnessing Artificial Intelligence for the Earth', lists some recommendations. These recommendations include: delivering "responsible AI", collaborating for interdisciplinary solutions, and directing finance for innovation. Recommendations for companies include establishing board-level AI advisory units to ensure that companies' boards understand AI, including safety, ethics, values and governance considerations, embedding environmental considerations into design principles, assuming a leadership role in embedding sustainability principles alongside wider AI safety, ethics, values and governance considerations, and others.

Rights, including Robot Rights

From a legal perspective, it seems that no reference has even been made to rights being directly granted to advanced SIS entities and/or robots, neither in the EU nor nationally in the Member States. The closest the Union has come to this is during the exploration of the possibility of granting such rights in 2017 with the Resolution adopted by the European Parliament on Civil Law Rules on Robotics (European Parliament, 2015).

Legal personhood would allow SIS to hold rights and obligations, be insured individually and even be held liable for damages. By contrast, the idea of 'electronic personality' does not refer to giving human rights to SIS but rather to ensuring that an SIS is recognized as a machine with a human supporting itand/or assuming responsibility. Granting legal personhood would not make SIS virtual people who can get married and benefit from human rights; it would merely put them on an equal footing with corporations, which already have status as 'legal persons' and are treated as such by the courts around the world (European Parliament, 2015).

Recommendation

Given the accelerating deployment of SIS in almost all areas of human life there is an urgent need to develop a special framework of SIS-enabled entities' rights, in accordance with the legal and ethical issues arising, to ensure smooth and effective integration of SIS into society as a whole, including workplaces, schools and hospitals, the economy, police forces, the military and national security.

Rights of SIS could include (i) the right to exist; (ii) the right to integrity; (iii) the right to function and perform one's mission; (iv) right to remedies. Any developing SIS right would need to be placed within the existing and any future legal framework, currently having at its heart the protection of individuals through human rights, while SIS rights in their majority must be granted qualified and not absolute status. Therefore, humans should always come first until time and technology allow the law to provide an equal status to SIS.

Conclusion

A multitude of organisations, researchers and conferences are trying to answer human rights and ethics questions about artificial intelligence and big data (i.e. SIS). From the UK House of Lords to the UNESCO, to the European Commission, activities are manifold.

D1.5 of SHERPA identified and examined 11 pressing challenges created by SIS with regards to human rights. Each challenge was introduced with an overview which demonstrated how the thematic area is relevant and is manifesting in the world and/or in our daily lives. After an overview of current relevant legal and human rights instruments (soft and hard, European and international), a discussion section outlined ideas and suggestions on how human rights can be respected through core legal principles, values and ethics.

The main value of this report are the concrete discussions presented in the light of human rights frameworks, each of which outlines the main positions taken on each challenge, a requirement for moving forward with solutions, some of which are suggested. For more info, see D1.5, p.82-83.

In response to this, SHERPA WP3 is developing a series of options for these next steps. Guidelines are being developed separately for the user and developer communities, along with implementation recommendations to see these incorporated into standard practice (Tasks 3.2 and 3.4). Regulatory options are similarly being explored include considerations regarding the creation of new regulatory bodies (Tasks 3.3 and 3.6). Technical options and interventions are also being explored (Task 3.5)

Taking the work of the human rights framework review forward, the SHERPA project will gain a deeper understanding of key human rights issues. This will in turn inform the development of practical suggestions to shape these communities so that SIS will become more ethical in shape and use in the coming years.

More information

- SHERPA Workbook on D1.5 is not currently available.

Conclusion

This appendix to D1.4 brought together brief overviews and key findings from the other four deliverables in Work Package 1. These deliverables were:

- D1.1 – Case Studies
- D1.2 – Scenarios
- D1.3 – Cybersecurity
- D1.5 – Human Rights Issues

Presenting these findings has clarified how each of these deliverables has contributed to the overview presented in the main body of D1.4. The findings of each of these prior deliverables was used in D1.4, but it is not always obvious where the input has been made. In conclusion to this appendix, this section synthesises the findings of these earlier deliverables.

Awareness of Issues

The deliverables demonstrate that there is a degree of awareness regarding ethical, societal and human rights concerns regarding SIS. No fundamentally new findings were made through the research, but it helped to solidify the appreciation that the existing issues were complete and reassure that practitioners are, as far as we are able to tell, aware of them. It is always possible that new issues may arise of which we or practitioners were hitherto unawares, but for now it appears that we have a solid base of understanding as to what the ethical and societal concerns are, and that this understanding is shared to at least some degree with the stakeholder community. What has been new as a result of this research has been the further appreciation that the degree to which an ethical issue is a worry for stakeholders is not always represented well in academic literature.

Management of Issues

The management of pressing ethical, societal and technical issues is a concern. Through the research it became clear that there is a lack of sophisticated ethical understanding within the technical community and a lack of technical understanding within the public. This is particularly obvious in D1.3 in which technical possibilities arising from SIS were discussed and the ethical aspects explored, but also in D1.1 and D1.2. Various responses have been tried, focusing on the developer community (organisational, technical, and human oversight methods) and on the user community (increasing stakeholder control of data and transparency of data use). The focus on informed consent alone as a means of guaranteeing ethical SIS is clearly insufficient. Regulation of some sort (guidelines or new laws) are needed for developer and user communities to clarify duties and boundaries of responsibility. Also needed are methods to ensure that these regulations are successfully implemented.

Cross-sector Applicability

While most regulators are sector specific, SIS crosses sectors. To be effective, guidelines or a regulator need to be able to cross industrial sectors and come equipped with the power to enforce decisions. New guidelines/regulators with a remit to challenge SIS practices in particular domains may lead to conflict with sector-specific regulators. Hence policymakers and legislators should take this into account when developing solutions. Human rights could be a helpful platform to achieve this cross-cutting exercise, coupled with the normative force that such rights are seen to carry. D1.5 pointed out

the strengths and weaknesses apparent in the human rights framework with respect to SIS as it is currently applied in Europe.

SHERPA Next Steps

Work Package 1 has provided a wealth of information to feed into later work packages. In particular, WP2 will carry out empirical research into public awareness of ethical, societal and technical issues. Identification of stakeholders has begun and continues (Task 2.1). Large-scale surveys (Task 2.3), Delphi studies (Task 2.4) and interviews with stakeholders (Task 2.2) are planned. This includes ongoing engagement with interviewees from the case studies (Task 2.2). We will hence develop a more complete picture of which issues are primary concern to the practitioner community.

Work Package 3 is developing a series of options for these next steps. Guidelines are being developed separately for the user and developer communities, along with implementation recommendations to see these incorporated into standard practice (Tasks 3.2 and 3.4). Regulatory options are similarly being explored include considerations regarding the creation of new regulatory bodies (Tasks 3.3 and 3.6). Based on the findings of WP1, we can say that any new guidelines, regulator or regulatory scheme will need to consider the inclusion of a wide range of competencies – technical, legal, ethical, organisational, economic, political, cultural – and contexts – medical, commercial, public, security – with enforcement powers across sectors and jurisdictions. Furthermore, any solution proposed in WP3 will need to have a trans-border, international dimension. Finally, SIS touch the lives of many (even most) consumers and citizens, hence any new regulatory scheme will need to raise public awareness about the benefits and dangers of SIS.

11. References

Adams R.L., 2017. 10 powerful examples of artificial intelligence in use today. Forbes. Retrieved from: https://www.forbes.com/sites/robertadams/2017/01/10/10-powerful-examples-of-artificial-intelligence-in-use-today/

Adler P, Falk C, Friedler SA, et al. (2016) Auditing black-box models by obscuring features. arXiv:1602.07043 [cs, stat]. Available at: http://arxiv.org/abs/1602.07043

Aggarwal, C. (2011). An Introduction to Social Network Data Analytics in Social Network Data Analytics. Springer. DOI: 10.1007/978-1-4419-8462-3_1.Agrawal R and Srikant R (2000) Privacy-preserving data mining. ACM Sigmod Record. ACM, pp. 439–450. Available at: http://dl.acm.org/citation.cfm?id=335438

Agrawal, N. & Tripathi, A. (2017). Big Data Security and Privacy Issues: A Review. International Journal of Innovative Computer Science & Engineering, 2(4), 12-15.

Alder, G. S. (1998). Ethical Issues in Electronic Performance Monitoring: A Consideration of Deontological and Teleological Perspectives, Springer Business Ethics, 7 (17), pp. 729 – 743.

Alexiou, A., Theocharopoulou, G., Vlamos, P., 2013. Ethical Issues in Neuroinformatics, in: Papadopoulos, H., Andreou, A.S., Iliadis, L., Maglogiannis, I. (Eds.), Artificial Intelligence Applications and Innovations. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 700–705. https://doi.org/10.1007/978-3-642-41142-7_71.

Ananny M (2016) Toward an ethics of algorithms convening, observation, probability, and timeliness. Science, Technology & Human Values 41(1): 93–117.

Anagnostopoulos, I., Zeadally, S., Exposito, E., 2016. Handling big data: research challenges and future directions. J. Supercomput. 72, 1494–1516. https://doi.org/10.1007/s11227-016-1677-z.

Anjuwa, I., Crawford, K. & Schultz, J. (2017). Limitless Worker Surveillance. California Law Review, 105(3), 735-776.

Antle, John, Susan Capalbo, and Laurie Houston, "Using Big Data to Evaluate Agro-Environmental Policies", Choices Vol. 30, Issue 3, Autumn 2015, pp. 1-8.

Ajunwa, I. et al. (2016). Hiring by Algorithm: Predicting and Preventing Disparate Impact. Retrieved from: http://friedler.net/papers/SSRN-id2746078.pdf.

Applin, S.A., and Fischer, M.D. (2015) New technologies and mixed-use convergence: How humans and algorithms are adapting to each other. In: *2015 IEEE international symposium on technology and society (ISTAS)*. Dublin, Ireland: IEEE, pp. 1–6.

Aronson, J. D. (2016). Mobile Phones, Social Media and Big Data in Human Rights Fact-Finding. Possibilities, Challenges, and Limitations. In the Transformation of Human Rights Fact-Finding, ed. by P. Alston & S. Knuckey, 442-459.

Asta, T. A. (2017). Guardians of the Galaxy of Personal Data: Assessing the Threat of Big Data and Examining Potential Corporate and Governmental Solutions. 45 Fla. St. U. L. Rev. 261, 262-312. Retrieved from: https://heinonline.org/HOL/Page?handle=hein.journals/flsulr45&collection=journals&id=274&startid=&endid=325.

Augur, H. (2016). Data tracking in the workplace: Is big data hurting employees? *Dataconomy*. Retrieved from https://dataconomy.com/2016/02/data-tracking-in-the-workplace-is-big-data-hurting-employees/.

Auschitzky, E., Hammer, M., & Rajagopaul, A. (2014). How big data can improve manufacturing. McKinsey & Company, 822.

Ayadi, Moataz El, Mohamed S. Kamel, and Fakhri Karray, “Survey on Speech Emotion Recognition: Features, Classification Schemes, and Databases,” Pattern Recognition, Vol. 44, No. 3, 2011, pp. 572–587., p. 572

Badri, A., Boudreau-Trudel, B. & Souissi, A. S. (2018). Occupational health and safety in the industry 4.0 era: A cause for major concern? Safety Science, 109, 403–411.
Bali, C. (2014). 'The cultural environment: measuring culture with big data' in Theoretical Society. DOI: 10.1007/s11186-014-9216-5.
Balkin, J. M. (2018). Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation. Regulation, 51 U.C.D. L. Rev. 1149.
Balkin, J. M. (2017). 2016 Sidley Austin Distinguished Lecture on Big Data Law and Policy: The Three Laws of Robotics in the Age of Big Data. *Ohio St. LJ*, *78*, 1217.
Barbero, M., Counter, J., Jackers, R., Moueddene, K., Render, E, Stevens, W., … & Versteele, D. (2016). Big data analytics for policy making. A study prepared for the European Commission DG INFORMATICS by Deloitte.
Barbier, G. and Liu, H., (2011). Data Mining in Social Media in Social Network Data Analytics by Charu C. Aggarwal. Springer. DOI: 10.1007/978-1-4419-8462-3_12.
Barnet, B. (2009). Idiomedia: the rise of personalized, aggregated content. Continuum: the Journal of Media & Cultural Studies, 23 (1), 93–99.
Barocas, S., & Nissenbaum, H. (2014). Big Data's end run around procedural privacy protections. Communications of the ACM 57(11): 31-33. DOI: 10.1145/2668897.
Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. California Law Review, 104(671): 671-732. DOI: http://dx.doi.org/10.15779/Z38BG31.
Barocas, S. (2014) Data mining and the discourse on discrimination. Available at: https://dataethics.github.io/proceed ings/DataMiningandtheDiscourseOnDiscrimination.pdf.
Barocas, S and Selbst, A.D. (2015) Big data's disparate impact. SSRN Scholarly Paper, Rochester, NY: Social Science Research Network. Available at: http://papers.ssrn.com/ abstract=2477899.
Bartoli, A., J. Hernández-Serrano, M. Soriano, M. Dohler, A. Kountouris, and D. Barthel, "Security and Privacy in your Smart City", Proceedings of the Barcelona Smart Cities Congress, Vol. 292, 2011, pp. 1-6.
Bates, D.W., Saria, S., Ohno-Machado, L., Shah, A., Escobar, G., 2014. Big Data In Health Care: Using Analytics To Identify And Manage High-Risk And High-Cost Patients. Health Aff. (Millwood) 33, 1123–1131. https://doi.org/10.1377/hlthaff.2014.0041.
Batty, Michael, "Artificial Intelligence and Smart Cities", Environment and Planning B: Urban Analytics and City Science, Vol. 45, Issue 1, 2018, pp. 3-6.
Batty, Michael, Kay W. Axhausen, Fosca Giannotti, Alexei Pozdnoukhov, Armando Bazzani, Monica Wachowicz, Georgios Ouzounis, and Yuval Portugali, "Smart Cities of the Future", The European Physical Journal Special Topics, Vol. 214, Issue 1, 2012, pp. 481-518.
Bederson, B.B., Lee, B., Sherman, R.M., Herrnson, P.S., Niemi, R.G., 2003. Electronic Voting System Usability Issues, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '03. ACM, New York, NY, USA, pp. 145–152. https://doi.org/10.1145/642611.642638.
Bellazzi, R. (2014). 'Big Data and Biomedical Informatics: A Challenging Opportunity' in IMIA Yearbook of Medical Informatics 8-13.
Bentley, R. Alexander ; O'Brien, Michael J. & Brock, William A. (2014). Mapping collective behavior in the big-data era. Behavioral and Brain Sciences 37 (1):63-76.
Beresford, M. J. (2010). Medical reductionism: lessons from the great philosophers. QJM: An International Journal of Medicine, 103(9), 721-724.
Berlin, I. (1969). Two concepts of liberty. In I. Berlin (Ed.), *Four essays on liberty* (pp. 118–172). London: Oxford University Press.
Bibri, Simon Elias, "Data Science for Urban Sustainability: Data Mining and Data-Analytic Thinking in the Next Wave of City Analytics", Smart Sustainable Cities of the Future, Springer, Cham, 2018, pp. 189-246.
Bifet, A. (2013). 'Mining Big Data in Real Time' in Informatica 37: 15-20.
Binns, R. (2017). Algorithmic Accountability and Public Reason. Philosophy & Technology, 1-14.

Birrer, F. A. J. (2005) Data mining to combat terrorism and the roots of privacy concerns. Ethics and Information Technology 7(4): 211–220.
Bharadwaj, R. (2018). How Insurance Leaders Can Prepare for Artificial Intelligence Today -. [online] TechEmergence. Available at: https://www.techemergence.com/insurance-leaders-can-prepare-artificial-intelligence-today/ [Accessed 9 Nov. 2018].
Blodgett, Su Lin, Lisa Green, and Brendan O'Connor, "Demographic Dialectal Variation in Social Media: A Case Study of African-American English," Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016.
Bohn, J., V. Coroamă, M. Langheinrich, F. Mattern, and M. Rohs, "Social, Economic, and Ethical Implications of Ambient Intelligence and Ubiquitous Computing," Ambient Intelligence, 2005, pp. 5–29.
Bolukbasi, Tolga, Kai-Wei Chang, James Y. Zou, Venkatesh Saligrama, and Adam T. Kalai, "Man is to computer programmer as woman is to homemaker? debiasing word embeddings," In Advances in neural information processing systems, pp. 4349-4357, 2016.
Bovens, L. (2008). The ethics of nudge. In T. Grune-Yanoff and S. O. Hansson (Eds.), *Preference change: Approaches from philosophy, economics and psychology* (pp. 207–219). Dordrecht: Springer.
Boyd, D. (2016). Untangling research and practice: What Facebook's "emotional contagion" study teaches us. Research Ethics 12(1): 4-13. DOI: 10.1177/1747016115583379.
Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information, communication & society, 15(5), 662-679.
Bossmann, J., 2016. Top 9 ethical issues in artificial intelligence. World Economic Forum. Retrieved from: https://www.weforum.org/agenda/2016/10/top-10-ethical-issues-in-artificial-intelligence/
Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and Information Technology* 15(3): 209–227.
Bozdag, E., and Van den Hoven, J. (2015). Breaking the filter bubble: democracy and design. Ethics and Information Technology 17, no.4.
Brayne, S. (2017). Big Data Surveillance: The Case of Policing. American Sociological Review, 82(5), 977-1008.
Brewster, B. et al. (2015). Cybercrime: Attack Motivations and Implications for Big Data and National Security. In: B. Akhgar, G. B. Saathoff, H. R. Arabnia, R. Hill, A. Staniforth, P. S. Bayerl (eds.), Application of Big Data for National Security: A Practitioner's Guide to Emerging Technologies. Oxford, Butterworth Heinemann (Elsevier), 108-129.
Brin, S. & Page, L. (2000). The Anatomy of a Large-Scale Hypertextual Web Search Engine. Retrieved from: http://infolab.stanford.edu/~backrub/google.html.
Browniee, J., 2013. A Tour of Machine Learning Algorithms. Retrieved from: https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/
Bostrom, Nick. (2003) "Ethical issues in advanced artificial intelligence." *Science Fiction and Philosophy: From Time Travel to Superintelligence*: 277-284.
Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. The Cambridge handbook of artificial intelligence, 1, 316-334.
Bovens, L. (2008). The ethics of nudge. In T. Grune-Yanoff and S. O. Hansson (Eds.), Preference change: Approaches from philosophy, economics and psychology (pp. 207–219). Dordrecht: Springer.
Bozdag, E. (2013). Bias in algorithmic filtering and personalization. Ethics and Information Technology 15(3): 209–227.
Braun, A., & Garriga, G. (2018). Consumer Journey Analytics in the Context of Data Privacy and Ethics. In Digital Marketplaces Unleashed (pp. 663-674). Springer, Berlin, Heidelberg.
Brey, P. (2010). Values in Technology and Disclosive Computer Ethics. In L. Floridi (Ed.), *The Cambridge Handbook of Information and Computer Ethics* (pp. 41-58). Cambridge: Cambridge University Press.
Broeders, D. et al. (2017). Big Data and security policies: Towards a framework for regulating the phases of analytics and use of Big Data. Computer Law and Security Review, 33 (3), 309-323.

Brooker, P., Dutton, W., & Greiffenhagen, C. (2017). What would Wittgenstein say about social media?. Qualitative Research, 17(6), 610-626.
Bryson, J., & Winfield, A. (2017). Standardizing ethical design for artificial intelligence and autonomous systems. *Computer*, *50*(5), 116-119.
Burrel, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society, 3(1).
Burnett, S., Feamster, N., 2015. Encore: Lightweight Measurement of Web Censorship with Cross-Origin Requests, in: Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication, SIGCOMM '15. ACM, New York, NY, USA, pp. 653–667. https://doi.org/10.1145/2785956.2787485.
Busch, D. (2016) "MiFID II: Regulating High Frequency Trading, Other Forms of Algorithmic Trading and Direct Electronic Market Access" [online] Available at: https://www.law.ox.ac.uk/business-law-blog/blog/2016/07/mifid-ii-regulating-high-frequency-trading-other-forms-algorithmic [Accessed 19 February 2019].
Byarugaba Agaba, G, et al., "Big Data and Positive Social Change in the Developing World: A White Paper for Practitioners and Researchers", Rockefeller Foundation Bellagio Centre Conference, 2014. https://www.rockefellerfoundation.org/report/big-data-and-positive-social-change-in-the-developing-world/.
Byers, J.W., 2015. Encore: Lightweight Measurement of Web Censorship with Cross-Origin Requests–Public Review. Technical Report http://conferences. sigcomm. org/sigcomm/2015/pdf/reviews.
Calders T, Kamiran F and Pechenizkiy M (2009) Building classifiers with independency constraints. In: Data mining workshops, 2009. ICDMW'09. IEEE international confer- ence on, Miami, USA.
Calders T and Verwer S (2010) Three naive Bayes approaches for discrimination-free classification. Data Mining and Knowledge Discovery 21(2): 277–292.
Calude, C. S., & Longo, G. (2017). The deluge of spurious correlations in big data. Foundations of science, 22(3), 595-612.
Carbonell, Isabelle, "The Ethics of Big Data in Big Agriculture", Internet Policy Review, Vol. 5, Issue 1, March 2016, pp. 1-13.
Cardona B (2008) 'Healthy ageing' policies and anti-ageing ideologies and practices: On the exercise of responsibility. Medicine, Health Care and Philosophy 11(4): 475–483.
Cardullo, Paulo, and Rob Kitchin, "Being a 'Citizen' in the Smart City: Up and Down the Scaffold of Smart Citizen Participation", The Programmable City Working Paper 30, SocArXiv Website, 15th May 2017.
Carlini, Nicholas, Pratyush Mishra, Tavish Vaidya, Yuankai Zhang, Micah Sherr, Clay Shields, David Wagner, and Wenchao Zhou, "Hidden voice commands," 25th {USENIX} Security Symposium ({USENIX} Security 16), pp. 513-530, 2016.
Capgemini Consulting, "Unleashing the Potential of Artificial Intelligence in the Public Sector", Capgemini website, 2017, retrieved 27th July 2018: https://www.capgemini.com/consulting/wp-content/uploads/sites/30/2017/10/ai-in-public-sector.pdf.
Carolan, Michael, "Publicising Food: Big Data, Precision Agriculture, and Co-Experimental Techniques of Addition", Sociologia Ruralis Vol. 57, Issue 2, December 2017, pp. 135-54..
Cave, J. (2017) Get with the Program: Fintech Meets Regtech in the Light-Touch Sandbox. Available at SSRN: https://ssrn.com/abstract=2944249
Cawley, K., 2014. When to Use Supervised and Unsupervised Data Mining. Productive Analytics Times. Retrieved from: https://www.predictiveanalyticsworld.com/patimes/use-supervised-unsupervised-data-mining/4046/.
Chae, B. K. (2015). Insights from hashtag supply chain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research. International Journal of Production Economics, 165, 247-259.
Chandler, N., Hostmann, B., Rayner, N., & Herschel, G. (2011). Gartner's business analytics framework. Gartner, Inc.

Chang, F., Dean, J., Ghemawat, S., Hsieh, W. C., Wallach, D. A., Burrows, M., Chandra, T., Fikes, A., and Gruber, R. E. (2008). Bigtable: A distributed storage system for structured data. ACM Transactions on Computer Systems (TOCS), 26(2):4.
Chen, M., Mao, S., and Liu, Y. (2014). Big data: A survey. Mobile networks and applications, 19(2):171-209.
Cherepanov, A. and Cherepanov, A. (2016). BlackEnergy by the SSHBearDoor: attacks against Ukrainian news media and electric industry. [online] WeLiveSecurity. Available at: https://www.welivesecurity.com/2016/01/03/blackenergy-sshbeardoor-details-2015-attacks-ukrainian-news-media-electric-industry/ [Accessed 25 Jan. 2019].
Cherepanov, A. and Cherepanov, A. (2017). Industroyer: Biggest threat to industrial control systems since Stuxnet. [online] WeLiveSecurity. Available at: https://www.welivesecurity.com/2016/01/03/blackenergy-sshbeardoor-details-2015-attacks-ukrainian-news-media-electric-industry/ [Accessed 25 Jan. 2019].
Chow-White, P. A., MacAulay, M., Charters, A., & Chow, P. (2015). From the bench to the bedside in the big data age: ethics and practices of consent and privacy for clinical genomics and personalized medicine. Ethics and Information Technology, 17(3), 189-200.
Chory, R., Vela, L. and Avtgis, T. (2016). Organizational Surveillance of Computer-Mediated Workplace Communication: Employee Privacy Concerns and Responses. Employee Responsibilities and Rights Journal, 28(1), pp.23-43.
Chourabi, Hafedh, Taewoo Nam, Shawn Walker, J. Ramon Gil-Garcia, Sehl Mellouli, Karine Nahon, Theresa A. Pardo, and Hans Jochen Scholl, "Understanding Smart Cities: An Integrative Framework", 45th Hawaii International Conference on System Science (HICSS), 2012, pp. 2289-2297.
Çintiriz, H., Buhur, M. N. & Şensoy, E. (2015). Military Implications of Big Data. In: Ş. Çetin, K. Göztepe, A. Kayaalp (eds.), Proceeding of the International Conference on Military and Security Studies 2015. Istanbul, ICMSS, 55-60.
Citron, D. K., & Pasquale, F. (2014). The scored society: due process for automated predictions. Washington Law Review, 89(1): 1-33.
Clarke, A., and Margetts, H. (2014) Governments and citizens getting to know each other?
Open, closed, and big data in public management reform. *Policy & Internet* 6(4): 393–417.
Clayton, E.W. (2005). Informed consent and biobanks. Journal of Law, Medicine & Ethics33(11): 15-21.
Cleeff, A. van, Pieters, W., Wieringa, R.J., 2009. Security Implications of Virtualization: A Literature Study, in: 2009 International Conference on Computational Science and Engineering. Presented at the 2009 International Conference on Computational Science and Engineering, pp. 353–358. https://doi.org/10.1109/CSE.2009.267.
CNIL, "How can Humans Keep the Upper Hand? The Ethical Matters Raised by Algorithms and Artificial Intelligence", Report on the Public Debate Led by th French Data Protection Authority (CNIL) as Part of the Ethical Discussion Assignment Set by the Digital Republic Bill, December 2017, available here: https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf.
CNIL, "Connected Vehicles: A Compliance Package for a Responsible Use of Data", CNIL [website], February 13th, 2018, available here: https://www.cnil.fr/en/connected-vehicles-compliance-package-responsible-use-data.
Coble, Keith H, Ashok K. Mishra, Shannon Ferrell and Terry Griffin, "Big Data in Agriculture: A Challenge for the Future", Applied Economic Perspectives and Policy Vol. 40, Issue 1, May 2018, pp. 79-96.
Coll, S. (2014). Power, knowledge, and the subjects of privacy: understanding privacy as the ally of surveillance. Information, Communication & Society, 17(10), 1250-1263.
Condliffe, Jamie, "AI Has Beaten Humans at Lip-reading", Technology Review, November 21st 2016, https://www.technologyreview.com/s/602949/ai-has-beaten-humans-at-lip-reading/
Contissa, Giuseppe, Francesca Lagioia, and Giovanni Sartor, "The Ethical Knob: ethically-customisable automated vehicles and the law", Artificial Intelligence and Law, Vol. 25, Issue 3, 2017, pp. 365-378.

Cornet, G. (2013). 'Robot companions and ethics: A pragmatic approach of ethical design'. *Journal International de Bioéthique*, pp. 49-58.
Connolly, R. (2015). Dataveillance in the Workplace: Privacy Threat or Market Imperative? Retrieved from: https://www.researchgate.net/publication/290949985_Dataveillance_in_the_workplace_Privacy_threat_or_market_imperative.
Couldry, N., & Powell, A. (2014). 'Big Data from the bottom up' in *Big Data & Society*, pp. 1-5. DOI: 10.1177/2053951714539277.
Cohen IG, Amarasingham R, Shah A, et al. (2014) The legal and ethical concerns that arise from using complex pre- dictive analytics in health care. Health Affairs 33(7): 1139–1147.
Collectif, Cerna. "Research Ethics in Machine Learning," PhD diss., CERNA; ALLISTENE, 2018.
Cookson, C., 2016. AI and robots threaten to unleash mass unemployment, scientists warn. Retrieved from: https://www.ft.com/content/063c1176-d29a-11e5-969e-9d801cf5e15b.
Copeland, M. 2016. What's the Difference Between Artificial Intelligence, Machine Learning, and Deep Learning?. Financial Times. Retrieved from: https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/.
Costa, F. (2014). 'Big Data in Biomedicine' in Drug Discovery Today 19(4): 433-440.
Crawford, K., Gray, M. L., & Miltner, K. (2014). Big Data critiquing Big Data: Politics, ethics, epistemology special section introduction. International Journal of Communication, 8, 10.
Crawford, K. & Schultz, J. (2014). Big data and due process: toward a framework to redress predictive privacy harms. Boston College Law Review 55(93): 93-128.
D'Orazio Paola, (2017). Big data and complexity: Is macroeconomics heading toward a new paradigm? Journal of Economic Methodology 24 (4):410-429.
Danaher, J. (2016). The threat of algocracy: Reality, resistance and accommodation. Philosophy & Technology, 29(3), 245-268.
Danna A and Gandy OH Jr (2002) All that glitters is not gold: Digging beneath the surface of data mining. Journal of Business Ethics 40(4): 373–38.
Darr, Matt, "Big Data—the Catalyst for a Transformation to Digital Agriculture", Proceedings of the 26th Annual Integrated Crop Management Conference, 2014.
Datta, A., Tschantz M. C. & Datta, A. (2015). Automated Experiments on Ad Privacy Settings A Tale of Opacity, Choice, and Discrimination. Proceedings on Privacy Enhancing Technologies 2015 (1), 92–112.
Dean, J. and Ghemawat, S. (2008). Mapreduce: simpli_ed data processing on large clusters. Communications of the ACM, 51(1):107-113.
De Laat, P. B. (2017). Algorithmic Decision-Making Based on Machine Learning from BigData: Can Transparency Restore Accountability?. Philosophy & Technology, 1-17.
Deloitte Digital (2017). From mystery to mastery: Unlocking the business value of Artificial Intelligence in the insurance industry. [online] Www2.deloitte.com. Available at: https://www2.deloitte.com/content/dam/Deloitte/nl/Documents/financial-services/deloitte-nl-artificial-intelligence-in-insurance-whitepaper.pdf [Accessed 9 Nov. 2018].
Diakopoulos, N. (2017). 'Enabling Accountability of Algorithmic Media: Transparency as a Constructive and Critical Lens' in *Transparent Data Mining for Big and Small Data*, (Ed.) Tania Cerquitelli, Daniele Quercia and Frank Pasquale. Springer.
Diakopoulos N (2015) Algorithmic accountability: Journalistic investigation of computational power struc-tures. Digital Journalism 3(3): 398–415.
Dierksmeier, Claus and Peter Seele. "Cryptocurrencies and Business Ethics" Journal of Business Ethics, Vol. 152(1), 2016. pp.1-14.
DiMaggio, P. (2015). 'Adapting computational text analysis to social science (and vice versa)' in Big Data & Society pp. 1-5. DOI: 10.1177/2053951715602908.
DiResta, R., Little, J., Morgan, J., Neudert, L.M., Nimmo, B. (2017). The Bots That Are Changing Politics. *Motherboard*. Retrieved form https://motherboard.vice.com/en_us/article/mb37k4/twitter-facebook-google-bots-misinformation-changing-politics

Dorey, C. M. (2016). Rethinking the ethical approach to health information management through narration: pertinence of Ricœur's 'little ethics'. Medicine, Health Care and Philosophy, 19(4), 531-543.
Doshi-Velez, Finale, and Been Kim, "Towards a rigorous science of interpretable machine learning," arXiv preprint arXiv:1702.08608, 2017.
Doughty, A. (2014). Surveillance, Big Data Analytics and the Death of Privacy. College Quarterly, 17(3), 1-21.
Dubey, R. (2017). 'Can big data and predictive analytics improve social and environmental sustainability?' in *Technological Forecasting and Social Change*. DOI: 10.1016/j.techfore.2017.06.020.
Duhigg, C., 2012. How Companies Learn Your Secrets. The New York Times.
Dutt, R. (2018). Why artificial intelligence in health care is harder than you would think. [online] InfoWorld. Available at: https://www.infoworld.com/article/3269197/artificial-intelligence/why-artificial-intelligence-in-health-care-is-harder-than-you-would-think.html [Accessed 9 Nov. 2018].
Dwork C, Hardt M, Pitassi T, et al. (2011) Fairness through awareness. arXiv:1104.3913 [cs]. Available at: http://arxi- v.org/abs/1104.3913 (accessed 15 February 2016).
Dwork, Cynthia, and Aaron Roth, "The algorithmic foundations of differential privacy," Foundations and Trends® in Theoretical Computer Science, Vol. 9, No. 3–4, 2014, pp. 211-407.
Eastman, R., Versace, M. & Webber, A. (2015). Big Data and Predictive Analytics: On the Cybersecurity Front Line. IDC. Retrieved from: https://informationsecurity.report/Resources/Whitepapers/11e1f237-7156-4627-85f6-d781cf74ef0f_Big%20Data%20and%20Predictive%20Analytics%20On%20the%20CybersecurityFront%20Line.pdf.
Edwards, L., Martin, L. & Henderson, T. (2018). Employee Surveillance: The Road to Surveillance is Paved with Good Intention, SSRN Electronic Journal. DOI: 10.2139/ssrn.3234382. Retrieved from: https://www.researchgate.net/publication/328467820_Employee_Surveillance_The_Road_to_Surveillance_is_Paved_with_Good_Intentions.
Eder-Neuhauser, P., Zseby, T., Fabini, J. and Vormayr, G. (2017). Cyber attack models for smart grid environments. Sustainable Energy, Grids and Networks, 12, pp.10-29.
Einav, L. & Levin, J. (2014). Economics in the age of big data. Science, 346(6210).
Ekbia, H, Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., Suri, V. R., Tsou, A, Weingart, S., and Sugimoto, C. (2015). 'Big data, bigger dilemmas: A critical review' in *Journal of the Association for Information Science and Technology* 66.8 (2015): 1523-1545.
Elgendy, N. and Elragal, A. (2014). Big data analytics: a literature review paper. In Industrial Conference on Data Mining, pages 214-227. Springer.
Elmaghraby, Adel S., and Michael M. Losavio, "Cyber Security Challenges in Smart Cities: Safety, Security and Privacy", Journal of Advanced Research, Vol. 5, Issue 4, 2014, pp. 491-497.
Ensign, D., Friedler, S. A., Neville, S., Scheidegger, C., & Venkatasubramanian, S. (2018). Runaway feedback loops in predictive policing. Proceedings of Machine Learning Research 81, 1–12.
Esposito, E. (2017). 'Algorithmic memory and the right to be forgotten on the web' in *Big Data & Society*, pp. 1-11. DOI: 10.1177/2053951717703996
Fairfield, J., & Shtein, H. (2014). Big data, big problems: Emerging issues in the ethics of data science and journalism. Journal of Mass Media Ethics, 29(1), 38-51.
Fairweather, N., B. (1999). Surveillance in Employment: the Case of Teleworking, Journal of Business Ethics , 22(1). Pp. 39-49.
Fan, J. (2016). Threat Intelligence in the Era of Big Data. Library and Information Service, 3. Retrieved from: http://english.sass.org.cn:8001/u/cms/www/201703/311030295o1b.pdf.
Fan, W., and Gordon, M. D., (2014). 'Unveiling the Power of Social Media Analytics' in Communications of the ACM. DOI: 10.1145/2602574.
Faruqui, A. (2010) The Ethics of Dynamic Pricing Available at: http://gridsolar.com/smartgrid/docket2010-267/Attachment_11.pdf [Accessed 25 Jan. 2019].

Felt, M. (2016). 'Social media and the social sciences: How researchers employ Big Data analytics' in Big Data & Society pp. 1-15. DOI: 10.1177/2053951716645828.
Feki, M., Boughzala, I., & Wamba, S. F. (2016, January). Big Data Analytics-enabled Supply Chain Transformation: A Literature Review. In 2016 49th Hawaii International Conference on System Sciences (HICSS) (pp. 1123-1132). IEEE.
Ferris, Jody L, "Data Privacy and Protection in the Agriculture Industry: Is Federal Regulation Necessary." Minn. JL Sci. & Tech, Vol. 18, Issue 1, 2017, pp. 309-342.
Floridi, L. (2005). The ontological interpretation of informational privacy. Ethics and Information Technology, 7(4), 185-200.
Floridi, L. (2008) The method of levels of abstraction. Minds and Machines 18(3): 303–329.
Floridi, L. (2012). Big Data and their epistemological challenge. Philosophy & Technology 25(4): 435-437. DOI: 10.1007/s13347-012-0093-4.
Floridi, L. (2014). Open data, data protection, and group privacy. Philosophy & Technology 27(1): 1-3. DOI: 10.1007/s13347-014-0157-8.
Floridi L (2014) *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*. Oxford: OUP
Floridi, L. Taddeo , M. (2016). What is data ethics? Philosophical Transactions of the Royal Society A, 1-5. DOI: 10.1098/rsta.2016.0360.
Floridi, L and Sanders, J. W. (2004b) On the morality of artifi- cial agents. Minds and Machines 14(3). Available at: http:// dl.acm.org/citation.cfm?id=1011949.1011964 (accessed 1 August 2004).
Foggan, L. and Panagakos, E. (2018). AI in insurance: New opportunities come with new worries | PropertyCasualty360. [online] PropertyCasualty360. Available at: https://www.propertycasualty360.com/2018/05/08/ai-in-insurance-new-opportunities-come-with-new-wo/?slreturn=20181009024138 [Accessed 9 Nov. 2018].
Foster, V., Young, A., 2011. The use of routinely collected patient data for research: A critical review. Health 16, 448–463. https://doi.org/10.1177/1363459311425513.
Foth, Marcus, "The Software-Sorted City: Big Data & Algorithms", Odendaal, Nancy and Alessandro Aurigi (Eds.), Digital Cities 10: Towards a Localised Socio-Technical Understanding of the 'Real' Smart City, 26 June 2017, Troyes, France, 2017.
Frey C. B., Osborne M. A., 2013. The Future of Employment: how susceptible are jobs to computerisation? Retrieved from: https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf.
Friedler, S. A., Scheidegger, C., & Venkatasubramanian, S. (2016). On the (im) possibility of fairness. arXiv preprint arXiv:1609.07236.
Friedman, B. & Nissenbaum, H. (1996). Bias in Computer Systems. ACM Transactions on Information Systems, 14(3), 330–347.
Fuchs, C. (2011). New Media, Web 2.0 and Surveillance. Sociology Compass, 5(2), 134–147.
Fule P and Roddick JF (2004) Detecting privacy and ethical sensitivity in data mining results. In: Proceedings of the 27th Australasian conference on computer science – Volume 26, Dunedin, New Zealand, Australian Computer Society, Inc., pp. 159–166. Available at: http://dl.acm.org/citation.cfm?id=979942.
Fuller, Michael (2015). Big data: New science, new challenges, new dialogical opportunities. Zygon 50 (3):569-582.
Galdon-Clavell, Gemma, "(Not So) Smart Cities? The Drivers, Impact and Risks of Surveillance-Enabled Smart Environments", Science and Public Policy, Vol. 40, 2013, p. pp. 717-723.
Gams, Matjaz, Irene Yu-Hua Gu, Aki Härmä, Andrés Muñoz, and Vincent Tam, "Artificial Intelligence and Ambient Intelligence," Journal of Ambient Intelligence and Smart Environments, Vol. 11, No. 1, 2019, pp. 71-86.
Gandomi, A. and Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2):137-144.
Garattini, C., Raffle, J., Aisyah, D. N., Sartain, F., & Kozlakidis, Z. (2017). Big Data Analytics, Infectious Diseases and Associated Ethical Impacts. Philosophy &

Technology, 1-17.
Ghemawat, S., Gobio_, H., and Leung, S.-T. (2003). The Google file system, volume 37. ACM.
Gillespie, T. (2014). The relevance of algorithms. Media technologies: Essays on communication, materiality, and society, 167.
Gitelman, L., & Jackson, V. (2013). Introduction. In Gitelman, L. (Ed.). "Raw Data" is an oxymoron. (1-14). USA: Massachusetts Institute of Technology.
Glaeser, Edward L., Scott Duke Kominers, Michael Luca, and Nikhil Naik, "Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life", Economic Inquiry, Vol. 56, Issue 1, 2018, pp. 114-137.
Glasmeier, Amy, and Susan Christopherson, "Thinking About Smart Cities", Cambridge Journal of Regions, Economy and Society, Vol. 8, 2015, pp. 3-12.
Glass, A., McGuinness, D.L., Wolverton, M., 2008. Toward establishing trust in adaptive agents, in: Proceedings of the 13th International Conference on Intelligent User Interfaces. ACM, pp. 227–236.
Gluckman, P. (2017). Using evidence to inform social policy: The role of citizen-based analytics. *Office of the Prime Minister's chief Science Advisor*.
Goodman, Bryce, and Seth Flaxman, "European Union regulations on algorithmic decision-making and a "right to explanation"," AI Magazine, Vol. 38, No. 3, 2017, pp. 50-57.
Grafanaki, S. (2017). Autonomy Challenges in the Age of Big Data. Fordham Intellectual Property, Media and Entertainment Law Journal, 27 (4), 803-868.
Gray, J. (2019). Pricing and trust: a utilities conundrum - Utility Week. [online] Utility Week. Available at: https://utilityweek.co.uk/pricing-trust-utilities-conundrum/ [Accessed 25 Jan. 2019].
Golder, S. A. & Macy, M. W., (2014). 'Digital Footprints: Opportunities and Challenges for Online Social Research' in *The Annual Review of Sociology* 40: 129-152. DOI: 10.1146/annurev-soc-071913-043145.
Goldman E (2006) Search engine bias and the demise of search engine utopianism. Yale Journal of Law & Technology 8: 188–200.
Granka LA (2010) The politics of search: A decade retro- spective. The Information Society 26(5): 364–374.
Guillermin, M., & Magnin, T. (2017). Big Data for Biomedical Research and Personalised Medicine: an Epistemological and Ethical Cross-Analysis. Human and Social Studies, 6(3), 13-36.
Guiora, A.N., 2017. Cybersecurity: Geopolitics, Law, and Policy, 1 edition. ed. Routledge, Boca Raton, FL.
Gupta, A., 2018. The Evolution of Fraud: Ethical Implications in The Age Of Large-Scale Data Breaches And Widespread Artificial Intelligence Solutions Deployment 7.
Hacker, P., 2018. Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination Under EU Law (SSRN Scholarly Paper No. ID 3164973). Social Science Research Network, Rochester, NY.
Hacking, I. (2001). An Introduction to Probability and Inductive Logic. Cambridge University Press.
Hampton, S., Strasser, C., Tewksbury, J., Gram, W. K., Budden, A. E., Batcheller, A. L., Duke, C. S., & Porter, J. H. (2013). 'Big data and the future of ecology' in *Frontiers in Ecology and the Environment* 11(3): 156-158.
Hansson, S.O., 2013. The Ethics of Risk: Ethical Analysis in an Uncertain World. Palgrave Macmillan.
Haridas, M. (2015). Redefining Military Intelligence Using Big Data Analytics. Scholar Warrior, 72-78. Retrieved from: https://www.claws.in/images/journals_doc/1511401708_RedefiningMilitaryIntelligenceUsingBigDataAnalytics.pdf.
Harman, G. (1999). Moral philosophy meets social psychology: Virtue ethics and the fundamental attribution error. Proceedings of the Aristotelian society , 99, 315-331.
Harris, M. (2017) "Impact Of Artificial Intelligence And Machine Learning on Trading And Investing". Available at: https://towardsdatascience.com/impact-of-artificial-intelligence-and-machine-learning-on-trading-and-investing-7175ef2ad64e [Accessed 19 February, 2019]

Harrison, P., & Gray, C. (2010). The ethical and policy implications of profiling 'vulnerable' customers. International journal of consumer studies , 34 (4), 437-442.
Harwell, Drew, "Google bans development of artificial intelligence used in weaponry", The Washington Post, 7 June 2018.
Hashem, Ibrahim Abaker Targio, Victor Chang, Nor Badrul Anuar, Kayode Adewole, Ibrar Yaqoob, Abdullah Gani, Ejaz Ahmed, and Haruna Chiroma, "The Role of Big Data in Smart City", International Journal of Information Management, Vol. 36, Issue 5, 2016, pp. 748-758.
Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., and Khan, S. U. (2015). The rise of "big data" on cloud computing: Review and open research issues. Information Systems, 47:98-115.
Helbing, D. (2015). Societal, Economic, Ethical and Legal Challenges of the Digital Revolution: From Big Data to Deep Learning, Artificial Intelligence, and Manipulative Technologies, 47-72. Retrieved from: https://arxiv.org/abs/1504.03751v1.
Helbing, D. et al. (2018). Will Democracy Survive Big Data and Artificial Intelligence? Towards Digital Enlightenment, 73-98.
Helbing, D., Frey, B. S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., ... & Zwitter, A. (2019). Will democracy survive big data and artificial intelligence?. In Towards Digital Enlightenment (pp. 73-98). Springer, Cham.
Hilbert, M. (2016). 'Big Data for Development: A Review of Promises and Challenges' in *Development Policy Review*, 34(1): 135-174. http://doi.org/10.1111/dpr.12142
Hildebrandt, M. (2011). Who Needs Stories if You Can Get the Data? ISPs in the Era of Big Number Crunching. Philosophy and Technology 24 (4):371-390.
Holm, Søren & Ploug, Thomas (2017). Big Data and Health Research—The Governance Challenges in a Mixed Data Economy. Journal of Bioethical Inquiry 14 (4):515-525.
Hollands, Robert G. "Critical Interventions into the Corporate Smart City", Cambridge Journal of Regions, Economy and Society, Vol. 8, Issue 1, 2015, pp. 61-77.
Horvitz, E., 2017. AI, people, and society. Science 357, 7–7. https://doi.org/10.1126/science.aao2466.
Horvitz, E., Mulligan, D., 2015. Data, privacy, and the greater good. Science 349, 253–255.
Johnson, M.L., Bellovin, S.M., Kromytis, A.D., 2012. Computer Security Research with Huma Subjects: Risks, Benefits and Informed Consent, in: Danezis, G., Dietrich, S., Sako, K. (Eds.), Financial Cryptography and Data Security. Springer, Berlin, pp. 131–3
Hosni, Hykel, & Vulpiani, A., (2017). Forecasting in Light of Big Data. Philosophy and Technology, 1-13. DOI 10.1007/s13347-017-0265-3
Howard, P. N., Woolley, S., & Calo, R. (2018). Algorithms, bots, and political communication in the US 2016 election: The challenge of automated political communication for election law and administration. *Journal of Information Technology & Politics*, *15*(2), 81-93.
Hovy, Dirk, "Demographic Factors Improve Classification Performance," Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2015., p. 752
Hovy, Dirk, and Anders Søgaard, "Tagging Performance Correlates with Author Age," Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), 2015.
Hovy, Dirk, and Shannon L. Spruit, "The Social Impact of Natural Language Processing," Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2016.
Hull, G. (2015). Successful failure: what Foucault can teach us about privacy self-management in a world of Facebook and big data. Ethics and Information Technology, 17(2), 89-101.
Hutton, L. & Henderson, T. (2017). 'Beyond the EULA: Improving Consent for Data Mining' in *Transparent Data Mining for Big and Small Data*, (Ed.) Tania Cerquitelli, Daniele Quercia and Frank Pasquale. Springer.

Illari PM and Russo F (2014) Causality: Philosophical Theory Meets Scientific Practice. Oxford: Oxford University Press.
Ioannidis, J.P.A. (2013). Informed consent, Big Data, and the oxymoron of research that is not research. The American Journal of Bioethics 13(4). 40-42.
Ioannidis JPA (2005) Why most published research findings are false. PLoS Medicine 2(8): e124.
Jackson, J. R. (2018). Algorithmic Bias. Journal of Leadership, Accountability and Ethics, 15(4), 55-65.
James G, Witten D, Hastie T, et al. (2013) An Introduction to Statistical Learning. Vol. 6, New York: Springer.
Janssen, M. & Kuk, G. (2016). 'The challenges and limits of big data algorithms in technocratic governance' in *Government Information Quarterly* 33: 371-377.
Jha, A. (2016). An Overview of Security and Privacy Issues in Big Data. Retrieved from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2956050.
Johnson, J.A. (2014). The ethics of big data in higher education. International Review of Information Ethics 21(21), 3-9.
Johnson, J. A. (2014). From open data to information justice. Ethics and Information Technology, 4(16), 263-274.
Johnson, J. A. (2018). Open Data, Big Data, and Just Data. In Toward Information Justice, ed. by J. A. Johnson, 23-49.
Johnson, J. A. (2013) Ethics of data mining and predictive analytics in higher education. SSRN Scholarly Paper, Rochester, NY: Social Science Research Network. Available at: http://papers.ssrn.com/abstract=2156058.
Johnson, M. (2013). Cyber Crime, Security and Digital Intelligence. Edited by Routledge, New York: Taylor & Francis.
Jones, S., and Gupta, O. K., (2006). 'Web Data Mining: A Case Study' in Communications of the IIMA 6(4): 59-64.
Joshi, N. 2018. 4 Ways Global Defense Forces Use AI. Forbes. Retrieved from: https://www.forbes.com/sites/cognitiveworld/2018/08/26/4-ways-the-global-defense-forces-are-using-ai/#32117bb1503e.
Kamilaris, Andreas, Andreas Kartakoullis, and Francesc X Prenafeta-Boldú, "A Review on the Practice of Big Data Analysis in Agriculture." Computers and Electronics in Agriculture Vol. 143, October 2017, pp. 23-37.
Kamishima T, Akaho S, Asoh H, et al. (2012) Considerations on fairness-aware data mining. In: IEEE 12th International Conference on Data Mining Workshops, Brussels, Belgium, pp. 378–385. Available at: http://ieeexplore.ieee.org/ lpdocs/epic03/wrapper.htm?arnumber=6406465.
Karahisar, T. (2014). Developments in Communication Technologies and Employee Privacy in the Workplace. Journal of Media Critiques, 1(3), pp.221-234.
Kennedy, H., Hill, R. L., Aiello, G., & Allen, W. (2016). The work that visualisation conventions do. Information, Communication & Society, 19(6), 715-735.
Kennedy, S. 2018. Artificial Intelligence and Machine Learning. What are they and why are they important. Retrieved from: https://mapr.com/blog/artificial-intelligence-and-machine-learning-what-are-they-and-why-are-they-important/
Keeso, A. (2014). 'Big Data and Environmental Sustainability: A Conversation Starter' in *Smith School Working Paper Series*. Smith School of Enterprise and the Environment.
Kim, M. K. (2016). Algorithmic Opportunity: Digital Advertising and Inequality in Political Involvement. The Forum, 14(4), 471–484.
Kim H, Giacomin J and Macredie R (2014) A qualitative study of stakeholders' perspectives on the social network service environment. International Journal of Human– Computer Interaction 30(12): 965–976.
Kim, P. T. (2018). Big Data and Artificial Intelligence: New Challenges for Workplace Equality. Legal studies research paper series. Retrieved from: https://ssrn.com/abstract=3296521.

Kitchin, R. (2013). 'Big data and human geography: Opportunities, challenges and risks' in Dialogues in Human Geography 3(3): 262-267. DOI: 10.1177/2043820613513388
King, A. G. & Mrkonich, M. J. (2016). Big Data and the Risk of Employment Discrimination. Oklahoma Law Review, 68(3), 555-584.
Kitchin, R. (2015a). "Data-Driven Networked Urbanism", The Programmable City Working Paper 14, 10th August 2015, 2015a.
Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. Big Data & Society, 1(1):1-12. DOI: 10.1177/2053951714528481.
Kitchin, R. (2017). Thinking critically about and researching algorithms. Information, Communication & Society, 20(1), 14-29.
Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in human geography*, *3*(3), 262-267.
Kitchin, R. (2016a). Reframing, Reimagining and Remaking Smart Cities, *The Programmable City Working Paper 20.*
Kitchin, R. (2016b). The ethics of smart cities and urban science. *Phil. Trans. R. Soc. A* 374: 20160115.
Kitchin, R., & Lauriault, T.P. (2015). Small data in the era of big data. GeoJournal 80(4): 463-475. DOI: 10.1007/s10708-014-9601-7.
Kitchin, R. (2015b). "The Promise and Perils of Smart Cities", Society for Computers & Law, Vol. 26, Issue 2, pp. 1-5.
Kitchin, R., and McArdle, G. (2016). 'What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets' in *Big Data & Society*, pp. 1-10. DOI: 10.1177/2053951716631130.
Kitchin, R. (2014). "Making Sense of Smart Cities: Addressing Present Shortcomings", Cambridge Journal of Regions, Economy and Society, Vol. 8, pp. 131-136.
Kitchin, R. (2014). "The Real-Time City? Big Data and Smart Urbanism", GeoJournal, Vol. 79, pp. 1-14.
Kitchin, R. (2016). "Getting Smarter about Smart Cities: Improving Data Privacy and Data Security", Data Protection Unit, Department of the Taoiseach, Dublin, Ireland.
Kitchin, R. (2018). "The Realtimeness of Smart Cities", TECNOSCIENZA: Italian Journal of Science & Technology Studies, Vol. 8, Issue 2, pp. 19-42.
Kitchin, R., Claudio, C., Leighton, E., Liam, H. and Donncha, D. (2017). "Smart Cities, Urban Technocrats, Epistemic Communities and Advocacy Coalitions", The Programmable City Working Paper 26, 8th March 2017.
Kitchin, R., Tracey, P. L., and Gavin, M. (2015). "Smart Cities and the Politics of Urban Data", Smart Urbanism: Utopian Vision or False Dawn? Routledge, London, pp. 16-33. ISBN 9781138844223.
Kleissner, C. (1998). 'Data Mining for the Enterprise'. IEEE
Knapman, H. (2018). Households pressured into getting smart meters. [online] Available at: https://www.moneywise.co.uk/news/2018-01-30/households-pressured-getting-smart-meters [Accessed 25 Jan. 2019].
Knight, Kevin, and Irene Langkilde, "Preserving ambiguities in generation via automata intersection," AAAI/IAAI, pp. 697-702, 2000.
KnowledgeHut Editor. (2018). Types of Big Data. *KnowledgeHut Blog*. Retrieved from https://www.knowledgehut.com/blog/big-data/types-of-big-data.
Kohli, Devika, "How Smart Cities Will Force the Poor Out", Youth Ki Awaaz [website], 2014, available here: https://www.youthkiawaaz.com/2015/07/smart-cities-keep-the-poor-out/.
Kosior, Katarzyna, "Agricultural Education and Extension in the Age of Big Data", European Seminar on Extension and Education, 2017.
Kroll, J. A., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2016). Accountable algorithms. University of Pennsylvania Law Review, 165(633): 633-705.
Kraemer, F., van Overveld, K. and Peterson, M. (2011). Is there an ethics of algorithms? *Ethics and Information Technology* 13(3): 251–260.
Kshetri, N. (2014). "The Emerging Role of Big Data in Key Development Issues: Opportunities, Challenges, and Concerns", Big Data & Society Vol. 1, Issue 2.

Kuriakose, F. & Iyer, D. (2018). Human Rights in the Big Data World. SSRN Electronic Journal, 1-25.
La Torre, M., Dumay, J. C. & Rea, M. (2018). Breaching intellectual capital: Critical reflections on Big Data security. Meditari Accountancy Research, 26(3), 463-482.
Langheinrich, Marc, "Privacy by Design — Principles of Privacy-Aware Ubiquitous Systems," Ubicomp 2001: Ubiquitous Computing Lecture Notes in Computer Science, 2001, pp. 273–291.
Landau, S. (2015). 'Control Use of Data to Protect Privacy' in *Science* Vol. 347, no. 6221. DOI: 10.1126/science.aaa4961.
Lake, R. W. (2017). Big data, urban governance and the ontological politics of hyperindividualism. Big Data & Society 4, no.1.
Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind.* University of Chicago Press.
LASPROGATA, G., J. KING, N. and Pillay, S. (2004). Regulation of Electronic Employee Monitoring: Identifying Fundamental Principles of Employee Privacy through a Comparative Study of Data Privacy Legislation in the European Union, United States and Canada. Stanford Technology Law Review, [online] 4. Available at: https://www.sukanyapillay.com/wp-content/uploads/Regulation-of-Electronic-Employee-Monitoring.pdf [Accessed 1 Nov. 2018].
Latonero, M. (2018). Big Data Analytics and Human Rights. Privacy Considerations in Context. In New Technologies for Human Rights Law and Practice, ed. by M. K. Land and J. K. Aronson, 149-161.
LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., and Kruschwitz, N., (2012). 'Big Data, Analytics and the Path From Insights to Value' in MIT Sloan: Management Review 52(2): 21-31.Lazer D, Kennedy R, King G, et al. (2014) The parable of Google flu: Traps in big data analysis. Science 343(6176): 1203–1205.
Lee, J., Kao, H. A., & Yang, S. (2014). Service innovation and smart analytics for industry 4.0 and big data environment. Procedia Cirp, 16, 3-8.
Leese M (2014) The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union. Security Dialogue 45(5): 494–511.
Lei, Tao, Regina Barzilay, and Tommi Jaakkola, "Rationalizing Neural Predictions," Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016.
Leidner, Jochen L., and Vassilis Plachouras, "Ethical by Design: Ethics Best Practices for Natural Language Processing," Proceedings of the First ACL Workshop on Ethics in Natural Language Processing, 2017.
Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2017). Fair, Transparent, and Accountable Algorithmic Decision-making Processes. Philosophy & Technology, 1- 17.
Lepri, B., Staiano, J., Sangokoya, D., Letouze, E., and Oliver, N., (2017). 'The Tyranny of Data? The Bright and Dark Sides of Data-Driven Decision-Making for Social Good' in *Transparent Data Mining for Big and Small Data*, (Ed.) Tania Cerquitelli, Daniele Quercia and Frank Pasquale. Springer.
Leveling, J., Edelbrock, M., & Otto, B. (2014, December). Big data analytics for supply chain management. In Industrial Engineering and Engineering Management (IEEM), 2014 IEEE International Conference on (pp. 918-922). IEEE.
Lewis, S. C., and Westlund, O., (2015). 'Big Data and Journalism: Epistemology, expertise, economics, and ethics' in *Digital Journalism*. DOI: http://dx.doi.org/10.1080/21670811.2014.976418
Li, J., Tao, F., Cheng, Y., & Zhao, L. (2016). Big data in product lifecycle management. The International Journal of Advanced Manufacturing Technology, 81(1-4), 667-684.
Li, Yibin, Wenyun Dai, Zhong Ming, and Meikang Qiu, "Privacy Protection for Preventing Data Over-Collection in Smart City", IEEE Transactions on Computers, Vol. 65, Issue 5, 2016, pp. 1339-1350.
Lipton, Z. C. (2018). The mythos of model interpretability. Communications of the ACM, 61(10), 36-43.
Lin, T. C. W. (2017) "The New Market Manipulation". Temple University Legal Studies Research Paper No. 2017-20, Emory Law Journal, Vol. 66, p. 125.
Lipton, Zachary C., "The Mythos of Model Interpretability," Communications of the ACMVol. 61, No. 10, 2018, pp. 36–43.
Lipworth, W., Mason, P. H., & Kerridge, I. (2017). Ethics and epistemology of big

data. Journal of bioethical inquiry, 14(4), 485-488.
Lipworth, W., Mason, P. H., Kerridge, I., & Ioannidis, J. P. (2017). Ethics and epistemology in big data research. Journal of bioethical inquiry, 14(4), 489-500.
Lokers, R., Knapen, R., Janssen, S., Randen, Y., Jansen, J. "Analysis of Big Data Technologies for Use in Agro-Environmental Science", Environmental Modelling & Software, Vol. 84, 2-16, pp. 494-504.
Luo, J., Wu, M., Gopukumar, D., and Zhao, Y., (2016). 'Big Data Application in Biomedical Research and Health Care: A Literature Review' in Biomedical Informatics Insights 8: 1-10.
Lustig, I., Dietrich, B., Johnson, C., & Dziekan, C. (2010). The analytics journey. Analytics Magazine, 3(6), 11-13.
Lyon, D. (2014). Surveillance, Snowden, and big data: Capacities, consequences, critique. Big Data & Society, 1(2), 2053951714541861.
Maciejewski, M. (2017). To do more, better, faster and more cheaply: Using big data in public administration. *International Review of Administrative Sciences*, *83*(1_suppl).
Macnish, K. (2014). An Eye for an Eye: Proportionality and Surveillance. Ethical Theory and Moral Practice, 18(3), 529-548.
Macnish, K., van der Ham, J., 2019. Ethics and Cybersecurity Research. Journal of Science and Engineering Ethics.
Macnish, K., 2012. Unblinking eyes: the ethics of automating surveillance. Ethics and Information Technology 14, 151–167. https://doi.org/10.1007/s10676-012-9291-0
Macnish, K., van der Ham, J., 2019. Ethics and Cybersecurity Research. Journal of Science and Engineering Ethics.
MacQuillan, D. (2017). Data Science as Machinic Neoplatonism. Philosophy & Technology.
Making policy with big data, (2017, November 14). Retrieved from https://www.universiteitleiden.nl/en/news/2017/11/making-policy-with-big-data
Macnish et al. (2019) D1.1 Case Studies, SHERPA Project, available here: https://dmu.figshare.com/articles/D1_1_Case_studies/7679690
Madden, M. et al. (2017). Privacy, Poverty, and Big Data: A Matrix of Vulnerabilities for Poor Americans. Washington University Law Review, 95(1), 53-125.
Mah, Alice (2017). Environmental justice in the age of big data: challenging toxic blind spots of voice, speed, and expertise. Environmental Sociology, 3(2), 122-133.
Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A.H. (2011) Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute* (June).
Marabelli, M. & Markus, M. L. (2017). 'Researching Big Data Research: Ethical Implications for IS Scholars' in *Ethical Implications of Big Data Research*, 1-5.
Maruti Techlabs. (2017). Big Data Analytics: Its Technologies and Tools. *Towards Data Science*. Retrieved from https://towardsdatascience.com/big-data-analytics-its-technologies-and-tools-e77f9bd0d37c.
Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology* 6(3): 175–183.
Mattioli, D., 2012. On Orbitz, Mac Users Steered to Pricier Hotels. Wall Street Journal.
Matturdi, B. et al. (2014). Big Data Security and Privacy: A Review. Big Data, Cloud & Mobile Computing, 11(2), 135-145.
Matzner, (2016). Beyond data as representation: The performativity of Big Data in surveillance. Surveillance & Society, 14(2), 197-210.
McCosker, A., & Wilken, R. (2014). Rethinking 'big data' as visual knowledge: The sublime and the diagrammatic in data visualisation. Visual Studies, 29(2), 155-164.
Meira, W., Jr., 2017. Fairness, Accountability, and Transparency While Mining Data from the Web and Social Networks, in: Proceedings of the 23rd Brazillian Symposium on Multimedia and the Web, WebMedia '17. ACM, New York, NY, USA, pp. 17–17. https://doi.org/10.1145/3126858.3133314.
Metcalf. J and Crawford, K. (2016). 'Where are the human subjects in Big Data research? The emerging ethics divide' in *Big Data & Society*. DOI: 10.1177/2053951716650211.

Micheni, Muthoni, E. (2015) "Diffusion of Big Data and Analytics in Developing Countries", The International Journal of Engineering and Science Vol. 4, Issue 8, pp. 44-50.
Mieskes, Margot, "A Quantitative Study of Data in the NLP Community," Proceedings of the First ACL Workshop on Ethics in Natural Language Processing, 2017.
Miller, B. and Record, I. (2013) Justified belief in a digital age: On the epistemic implications of secret Internet technologies. Episteme 10(2): 117–134.
Miller, F., Wertheimer, A. (Eds.), 2009. The Ethics of Consent: Theory and Practice, 1 edition. ed. OUP USA, Oxford ; New York.
Millar, J. (2009). Core privacy: a problem for predictive data mining. Lessons from the identity trail: Anonymity, privacy and identity in a networked society, pages 103-119.
Min Chen, Mao, S., and Liu, Y. (2014). 'Big Data: A Survey' in Mobile Network Applications 19: 171-209. DOI: 10.1007/s11036-013-0489-0.
Mishra, J. and Crampton, S. (1998). Employee monitoring: Privacy in the workplace?. Advanced Management Journal, [online] 63(3), pp.4-14. Available at: http://faculty.bus.olemiss.edu/breithel/final%20backup%20of%20bus620%20summer%202000%20from%20mba%20server/frankie_gulledge/employee_workplace_monitoring/employee_monitoring_privacy_in_the_workplace.htm [Accessed 4 Nov. 2018].
Mitrou, L. et al. (2014). Social Media Profiling: A Panopticon or Omniopticon Tool? Retrieved from: https://www.semanticscholar.org/paper/SOCIAL-MEDIA-PROFILING-%3A-A-PANOPTICON-OR-TOOL-Mitrou-Kandias/4fe7dde0066940748f3822d88db05014546e8d49
Mittelstadt, B. (2017). Designing the health-related internet of things: ethical principles and guidelines. Information, 8(3), 77. DOI: 10.3390/info8030077.
Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2).
Misselhorn, C., Pompe, U., and Stapleton, M. (2013). Ethical considerations regarding the use of social robots in the fourth age. *The Journal of Gerontopsychology and Geriatric Psychiatry*, 26(2), pp. 121-133.
Mittelstadt, B. (2017). From individual to group privacy in big data analytics. Philosophy & Technology, 30(4), 475-494.
Mittelstadt, B., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The Ethics of Algorithms: Mapping the Debate. Big Data and Society 3 (2): 1-21.
Mittelstadt, B. D., & Floridi, L. (2016). The ethics of big data: Current and foreseeable issues in biomedical contexts. Science and Engineering Ethics, 22(2), 303-341.
Mittelstadt, B. D. et al. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 1-21.
Moorthy, J., Lahiri, R., Biswas, N., Sanyal, D., Ranjan, J., Nanath, K., & Ghosh, P. (2015). Big data: prospects and challenges. Vikalpa , 40 (1), 74-96.
Muehlhauser, L., & Helm, L. (2012). The singularity and machine ethics. In *Singularity Hypotheses* (pp. 101-126). Springer, Berlin, Heidelberg.
Mujtaba, B. (2004). Ethical Implications of Employee Monitoring: What Leaders Should Consider' Journal of Applied Management and Entrepreneurship. The Journal of Applied Management and Entrepreneurship, 8(3), pp.22-47.
Munoz, Mark J., Al Naqvi, "Artificial Intelligence and Urbanization: The rise of the Elysium City", Economics and Political Economy, Vol. 4, Issue 1, March 2017, pp. 1-13.
Negnevitsky, M. (2005). Artificial intelligence: a guide to intelligent systems. Pearson Education.
Newell, S. & Marabelli, M. (2015). Strategic Opportunities (and Challenges) of Algorithmic Decision-making: A Call for Action on the Long-term Societal Effects of 'Datification'. Journal of Strategic Information Systems, 24 (1), 3-14.
Newells and Marabelli M (2015) Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datification'. The Journal of Strategic Information Systems 24(1): 3–14.

Newman, N. (2014). The Costs of Lost Privacy: Consumer Harm and Rising Economic Inequality in the Age of Google. William Mitchell Law Review, 40(2), 849-889.
Newton Media (2018). Revolutionising claims handling and fraud detection with AI. [online] Intelligent Insurer. Available at: https://www.intelligentinsurer.com/news/revolutionising-claims-handling-and-fraud-detection-with-ai-13318 [Accessed 9 Nov. 2018].
Nichols, S., 2016. St Jude sues short-selling MedSec over pacemaker "hack" report [WWW Document]. The Register. URL https://www.theregister.co.uk/2016/09/07/st_jude_sues_over_hacking_claim/ (accessed 7.4.18).
Nissenbaum, H. (2011). A contextual approach to privacy online. Daedalus, 140(4): 32-48.
Nunan, D., & Di Domenico, M. (2013). Market research & the ethics of big data. International Journal of Market Research, 55(4): 505-520.
Nunan, D., & Di Domenico, M. (2017). Big data: a normal accident waiting to happen?. Journal of Business Ethics, 145(3): 481-491.
O'Doherty, K. C., Christofides, E., Yen, J., Bentzen, H. B., Burke, W., Hallowell, N., Koenig, B.A. & Willison, D. J. (2016). If you build it, they will come: unintended future uses of organised health data collections. BMC medical ethics, 17(54): 1-16.
DOI 10.1186/s12910-016-0137-x
O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. New York: Crown Publishers, 2016.
O'Grady, Michael J, and Gregory MP O'Hare, "Modelling the Smart Farm", Information Processing in Agriculture, Vol. 4, 2017, pp. 179-187.
O'Neil, C., 2016. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown/Archetype.
Olson, D. (2008). 'Ethical Aspects of Web Log Data Mining' in the International Journal of Information Technology and Management 7(2): 190-201
Oostveen, M., & Irion, K. (2018). 'The Golden Age of Personal Data: How to Regulate an Enabling Fundamental Right?' in M. Bakhoum, B. Conde Gallego, M-O. Mackenrodt, & G. Surblytė-Namavičienė (Eds.), *Personal Data in Competition, Consumer Protection and Intellectual Property Law: Towards a Holistic Approach*? (pp. 7-26). (MPI Studies on Intellectual Property and Competition Law; Vol. 28). Berlin: Springer. DOI: 10.1007/978-3-662-57646-5_2.
Oussous, A., Benjelloun, F.-Z., Lahcen, A. A., and Belfkih, S. (2017). Big data technologies: A survey. Journal of King Saud University-Computer and Information Sciences.
Pariser, E. (2011) The Filter Bubble: What the Internet is Hiding from You. London: Viking.
Pan, S. B. (2016). Get to Know Me: Protecting Privacy and Autonomy under Big Data's Penetrating Gaze. Harvard Journal of Law & Technology. 30 (1), 239-261.
Pariser, E. (2011). The filter bubble: What the Internet is hiding from you. Penguin UK.
Pauleen, D.J., Rooney, D. & Intezari, A. (2017). Big data, little wisdom: trouble brewing? Ethical implications for the information systems discipline. Social Epistemology 31 (4):400-416.
Pedreshi, D., Ruggieri, S., & Turini, F. (2008, August). Discrimination-aware data mining. In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, 560-568. ACM.
Peek, N., Holmes, J. H., and Sun, J., (2014). 'Technical Challenges for Big Data in Biomedicine and Health: Data Sources, Infrastructure, and Analytics' in IMIA Yearbook of Medical Informatics 42-47.
Pencheva, I., Esteve, M., & Mikhaylov, S. J. (2018). Big Data and AI–A transformational shift for government: So, what next for research?. Public Policy and Administration.
Petersen, C. J. (2017). Big Data, Health Care, and International Human Rights Norms. Asia Pacific Journal of Health Law & Ethics, 11(1), 1-22.
Petersen, C., 2018. Through Patients' Eyes: Regulation, Technology, Privacy, and the Future. Yearb. Med. Inform. 27, 010–015. https://doi.org/10.1055/s-0038-1641193.

Pieters, W., 2011. Explanation and trust: what to tell the user in security and AI? Ethics Inf Technol 13, 53–64. https://doi.org/10.1007/s10676-010-9253-3.
Pietsch, W. (2016). The causal nature of modeling with big data. Philosophy & Technology, 29(2), 137-171.
Pietsch, W. (N.d.) Big Data – The New Science of Complexity.
Politou, E., Alepis, E. & Patsakis, C. (2018). 'Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions' in *Journal of Cybersecurity*, pp. 1-20. DOI: 10.1093/cybsec/tyy001.
Poppe, Krijn, Sjaak Wolfert, and Cor Verdouw, How Ict Is Changing the Nature of the Farm: A Research Agenda on the Economics of Big Data. 11th European IFSA Symposium, Farming Systems Facing Global Challenges: Capacities and Strategies, Proceedings, Berlin, Germany, 1-4 April 2014. 2014.
Portmess, L., & Tower, S. (2015). Data barns, ambient intelligence and cloud computing: the tacit epistemology and linguistic representation of Big Data. Ethics and Information Technology, 17(1): 1-9.
Pryzant, Reid, Kelly Shen, Dan Jurafsky, and Stefan Wagner, "Deconfounded Lexicon Induction for Interpretable Social Science," Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), 2018.
Purtova, N. (2017). Do Property Rights in Personal Data Make Sense after the Big Data Turn? Individual Control and Transparency. Journal of Law and Economic Regulation, 10(2), 64-78.
Qiu, J., Wu, Q., Ding, G., Xu, Y., & Feng, S. (2016). A survey of machine learning for big data processing. EURASIP Journal on Advances in Signal Processing, 2016(1), 67.
Rader, E., Cotter, K. & Cho, J. (2018). 'Explanations as Mechanisms for Supporting Algorithmic Transparency' in *CHI*, Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems Paper No. 103. DOI: http://dx.doi.org/10.1145/3173574.3173677
Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. Health information science and systems, 2(3): 1-10.
Raisinghani, Mahesh S., Ally Benoit, Jianchun Ding, Maria Gomez, Kanak Gupta, Victor Gusila, Daniel Power and Oliver Schmedding, "Ambient intelligence: Changing forms of human-computer interaction and their social implications," Journal of digital information, Vol. 5, No. 4, 2004.
Raub, M. (2018). Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices. Arkansas Law Review, 71(2), 529-570.
Raymond, A. H. (2015). The Dilemma of Private Justice Systems: Big Data Sources, the Cloud and Predictive Analytics. Northwestern Journal of International Law & Business, 35 (4), 1-45.
Reiter, Ehud, and Robert Dale. Building natural language generation systems. Cambridge university press, 2000.
Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin, "Semantically equivalent adversarial rules for debugging nlp models," Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 856-865, 2018.
Richards, N. M., & King, J. H. (2014). Big data ethics. Wake Forest L. Rev., 49, 393.
Rieder, G. & Simon, J. (2016). Datatrust: Or, the political quest for numerical evidence and the epistemologies of Big Data, Big Data & Society 3(1): 1-6.
Rieder, B. (2016) 'Big Data and the Paradox of Diversity' in *Digital Culture and Society* Vol. 2, issue 2. DOI: 10.14361/dcs-2016-0204.
Romei A and Ruggieri S (2014) A multidisciplinary survey on discrimination analysis. The Knowledge Engineering Review 29(5): 582–638.
Rommelfanger, K.S., Jeong, S.-J., Ema, A., Fukushi, T., Kasai, K., Ramos, K.M., Salles, A., Singh, I., Amadio, J., Bi, G.-Q., Boshears, P.F., Carter, A., Devor, A., Doya, K., Garden, H., Illes, J., Johnson, L.S.M., Jorgenson, L., Jun, B.-O., Lee, I., Michie, P., Miyakawa, T., Nakazawa, E., Sakura, O., Sarkissian, H., Sullivan, L.S., Uh, S., Winickoff, D., Wolpe, P.R., Wu, K.C.-C., Yasamura, A., Zheng, J.C., 2018.

Neuroethics Questions to Guide Ethical Research in the International Brain Initiatives. Neuron 100, 19–36. https://doi.org/10.1016/j.neuron.2018.09.021.
Rosenheim, Jay A, and Claudio Gratton, "Ecoinformatics (Big Data) for Agricultural Entomology: Pitfalls, Progress, and Promise." Annual review of entomology, Volume 62, 2017, pp. 399-417.
Rumbold, J. M., & Pierscionek, B. K. (2017). A critique of the regulation of data science in healthcare research in the European Union. BMC medical ethics, 18(27): 1-11.
DOI 10.1186/s12910-017-0184-y.
Russom, P. (2011). 'Big Data Analytics' from The Data Warehouse Institute Best Practices Report. TDWI.
Ryan, M. (2019). Ethics of Public Use of AI and Big Data. ORBIT Journal, 2(2). https://doi.org/10.29297/orbit.v2i1.101.
Sagiroglu, S. and Sinanc, D. (2013). Big data: A review. In Collaboration Technologies and Systems (CTS), 2013 International Conference on, pages 42-47. IEEE.
Salganik, M. J., & Watts, D. J. (2008). Leading the herd astray: An experimental study of self-fulfilling prophecies in an artificial cultural market. Social psychology quarterly, 71(4), 338-355.
Salvini, 2015. On Ethical, Legal and Social Issues of Care Robots. In: S. Mohammed, J. C. Moreno, K. Kong & Y. Amirat , eds. *Intelligent Assistive Robots*. Springer Tracts in Advanced Robotics Volume 106. Cham (ZG) Switzerland: Springer International Publishing, pp. 431-445.
Sandvig C, Hamilton K, Karahalios K, et al. (2014) Auditing algorithms: Research methods for detecting discrimination on internet platforms. Data and Discrimination: Converting Critical Concerns into Productive Inquiry. Available at: http://social.cs.uiuc.edu/papers/pdfs/ICA2014.
Sample, Ian, "Thousands of leading AI researchers sign pledge against killer robots", The Guardian, 18 Jul 2018.
Sanders, N. R. (2016). How to use big data to drive your supply chain. California Management Review, 58(3), 26-48.
Sarfaty, G. A. (2017). Can Big Data Revolutionize International Human Rights Law. University of Pennsylvania Journal of International Law, 39(1), 73-101.
Saxena, M., Ali, Z., and Singh, V. K. (2014). Nosql databases-analysis, techniques, and classification. Journal of Advanced Database Management & Systems, 1(2):13-24.
Sen, Mourjo, Anuvabh Dutt, Shalabh Agarwal, and Asoke Nath, "Issues of Privacy and Security in the Role of Software in Smart Cities", International Conference on Communication Systems and Network Technologies (CSNT), 2013, pp. 518-523.
Shen, W. & Wen, H. (2016). Impacts of Big Data Marketing upon Consumption Freedom and its Reflections. Atlantic Press, 68-71.
Shelton, Taylor, Matthew Zook, and Alan Wiig, "The 'Actually Existing Smart City'", Cambridge Journal of Regions, Economy and Society, Vol. 8, Issue 1, 2015, pp. 13-25.
Schönfeld, Max v, Reinhard Heil, and Laura Bittner, "Big Data on a Farm—Smart Farming", Big Data in Context. Springer, 2018. pp. 109-20.
Sax, M. (2016). Big data: Finders keepers, losers weepers? Ethics and Information Technology, 18(1), 25-31.
Schradie, J. (2017). Big Data is Too Small: Research Implications of Class Inequality for Online Data Collection. Media and Class: TV, Film and Digital Culture. Edited by June Deery and Andrea Press. Abingdon, UK: Taylor & Francis.
Schradie, J. (2017). Big Data is Too Small: Research Implications of Class Inequality for Online Data Collection. Media and Class: TV, Film and Digital Culture. Edited by June Deery and Andrea Press. Abingdon, UK: Taylor & Francis.
Sedayao, J., Bhardwaj, R. and Gorade, N., (). 'Making Big Data, Privacy, and Anonymization work together in the Enterprise: Experiences and Issues'
Singer, P.W., and Allan Friedman, Cybersecurity and Cyber War: what everyone needs to know, Oxford University Press, 2014.

Singer, P.W., and Emerson T. Brooking, LikeWar: The Weaponisation of Social Media, Eamon Dolan/Houghton Mifflin Harcourt, New York, NY, 2018, p. 18.
Singh, D. and Reddy, C. K. (2015). A survey on platforms for big data analytics. Journal of Big Data, 2(1):8.
Sleeman, S. P. & Rademan B. (2017). Freedom from Social Echo Chambers: Policy Implications of an Algorithmic Bias. SSRN Electronic Journal. Retrieved from: https://ssrn.com/abstract=3044265.
Sharkey, 2014. 'Robots and human dignity: a consideration of the effects of robot care on the dignity of older people'. *Ethics and Information Technology*, 16(1): 63-75.
Sholla, Sahil, Roohie Naaz, and Mohammad Ahsan Chishti. "Ethics Aware Object-Oriented Smart City Architecture", China Communications, Vol. 14, Issue 5, 2017, pp. 160-173.
Smith, P.T., 2018. Cyberattacks as Casus Belli: A Sovereignty-Based Account. Journal of Applied Philosophy, 35(2), 222–241.
Someh, I. A., Breidbach, C. & Davern, M. (2016). Ethical Implications of Big Data Analytics. Twenty-Fourth European Conference on Information Systems (ECIS), İstanbul, Turkey. Retrieved from: https://www.researchgate.net/publication/308024119_ETHICAL_IMPLICATIONS_OF_BIG_DATA_ANALYTICS.
Song, M, cen, L, zheng, Z, fisher, R, Liang, X, wang, Y & Huisingh, D (2016). 'How would big data support societal development and environmental sustainability? Insights and practices' *Journal of Cleaner Production*, vol. 142, pp. 489-500. DOI: 10.1016/j.jclepro.2016.10.091
Someh, I.A., Davern, M., & Breidbach, C. (2016). Ethical implications of big data analytics. Research in Progress Papers, 24.
Soni, D. 2018. Supervised vs. Unsupervised Learning: Understanding the differences between the two main types of machine learning methods. Towards Data Science. Retrieved from: https://towardsdatascience.com/supervised-vs-unsupervised-learning-14f68e32ea8d.
Sorell & Draper, 2014. 'Robot carers, ethics, and older people'. *Ethics and Information Technology*,16(3): 183-195.
Souto-Otero, M. and Benito-Montagut, R. (2016). 'From governing through data to governmentality through data: Artefacts, strategies and the digital turn' in *European Educational Research Journal* Vol. 15(1): 14-33
Soraker, Johnny Hartz, and Philip Brey, "Ambient intelligence and problems with inferring desires from behaviour," International Review of Information Ethics, Vol. 8, no. 1, 2007
Sparrow & Sparrow, 2006. 'In the hands of machines? The future of aged care'. *Minds and Machines*, 16(2): 141-161.
Spring, T., 2016. Researchers: MedSec, Muddy Waters Set Bad Precedent With St. Jude Medical Short. The first stop for security news | Threatpost.
Stark, M., and Fins, J.J. (2013). *Engineering medical decisions*. Cambridge Quarterly of Healthcare Ethics 22(4): 373–381.
Stankovic, Mirjana. Ravi Gupta, Bertrand Andre Rossert, Gordon Myers and Marco Nicoli, "Exploring Legal, Ethical and Policy Implications of Artificial Intelligence" *Tambourine*.
Stoeklé, H. C., Mamzer-Bruneel, M. F., Frouart, C. H., Le Tourneau, C., Laurent-Puig, P.,
Vogt, G., & Hervé, C. (2018). Molecular Tumor Boards: Ethical Issues in the New Era of Data Medicine. Science and engineering ethics, 24(1), 307-322.
Stone, N., 2017. The Yahoo Cyber Attack & What should you learn from it? [WWW Document]. Cashfloat. URL https://www.cashfloat.co.uk/blog/technology-innovation/yahoo-cyber-attack/ (accessed 12.17.18).
Suster, Simon, Stephan Tulkens, and Walter Daelemans, "A Short Review of Ethical Challenges in Clinical Natural Language Processing," Proceedings of the First ACL Workshop on Ethics in Natural Language Processing, 2017.
Sweeney L (2013) Discrimination in online ad delivery. Queue 11(3): 10:10–10:29.

Sykuta, Michael E. "Big Data in Agriculture: Property Rights, Privacy and Competition in Ag Data Services", The International Food and Agribusiness Management Review, Vol. 19, Issue A, June 2016, pp. 57-74.
Symeonidis, Andreas L., Kehagias, D. D., and Mitkas, P. A., (2003). 'Intelligent policy recommendations on enterprise resource planning by the use of agent technology and data mining techniques' in Expert Systems with Applications 25: 589-602. DOI: 10.1016/S0957-4174(03)00099-X.
Talavera, Jesús Martín, Luis Eduardo Tobón, Jairo Alejandro Gómez, María Alejandra Culman, Juan Manuel Aranda, Diana Teresa Parra, Luis Alfredo Quiroz, Adolfo Hoyos, and Luis Ernesto Garreta, "Review of Iot Applications in Agro-Industrial and Environmental Fields", Computers and Electronics in Agriculture, Vol. 142, September 2017, pp. 283-97.
Tatman, Rachael, "Gender and Dialect Bias in YouTubes Automatic Captions," Proceedings of the First ACL Workshop on Ethics in Natural Language Processing, 2017.
Taylor, L. (2017). 'What is data justice? The case for connecting digital rights and freedoms globally' in *Big Data & Society*, pp. 1-14. DOI: 10.1177/2053951717736335.
Taylor, Linnet, "Safety in Numbers? Group Privacy and Big Data Analytics in the Developing World", in Linnet Taylor B. van der Sloot, and L. Floridi, Group Privacy: The Challenges of New Data Technologies, Springer, 2017. pp. 13-36.
Taylor, Linnet, and Dennis Broeders, "In the Name of Development: Power, Profit and the Datafication of the Global South", Geoforum, Vol. 64, 2015, pp. 229-37.
Tene, O., & Polonetsky, J. (2012). Big data for all: Privacy and user control in the age of analytics. Northwestern Journal of Technology & Intellectual Property, 11(5): 238-273.
The Cyclance Data Science Team. 2017. Introduction to AI for security. Retrieved from: https://threatvector.cylance.com/en_us/home/introduction-to-ai-for-security.html.
Terzi, D.S., Terzi, R., Sagiroglu, S., 2015. A survey on security and privacy issues in big data, in: 2015 10th International Conference for Internet Technology and Secured Transactions (ICITST). Presented at the 2015 10th International Conference for Internet Technology and Secured Transactions (ICITST), pp. 202–207. https://doi.org/10.1109/ICITST.2015.7412089.
Tittel, S. (2019). Top 5 ways AI is changing traditional finance. [online] Available at: https://www.worldfinance.com/markets/top-5-ways-ai-is-changing-traditional-finance [Accessed 19 Feb. 2019].
Tiwari, S., Wee, H. M., & Daryanto, Y. (2018). Big data analytics in supply chain management between 2010 and 2016: Insights to industries. Computers & Industrial Engineering, 115, 319-330.
Tramer, F., Atlidakis, V., Geambasu, R., Hsu, D., Hubaux, J., Humbert, M., Juels, A. and Lin, H. (2016). 'FairTest: Discovering Unwarranted Associations in Data-Driven Applications' from *2017 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE.
Tractenberg, Rochelle E. ; Russell, Andrew J. ; Morgan, Gregory J. ; FitzGerald, Kevin T. ; Collmann, Jeff ; Vinsel, Lee ; Steinmann, Michael & Dolling, Lisa M. (2015). Using Ethical Reasoning to Amplify the Reach and Resonance of Professional Codes of Conduct in Training Big Data Scientists. Science and Engineering Ethics 21 (6):1485- 1507.
Tsou, M. (2015). 'Research challenges and opportunities in mapping social media and Big Data' in Cartography and Geographic Information Science 42(S1): 70-74.
Tufekei, Z. (2014). 'Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls' from the Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media.
Turkle, S. (2007). 'Authenticity in the age of digital companions' in *Interaction Studies* 8(3): 501-517.
Tung, Liam, "Google AI Can Pick Out a Single Speaker in a Crowd: Expect To See it in Tons of Products", ZDNet [website], April 13th 2018, https://www.zdnet.com/article/google-ai-can-pick-out-a-single-speaker-in-a-crowd-expect-to-see-it-in-tons-of-products/.
Turilli M (2007) Ethical protocols design. Ethics and Information Technology 9(1): 49–62.
Tutt A (2016) An FDA for algorithms. SSRN Scholarly Paper, Rochester, NY: Social Science Research Network. Available at: http://papers.ssrn.com/abstract=2747994 (accessed 13 April 2016).

Tzounis, Antonis, Nikolaos Katsoulas, and Thomas Bartzanas, "Internet of Things in Agriculture, Recent Advances and Future Challenges", Biosystems Engineering, Vol. 164, September 2017, pp. 31-48.
Vallor, 2011. 'Carebots and caregivers: Sustaining the Ethical Ideal of Care in the Twenty-First Century' in *Philosophy of Technology* 24: 251-268.
Upadhyaya, S. and Kyn_clov_a, P. (2017). Big data-its relevance and impact on industrial statistics.
Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. Surveillance & Society, 12(2), 197.
Van Dijck, J., & Poell, T. (2013). Understanding social media logic. Media and Communication, 1(1): 2-14.
van den Hoven, J. and Rooksby, E. (2008) Distributive justice and the value of information: A (broadly) Rawlsian approach. In: van den Hoven J and Weckert J (eds) Information Technology and Moral Philosophy. Cambridge: Cambridge University Press, pp. 376–396.
Van den Hoven, J., Vermaas, P. & Van de Poel, I. (Eds.), *Handbook of Ethics, Values, and Technological Design. Sources, Theory, Values and Application Domains*. Dordrecht: Springer.
van Otterlo, M. (2017). From Algorithmic Black Boxes to Adaptive White Boxes: Declarative Decision-Theoretic Ethical Programs as Codes of Ethics. *arXiv preprint arXiv:1711.06035*.
van Wel, L. and Royakkers, L. (2004) Ethical issues in web data mining. Ethics and Information Technology 6(2): 129–140.
Vayena, E. & Tasioulas, J. (2016). The dynamics of big data and human rights: the case of scientific research. Phil.Trans.R. Soc. A, 374(2083), 1-14.
Vayena, E., & Blasimme, A. (2017). Biomedical big data: New models of control over access, use and governance. Journal of bioethical inquiry, 14(4), 501-513.
Vedder, A. (1999). KDD: The challenge to individualism. Ethics and Information Technology, 1(4), 275-281.
Venkatram, K. and Geetha, M. A. (2017). Review on big data & analytics-concepts, philosophy, process and applications. Cybernetics and Information Technologies, 17(2):3-27.
Voda, Ana Iolanda, and Laura Diana Radu, "Artificial Intelligence and the Future of Smart Cities", Broad Research in Artificial Intelligence and Neuroscience, Vol. 9, Issue 2, 2018, pp. 110-127.
Voinea, C. & Uszkai, R. (2018). 'An Assessment of Algorithmic Accountability Methods' from the Proceedings of the 12th International Management Conference, "Management Perspectives in the Digital Era'.
Wachter, S., Mittelstadt, B., Floridi, L., 2017b. Transparent, explainable, and accountable AI for robotics. Sci. Robot. 2, eaan6080. https://doi.org/10.1126/scirobotics.aan6080.
Wan, J., Tang, S., Li, D., Wang, S., Liu, C., Abbas, H., & Vasilakos, A. V. (2017). A manufacturing big data solution for active preventive maintenance. IEEE Transactions on Industrial Informatics, 13(4), 2039-2047.
Wang, H., Jiang, X. & Kambourakis, G. (2015). Special issue on Security, Privacy and Trust in network-based Big Data. Information Sciences, 318, 48-50.
Wang, H., Xu, Z., Fujita, H., and Liu, S., (2016). 'Towards felicitous decision making: An overview on challenges and trends of Big Data' in *Information Sciences* 367-368: 747-765
Wang, T. (2013). Big data needs thick data. Ethnography Matters, 13. Retrieved from: https://medium.com/ethnography-matters/why-big-data-needs-thick-data-B4b3e75e3d7.
Ward, J. S. and Barker, A. (2013). Undefined by data: a survey of big data definitions. arXiv preprint arXiv:1309.5821.
Watson, H. J. (2014). Tutorial: Big data analytics: Concepts, technologies, and applications. CAIS, 34:65.
van Wel, L., and Royakkers, L., (2004). 'Ethical Issues in Web Data Mining' in Ethics and Information Technology 6: 129-140.
Werhane, P. H. (1998). Moral imagination and the search for ethical decision-making in management. The Ruffin Series of the Society for Business Ethics, 1, 75-98.
Williamson, A. (2014). Big data and the implications for government. *Legal Information*

Management 14(4): 253–257.
Wilson, R. J., Belliveau, K. M. & Gray, L. E. (2017). Busting the Black Box: Big Data, Employment and Privacy. Defence Counsel Journal, 84(3), 2-34.
Wiseman, Sam, and Alexander M. Rush, "Sequence-to-Sequence Learning as Beam-Search Optimization," Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016.
Wolf, B. (2015). Big data, small freedom? Informational surveillance and the political. Radical Philosophy, 191, 13-20.
Wolf, S. (1987). Sanity and the Metaphysics of Responsibility. In Ferdinand David Schoeman (ed.), Responsibility, Character, and the Emotions: New Essays in Moral Psychology. Cambridge University Press. pp. 46-62.
Wolfert, Sjaak, et al. "Big Data in Smart Farming–a Review", Agricultural Systems Vol. 153, 2017, pp. 69-80.
Wolff, J., 2010. Five Types of Risky Situation. Law, Innovation and Technology 2, 151–163. https://doi.org/10.5235/175799610794046177
Woolley J. Patrick, (2017). Trust and Justice in Big Data Analytics: Bringing the Philosophical Literature on Trust to Bear on the Ethics of Consent. Philosophy and Technology:1-24.
Wright, David, "The Dark Side of Ambient Intelligence," Info, Vol. 7, No. 6, 2005, pp. 33–51., p. 34; Brey, Philip, "Freedom and Privacy in Ambient Intelligence," Ethics and Information TechnologyVol. 7, No. 3, 2005
Wu, J., Guo, S., Li, J., and Zeng, D. (2016). 'Big Data Meet Green Challenges: Greening Big Data' in *IEEE Systems Journal* 10(3):873-887.
Yampolskiy, R., & Fox, J. (2013). Safety engineering for artificial general intelligence. *Topoi*, *32*(2), 217-226.
Yang, M. et al. (2016). Challenges and Solutions of Information Security Issues in the Age of Big Data. Security Schemes and Solutions, 13(3), 193-202.
Yeung, K. (2016). 'Hypernudge': Big Data as a mode of regulation by design. Information, Communication & Society, 20 (1), 118-136.
Yeung, K. (2012). *Nudge as fudge*. The Modern Law Review, 75(1), 122–148.
Yeung, K. (2017). 'Hypernudge': Big Data as a mode of regulation by design. *Information, Communication & Society*, *20*(1), 118-136.
Zarsky T (2016) The trouble with algorithmic decisions an analytic road map to examine efficiency and fairness in automated and opaque decision making. Science, Technology & Human Values 41(1): 118–132.
Zhang, Peng, Xiang Shi and Samee Ullah Khan (2017). "Can quantitative finance benefit from IoT?" SmartIoT@SEC.
Zhang, Haoran, Xuyang Wei, Tengfei Zou, Zhongliang Li, and Guocai Yang. (2014). Agriculture Big Data: Research Status, Challenges and Countermeasures, International Conference on Computer and Computing Technologies in Agriculture, Springer.
Zhang, Kuan, Jianbing Ni, Kan Yang, Xiaohui Liang, Ju Ren, and Xuemin Sherman Shen. (2017). "Security and Privacy in Smart City Applications: Challenges and Solutions", IEEE Communications Magazine, Vol. 55, Issue 1, pp. 122-129.
Zhang, Q., and Segall, R. (2008). 'Web Mining: A Survey of Current Research, Techniques, and Software' in International Journal of Information Technology & Decision Making 7(4): 683-720.
Zheng, Wujie, Wenyu Wang, Dian Liu, Changrong Zhang, Qinsong Zeng, Yuetang Deng, Wei Yang, Pinjia He, and Tao Xie, "Testing untestable neural machine translation: An industrial case," Proc. 41st International Conference on Software Engineering: Companion, Poster, 2019.
Zhao, R., Liu, Y., Zhang, N., & Huang, T. (2017). An optimization model for green supply chain management by using a big data analytic approach. Journal of Cleaner Production, 142, 1085-1097

Zhong, R. Y., Newman, S. T., Huang, G. Q., & Lan, S. (2016). Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives. Computers & Industrial Engineering, 101, 572-591.
Zou, H. (2016). Protection of Personal Information Security in the Age of Big Data. 2016 12th International Conference on Computational Intelligence and Security (CIS), 586-589. Retrieved from: https://ieeexplore.ieee.org/abstract/document/7820533.
Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. Journal of Information Technology, 30(1), 75–89.
Singapore Statement on Research Integritym(2010). Accessed at: https://www.jsps.go.jp/english/e-kousei/data/singapore_statement_EN.pdf
Elliot, Joshua "Automating Scientific Knowledge Extraction (ASKE)" Accessed at: https://www.darpa.mil/program/automating-scientific-knowledge-extraction

Bainbridge, William Sims (2007) The Scientific Research Potential of Virtual Worlds Accessed at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.564.112&rep=rep1&type=pdf
Brownlee, Jason (2019) what is a hypothesis in machine learning? accessed at: https://machinelearningmastery.com/what-is-a-hypothesis-in-machine-learning/
Bussen, Lissie (2009) "Robot makes Scientific Discovery all by itself. Accessed at: https://www.wired.com/2009/04/robotscientist/

Cockburn, Iain M., Rebecca Henderson and Scott Stern (2017). "The Impact of Artificial Intelligence on Innovation" Conference on Research Issues in Artificial Intelligence, Toronto, September 2017

Epstein, Joshua M (2007) Generative Social Science: Studies in Agent-Based Computational Modeling. Princeton University Press, Princeton, NJ, 2007, 352 pp.

European Committee For Standardization (2017) Ethics assessment for research and innovation, CWA 17145-1:2017 (E)
European Science Foundation (ESF), and All European Academics (ALLEA), "The European Code of Conduct for Research Integrity" (2011) accessed at: https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020-ethics_code-of-conduct_en.pdf

Falk, Dan (2019) "How Artificial Intelligence Is Changing Science" Accessed at: https://www.quantamagazine.org/how-artificial-intelligence-is-changing-science-20190311/

Jansen, Phillip, Wessel Reijers, David Douglas, Faridun Sattarov, Agata Gurzawska, Alexandra Kapeller & Philip Brey, Rok Benčin, Zuzanna Warso, Robert Braun. (2017) Satori Deliverable 4.1, Accessed at: http://satoriproject.eu/media/D4.1_Proposal_Ethics_Assessment_Framework.pdf

Kitano, Hiroaki (2016) "Artificial Intelligence to Win the Nobel Prize and Beyond: Creating the Engine for Scientific Discovery" Association for the Advancement of Artificial Intelligence. All rights reserved. ISSN 0738-4602
Mosaic, Tom Chivers, (2018) How big data is changing science Accessed at: https://phys.org/news/2018-10-big-science.html
Shamoo, Adil E., and David B. Resnik, (2015) *Responsible Conduct of Research*, 2nd ed., Oxford University Press, Oxford.

Shillabeer, Anna and John F. Roddick (2006) "Towards Role-based Hypothesis Evaluation for Health Data Mining." Electronic Journal of Health Informatics. Vol1(6) Accessed at: https://www.researchgate.net/profile/John_Roddick/publication/215744310_Towards_Role-based_Hypothesis_Evaluation_for_Health_Data_Mining/links/09e415074b5fb67cfa000000.pdf

Wings portal Accessed at: http://wings-workflows.org/wings-portal/