LEVERAGING BIG DATA FOR MANAGING TRANSPORT OPERATIONS

Deliverable D2.3 **Report on Ethical and Social Issues**

Julien Debussche[1], Jasmien César[1], Min-Sung Hong[2], Zuzana Nordeng[2], Rut Waldenfels[3]

Bird & Bird[1], Western Norway Research Institute [2], Goethe-University Frankfurt [3]

August 2018

Work Package 2

Project Coordinator

Prof. Dr. Rajendra Akerkar (Western Norway Research Institute)

Horizon 2020 Research and Innovation Programme

MG-8-2-2017 - Big data in Transport: Research opportunities, challenges and limitations

Distribution level	Public (P)
Due date	31/08/2018 (M10)
Sent to coordinator	27/08/2018
No. of document	D2.3
Title	*Report on Ethical and Social Issues*
Status & Version	*Final*
Work Package	*2: Institutional and governmental issues and barriers*
Related Deliverables	*D2.1, D2.2, D2.4*
Leading Partner	*Bird & Bird*
Leading Authors	*JULIEN DEBUSSCHE, Bird & Bird, (Chapters 1, 2, 3, 4)* *JASMIEN CÉSAR, Bird & Bird, (Chapters 1, 2, 3, 4)* *MINSUNG HONG, WNRI, (Chapters 3.1, 3.4, 3.7)* *RUT WALDENFELS, GUF (Chapter 3.6)*
Contributors	*ZUZANA NORDENG, WNRI, (Chapters 3.2)* *OTTO ANDERSEN, WNRI, (Chapter 3.7)* *RAJENDRA AKERKAR, WNRI, (Chapters 3.1, 3.2, 3.4, 3.6)* *ROBERTO ZICARI, GUF (Chapter 2)*
Reviewers	*REMY RUSSOTTO, CORTE* *MARUSA BENKIC, CORTE*
Keywords	*Big data, Transportation, Ethics, Social issues*

Disclaimer:

This report is part of the LeMO project which has received funding by the European Union's Horizon 2020 research and innovation programme under grant agreement number 770038.

The content of this report reflects only the authors' view. The European Commission and Innovation and Networks Executive Agency (INEA) are not responsible for any use that may be made of the information it contains.

Executive summary

This Deliverable identifies and examines various societal and ethical issues that are relevant to the production of, access to, linking of and re-use of big data in the transport sector.

Chapter 2, on the one hand, discusses the concept of big data, its particular characteristics, and its possible use in the transport sector. On the other hand, it delves into the interaction between ethical and social issues and the ways to integrate these into the existing policy framework. In Chapter 3, the authors examine the various identified ethical and social issues and discuss the challenges and opportunities that may arise in this respect, coming up with notably the following findings:

- **Trust**: Although the research in trust has already become relatively mature, the huge amount and diversity of data and data sources provides lots of new opportunities but at the same time poses many challenges for online trust, notably in the context of transportation.
- **Surveillance**: Considering the role of government agencies and their increasing requests of information to the private sector for public security purposes, it appears necessary to adopt specific rules to regulate the information flow, to define the rights over data and to ensure adequate enforcement.
- **Privacy** (including **transparency**, **consent** and **control**): The advent of the GDPR has had a considerable impact in the domains of privacy, transparency, consent and control. This strengthened legal framework is likely to respond to several ethical issues and thus improve end users' trust in the use of personal data in a big data context.
- **Free will**: Although big data-driven profiling practices can limit free will, a huge part of what we know about the world comes from data analysis. Careful and appropriate information analysis can open up plenty of chances and might reduce the limitations and problems for free will.
- **Personal data ownership**: This Deliverable concludes that a claim of ownership by a data subject in its personal data would be hard to sustain. Nevertheless, in a big data context, different third-party entities may try to claim ownership in (parts of) a dataset, which may hinder the use of big data, including in the transport sector.
- **Discrimination**: Using big data analytics to improve business processes or provide personalised services may lead to discrimination of certain groups of people. Also, the "Digital Divide", i.e. the social differences in access to technology and education or skills to use it, may lead to data-driven discrimination.
- **Environmental**: There are trade-off or rebound effects from the use of big data in transport, which limits the effect of big data exploitation or creates unintended consequences. Such trade-off or rebound effects will be further assessed in Deliverable D2.4.

Finally, the last Chapter serves as a conclusion and introduces possible ways of moving forward to encourage the production of, access to, linking of and re-use of big data in the

transport sector, with a particular focus on the EU. Particularly, Chapter 4 examines whether regulatory intervention or ethics-by-design are appropriate solutions to the challenges caused by ethical and social issues in relation to big data. The conclusion is that regulatory intervention is not desirable. Instead, the authors advocate an approach whereby ethics-by-design is recognised as an EU legal principle, similarly to privacy-by-design, and is supplemented by self-regulation and soft law. Section 4.3.1 aims to provide inspiration for ethics-by-design core implementation principles to be developed further in working groups at EU level.

Table of contents

List of Figures

List of Tables

Glossary

Abbreviation	Expression
AI	Artificial Intelligence
EDPS	European Data Protection Supervisor
ENISA	European Union Agency for Network and Information Security
EU	European Union
GDPR	General Data Protection Regulation
IoT	Internet of Things
ITS	Intelligent Transportation Systems
LeMO	Leveraging Big Data for Managing Transport Operations
PET	Privacy Enhancing Technologies
PIMS	Personal Information Management Systems
RAMS	reliability, availability, maintainability and safety
RFID	Radio Frequency Identification

1 Introduction

1.1 Abstract

This Deliverable identifies and examines various societal and ethical issues that are relevant to the production of, access to, linking of and re-use of big data in the transport sector.

The social and ethical issues that have been identified as most relevant in light of the performed research, and which are further examined in this Deliverable, are as follows:

- Trust
- Surveillance
- Privacy
- Transparency
- Consent
- Control
- Free will
- Personal data ownership
- Discrimination
- Environmental

In a first chapter (Chapter 2), this Deliverable sets the scene by reiterating the concept of big data with its particular characteristics and highlighting the interaction between ethics and social issues.

Chapter 3 examines the various identified ethical and social issues and discusses the challenges and opportunities related thereto. Whenever possible, the various Chapters and Sections provide illustrations in the transport sector.

Finally, the conclusion of this Deliverable (Chapter 4) introduces possible ways to move forward and highlights initiatives that may already exist, focusing mainly on the EU.

1.2 Purpose of the document

Task 2.3 aims to identify and examine the different societal and ethical issues that are relevant to big data currently, and which may be relevant to big data as opportunities for the production of, access to, linking of and re-use of big data in the transport sector that further develops. As part of this task, partners of the LeMO Project examined literature[1] and produce a high-level discussion of ethical and societal issues, including both constructive findings and recommendations, which will serve as a baseline for further development in relation to other Work Packages of the Project. Particular social and ethical issues to be examined may include trust, exploitation, discrimination, as well as any other ethical or social issues we identified from the literature.

[1] Academic journal articles, project reports, media materials, materials from industry and any other relevant information.

More particularly, this Deliverable provides an overview of the context (big data, with a particular focus on the transport sector) and identified, where relevant, issues (challenges) and opportunities. Whenever possible, the various Chapters and Sections provide illustrations with practical considerations and introduce possible ways to move forward and initiatives that may exist, focusing mainly on the EU.

1.3 Target audience

This document will be made publicly available. The results of the research are especially interesting for people working for organisations in the public sector (e.g. politicians, policy makers, policy consultants etc.) and in the private sector (e.g. managers, directors, and consultants).

More specifically, the target audience for this deliverable includes:

- Partners and Advisory & Reference Group in the LeMO project
- European Commission
- EU Parliament
- Horizon 2020 projects and related transport projects (cf. clustering activities)
- Organisations and experts involved in the LeMO case studies
- Public and private transport organisations
- Authorities (regional and national level) that develop and enforce policies and legislation

1.4 Methodology

The present Deliverable focuses on the ethical and social challenges and opportunities relevant to big data in the transport sector. The challenges and opportunities analysed in this Deliverable have been identified by conducting a desk research and examining relevant literature. The conducted desk research consists of the following four pillars:

The first two pillars are online research and the analysis of public journals and magazines. The Internet was searched for articles in reputable magazines and content on relevant websites (e.g., European Union websites) in order to become familiar with the topic. The following key words were used to conduct our research: (i) “Big Data” + “Ethics” / “Ethical/Social issues” / “Ethical/Social challenges” /

"Ethical/Social aspects" / "Social impact"; and (ii) "Transport" or "Transportation" + "Ethics" / "Ethical/Social issues" / "Ethical/Social challenges" / "Ethical/Social aspects" / "Social impact"

The third pillar consists in the review of scientific journals and publications to find relevant and accurate information about the ethical and social aspects of big data, and notably in the transport sector.

The fourth pillar refers to the existing experience of the authors of this Deliverable. Their knowledge of and their experience in big data research and analytics and/or the transport sector have created a general awareness of the challenges and opportunities relevant to this field, and have allowed them to lend a critical eye to the literature.

In addition to these four pillars, whenever relevant, this Deliverable relies on and quotes responses provided by key individuals, mainly from the industry, in the context of interviews.

The list of key ethical and social issues covered by this Deliverable was discussed during a workshop with the whole project consortium as well as with the members of the Advisory & Reference Group to exchange ideas, validate the list, and add additional insights.

The list is indubitably not exhaustive, but aims to cover the ethical and social challenges and opportunities of big data that may be particularly relevant to the transport sector.

This Deliverable therefore identifies, analyses and summarises the challenges and opportunities related to big data from a social and ethical perspective and exemplifies their importance in the transport sector.

The starting point of the analysis of each issue is the EU Charter of Fundamental Rights.[2] Even if this document is also legally binding, it aims at codifying and protecting the common ethical values and principles of the European Union. Accordingly, the recitals of the Charter state that "*conscious of its spiritual and moral heritage, the Union is founded on the indivisible, universal values of human dignity, freedom, equality and solidarity*." In the context of an EU-funded project, it is deemed particularly relevant and appropriate to link the analysis provided in this Deliverable to an EU instrument of such paramount importance.

The Deliverable then gives a general overview of the issues in its conclusion, and suggests possible ways to move forward to strengthen opportunities and/or limit challenges.

[2] Charter of Fundamental Rights of the European Union [2012] OJ C 326/391

2 Setting the scene

2.1 The concept of "big data"

Although this Deliverable does not aim to delve into the technical aspects of big data, it nonetheless emphasises, where needed, some of the particularities of big data and will consider the ethical and social issues having big data analytics technologies in mind.

More concretely, what is to be understood by "big data"? While there is no real consensus on a definition, the initial logical observation is that it is often described as a large dataset comprising different types of data that have grown beyond the ability to manage and analyse them with traditional tools.[3] Hence, handling variable (un-)structured data in real-time requires the adoption and use of new methods and tools (e.g., processors, software, algorithms, etc.).[4]

One could however not discuss the notion of big data without highlighting some of the key characteristics of big data, usually expressed with a series of "V's", and in particular:

- **Volume**: refers to the vast amount of data acquired, stored, searched, shared, analysed, visualised, generated and/or managed. Big data technologies have notably enabled the storage and use of large datasets with the help of distributed systems, where parts of the data are stored in different locations, connected by networks and brought together by software.[5]
- **Velocity**: refers to the speed, which is of essence in a big data context. More particularly, it refers to the speed with which data is stored and analysed, as well as the speed at which new data is generated.[6]
- **Variety**: refers to the heterogeneous types of data that can be analysed, combining structured but also unstructured datasets. There are unanimous findings that most of the data being generated and analysed today is unstructured.

In addition to these three key features, several authors also refer to "**Veracity**" which relates to the ability of analysing datasets that comprise less controllable and accurate data.[7] The accuracy principle is being challenged by some key features of big data. Indeed, "*big data applications typically tend to collect data from diverse sources, and without careful verification of the relevance or accuracy of the data thus collected*."[8] This typically poses legal issues but

[3] Frank J. Ohlhorst, *Big Data Analytics: Turning Big Data into Big Money* (John Wiley & Sons 2012) 3

[4] Commission, 'Towards a Thriving Data-Driven Economy' (Communication) COM(2014) 442 final, 4

[5] Bernard Marr, 'Why only one of the 5 Vs of Big Data really Matters' (*IBM Big Data & Analytics Hub*, 19 March 2015) <http://www.ibmbigdatahub.com/blog/why-only-one-5-vs-big-data-really-matters> accessed 21 August 2018

[6] James R. Kalyvas and Michael R. Overly, *Big Data: A Business and Legal Guide* (Auerbach Publications 2014) 5

[7] Frank J. Ohlhorst, *Big Data Analytics: Turning Big Data into Big Money* (John Wiley & Sons 2012) 3

[8] European Data Protection Supervisor, 'Opinion 7/2015 Meeting the Challenges of Big Data: A Call for Transparency, User Control, Data Protection by Design and Accountability' (EDPS 2015) <https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf> accessed 23 August 2018

also ethical ones such as trust, privacy or transparency. Such issues are further elaborated in this Deliverable.

The "V" of "**Value**" has also been highlighted to refer to the ability of turning data into value.[9] While it could be argued that data, per se, has no value, processing it creates value. In other words, all data collected and stored would not likely generate any value unless it is used by some "intelligent" software algorithms, which analyse data, learn from data, and make/suggest decisions or predictions. Moreover, the value related to data may also lie with the time spent by humans to create the algorithms, to train such algorithms with human-generated examples and answers, or to organise the data. Similarly, the (personal) data provided by individuals in their day-to-day life (by using social media platforms or using an itinerary application for instance), also has a value. In fact, the European Commission explicitly recognised in a proposed Directive in 2015 concerning contracts for the supply of digital content that "*information about individuals is often and increasingly seen by market participants as having a value comparable to money*."[10] It further finds that "*Digital content is often supplied not in exchange for a price but against counter-performance other than money i.e. by giving access to personal data or other data*."[11] On such basis, the Commission proposed to harmonise certain aspects concerning contracts for supply of digital content, taking as a base a high level of consumer protection.[12]

Deliverable D2.1 of the LeMO Project will notably look into the "value" aspect of data in the context of a broader examination of the different economic issues that are relevant to big data currently, and which may be relevant to big data as opportunities for the production of, access to, linking of and re-use of big data in the transport sector.

Finally, when looking into the social and ethical issues related to big data, it is worth considering other disruptive technologies such as Artificial Intelligence ("AI") and its sub-branches, including Machine Learning, Deep Learning, or Neutral Networks, which are all algorithm-based. Such algorithmic methods rely on a vast amount of data (big data) to produce desired results and to find trends, patterns and predictions. In some instances, this Deliverable will elaborate on such other technologies.

2.2 Big data in the transport sector

In the transportation industry, each day vast volumes of data are generated, for example through sensors in passenger counting and vehicle locator systems and ticketing and fare collection systems, just to name a few.

---

[9] Bernard Marr, 'Why only one of the 5 Vs of Big Data really Matters' (*IBM Big Data & Analytics Hub*, 19 March 2015) <http://www.ibmbigdatahub.com/blog/why-only-one-5-vs-big-data-really-matters> accessed 21 August 2018

[10] Commission, 'Proposal for a Directive of the European Parliament and of the Council on certain aspects concerning contracts for the supply of digital content' COM (2015) 634 final

[11] COM (2015) 634 final, Recital 13. See also Gianclaudio Malgieri and Bart Custers, 'Pricing Privacy – The Right to Know the Value of Your Personal Data' (2017) CLSR 289-303

[12] COM (2015) 634 final, Recital 2

Big data opens up new opportunities to define "intelligent" mobility and transportation solutions. Using data analytics, leveraging big data tools and predictive analytics, one can help transportation stakeholders, to make better decisions, improve operations, reduce costs, streamline processes, and eventually better serve travellers and customers.

Deliverable D1.1 of the LeMO Project, entitled "Understanding and mapping big data in transport sector", offers an introduction to big data in the transport sector. It notably identifies untapped opportunities and challenges and describes numerous data sources. Deliverable D1.1 covers six transportation modes (i.e. air, rail, road, urban, water and multimodal) as well as two transportation sectors (passenger and freight). It further identifies several opportunities and challenges of big data in transportation, based on several subject matter expert interviews, applied cases, and a literature review. Finally, it concludes that the combination of different means and approaches will enhance the opportunities for successful big data services in the transport sector, and presents an intensive survey of the various data sources, data producers, and service providers.

2.3 Classification of and interaction between ethical and social issues

The present Deliverable aims to examine the ethical and social aspects of the use of big data in the transport sector. It is however difficult to define and distinguish between what is "social" and what is "ethical". For the purpose of this Deliverable we have attempted to provide the following classification:

- Ethics relates to what actions are perceived as right or wrong within the community one forms part of. It is generally seen as an aspect attached to the individuals within society as well as to the organisations (e.g. public authorities and companies) active within this society, pointing towards their moral conduct. As such, it relates to the accepted behavioural patterns in society. Most Western civilisations have adopted laws and regulations to combat the most conspicuous and persistent ethical issues. In practice, this entails that each individual or organisation going against what is ethically acceptable may be legally punished.
- Social issues relate to problems or matters that affect (a part of) society at large. They may be caused by several societal factors, such as geography, education, environment, economy or politics. Social issues thus affect a large group of individuals, who generally have no power of their own to influence the social issue at stake. Usually, a collective movement, often including government intervention, is necessary in order to change the situation.

Keeping in mind these classifications, it is apparent that ethical and social issues are closely related and, in some situations, even interdependent. For instance, some ethical issues (e.g. discrimination) may also affect society as a whole, thus becoming a social issue (e.g. racism and gender issues). In addition, the use of a particular big data application (e.g. in automated vehicles) may engender both ethical and social issues that are not necessarily interrelated, such as privacy concerns and workforce disruption respectively.

This Deliverable will therefore not depart from a strict distinction between ethical and social aspects when looking into the particularities of the use of big data in the transport sector. Nevertheless, where possible, each issue will be labelled as ‘ethical’ and/or ‘social’.

2.4 Policy framework

The legislator, at EU and/or national level, has adopted policies in order to regulate several aspects related to (big) data, the transport sector, but also – as indicated above – to combat the most conspicuous and persistent ethical issues or to set social norms.

While there are no policies specific to big data, lawmakers have adopted some legislations aimed at protecting the privacy of its citizens, encouraging data sharing among private and public sector entities, and developing policies that support the digitalisation of the transport sector. Some of the key areas of recent policies in the transport sector are for instance the implementation of Intelligent Transport Systems, the increased Open Data policies, Automated Driving, and Smart Mobility.[13]

In a recent interview, Gerhard Kress, who is heading Data Services globally for the Rail business at Siemens and who is part of the Advisory & Reference Group of the LeMO Project, is of the opinion that part of the legal framework is no longer fit for purpose. He states that *“the biggest challenge is that in the rail business we have a very large set of old and country specific regulations that date back many decades. These regulations are meant to protect passengers, but some of them are not anymore fitting to the modern capabilities of technology and instead drive cost and slow innovation down dramatically.”*[14]

In addition to public policies, companies – including in the transport sector – have adopted or adhered to private sector policies. More particularly, the private sector has moved ahead to incorporate policies on the use of big data techniques into their own business models as process or product innovations. The potential applications in the transport sector are diverse, as digitalisation is a major trend of the transport sector.[15]

Despite the existence of public and private policies, the use of new technologies, such as in this case big data-driven technologies, creates new ethical and policy issues that require adopting new or replacing existing policies. “*Normative assessments of transportation plans and policies invoked by policy-makers, researchers and activists often use concepts such as equality, equity, fairness and justice, which are informed by ethical views. Despite the increased interest in these issues in policy debates and research, there are few examples of*

[13] Deliverable D1.2 of the LeMO project reviews current public policies implemented in the EU, its Member States and internationally, which support or restrict the (re-)use, linking of and sharing of data, in the context of big data techniques and in the transport sector.

[14] Interview with Gerhard Kress, Responsible for Data services, Siemens (31 July 2018) http://www.odbms.org/blog/2018/07/on-ai-and-data-technology-innovation-in-the-rail-industry-interview-with-gerhard-kress/

[15] Deliverable D1.2 of the LeMO project illustrates in selected examples of transport-related private companies, the types of private sector policies that have been adopted or promoted.

actual attempts to explicitly address them in transport planning."[16] This Deliverable will however attempt to propose, where possible, possible ways to move forward and determine whether regulatory intervention and/or soft law measures are desirable.

2.5 Identifying ethical and social issues related to big data in the transport sector

The initial use of big data technologies started with marketing. One stage in the life cycle of an emerging science, marketing is low-risk – and, yes, lucrative: "*In marketing and advertising, a decision that is better on average is plenty good enough. You've increased sales and made more money. You don't really have to know why."*[17] But as soon as big data technology moved beyond increasing the odds of making a sale, to being used in higher-stakes decisions like medical diagnosis, loan approvals, hiring and crime prevention, social and ethical implications arose.

The discussions related to ethical (and social) issues in transportation are not new. In 1996 already, Professor Barbara Richardson suggested the need for the establishment of a new field of study and method of analysis to be known as "Transportation Ethics". This new discipline was needed to "*recognize the impact of proposed changes in the transportation system upon elements without our society, and to ensure equity in the distribution of the benefits and the allocation of the harms that together make up that impact."*[18]

What has changed in the transport sector since 1996 is the huge technological development, notably in big data and artificial intelligence.

It follows that, today, there is more than ever a need to look at the ethical and social implications of the use of data-driven technologies, including big data and AI, in the transportation sector.[19]

This Deliverable addresses how the use of (big) data and the deployment of new data-driven technologies may have a strong impact on the ethical and soci(et)al discussions, putting a

[16] Paulo Rui Anciaes and Nikolas Thomopoulos, 'Ethical Issues in Transportation' in Mark Garrett (ed.), *Encyclopedia of Transportation: Social Science and Policy* (SAGE Publications: Thousand Oaks, California, USA 2014)

[17] Interview with Steve Lohr, Reporter, New York Times (19 December 2016) http://www.odbms.org/blog/2016/12/big-data-and-the-great-a-i-awakening-interview-with-steve-lohr/

[18] Larue Tone Hosmer, 'The Call for Transportation Ethics' (1996) 50(1) Transportation Quarterly 22

[19] Rob Smith, '5 Core Principles to Keep AI Ethical' (World Economic Forum, 19 April 2018) <https://www.weforum.org/agenda/2018/04/keep-calm-and-make-ai-ethical/> accessed 22 August 2018

particular emphasis on big data in the transport sector. The issues presented here may nonetheless be valid for other domains.[20]

More specifically, the research conducted in the context of the LeMO Project and this Deliverable has enabled identifying the following key ethical and social issues, deemed to be particularly relevant to big data, including in the transport sector:

- Trust
- Surveillance
- Privacy (including transparency, consent and control)
- Free will
- Personal data ownership
- Data-driven social discrimination and equity
- Environmental

This Deliverable focuses on the above issues, which are detailed in the following Chapters.

This, however, does not mean that other ethical and social issues are not relevant. Indeed, the development of new services in the transport sector that rely on data-driven technologies raise a myriad of technical, economic, legal, ethical and social issues.

The development of "Amazon Prime Air" allows illustrating the possible high number of ethical and social issues that may arise from one particular service. The Seattle-based multinational started Amazon Prime Air as their future delivery system designed to get packages to customers using unmanned aerial vehicles (also known as drones).[21] Prime Air is an example of a data-driven service implemented using advanced AI technologies.

On the one hand, the potential business benefits of Prime Air is to allow decreasing delivery and waiting time (which is positive towards customers) and delivery costs (which is positive towards both Amazon itself and the customers). On the other hand, Prime Air is also a good example of how a new technology-driven innovation raises numerous questions, notably highlighting social and ethical concerns:

- Is it safe to have drones delivering packages?
- Can drones fly over homes and properties, and when?
- Who would be responsible in case of a crash or a collision?

---

[20] A possible example in healthcare: where predictive analytics are used for example to forecast if patient will develop a particular disease. The ethical issues arising are illustrated by Claudia Perlich- a well-known data scientist- ex Chief scientist at Dstillery, which poses this controversial question: "What happens if my algorithm is wrong? Someone sees the wrong ad. What's the harm? It's not a false positive for breast cancer." Claudia Perlich, quoted by Steve Lohr in 'Civility in the Age of Artificial Intelligence' (ODBMS, 6 February 2016) <http://www.odbms.org/2016/02/civility-in-the-age-of-artificial-intelligence/> accessed 22 August 2018

[21] From the web site of Amazon: "*We're excited about Prime Air — a delivery system from Amazon designed to safely get packages to customers in 30 minutes or less using unmanned aerial vehicles, also called drones. Prime Air has great potential to enhance the services we already provide to millions of customers by providing rapid parcel delivery that will also increase the overall safety and efficiency of the transportation system*." https://www.amazon.com/Amazon-Prime-Air/b?ie=UTF8&node=8037720011

- Will drones replace people?[22]
- Is it justifiable for companies to replace a deliveryman by a machine?
- Will the workforce change by hiring more qualified individuals possibly at a higher pay (e.g. engineer, data scientist, etc.)?
- Is the use of drones more ecological?
- Will the drone-based delivery service concentrate on certain urban or rural areas?
- If a drone delivery is faster, will it entail extra costs affordable only by a higher social class?
- Can drones serve multiple purposes and collect additional information (including personal data) while flying over areas and delivering packages?

2.6 Assigning responsibilities

The data value cycle can be rather complex and involves numerous stakeholders. Many of such stakeholders are likely to have some kind of responsibility because, for instance, they create or generate data or algorithms, or because they use, compile, select, structure, re-format, enrich, analyse, purchase, take a licence on, or add value to the data.

This complexity increases the difficulties in determining who could be ethically and socially responsible and consequently legally liable for any wrongdoing and damage, or required to integrate ethical and social principles in their processes. Are computer system designers (e.g. software developers, software engineers, data scientists, data engineers), data providers (e.g. data brokers and marketplaces, individuals, public authorities), or other actors responsible?

To illustrate the intracies of the question, let‘s consider an autonomous car that relies entirely on an algorithm that had taught itself to drive by watching a human doing it. What if one day the car crashes into a tree, or even worse kills a pedestrian? "*If the learning took place before the car was delivered to the customer, the car's manufacturer would in all likelohood be liable, just as with any other machinery. The more interesting problem is if the car learned from its driver. Did the driver set a bad example, or did the car not learn properly?*"[23] This is referred to in the literature as the crash assignment, especially between automated vehicles and non-automated vehicles. Some researchers have indicated that automated vehicles will need to be programmed with some sort of ethical system in order to make decisions on how to crash. Few studies, however, have been conducted on how particular ethical theories will actually make crash decisions and how these ethical paradigms will affect automated vehicle programming.[24]

---

[22] Since Data is useful only when it is used, we can distinguish between two approaches of using data: You can use data technologies (such as AI) either to automate or to augment humans. In the first case, machines replace people, in the second case machine complements people (at least in theory) – Ugo Pagallo, 'The Legal Challenges of Big Data: Putting Secondary Rules First in the Field of EU Data Protection' (2017) 3(1) EDPL 36

[23] Interview with Pedro Domingos, Professor of computer science, University of Washington (18 June 2018) http://www.odbms.org/blog/2018/06/on-artificial-intelligence-machine-learning-and-deep-learning-interview-with-pedro-domingos/

[24] Wesley Kumfer and Richard Burgess, 'Investigation into the Role of Rational Ethics in Crashes of Automated Vehicles' (2015) 2489 Transportation Research Record 130

It follows that, in many instances, the responsibility will be shared between the many actors. What seems however clear is that system designers have a responsibility to integrate ethical and social principles when developing, designing, selecting and using applications, services and products that are based on the processing of data or process data to fulfil their task. This requires them, when developing and designing such products, services and applications, to consider the state of the art, to make sure that they are able to incorporate ethical and social principles:

- Oren Etzioni (CEO of the Allen Institute for Artificial Intelligence) affirmed that *"we have a profound ethical responsibility to design systems that have a positive impact on society, obey the law, and adhere to our highest ethical standards."*[25]
- John Markoff (Journalist, specialized in IT, The New York Times) believes that *"the most important aspect of this question is the simple acknowledgement that intelligent system designers do have ethical responsibilities."*[26]

2.7 Integrating ethical and social principles

Assuming that ethical and social norms should be taken into account, and despite the complexity of determining who should be responsible for their integration, questions nonetheless arise as to whether it is possible to design intelligent software that uses algorithms and data with ethical and social principles and where such principles should be laid down.

Murphy et al., developed propositions — based on both previous ethics research as well as the larger organisational behaviour literature — examining the impact of attitudes, leadership, presence/absence of ethical codes and organisational size on corporate ethical behaviour. The results, which come from a mail survey of 149 companies in a major U.S. service industry, indicate that attitudes and organisational size are the best predictors of ethical behaviour. Leadership and ethical codes contribute little to predicting ethical behaviour.[27]

In order to further elaborate on the topic, one may wonder how it is possible to define incentives to adopt an ethical approach to service, software and algorithm development, especially in the context of AI. Indeed, "*if we just let machines learn ethics by observing and emulating us, they will learn to do lots of unethical things. So maybe AI will force us to confront what we really mean by ethics before we can decide how we want AIs to be ethical*."[28] In the same vein, we currently do not exactly understand how advanced AI

[25] Interview with Oren Etzioni, CEO, Allen Institute for Artificial Intelligence (15 January 2016) http://www.odbms.org/blog/2016/01/on-artificial-intelligence-and-society-interview-with-oren-etzioni/

[26] Interview with John Markoff, Reporter, The New York Times (11 August 2016) http://www.odbms.org/blog/2016/08/machines-of-loving-grace-interview-with-john-markoff/

[27] Paul R. Murphy, Jonathan E. Smith, James M. Daley, 'Executive Attitudes, Organizational Size and Ethical Issues: Perspectives on a Service Industry' (1992) 11(1) Journal of Business Ethics 11

[28] Interview with Pedro Domingos, Professor of computer science, University of Washington (18 June 2018) http://www.odbms.org/blog/2018/06/on-artificial-intelligence-machine-learning-and-deep-learning-interview-with-pedro-domingos/

techniques (such as used in deep learning and neural networks) work as it is a trial and error. This is due to the technical complexity of such advanced technologies and neural networks, which need a huge amount of data to learn properly. This naturally poses ethical and societal issues as we do not know how and why the most advanced algorithms do what they do and can therefore not explain how (non-ethical) decisions are made.[29]

Despite the technical difficulties as well as the uncertainties as to where ethical and social principles should be incorporated and as to whether they should be codified, by whom, and how, there have been initiatives to define such principles, specifically in the area of autonomous vehicles, by both independent associations and political bodies:

- For instance, t*he IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*, which aims *"to ensure every stakeholder involved in the design and development of autonomous and intelligent systems is educated, trained, and empowered to prioritize ethical considerations so that these technologies are advanced for the benefit of humanity."*[30]
- Similarly, Microsoft established an AI and Ethics in Engineering and Research (AETHER) Committee, bringing together senior leaders from across the company to focus on proactive formulation of internal policies and how to respond to specific issues in a responsible way. AETHER aims to ensure that Microsoft's AI platform and experience efforts are deeply grounded within its core values and principles and benefit the broader society.[31]
- The German *Ethics Commission on Automated Driving*, a political body, was set up by the German Federal Minister of Transport and Digital Infrastructure.[32] It recently presented a report developing guidelines for the programming of automated driving systems.[33] More particularly, "*the Ethics Commission on Automated and Connected Driving has developed initial guidelines for policymakers and lawmakers that will make it possible to approve automated driving systems but that set out special requirements*

[29] "Since the algorithms learn from data, it's not as easy to understand what they do as it would be if they were programmed by us, like traditional algorithms. But that's the essence of machine learning: that it can go beyond our knowledge to discover new things. A phenomenon may be more complex than a human can understand, but not more complex than a computer can understand. And in many cases we also don't know what humans do: for example, we know how to drive a car, but we don't know how to program a car to drive itself. But with machine learning the car can learn to drive by watching video of humans drive." Interview with Pedro Domingos, Professor of computer science, University of Washington (18 June 2018) http://www.odbms.org/blog/2018/06/on-artificial-intelligence-machine-learning-and-deep-learning-interview-with-pedro-domingos/

[30] https://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

[31] https://news.microsoft.com/2018/03/29/satya-nadella-email-to-employees-embracing-our-future-intelligent-cloud-and-intelligent-edge/

[32] The commission was set up by Federal Minister Alexander Dobrindt.

[33] Written by a body of experts, headed by Professor Udo Di Fabio, a former Federal Constitutional Court judge.

in terms of safety, human dignity, personal freedom of choice and data autonomy."[34] The report presented in August 2017 comprises twenty propositions. The key elements are summarized as follows:[35]

- Automated and connected driving is an ethical imperative if the systems cause fewer accidents than human drivers (positive balance of risk).
- Damage to property must take precedence over personal injury. In hazardous situations, the protection of human life must always have top priority.
- In the event of unavoidable accident situations, any distinction between individuals based on personal features (age, gender, physical or mental constitution) is impermissible.
- In every driving situation, it must be clearly regulated and apparent who is responsible for the driving task: the human or the computer.
- It must be documented and stored who is driving (to resolve possible issues of liability, among other things).
- Drivers must always be able to decide themselves whether their vehicle data are to be forwarded and used (data sovereignty).

[34] Prof. Di Fabio in a press release https://www.bmvi.de/SharedDocs/EN/PressRelease/2017/084-ethic-commission-report-automated-driving.html. See Ethics Commission's complete report https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile

[35] Federal Ministry of Transport and Digital Infrastructure, 'Ethics Commission, Automated and Connected Driving' (BMVI 2017) <https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile> accessed 22 August 2018

3 Analysis of key ethical and social issues

3.1 Trust

3.1.1 Introduction

The Oxford English Dictionary defines trust as the "*firm belief in the reliability, truth, or ability of someone or something*". As such, trust is not recognised as a fundamental right in the EU Charter of Fundamental Rights. However, together with the concept of surveillance, it may be discerned in the right to liberty and security acknowledged by Article 6 of the Charter.

Big data refers to huge datasets that are challenging to store, search, share, visualise and analyse, regarding the intrinsic characteristics of big data (i.e. the 5 V's - see also Section 2.1 above). Veracity is one of the main dimensions of big data, describing consistency and trustworthiness.[36] It is quite difficult if not impossible to outline one general shared understanding of trust in relation to big data, as Vallor[37] points out, since the trustworthiness of information can change depending on whom we speak to, where the data is gathered or how it is presented.

"Trust" related to big data can be distinguished as follows: (i) trust is the result of the belief in the 'honesty of stakeholders' in the process of collecting, processing, analysing the big data; and (ii) 'trust with big data', meaning research of relation between trust and big data.

Stakeholders' honesty: Veracity is in principle a moral requirement according to which big data users (collectors, analysts, brokers, etc.) should respect the individual citizen as a data provider, for instance, by facilitating his or her informed consent. The right not to be subject to profiling or the right to information for individual citizens (as enshrined in the General Data Protection Regulation) seem to be constantly under pressure when veracity, honesty and consequently trust are not upheld.[38]

Trust with big data: One can observe two sides to research in trust and big data, namely 'big data for trust' and 'trust in big data'. The former addresses questions of how to use big data for trust assessment (i.e. to measure trust) whereas the latter discusses possibilities and challenges to create trust in big data.[39] This Chapter focuses on the technologies of trust with big data, providing practical illustrations in the transport sector.

---

[36] Akhil Mittal, 'Trustworthiness of Big Data' (2013) 80(9) International Journal of Computer Applications 35

[37] Shannon Vallor, *Technology and the virtues: A philosophical guide to a future worth wanting* (Oxford University Press 2016)

[38] Bart Custers and others, 'Deliverable 2.2 Lists of Ethical, Legal, Societal and Economic Issues of Big Data Technologies. Ethical and Societal Implications of Data Sciences' (e-SIDES, 2017) 33-34 <https://e-sides.eu/resources/deliverable-22-lists-of-ethical-legal-societal-and-economic-issues-of-big-data-technologies> accessed 22 August 2018

[39] Johannes Sänger and others, 'Trust and Big Data: A Roadmap for Research' in Morvan F, Wagner R R and Tjoa A M (eds) *2014 25th International Workshop on Database and Expert Systems Applications* (IEEE, 2014) 278-282, DOI: 10.1109/DEXA.2014.63

3.1.2 Description of the challenges in relation to trust

3.1.2.1 Big data for trust

The variety of big data may bring opportunities but may also be challenging in relation to its quality (e.g. heterogeneous and unstructured data). Trust and reputation systems are closely related to the 'trust for big data'. The general process of reputation systems can be separated into three steps as follows: (i) collection and preparation; (ii) computation; and (iii) storage and communication.

In the first step, data or information about the past behaviour of a trustee are gathered and prepared for subsequent computing. The vast number of web applications such as e-commerce platforms, online social networks or content communities has led to huge amounts of reputation data being created from big data. In order to collect and integrate data from different sources, several cross-community reputation systems have already been made by research groups of Pingel[40] and Gal-Oz[41]. Particularly, users who have invested a lot of time in shaping their good reputation are interested to know how their reputation profiles are transformed into different platforms. While Steinbrecher[42] points out the need for interoperable reputation systems, Sänger et al. further highlight the need to integrate both explicit and implicit information.[43] To extract implicit reputation information from big data, most of data that is semi- or unstructured according to the variety property of big data, is handled to get the implicit information through natural language processing and machine learning.

In the second step, both the explicit and implicit reputation information are used in the computation phase to calculate a reputation value as its output. This phase consists of filtering, weighting and aggregation processes. Filtering of answers is useful for further processing. The second process-step concerns the question of how relevant the information is used for the specific situation. Finally, the reputation values are combined to generate one or several scores. Regarding the computation step, following issues might be raised: (i) reducing big data for trust assessment; (ii) differentiating between fast changed and important information in the long term; and (iii) designing the computation process considering rapid increment of data.

---

[40] Franziska Pingel and Sandra Steinbrecher, 'Multilateral Secure Cross-community Reputation Systems for Internet Communities' in Steven Furnell, Sokratis K. Katsikas, Antonio Lioy (eds), *Trust, Privacy and Security in Digital Business* (TrustBus 2008, Lecture Notes in Computer Science, vol 5185, Springer, Berlin, Heidelberg)

[41] Nurit Gal-Oz, Tal Grinshpoun, and Ehud Gudes, 'Privacy Issues with Sharing Reputation across Virtual Communities' in Traian Marius Truta and others (eds) *Proceedings of the 4th International Workshop on Privacy and Anonymity in the Information Society* (ACM, 2011) DOI: 10.1145/1971690.1971693

[42] Sandra Steinbrecher, 'The need for Interoperable Reputation Systems' in Jan Camenisch, Valentin Kisimov, Maria Dubovitskaya (eds) in *Open Research Problems in Network Security* (Lecture Notes in Computer Science, vol 6555, Springer, Berlin, Heidelberg 2011) 159

[43] Johannes Sänger and others, 'Trust and Big Data: A Roadmap for Research' in Morvan F, Wagner R R and Tjoa A M (eds) *2014 25th International Workshop on Database and Expert Systems Applications* (IEEE, 2014) 278-282, DOI: 10.1109/DEXA.2014.63

The final storage and communication step stores the predicted reputation scores and provides them with extra information to support the end users in understanding the meaning of a score-value. In this regard, we may encounter challenges about the reusability of reputation information and the transparency of communication.

3.1.2.2 Trust in big data

The goal is to measure the trustworthiness and accuracy of big data to create high values of data which are coming in large volume from a wide variety of applications/interfaces in different formats and are rapidly changing. Organisation of data contains valuable information and this valuable information is not limited to internal data. For effective analytical business intelligence, it should be understood by organisations.

In this regard, analysing this valuable information is not as easy as it seems. There are tools available to extract and process this valuable information from disparate sources, but the real challenge is to know whether the data processed are trustworthy, accurate and meaningful.[44]

With respect to the trust in big data, there are several trust issues as follows: (i) trust in data quality; (ii) measuring trust in big data; and (iii) trust in information sharing.[45] As mentioned above, verifying their trustworthiness and especially evaluation of the quality of the input data is essential due to the higher volume of data sources than ever before. In order to ensure the quality of input data, detecting manipulations of the data should be conducted before processing it.[46] To ensure data quality, several data mining approaches such as feature selection and unsupervised learning methods for sparsity, error-aware data mining for uncertainty, and data imputation methods for incompleteness has been studied[47] as well as the authentic synthetic data benchmarking of different big data solutions has been generated[48].

Big data has also been used to overcome the lack of data quality with data quantity. Tien[49] argues that this approach may even be more realistic than the optimal focus of traditional methods. Finding ways to quantify trust is vital to enable the comparability of data sources and help to make the notion of trust more tangible. In one study, three dimensions of

[44] Akhil Mittal, 'Trustworthiness of Big Data' (2013) 80(9) International Journal of Computer Applications 35

[45] Johannes Sänger and others, 'Trust and Big Data: A Roadmap for Research' in Morvan F, Wagner R R and Tjoa A M (eds) *2014 25th International Workshop on Database and Expert Systems Applications* (IEEE, 2014) 278-282, DOI: 10.1109/DEXA.2014.63

[46] Cloud Security Alliance Big Data Working Group, *Expanded top ten big data security and privacy challenges* (White Paper, 2013)

[47] Xindong Wu and others, 'Data Mining with Big Data' (2014) 26(1) IEEE Transactions on Knowledge and Data Engineering 97

[48] Zijian Ming and others, 'BDGS: A Scalable Big Data Generator Suite in Big Data Benchmarking' in Tilmann Rabl and others (eds) *Advancing Big Data Benchmarks* (WBDB 2013, Lecture Notes in Computer Science, vol 8585, Springer, Cham)

[49] James M. Tien, 'Big Data: Unleashing Information' (2013) 22(2) Journal of Systems Science and Systems Engineering 127

subjectivity, deception and implausibility were combined to measure trust in big data.[50] Additionally, an automated system for trust assessment of online open media which is based on both the assessment of the source and the content was proposed.[51] Albanese[52] tried to quantify the trustworthiness of both the data source and data items, based on Google's PageRank method.

Two directions of trust need to be considered in connection with information sharing, for partnering organisations. One relates to the adequate quality of providing data by partners and currently focuses more on the quality of data that has already been processed by another organisation rather than raw data. On the other hand, the latter concerns processing data only in appropriate and agreed ways and has attracted the attention of researchers in the research field of de-identification.

3.1.3 Description of the opportunities in relation to trust

To derive business value from non-traditional, unstructured and social media data, organisations need to adopt the right technology and infrastructure to analyse the data to get new insights and business intelligence analysis. It can be feasible with the completeness, trustworthiness, consistency and accuracy of big data.

In addition, as mentioned above, big data can provide a lot of opportunities by providing trustworthiness such as reputation. In this regard, the challenges and opportunities are raised for applications using big data, as follows: workload provisioning, simulate production behaviour, understanding of complex data architecture, processing and analytics, lack of realistic data sets, regulatory compliance and early involvement during algorithm definition phase.[53]

[50] Tatiana Lukoianova and Victoria Rubin, 'Veracity Roadmap: Is Big Data Objective, Truthful and Credible?' (2013) 24(1) Advances In Classification Research Online 4 DOI: 10.7152/acro.v24i1.14671

[51] Nadya Belov and others, 'Computational Trust Assessment of Open Media Data' in *2013 IEEE International Conference on Multimedia and Expo Workshops* (IEEE, 2013) DOI**:** 10.1109/ICMEW.2013.6618392

[52] Massimiliano Albanese, 'Measuring Trust in Big Data' in Aversa R and others (eds) *Algorithms and Architectures for Parallel Processing* (ICA3PP, Lecture Notes in Computer Science, vol 8286, Springer, Cham 2013) https://doi.org/10.1007/978-3-319-03889-6_28

[53] Akhil Mittal, 'Trustworthiness of Big Data' (2013) 80(9) International Journal of Computer Applications 35

3.1.4 Examples of trust in the transport sector

Trust in the transport sector – Example 1

One example of big data use with regard to trust was proposed by Melis et al. They aimed at improving the quality of urban life by collecting personal data, tracking citizens' movements, correlating them with many other sources of information, and making the results widely available. Figure 1 below shows a development scenario of mobility-as-a-service in an urban area.

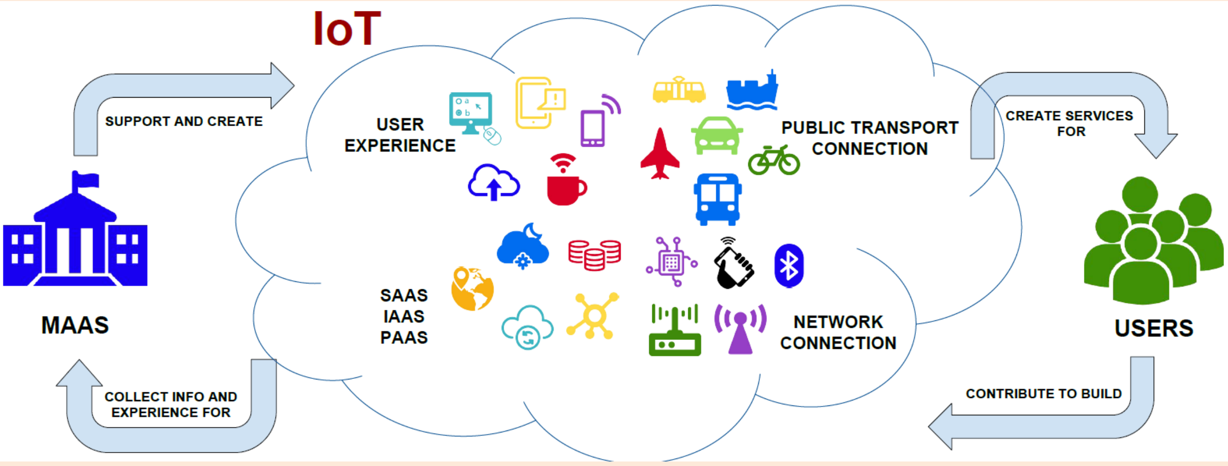

Figure 1 Mobile as a service, Internet of Things and User work-flow for Urban area[54]

The central cloud is the Internet of Things (IoT) architecture understood as a heterogeneous set of networks, technologies and experiences. With the central infrastructure, the user enjoys the huge power of the IoT infrastructure and its vast datasets to improve the quality of reliable urban life. The role of mobility-as-a-service is twofold: (i) using infrastructural functions for high quality data; and (ii) creating many layers of value-added applications based on the trusted data. The principle of trust presents a transparent way of big data collection and the big data analysis results sharing with wide population without restrictions. This way we observe both aspects of trust with data.

[54] Andrea Melis, 'Public Transportation, IoT, Trust and Urban Habits' in Franco Bagnoli and others (eds) *Internet Science* (INSCI 2016, Lecture Notes in Computer Science, vol 9934, Springer, Cham, 2016) 318.

Trust in the transport sector – Example 2

Intelligent transportation systems (ITS) are arguably the most anticipated smart city services.[55] Ferdowsi et al. proposed an edge analytics architecture for ITS in which data is processed at the vehicle or roadside smart sensor level to overcome the ITS reliability challenges.[56] The architecture is illustrated in Figure 2 below.

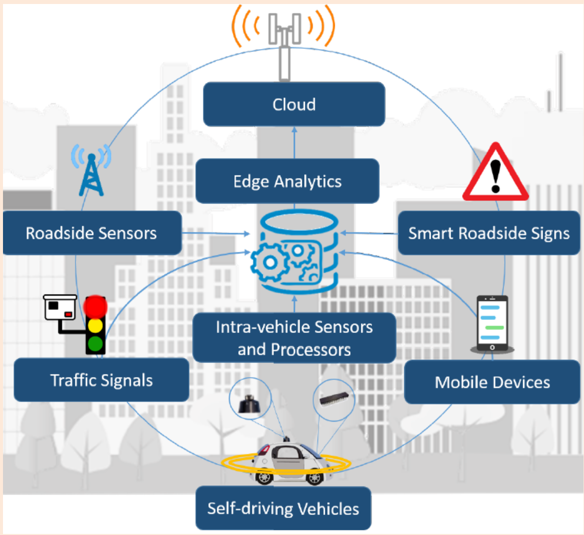

Figure 2 ITS edge analytics architecture and components

The proposed ITS edge analytics architecture exploits deep learning techniques running at the level of passengers' mobile devices and intra-vehicle processors to process large datasets and enable a truly smart transportation system operation. This architecture improves the performance of ITS in terms of reliability and latency. This example also presents a transparent way of big data collection and the immediate use of processed data in traffic management.

[55] Junping Zhang and others, 'Data-driven Intelligent Transportation Systems: A Survey' (2011) 12(4) IEEE Transactions on Intelligent Transportation Systems 1624

[56] Aidin Ferdowsi, Ursula Challita and Walid Saad, 'Deep Learning for Reliable Mobile Edge Analytics in Intelligent Transportation Systems' (2017) abs/1712.04135 CoRR arXiv:1712.04135

3.1.5 Summary

The research in trust has already become relatively mature.[57] However, the huge amount and diversity of data and data sources provides lots of new opportunities but at the same time poses many challenges for online trust. This research is relevant to academia as well as to industry and the ordinary citizen specialised in the transportation area.

Dynamic technological developments where individuals do not have enough time to adapt, may lead to significant trust issues.[58] When it comes to mobility in the field of public transportation, safety and security are linked to trust: people need to feel safe when using transportation means which are new (e.g. self-driving cars) or which could be perceived as threatening (e.g. car-sharing with unknown drivers). As an example, gender or age perspectives could help in designing mobility-as-a-service, considering different needs in terms of easiness and sense of safety in public spaces.

Opportunities in relation to trust in the context of big data in the transport sector	Challenges in relation to trust in the context of big data in the transport sector
Higher volume in big data leads to trustworthiness, such as reputation.	The huge amount and diversity of data and data sources reduce the average quality of overall data.
A solid partnership between organisations can lead to trust in information sharing.	Trust related to big data can be easily polluted by various kinds of people who act against a moral requirement.

Table 1 Summary table of opportunities and challenges in relation to trust in the context of big data in the transport sector

[57] Audun Josang, 'Robustness of Trust and Reputation Systems' in *2010 Fourth IEEE International Conference on Self-Adaptive and Self-Organizing Systems Workshop* (IEEE, 2010) DOI: 10.1109/SASOW.2010.33

[58] Akhil Mittal, 'Trustworthiness of Big Data' (2013) 80(9) International Journal of Computer Applications 35

3.2 Surveillance

3.2.1 Introduction

The Oxford English Dictionary describes surveillance as "*close observation, especially of a suspected spy or criminal*". In the context of this Deliverable, however, the term surveillance will be used to describe the close observation of all humans in general, irrespective of their criminal tendencies. In light of the EU Charter of Fundamental Rights, surveillance can be linked to Article 6 on the right to liberty and security, Article 7 on the respect for private and family life, and Article 8 on the protection of personal data (see also Section 0 on privacy 아래).

According to Andrejevic, in a big data context, monitoring and surveillance has the following six key characteristics[59]:

1. Tracking is "populational":	Big data has as a result that tracking relates to a group of people rather than being targeted at specific individuals.
2. Correlation and predictability are no longer needed:	When the necessary conditions are fulfilled, big data analytics can provide a reliable and veracious outcome, thus rendering the drawing of assumptions on the basis of correlation and predictions redundant. This is also demonstrated by Anderson's diagnosis of the fate of theory in the so-called "petabyte" era: "*Out with every theory of human behaviour, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves.*"[60]
3. Monitoring is pre-emptive:	As Bogard[61] notes in his analysis of the simulation of surveillance, the goal of predictive analytics is not simply predicting outcomes, but devising ways of altering them. In policing terms, the goal of predicting the likelihood of criminal behaviour is to deter it.
4. Tracking is interventionist:	In the future, we can expect that predictive analytics will become more sophisticated and will be deployed across

[59] Mark Andrejevic, 'Surveillance in the Big Data Era' in Kenneth D. Pimple (ed) *Emerging Pervasive Information and Communication Technologies (PICT)* (Law, Governance and Technology Series, vol 11, Springer, Dordrecht, 2014) 55

[60] Chris Anderson, 'The End of Theory: The Data Deluge Makes the Scientific Method Obsolete' *Wired Magazine* (16.07, 2008) <https://www.wired.com/2008/06/pb-theory/> accessed 23 August 2018

[61] William Bogard, *The Simulation of Surveillance: Hyper-control in Telematic Societies* (Cambridge University Press 1996)

a broad range of social life to shape and sort consumer behaviour and opportunities.

5. All information is relevant: Because predictive analytics is, as it were, model-agnostic, it does not rule out in advance the relevance of any kind of information.

6. "Privacy" is irrelevant: Any attempt to build a protective bulwark against big data surveillance on the foundation of privacy must confront the fact that much of the tracking is anonymous.

Digging a bit more into the sixth characteristic, some scholars claim that - in principle - data miners are not interested in the details of specific individuals so much as in the way these individuals fit into patterns of correlation. This being said, in some cases, data patterns are linked to individuals even if these people are not identified.[62] Similarly, while some only need to rely on "scraping" anonymous data off the internet, other forms of monitoring necessarily require tracking an individual in ways that can easily lead to identification.

At the same time, much of the data that qualifies as 'personal data' is not generally considered particularly sensitive, in part because we are not yet cognizant of how it can be used. In other words, its collection may not feel like the kind of intrusion that we associate with an invasion of privacy. For example, we may not think of our patterns of consumption or our movements throughout the day as intimate affairs to be protected by a veil of privacy. After all, they are often publicly visible, but they can nevertheless be used to influence us in unanticipated and unobtrusive ways. The invocation of traditional conceptions of privacy misses its mark in such cases.[63]

Furthermore, just as database monitoring focuses on the population and not on targeted individuals, the results it yields are probabilistic. Mining big data does not provide any definite answers about what individuals will do or how they will react. Rather, it provides probabilities about what someone in a category is likely to do.[64]

3.2.2 Description of the challenges in relation to surveillance

Two main issues arise in relation to surveillance: (i) risks of asymmetries in the control over information; and (ii) privacy.

The first one points to the availability of big data giving a competitive advantage to those who own them in terms of capability to predict new economic, social and political trends. The information and knowledge deriving from big data is not accessible to everyone, as it is based

[62] Mark Andrejevic, 'Surveillance in the Big Data Era' in Kenneth D. Pimple (ed) *Emerging Pervasive Information and Communication Technologies (PICT)* (Law, Governance and Technology Series, vol 11, Springer, Dordrecht, 2014) 55

[63] Ibid

[64] Ibid

on the availability of large datasets, expensive technologies and specific human skills to develop sophisticated systems of analyses and interpretation. For these reasons, governments and big businesses are in the best position to take advantage of big data: they have large amounts of information on citizens and consumers and enough human and computing resources to manage it.[65]

When it comes to the privacy issues, according to Andrejevic[66], the shift from targeted to "populational" monitoring is facilitated by the advent of interactive, networked forms of digital communication that generate easily collectible and storable meta-data. But the logic is self-stimulating and recursive: once the switch to an inductive, data-driven form of monitoring takes place, the incentive exists to develop the technology to collect more and more information and to "cover" as much of everyday life as possible.

3.2.3 Description of the opportunities in relation to surveillance

An opportunity that lies in big data in relation to surveillance is that comprehensiveness replaces comprehension. In other words, big data replaces detection with collection and lets the algorithm do the work. The whole population is monitored through allowing computers to detect anomalies or other patterns that correlate with suspicious activity. It is important to remember that the purpose of monitoring is not eavesdropping on everyone.

Surveillance in big data contributes to preventive policing. Rather than starting with a suspect and then monitoring him or her, the goal is to start from generalised surveillance and then generate suspects.[67] Building on the work of Baudrillard[68], Bogard[69] has described this form of monitoring as "the simulation of surveillance" – not just monitoring as deterrent (the placement of a surveillance camera in a notorious crime spot, for example) but as a strategy for intervening in the future by modelling it. As Bogard puts it, "*the goal of information and communication management technologies is simply to control as perfectly and seamlessly as possible all conceivable outcomes in advance. This is the logic behind data mining, profiling and the like*."[70] Take the example of predictive policing – the use of statistics about everything from past patterns of criminal behaviour to the weather to predict when and where crimes are going to take place before they happen. Yet if predictive analytics are not simply predicting outcomes but devising ways of altering them, the ethical and social question is not whether and how big data controls 'all conceivable outcomes' (Bogard) but rather to what

[65] Alessandro Mantelero and Giuseppe Vaciago, 'The "Dark Side" of Big Data: Private and Public Interaction in Social Surveillance' (2013) 14(6) Computer Law Review International 161

[66] Andrejevic, 'Surveillance in the Big Data Era' in Kenneth D. Pimple (ed) *Emerging Pervasive Information and Communication Technologies (PICT)* (Law, Governance and Technology Series, vol 11, Springer, Dordrecht, 2014) 55

[67] Ibid

[68] Jean Baudrillard, *Simulacra and simulation* (University of Michigan Press 1994)

[69] William Bogard, *The Simulation of Surveillance: Hyper-control in Telematic Societies* (Cambridge University Press 1996)

[70] William Bogard, 'Welcome to the Society of Control' in Kevin D. Haggerty and Richard V. Ericson (eds) *The New Politics of Surveillance and Visibility* (University of Toronto Press 2006) 4

extent it selects and directs one (or multiple) favoured outcome. The operations of surveillance may thereby impact the use of free will (see Section 3.4 below).

Applying this to the transport sector, ITS developments have given rise to car and in-car surveillance. This generates trails that are closely associated with individuals and which are available to various organisations. Crash cameras in cars, for example, could be imposed as a condition of purchase, insurance, or rental. Like so many other data trails, the data can be used for other purposes than originally intended (accident investigation), and with or without informed, freely given, and granular consent. In the UK and some other countries, automated number plate recognition (ANPR) has exceeded its nominal purpose of traffic management to provide vast mass transport surveillance databases.[71] These issues were already debated when RFID[72]-based smart passports were introduced, but the international bodies promoting and agreeing to their use only conditionally acknowledged these issues because the international agreements were restricted to border crossing. The threats involved have however penetrated far enough into public consciousness such that wallets providing shielding of Radio Frequency Identification (RFID) chips are now readily procurable.

3.2.4 Examples of surveillance in the transport sector

Surveillance in the transport sector – Example 1
Road traffic surveillance systems aid commuters, provide valuable data for traffic police and traffic infrastructure managers, enforce laws and encourage safe driving. Despite all these things, the leverage of software systems for traffic management has emerged but the focus is narrowed down to develop a core video processing algorithm, which serves the purpose of traffic monitoring.[73] These Integrated Video Security Systems (IVSS) support the traffic infrastructure managers with an unblinking eye support both in terms of monitoring and decision making. Figure 3 and Figure 4 show the proposed framework for and the surveillance part of the IVSS.

[71] Marcus R. Wigan, and Roger Clarke, 'Big Data's Big Unintended Consequences' (2013) 46(6) Computer 46
[72] Radio Frequency Identification
[73] Chunchu Mallikarjuna, A. Phanindra, and K. Ramachandra Rao, 'Traffic Data Collection under Mixed Traffic Conditions Using Video Image Processing' (2009) 135(4) Journal of Transportation Engineering 174

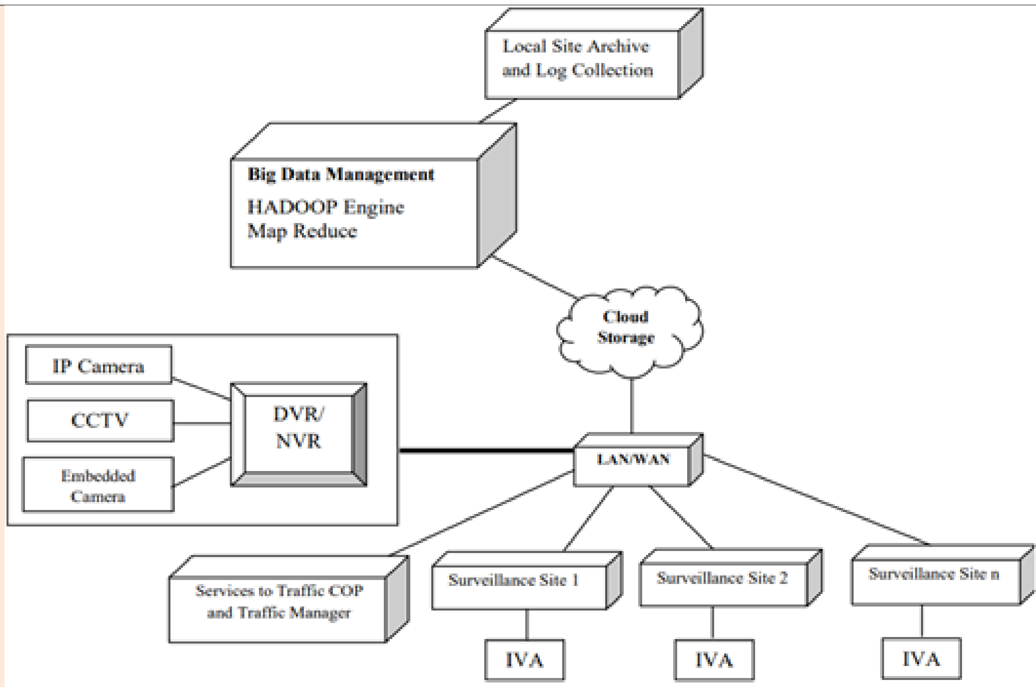

Figure 3 The framework of IVSS

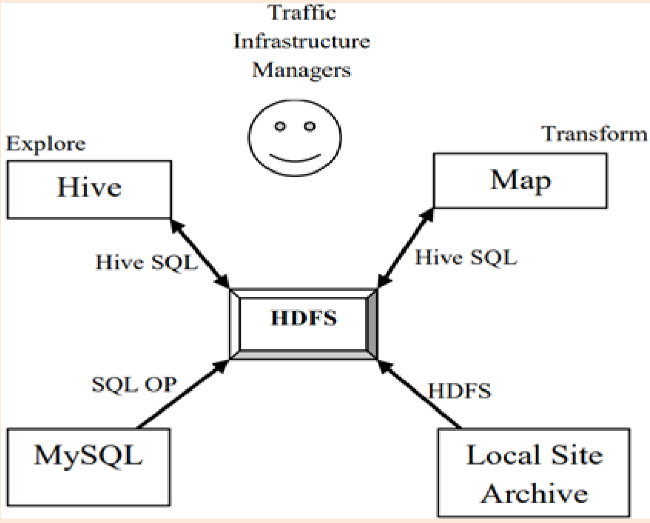

Figure 4 Implementation for delivery and access of surveillance sites

It also encourages commuters to follow traffic laws and drive safely by providing advanced information for decision-making on the travel time for path prediction, travel time forecasting, and emergency vehicle path selection.[74]

[74] Balaji R. Ganesh and S. Appavu, 'An Intelligent Video Surveillance Framework with Big Data Management for Indian Road Traffic System' (2015) 123(10) IJCA 12

Surveillance in the transport sector – Example 2

Smart Railway ensures train service reliability, availability, maintainability and safety (RAMS). This entails the monitoring of train speed and position via capabilities of GPS and RFID[75], employing sensors' intelligence[76] to monitor temperature of rolling stock axle counter[77], rail integrity and even equipment rooms.[78] In this regard, Priscoli et al. introduced the SHIELD framework for a smart railway surveillance system. Figure 5 illustrates an architecture of the SHIELD framework consisting of three layers with the technological enablers: metrics, ontologies and overlay.[79]

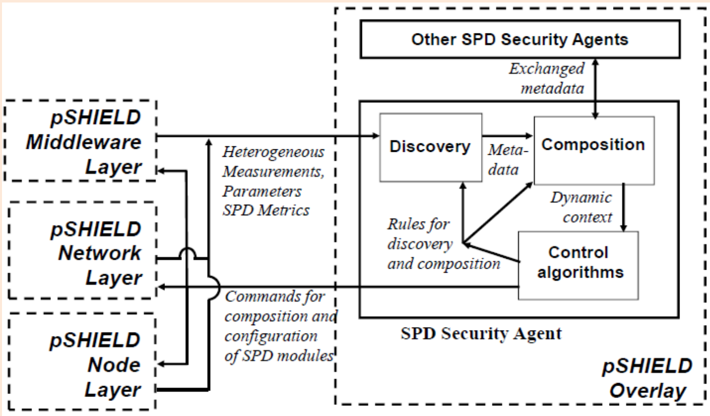

Figure 5 SHIELD functional architecture

[75] Wai Pan Tam, Shing-kai Chan and Sum Chan, 'Smart Light Rail: Integrated Speed and Position Supervision System (2017) 24(4) HKIE Transactions 237

[76] Peng Wenlong and others, 'Access Point Research in Rail Train Safety Monitoring Sensor Network' in *2012 Third International Conference on Digital Manufacturing and Automation* (IEEE, 2012) 157-160, DOI: 10.1109/ICDMA.2012.38

[77] Jianqin Qian and others, 'A Passive UHF Tag for RFID-based Train Axle Temperature Measurement System' in *2011 IEEE Custom Integrated Circuits Conference* (IEEE, 2011) 1-4, DOI: 10.1109/CICC.2011.6055420

[78] Tony Lee and May Tso, 'A Universal Sensor Data Platform Modelled for Realtime Asset Condition Surveillance and Big Data Analytics for Railway Systems: Developing a "Smart Railway" Mastermind for the Betterment of Reliability, Availability, Maintainability and Safety of Railway Systems and Passenger Service' in *2016 IEEE SENSORS* (IEEE, 2016) 1-3, DOI: 10.1109/ICSENS.2016.7808734

[79] Francesco Delli Priscoli and others, 'Ensuring Cyber-security in Smart Railway Surveillance with SHIELD' (2017) 7(2) International Journal of Critical Computer-Based Systems 138

Surveillance in the transport sector – Example 3

As described by Lederman, Taylor and Garrett in their article[80], more and more traffic control tools are becoming smart digital devices able to collect and process a high amount of (personal) data. Some of them, such as red-light cameras or speed detectors, are used to enforce the law and to legally punish those who commit violations. However, the increasing number of traffic control tools allow for a much more substantial collection of personal data (e.g. parking meters, smart parking applications, automatic tolling systems, etc.). The processing and use of such personal data, for other purposes, would allow tracking individuals. Those data combined with personal data from other sources might give a very accurate image of individuals' social habits who might not have given their consent to those processing activities nor even be aware of them (See also Section 3.3.2.1 on transparency below).

3.2.5 Summary

Considering the role of government agencies and their increasing requests of information to the private sector for public security purposes, it appears necessary to adopt specific rules to regulate the information flow, to define the rights over data and to ensure adequate enforcement. If it is true that information is often publicly available, it is also true that the line between the public and private sphere will become even more blurred in the big data era.[81]

The complexity of data processes and the power of modern analytics drastically limit the awareness of individuals, their capability to evaluate the various consequences of their choices and the expression of a real free and informed consent.[82] This lack of awareness is usually not avoided by giving adequate information to the individuals or by privacy policies, due to the fact that these notices are read only by a very limited number of users who, in many cases, are not able to understand part of the legal terms usually used in these notices, or the consequences of consenting.[83]

[80] Jaimee Lederman, Brian D. Taylor and Mark Garrett, 'A Private Matter: The Implication of Privacy Regulations for Intelligent Transportation Systems' (2016) 39(2) Transportation Planning and Technology 115

[81] Alessandro Mantelero and Giuseppe Vaciago, 'The "Dark Side" of Big Data: Private and Public Interaction in Social Surveillance' (2013) 14(6) Computer Law Review International 161

[82] Laura Brandimarte, Alessandro Acquisti and George Loewenstein, 'Misplaced Confidences: Privacy and the Control Paradox' (2013) 4(3) Social Psychological and Personality Science 340

[83] Joseph Turow and others, 'The Federal Trade Commission and Consumer Privacy in the Coming Decade' (2007) 3 ISJLP 723

<table>
<tr><th>Opportunities in relation to surveillance in the context of big data in the transport sector</th><th>Challenges in relation to surveillance in the context of big data in the transport sector</th></tr>
<tr><td rowspan="3">As the degree of sophistication increases, the centralised management of data, such as for instance traffic data, needs to be enabled, combined with an optimisation of massive data usage. If this is successful, a high degree of precious data can be gathered to support decision-making. This way, big data can facilitate smart transport. In other words, it can encourage:
• Reliability;
• Availability;
• Maintainability;
• Safety; but also
• Efficiency.</td><td>There are serious ethical issues posed by the emerging regimes of population-level monitoring, whereas recent privacy-protection initiatives fall short of addressing the challenge to democracy posed by big data surveillance.</td></tr>
<tr><td>If predictive analytics succeed in altering behaviours, surveillance cannot be understood as a passive technique but as a dynamic shaping of behaviours (impacting for instance free will).</td></tr>
<tr><td>Various commentators consider that the privacy risks related to big data analytics are low, pointing out the large amount of data processed by analytics and the de-identified nature of most of this data. This conclusion is likely to be wrong in practice[84], including from a legal perspective. This is notably due to the fact that anonymity by de-identification is a difficult goal to achieve, as demonstrated by different studies.[85]</td></tr>
</table>

Table 2 Summary table of opportunities and challenges in relation to surveillance in the context of big data in the transport sector

[84] Alessandro Mantelero and Giuseppe Vaciago, 'The "Dark Side" of Big Data: Private and Public Interaction in Social Surveillance' (2013) 14(6) Computer Law Review International 161

[85] Paul Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2009) 57 UCLA Law Review 1701; US General Accounting Office, 'Record Linkage and Privacy: Issues in Creating New Federal Research and Statistical Information (GAO, 2001) <https://www.gao.gov/assets/210/201699.pdf> accessed 23 August 2018; Hui Zang and Jean Bolot, 'Anonymization of Location Data Does not Work: A Large-scale Measurement Study' in Parameswaran Ramanathan (ed) *Proceedings of the 17th Annual International Conference on Mobile Computing and Networking* (ACM, 2011) 145-156 DOI: 10.1145/2030613.2030630; Philippe Golle, 'Revisiting the Uniqueness of Simple Demographics in the US Population' in Ari Juels (ed) *Proceedings of the 5th ACM Workshop on Privacy in Electronic Society* (ACM, 2006) DOI: 10.1145/1179601.1179615; Latanya Sweeney, 'Simple Demographics often Identify People Uniquely' (2000) Carnegie Mellon University, Data Privacy Working Paper 3 <https://dataprivacylab.org/projects/identifiability/paper1.pdf> accessed 24 August 2018; Latanya Sweeney, 'Foundations of Privacy Protection from a Computer Science Perspective' (Proceedings Joint Statistical Meeting, AAAS, Indianapolis, 2000) <https://dataprivacylab.org/projects/disclosurecontrol/paper1.pdf> accessed 23 August 2018; Omer Tene, and Jules Polonetsky, 'Big Data for All: Privacy and User Control in the Age of Analytics' (2013) 11(5) Northwestern Journal of Technology and Intellectual Property 239

3.3 Privacy

3.3.1 Introduction

Over the years, the concept of 'privacy' has covered many different dimensions, which makes grasping the concept rather difficult. The Oxford English Dictionary defines privacy as "*a state in which one is not observed or disturbed by other people*" or also "*the state of being free from public attention*".

The EU Charter of Fundamental Rights codifies the concept as a fundamental right in its Article 7, according to which: "*Everyone has the right to respect for his or her private and family life, home and communications.*"

Article 8 of the Charter provides specific fundamental rights and principles in relation to the protection of one's personal data in the following terms:

1. Everyone has the right to the protection of personal data concerning him or her.
2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.
3. Compliance with these rules shall be subject to control by an independent authority.

As indicated in the first Recital of the GDPR, the new EU privacy legal framework, it further elaborates Article 8 of the Charter. Nevertheless, Recital 4 of the GDPR clearly favours a balanced approach by stating that "*the right to the protection of personal data is not an absolute right; it must be considered in relation to its function in society and be balanced against other fundamental rights, in accordance with the principle of proportionality*".

Privacy is probably the most recurrent topic in the debate on ethical issues surrounding big data, which is not illogical given that the concepts of big data and privacy are *prima facie* mutually inconsistent.[86] Indeed, big data refers to the analysis of extremely large data sets, which may include personal data. The more personal information included in the analytics,

[86] European Data Protection Supervisor, 'Opinion 4/2015 Towards a New Digital Ethics' (EDPS 2015) <https://edps.europa.eu/sites/edp/files/publication/15-09-11_data_ethics_en.pdf> accessed 23 August 2018; European Data Protection Supervisor, 'Opinion 7/2015 Meeting the Challenges of Big Data: A Call for Transparency, User Control, Data Protection by Design and Accountability' (EDPS 2015) <https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf> accessed 23 August 2018; European Union Agency for Network and Information Security, 'Privacy by Design in Big Data – An Overview of Privacy Enhancing Technologies in the Era of Big Data Analytics' (ENISA 2015) <https://www.enisa.europa.eu/publications/big-data-protection> accessed 23 August 2018; Evodevo, 'The Ethics of Big Data: Balancing Economic Benefits and Ethical Questions of Big Data in the EU Policy Context' (European Economic and Social Committee 2017) <https://www.eesc.europa.eu/resources/docs/qe-02-17-159-en-n.pdf> accessed 23 August 2018

the more it might interfere with the privacy of the individuals concerned.[87] This Chapter will therefore examine the challenges and opportunities related to the use of big data in the transport sector.

Apart from looking into the challenges and opportunities of privacy in general, this Chapter will also dig deeper into three particular ethical aspects linked to privacy that are often discussed in relation to big data and which will also have an impact on the use of big data in the transport sector, i.e. transparency, consent and control.

3.3.1.1 Transparency

The concept of 'transparency' is defined by the Oxford English Dictionary as something "*easy to perceive or detect*".

This concept is indirectly included in Article 8 of the EU Charter of Fundamental Rights, which states that "*Everyone has the right of access to data which has been collected concerning him or her*". This entails that individuals have the right to be informed about any processing activities of their personal data, i.e. what personal data are being processed, how, for what purposes, for how long, etc.[88] In a big data context, the concept of transparency also applies in the sense of transparency of the big data analytics, i.e. the entire ecosystem of big data analytics, the algorithms used to make predictions about individuals, and the decision-making process. Individuals often have very few knowledge and details about this.[89]

Transparency is essential in order for individuals to understand the processing of their personal data and to consciously control such processing activities.

3.3.1.2 Consent

The concept of 'consent' is defined by the Oxford English Dictionary as "*the permission for something to happen or agreement to do something*".

This concept has been foreseen in the EU Charter of Fundamental Rights stating that "*Such* [personal] *data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law*."[90] This means that any processing of personal data should be based on individuals' consent or on another legitimate ground.

The GDPR defines consent as "*any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her*."[91] This clearly

---

[87] Tobias Holstein T, Dodig-Crnkovic G and Pelliccione P, 'Ethical and Social Aspects of Self-Driving Cars' (2018) abs/1802.04103 CoRR arXiv:1802.04103

[88] EDPS, Opinion 7/2015

[89] EDPS, Opinion 7/2015; Tobias Holstein T, Dodig-Crnkovic G and Pelliccione P, 'Ethical and Social Aspects of Self-Driving Cars' (2018) abs/1802.04103 CoRR arXiv:1802.04103

[90] EU Charter of Fundamental Rights, Art. 8

[91] GDPR, Art. 4(11)

highlights the importance of transparency in the process of consenting to the processing of personal data.

Collecting individuals' consent does not solve everything. It does not mean that the organisations processing the data are free to process the data as they wish. They are still accountable and have to meet the privacy standards (ethical, legal, etc.).[92] It is also worth noting that if individuals have given their consent for a particular personal data processing activity, they also have the right to withdraw their consent.

3.3.1.3 Control

The concept of 'control' is defined by the Oxford English Dictionary as "*the power to influence or direct people's behaviour or the course of events*".

With the increase of digital tools relying on personal data, the concept of control of personal data has gained importance. It is implied in Article 8 of the EU Charter of Fundamental Rights as follows: "*Such* [personal] *data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified*."

Several aspects of the GDPR, such as transparency, consent, and data subjects' rights, allow individuals to retain control over their personal data, including in a big data environment.

3.3.2 Description of the challenges in relation to privacy

As mentioned above, Article 8 of the EU Charter of Fundamental Rights has been further elaborated by the General Data Protection Regulation (the "GDPR"), which became directly applicable in all EU Member States on 25 May 2018.

This development has not only increased protection of personal data in principle, it has also changed the way in which privacy is perceived by individuals. The GDPR has empowered – and continues to empower – data subjects by more clearly stipulating their rights and the obligations that companies processing personal data are subject to, by strengthening the rules on transparency and consent, and by giving various and stronger means of enforcement to data subjects. In doing so, the GDPR (and the media attention surrounding it) has raised the public's awareness in relation to privacy and data protection.

Although this empowerment can be qualified mostly as a positive evolution, it has also had some undesirable side effects, mainly due to incorrect reports on the GDPR's exact content. For instance, some authors and popular media have placed considerable emphasis on the consent requirements, even though the GDPR[93] allows companies to base their personal data processing on other legal grounds (see also Section 3.3.2.2 below). Another example of such "misunderstanding" is that data subjects now have a better understanding of their rights but they do not seem to be aware that in order to be able to exercise those rights certain

[92] EDPS, Opinion 4/2015

[93] GDPR, Art. 6(1)

conditions need to be fulfilled. Instead, they seem to believe – and sometimes claim so when exercising one of their rights – that the data subject rights enumerated in the GDPR are absolute. Even if an exercise of rights by a data subject does not fulfil the necessary conditions, the GDPR still requires companies to inform the data subject within one month about the reasons for not addressing his/her request.[94]

This evolution is not exactly encouraging the use of big data, be it in the transport sector or not, as it may either hold people back from sharing their personal data for big data analytics purposes or, where data has been made available, subject organisations carrying out big data analytics to unfounded data subject requests. Although one of the underlying reasons to adopt the GDPR was to update the data protection rules applicable throughout the EU and to bring them in line with the technological reality[95], it appears that this intention has only been met partially.

Furthermore, even though one single set of rules and principles is now applicable throughout the EU, privacy desires may still vary between individuals or between situations.[96] Indeed, certain individuals may be willing to give up part of their privacy in return for the benefits certain big data applications may bring. Although the latter could be an opportunity, the real complexity lies in the fact that the different levels of desired privacy between people and situations would require an *ad hoc* approach, which is not often achievable in practice. A one-size-fits-all approach, on the contrary, may lead to the application of the greatest common divisor (i.e. the strongest privacy needs), which would eliminate the opportunity of performing farther-reaching big data analytics using personal data of people who are more permissive.

The following Sections will look into the challenges in relation to transparency, consent and control specifically.

3.3.2.1 Description of the challenges in relation to transparency

Today, personal data processing activities are common and numerous. People are however not always aware of the exact nature of the activities, particularly in a big data environment where many stakeholders are involved. Typically, individuals are unclear about the types of data processed, purposes of processing, data sharing with or data transfers to other actors who might also process their data, etc.[97] Furthermore, information about possible secondary

---

[94] GDPR, Art. 12(4)

[95] GDPR, Recital 6

[96] World Economic Forum in collaboration with Bain & Company, Inc., 'Personal Data – The Emergence of a New Asset Class' (World Economic Forum 2011) <http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf> accessed 23 August 2018

[97] European Union Agency for Network and Information Security, 'Privacy by Design in Big Data – An Overview of Privacy Enhancing Technologies in the Era of Big Data Analytics' (ENISA 2015) <https://www.enisa.europa.eu/publications/big-data-protection> accessed 23 August 2018

use of the personal data collected and the purposes of such activities are usually not communicated to individuals.[98]

A lack of transparency may lead to a decrease in individuals' power and free will as they do not know what is exactly happening to their data.[99] Individuals are indeed confronted to predictions and decisions they do not control and do not understand because the logic of algorithms and the way the decisions are made are not shared.[100] This might result in a decrease in the trust in big data and lead to individuals refusing to use certain technologies and/or services in order to avoid such unclear processing activities.[101]

The lack of transparency about personal data processing activities actually performed can be even more important, as it does not only concern the personal data an individual has knowingly communicated to an organisation. What individuals usually do not realise is that organisations also have access to publicly available personal data and to data from other sources. By combining all the data they have access to, including the data simply observed such as individuals' behaviour on the Internet or their locations, organisations are able to make more accurate and broader decisions/predictions that individuals do not expect.[102]

This reveals another challenge: citizens' limited knowledge about big data analytics. There is a general need to raise awareness and educate citizens about the digital world, including new technologies and the related personal data processing activities.[103]

From the perspective of organisations, transparency is also a challenge. Some of them are reluctant to be transparent, invoking business confidentiality or trade secrets protection. This should however not release them from informing individuals about the processing of their personal data. A balancing exercise is therefore needed to determine what information can be communicated. In this respect, it is worth noting that other means of protection of information exist, such as intellectual property rights.[104]

3.3.2.2 Description of the challenges in relation to consent

Some data are being processed without individuals' consent or another legal ground. This issue is addressed by the GDPR, which requires any personal data processing activities to be

---

[98] EDPS, Opinion 7/2015

[99] Evodevo, 'The Ethics of Big Data: Balancing Economic Benefits and Ethical Questions of Big Data in the EU Policy Context' (European Economic and Social Committee 2017) <https://www.eesc.europa.eu/resources/docs/qe-02-17-159-en-n.pdf> accessed 23 August 2018

[100] EDPS, Opinion 7/2015; Article 29 Data Protection Working Party, 'Opinion 3/2013 on purpose limitation' (WP29 2013) <http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf> WP203 accessed 23 August 2018

[101] Jaimee Lederman, Brian D. Taylor and Mark Garrett, 'A Private Matter: The Implication of Privacy Regulations for Intelligent Transportation Systems' (2016) 39(2) Transportation Planning and Technology 115

[102] EDPS, Opinion 7/2015

[103] Evodevo, 'The Ethics of Big Data: Balancing Economic Benefits and Ethical Questions of Big Data in the EU Policy Context' (European Economic and Social Committee 2017) <https://www.eesc.europa.eu/resources/docs/qe-02-17-159-en-n.pdf> accessed 23 August 2018

[104] EDPS, Opinion 7/2015

lawful, i.e. based on a legal ground.[105] This also means that, from a legal perspective, consent is not always needed. Individuals do not always have to consent to the processing of their personal data. Other legal grounds might be invoked: performance of a contract, compliance with a legal obligation, protection of the vital interests of individuals, performance of a task carried out in the public interest, or legitimate interests of the data controller.[106]

It however seems that individuals are not always aware of this legal reality. They tend to think that any processing is always subject to their consent and that no other legal grounds exist. Media reports indeed often focus on consent and wrongly interpret it, creating confusion. This perception also highlights the lack of awareness and transparency observed among individuals (see also Section 3.3.2.1).[107]

Furthermore, the GDPR imposes strict conditions when relying on consent.[108] For instance, consent should be given in an informed manner, which means that granular information about the processing activity should be given so that individuals consciously agree on the processing of their personal data and are aware of the details of the processing activity. However, it has often been pointed out that the notices accompanying consent forms are not sufficiently transparent and do not provide all details of the processing activity, particularly regarding the secondary use of personal data, and the transfer of personal data to third parties who will potentially reuse those data.[109] If they are well informed, individuals will be more inclined to consent to the processing of their personal data.[110]

In addition, where processing is based on consent, the data subject has the right to withdraw this consent whenever he or she pleases.[111] Such withdrawal of consent must be as easy as it was to give consent. Relying on consent in the context of big data analytics may therefore be risky, given that when an individual decides to withdraw consent the big data analytics process may be completely jeopardised.

---

[105] GDPR, Art. 5(1)(a)

[106] GDPR, Art. 6. Article 9 of the GDPR also provides legal grounds specific to the processing of special categories of personal data (i.e., data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, genetic data, biometric data, health data, data concerning a natural person's sex life or sexual orientation).

[107] Elizabeth Denham, 'Blog: Consent is not the "Silver Bullet" for GDPR Compliance' (Information Commissioner's Office 2017) <https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2017/08/blog-consent-is-not-the-silver-bullet-for-gdpr-compliance/> accessed 24 August 2018

[108] GDPR, Art. 7

[109] Evodevo, 'The Ethics of Big Data: Balancing Economic Benefits and Ethical Questions of Big Data in the EU Policy Context' (European Economic and Social Committee 2017) <https://www.eesc.europa.eu/resources/docs/qe-02-17-159-en-n.pdf> accessed 23 August 2018

[110] EDPS, Opinion 7/2015

[111] GDPR, Art. 7(3)

3.3.2.3 Description of the challenges in relation to control

In her article on the 'Philosophy of big data', Swan mentions an "*asymmetry of control with data having the upper hand*".[112] In a big data context, individuals indeed hardly control their personal data (including the storage, accuracy, sharing with other entities, etc.), and are sometimes not aware of the processing activities in which their data are involved. This is notably due to the high number of actors involved in big data analytics.[113] This lack of control may lead to decisions individuals do not understand.[114]

Being able to request deletion of one's personal data, and to exercise the right to erasure or the "right to be forgotten", is also a form of control.[115] However, even if this right exists, it does not systematically guarantee the actual deletion of the data. First of all, certain conditions need to be fulfilled in order for an erasure request to be honoured. Furthermore, some data may have been shared with, sold or transferred to other stakeholders who also process them, without individuals being aware of this. The request for erasure might not be transmitted to all the stakeholders involved in the processing activity, which makes it nearly impossible for individuals to ensure the complete erasure of their personal data.[116]

Another aspect of the concept of control refers to the control of individuals' identities. As stated in Opinion 7/2015 of the EDPS, there is a "*risk of 'dictatorship of data' where, according to one study by a European data protection authority, 'we are no longer judged on the basis of our actions, but on the basis of what all the data about us indicate our probable actions may be'*."[117] The massive processing of personal data creates a digital identity for each individual. This 'frozen' identity is sometimes taken into account to make predictions or decisions. Individuals themselves are not consulted anymore, nor taken into account in the decision-making process, which means that they might be discriminated without having the possibility to react.[118]

[112] Melanie Swan, 'Philosophy of Big Data: Expanding the Human-data Relation with Big Data Science Services' in *'15 Proceedings of the 2015 IEEE First International Conference on Big Data Computing Service and Applications* (IEEE, 2015) 468-477 DOI: 10.1109/BigDataService.2015.29

[113] European Union Agency for Network and Information Security, 'Privacy by Design in Big Data – An Overview of Privacy Enhancing Technologies in the Era of Big Data Analytics' (ENISA 2015) <https://www.enisa.europa.eu/publications/big-data-protection> accessed 23 August 2018

[114] EDPS, Opinion 7/2015; WP29, Opinion 3/2013

[115] GDPR, Art. 17

[116] Evodevo, 'The Ethics of Big Data: Balancing Economic Benefits and Ethical Questions of Big Data in the EU Policy Context' (European Economic and Social Committee 2017) <https://www.eesc.europa.eu/resources/docs/qe-02-17-159-en-n.pdf> accessed 23 August 2018

[117] EDPS, Opinion 7/2015; Norwegian Data Protection Authority, 'Big Data Report' (Datatilsynet 2013) 7 <https://www.datatilsynet.no/globalassets/global/english/big-data-engelsk-web.pdf> 23 August 2018

[118] Evodevo, 'The Ethics of Big Data: Balancing Economic Benefits and Ethical Questions of Big Data in the EU Policy Context' (European Economic and Social Committee 2017) <https://www.eesc.europa.eu/resources/docs/qe-02-17-159-en-n.pdf> accessed 23 August 2018

3.3.3 Description of the opportunities in relation to privacy

As mentioned above, the advent of the GDPR has empowered individuals with regard to the protection of their personal data. After years of wilful abuse or unintentional ignorance in respect of people's personal data, companies finally have to abide by a strict set of rules. Refraining from doing so could entail far-reaching sanctions, including administrative fines or the prohibition to process data any further (which often implies putting a stop to the provision of the service).[119] There is today more than ever an incentive for organisations to ensure an adequate protection of personal data

Rather than a compliance burden, organisations may see this development as an opportunity to guarantee high data protection standards to their customers and distinguish themselves from their competitors. Especially in a big data context, where considerable amounts of personal data may be processed, organisations should apply the GDPR's data protection principles in every aspect of the big data analytics lifecycle.[120] In particular, privacy and data protection requirements should be defined and adhered to in the development of big data analytics platforms and applications.

The following Sections go deeper into the opportunities related to transparency, consent and control specifically.

3.3.3.1 Description of the opportunities in relation to transparency

Transparency regarding personal data processing activities and big data analytics make individuals more inclined to communicate their personal data and to use digital tools. Individuals' awareness would result in a higher level of trust, which would make them more open to use other digital tools and to test new technologies. Organisations and governments would even more benefit from using big data due to increased data collection.[121]

Moreover, informing individuals about big data analytics and the algorithms used for decision-making allows them to verify the conclusions drawn, correct mistakes and avoid discrimination. This also ensures safety of the tools as the entire ecosystem and the related big data analytics might be properly verified, taking into account all the criteria.[122]

3.3.3.2 Description of the opportunities in relation to consent

On the one hand, organisations engaged in big data analytics should educate individuals about the possible lawful grounds for processing other than consent. In this respect, individuals

---

[119] GDPR, Art. 83

[120] Tobias Holstein T, Dodig-Crnkovic G and Pelliccione P, 'Ethical and Social Aspects of Self-Driving Cars' (2018) abs/1802.04103 CoRR arXiv:1802.04103

[121] Evodevo, 'The Ethics of Big Data: Balancing Economic Benefits and Ethical Questions of Big Data in the EU Policy Context' (European Economic and Social Committee 2017) <https://www.eesc.europa.eu/resources/docs/qe-02-17-159-en-n.pdf> accessed 23 August 2018; Jaimee Lederman, Brian D. Taylor and Mark Garrett, 'A Private Matter: The Implication of Privacy Regulations for Intelligent Transportation Systems' (2016) 39(2) Transportation Planning and Technology 115

[122] EDPS, Opinion 7/2015

should be reassured that the fact that they are not able to consent to the processing does not affect the exercise of their rights nor diminish the protection of their personal data.

On the other hand, giving the opportunity to individuals to consent to personal data processing activities may give them the impression that they have control over their personal data. This may increase their trust in big data technology and encourage them to use it.[123]

3.3.3.3 Description of the opportunities in relation to control

Giving more control to individuals allows them to make better choices and to benefit from the processing of their personal data in a big data context. Individuals are able to identify unfair decisions and rectify mistakes. They can also make sure that their personal data are not used for further processing activities if they do not agree with such processing.[124] Finally, giving individuals more control over their personal data may encourage them to participate in more data processing activities for big data purposes.

3.3.4 Examples of privacy in the transport sector

Privacy in the transport sector – Example 1
Self-driving cars will collect a high amount of personal data about the users but also about the environment of the car (neighbourhood, other drivers, etc.), and those data may be shared with many stakeholders. Users might be reluctant to give their consent to such massive processing of their personal data. However, without such processing activities, self-driving cars would not work properly and safely. Indeed, a high amount of partners will be involved in such ecosystem to make it function. It might be that individuals will have no other choice than to accept such processing.[125] From a legal perspective, the GDPR introduces different lawful bases for processing. Consent from individuals will therefore not always be necessary to process personal data (see Deliverable D2.2 for further information on data protection and the lawfulness of processing).

[123] Ibid

[124] Ibid

[125] Caitlin A. Surakitbanharn and others, 'Preliminary Ethical, Legal and Social Implications of Connected and Autonomous Transportation Vehicles' <https://www.purdue.edu/discoverypark/ppri/docs/Literature%20Review_CATV.pdf> accessed 23 August 2018

Privacy in the transport sector – Example 2

A smart parking application identifying available parking spaces, as described in the ENISA study[126], has several advantages: it can reduce congestion, decrease gas emissions, reduce fuel consumption, and improve traffic. To offer this type of service, the application requires the collection and processing of a high amount of personal data, allowing, if combined, to draw a detailed profile of the users. This type of service involves many stakeholders (e.g. city administration, telecommunication operators, banks, etc.). In such scenario, it is difficult to identify the responsibility of each stakeholder, especially regarding privacy issues, including transparency. ENISA suggests to designate a single point of contact responsible for the privacy issues who would be in charge of giving information to individuals about the service and the processing activities and to respond if they want to exercise their rights.

Privacy in the transport sector – Example 3

In their article, Bertot and Choi show the potential opportunities of big data, particularly for digital governments.[127] They give the example of sensors placed on smartphones allowing relevant organisations to manage public transportation capacity. The process is interesting, particularly in the context of a smart city. However, this will need to be analysed from ethical and legal perspectives to ensure the lawfulness of such processing and to make sure that the individuals agree with or, at least, are aware of such processing and all its implications, such as the sharing of their data with other stakeholders.

Privacy in the transport sector – Example 4

Civil drones collect data intentionally and unintentionally, especially pictures about individuals, which can give indications about their location, habits, physical characteristics, etc. Finn and Wright have conducted a survey about the use of civil drones and their related privacy, data protection and ethical implications.[128] In their article presenting their findings, they explain that in some instances the images captured by drones are recorded, stored and shared with other organisations. Individuals are not aware of such processing and have therefore no control over their data. Finn and Wright conclude their article by stating that drones operators and manufacturers do not have sufficient knowledge about legal and ethical

[126] European Union Agency for Network and Information Security, 'Privacy by Design in Big Data – An Overview of Privacy Enhancing Technologies in the Era of Big Data Analytics' (ENISA 2015) <https://www.enisa.europa.eu/publications/big-data-protection> accessed 23 August 2018

[127] John Carlo Bertot and Heeyoon Choi, 'Big Data and e-Government: Issues, Policies, and Recommendations' (2014) 19(12) Information Polity 5

[128] Rachel L. Finn and David Wright, 'Privacy, Data Protection and Ethics for Civil Drone Practice: A Survey of Industry, Regulators and Civil Society' (2016) 32(4) Computer Law & Security Review 577

standards in order to be able to raise and tackle those issues. Awareness and legal initiatives are necessary. They also mention other ethical and data protection issues that can be raised, such as data minimisation, consent, proportionality, and purpose limitation.

3.3.5 Summary

As mentioned above, the GDPR entered into application on 25 May 2018. In light of what has been described above, the GDPR notably provides for the following:

- a strengthened principle of transparency in relation to personal data processing, ensuring better information to individuals about the processing of their personal data[129]
- the requirement that any processing should be lawful, i.e. based on a legal ground[130]
- extended and strengthened rules on consent[131]
- new and reinforced rights for individuals aiming at giving individuals more control over their personal data, i.e. the rights of access, rectification, erasure, restriction of processing, data portability, objection and the right not to be subject to automated individual decision-making[132]

This strengthened legal framework is likely to respond to several ethical issues and thus improve end users' trust in the use of personal data by private and public organisations, notably in a big data context.

The legal aspects of the abovementioned topics will be addressed further in Deliverable D2.2 about legal issues.

Before entry into application, the EU observed a significant wave of GDPR compliance projects and programmes among the several companies and organisations operating in the EU. Many companies continue this exercise still today, e.g. by updating their policies and procedures in light of recent guidance or by carrying out internal audits to uncover remaining weaknesses. While writing the present Deliverable, it is still too premature to assess how strictly the GDPR will be applied and how the national data protection authorities will enforce it. We do know however that the EU Commission's first evaluation and review of the application and functioning of the GDPR is due by 25 May 2020.[133]

[129] GDPR, Art. 5.1(a) and 12
[130] GDPR, Art. 6
[131] GDPR, Art. 7
[132] GDPR, Chapter III, Art. 12-22
[133] GDPR, Art. 97

Opportunities in relation to privacy in the context of big data in the transport sector	Challenges in relation to privacy in the context of big data in the transport sector
If organisations involved in big data analytics are able to address the users' privacy needs, this could open the door to more people engaging in big data.	Difficulty of striking a balance between the right to privacy and the use of individuals' personal data in the context of big data applications used in the transport sector.
In certain situations, individuals may be willing to forsake part of their privacy in return for the benefits big data applications may bring.	As needs for privacy may vary between individuals or between situations (e.g. depending on the benefits the individual gets in return), it will be difficult for companies and developers to adopt a one-size-fits-all approach.
Providing transparent information to individuals whose personal data is involved in big data analytics may increase trust in the processing activities and the technology used. When people feel they can trust a technology, they tend to be more willing to engage in it.	The lack of transparency of personal data processing activities in a big data context negatively affects data subjects' trust in such activities and the related technology. Data subjects may be reluctant to use big data applications in the transport sector.
Both industry and government should take up responsibility to eliminate the misconceptions that exist regarding personal data protection. Data subjects should be educated, notably through transparent notices from industry, about the grounds for processing and the possible impacts on privacy.	The misconceptions regarding data protection concepts, such as consent, cause confusion both among data subjects and organisations. This general trend will also affect the use of big data in the transport domain.
Giving control to data subjects should not necessarily stifle the use of big data. Instead, a bigger involvement of data subjects may lead to improved analytics given that the data subjects can correct mistakes and detect unfair decisions.	There is an asymmetry of control of personal data between data subjects and the organisations processing the data. Data subjects may fear losing control over their digital identity by engaging in big data analytics.

Table 3 Summary table of opportunities and challenges in relation to privacy in the context of big data in the transport sector

3.4 *Free Will*

3.4.1 Introduction

'Free will' is defined by the Oxford English Dictionary as "*the power of acting without the constraint of necessity or fate*" or also "*the ability to act at one's own discretion*". It is an underlying principle to most, if not all, rights and freedoms enumerated in the EU Charter of Fundamental Rights. Of course, free will is no absolute given, just like the recitals of the EU Charter of Fundamental Rights state that enjoyment of the rights enshrined in the Charter "*entails responsibilities and duties with regard to other persons, to the human community and to future generations*".

Traditional deontological and utilitarian ethics place a strong emphasis on moral responsibility of the individual, often also called 'moral agency'. This idea of 'moral agency' very much stems from assumptions about individualism and free will.[134] These assumptions experience challenges in the era of big data, when it comes to the advancement of modern technology. In other words, big data as moral agency is being challenged on certain fundamental premises that most of the advancements in computer ethics took and still take for granted, namely free will and individualism.[135]

Usually, free will is considered as distinct from other concepts such as autonomy and authenticity. There are many ways of thinking about the nature of free will, and there are serious disagreements about what would constitute an adequate theory of free will. In general, free will can be defined as *a kind of power or ability to make decisions of the sort for which one can be morally responsible*.[136]

In this Chapter we discuss the positive and negative effects of big data in the transport sector with respect to free will.

3.4.2 Description of the challenges in relation to free will

With a hyper-connected era of big data, the concept of power, which is so crucial for ethics and moral responsibility, is changing into more networked fashion. Big data stakeholders such as big data collectors, big data utilisers and big data generators have *relational power* in the sense of a network.[137] In this regard, retaining the individual's agency (i.e. knowledge and ability to act) is one of the main and complex challenges for the governance of socio-technical epistemic systems[138]. However, with respect to the network nature of society, a dependent

[134] Alasdair MacIntyre, *A Short History of Ethics: A History of Moral Philosophy from the Homeric Age to the 20th Century* (Routledge 2003)

[135] Andrej Zwitter, 'Big Data Ethics' (2014) 1(2) Big Data & Society 1

[136] Merel Noorman, 'Computing and Moral Responsibility' (The Stanford Encyclopedia of Philosophy 2012) <https://stanford.library.sydney.edu.au/entries/computing-responsibility/> accessed 23 August 2018

[137] Robert A. Hanneman and Mark Riddle, *Introduction to Social Network Methods* (University of California 2005)

[138] Judith Simon, 'Distributed Epistemic Responsibility in a Hyperconnected Era' in Luciano Floridi (ed) *The Onlife Manifesto* (Springer, Cham, 2015)

agency might be a factor when judging the moral responsibility of the agent. On the contrary, with traditional ethics, hyper-networked ethics can be induced by big data and exacerbates the effect of network knock-on effects (i.e. effects on third mostly unrelated parties). This changes foundational assumptions about ethical responsibility by changing what power is and the extent of free will by reducing known outcomes of actions, while increasing unintended consequences.[139] Big data is the effect of individual actions, sensor data, and other real-world measurements creating a digital image of our reality, so-called "datafication."[140] The absence of knowledge about what data are collected or what they are used for might put the 'data generators' (e.g. online consumers and people owning handheld devices) at an ethical disadvantage in terms of free will.

Many researchers[141] believe that big data causes a loss of free will and autonomy of humans by applying deterministic knowledge to human behaviour. For instance, Mayer-Schoenberger and Cukier (2013) point out that big data can help to undermine the idea of personal responsibility, particularly the idea of free will.[142]

3.4.3 Description of the opportunities in relation to free will

With respect to supporting free will of humans, increasing accessibility and personalisation for passengers can provide benefit to people from more personalised or affordable services as organisations use some data like journey data through better understanding and serving their needs.[143] In contrast to opportunities of big data, even collection of anonymised data about individuals can lead to illegal behaviours in terms of free will of humans.[144] Indeed, aggregated and anonymised data can also be used to target individuals established on predictive models.[145]

On the other hand, a huge part of what we know about the world, especially about social and political phenomena comes from data analysis. This kind of insight can be extended into new domains by big data, which achieves greater accuracy in pinpointing individual behaviour, and the capability of generating this knowledge can be undertaken by new actors and more powerful tools.[146] Although a growing body of information being generated from big data

[139] Andrej Zwitter, 'Big Data Ethics' (2014) 1(2) Big Data & Society 1

[140] Kenneth Cukier, 'Big Data is Better Data' (TED 2014) <https://www.youtube.com/watch?v=8pHzROP1D-w> accessed 23 August 2018

[141] Chris Snijders, Uwe Matzat and Ulf-Dietrich Reips, '"Big Data": Big Gaps of Knowledge in the Field of Internet Science' (2012) 7(1) International Journal of Internet Science 1

[142] Kenneth Cukier and Viktor Mayer-Schoenberger, 'Rise of Big Data: How it's Changing the Way We Think about the World' in Mircea Pitici (ed) *The Best Writing on Mathematics 2014* (Princeton University Press 2015)

[143] Libby Young and others, 'Personal Data in Transport: Exploring a Framework for the Future' (Open Data Institute 2018) <https://theodi.org/wp-content/uploads/2018/06/OPEN-Personal-data-in-transport-.pdf> accessed 23 August 2018

[144] Sciencewise Expert Resource Centre 'Big Data: Public Views on the Collection, Sharing and Use of Personal Data by Government and Companies' (Sciencewise 2014)

[145] Charles Duhigg, 'How Companies Learn your Secrets' *The New York Times* (New York, 19 February 2012) 30

[146] Larisa Giber and Nikolai Kazantsev, 'The Ethics of Big Data: Analytical Survey' (2015) 2(3) Cloud of science 400

provides a level that is imperceptible to individuals[147], various fields of IT technologies such as information retrieval[148], user modelling and recommender system[149] have been studied to provide proper options for people.

With the changing role of data in transport, from data-poor to data-rich, big data in the field of transport is now accessible in new ways and at new scales. Companies are collecting higher volumes of this data, more frequently, and in real time – as technology makes this more feasible and viable – and using it to innovate. Customers expect more personalisation and communication as well as more real time data being shared. The transport data can be used and shared to benefit businesses, people and public services, potentially in ways that meet the needs of all three groups.

3.4.4 Example of free will in the transport sector

Free will in the transport sector - Example
Self-driving cars, by definition, monitor vehicles completely autonomously. This raises the ethical question of decision-making, especially in case of unavoidable impact.[150] The "Trolley Problem"[151] is often mentioned in this context: if a group of people is in the middle of the road and the self-driving car cannot stop because it is too fast, the car would have to choose between (i) driving into the group of people; (ii) driving into the pedestrian crossing the other lane (the pedestrian being for example an old lady or a young child); or (iii) ploughing into a wall and injuring or killing the driver and/or the passengers.[152] Such decisions can only be made by humans, and the decision-making results will differ between different persons. Creators will need to define algorithms to deal with these kinds of situations.[153] To address such types of moral dilemmas, the MIT Media Lab has developed a "Moral Machine"[154] to gather citizens' opinions about particular scenarios and share the results with car

[147] Ralph Schroeder and Josh Cowls, 'Big Data, Ethics, and the Social Implications of Knowledge Production' (Data Ethics Workshop, KDD@Bloomberg, New York, 2014) <https://pdfs.semanticscholar.org/5010/f3927ca8133a432ac1d12a8e57ac11cb3688.pdf> accessed 23 August 2018

[148] Beth Plale, 'Big Data Opportunities and Challenges for IR, Text Mining and NLP' in Xiaozhong Liu and others (eds) *Proceedings of the 2013 International Workshop on Mining Unstructured Big Data Using Natural Language Processing* (ACM, 2013) DOI: 10.1145/2513549.2514739

[149] Fatima EL Jamiy and others, 'The Potential and Challenges of Big Data - Recommendation Systems Next Level Application' (2015) arXiv:1501.03424v1 accessed 23 August 2018

[150] EDPS, Opinion 4/2015

[151] Molly Crocket, 'The Trolley Problem: Would you Kill one Person to Save many Others?' *The Guardian* (12 December 2016) <https://www.theguardian.com/science/head-quarters/2016/dec/12/the-trolley-problem-would-you-kill-one-person-to-save-many-others> accessed 23 August 2018

[152] Tobias Holstein T, Dodig-Crnkovic G and Pelliccione P, 'Ethical and Social Aspects of Self-Driving Cars' (2018) abs/1802.04103 CoRR arXiv:1802.04103

[153] Caitlin A. Surakitbanharn and others, 'Preliminary Ethical, Legal and Social Implications of Connected and Autonomous Transportation Vehicles' <https://www.purdue.edu/discoverypark/ppri/docs/Literature%20Review_CATV.pdf> accessed 23 August 2018

[154] http://moralmachine.mit.edu/

manufacturers and engineers who develop algorithms. This poll has reached millions of people who took part in the experiment. The research team is currently analysing the results. The first results show that the answers can vary a lot, and it seems that cultural values might play a role.[155]

There is however a debate on the usefulness of the Trolley Problem analysis.[156] For Holstein, the technological aspects mitigate the ethical questions and should be the main subject for discussion. Engineering will ultimately make the best decision possible using the information collected to avoid social consequences. Some technical factors should be considered in the discussion, such as the quality of the materials/technologies/sensors used to build the cars, and the sensors collecting data in the environment surrounding the cars (other cars, road infrastructures, etc.).[157]

Holstein mentions other ethical questions that can be raised: the liability in case of accidents, the security and safety of the vehicles, the car behaviour with non-self-driving cars, and the moral of the car (moral of the creator, or own moral?).[158] According to Holstein, transparency is of paramount importance to be able to deal with all the questions and challenges surrounding self-driving cars. It is the first answer, as all the actors involved can legitimately be informed, especially when it comes to the algorithms governing the cars.

[155] Oliver Smith, 'A Huge Global Study on Driverless Car Ethics Found the Elderly are Expendable' *Forbes* (21 March 2018) <https://www.forbes.com/sites/oliversmith/2018/03/21/the-results-of-the-biggest-global-study-on-driverless-car-ethics-are-in/#7b297b254a9f> accessed 23 August 2018

[156] Edmond Awad, 'Moral Machine – Perception of Moral Judgment Made by Machines' (Master's Thesis, Massachusetts Institute of Technology 2017)

[157] Tobias Holstein T, Dodig-Crnkovic G and Pelliccione P, 'Ethical and Social Aspects of Self-Driving Cars' (2018) abs/1802.04103 CoRR arXiv:1802.04103

[158] Ibid

3.4.5 Summary

Although big data-driven profiling practices can limit free will, a huge part of what we know about the world comes from data analysis, and this kind of insight can further lead into new domains by big data analysis. Careful and appropriate information analysis can open up plenty of chances and might reduce the limitations and problems for free will.

Opportunities in relation to free will in the context of big data in the Transport sector	Challenges in relation to free will in the context of big data in the Transport sector
Social and political phenomena can be extended into new domains by big data, it achieves greater accuracy in pinpointing individual behaviour to use as supporter for free will.	The rapidly increasing size and scope of information of big data technologies could lead to unfair use, in terms of the free will and autonomy of humans.
Increasing accessibility and personalisation for people may provide proper choices through a better understanding and serving people's needs.	

Table 4 Summary table of opportunities and challenges in relation to free will in the context of big data in the transport sector

3.5 Personal Data Ownership

3.5.1 Introduction

The concept of 'ownership' is defined by the Oxford English Dictionary as "*the act, state, or right of possessing something*", whereas the concept of 'property' is similarly defined as "*the right to the possession, use, or disposal of something*". Article 17 of the EU Charter of Fundamental Rights recognises the right to property or ownership in the following terms:

> "*Everyone has the right to own, use, dispose of and bequeath his or her lawfully acquired possessions. No one may be deprived of his or her possessions, except in the public interest and in the cases and under the conditions provided for by law, subject to fair compensation being paid in good time for their loss.*"

For many years already, the issue of ownership of data (whether it is personal or non-personal) has been heavily debated throughout the EU and in other parts of the world. While it could be labelled as a legal issue, given that ownership or property is traditionally a legal concept going back as far as the legal system of ancient Rome, the personal aspect of data ownership has an ethical connotation that is worth being looked into for the purpose of this Deliverable.

'Personal data' is defined by Article 4(1) of the GDPR as "*any information relating to an identified or identifiable natural person ('data subject')*", whereas an 'identifiable natural person' is defined as "*one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person*."

Individuals seem to have a general feeling that they own their personal data given that the data is about them or relates to them.[159] Moreover, where the personal data is particularly sensitive in nature, individuals even more vehemently tend to claim it as their own.

The recent developments at EU level, with the GDPR becoming applicable in May 2018, have increased the control (see also Section 3.3.1.3 above) individuals have over the collection, processing and sharing of their personal data.[160] This evolution seems to create a certain impression of personal data ownership.

For instance, some scholars highlight the fact that the GDPR "*recognises different levels of control rights to consumers in accordance with a 'proprietarian' approach to personal*

---

[159] World Economic Forum in collaboration with Bain & Company, Inc., 'Personal Data – The Emergence of a New Asset Class' (World Economic Forum 2011) 16 <http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf> accessed 23 August 2018

[160] Alex Howard, *Data for the Public Good* (O'Reilly Media Inc. 2012) 23

data."[161] More specifically, some have emphasised that in practice personal data is treated as property of the individual.[162]

We have also noted a similar phenomenon in some EU Member States, even before entry into application of the GDPR, taking a clearer stance on the issue of personal data ownership.

Germany

In Germany, the debates surrounding personal data also relate to the issues of ownership of that data. Some German scholars argue indeed that personal data should belong to the individual (i.e., the "data subject").[163] However, most commentators agree that data subjects have no general rights in their data.[164]

France

On 9 October 2016, the Digital Republic Act entered into force. It aims to incorporate support for innovation and new business models, more open data, better personal protection, heightened platform fairness and extended rollout of digital access.[165] The text notably imposed requirements on businesses, in anticipation of the GDPR. In such context, it is particularly interesting to note that the Digital Republic Act incorporates a principle into the French Data Protection Act (Article 1 of Act No. 78-17 of 6 January 1978), according to which "*every person shall have the right to decide and control the uses made of his or her personal data*." Although its terms have been criticised, this provision anchors the protection of personal data into the sphere of fundamental rights. Some commentators argue that in practice, such provision could be used as an additional legal ground to invalidate terms and conditions vesting in service providers a property right over individuals' personal data.[166]

[161] Gianclaudio Malgieri, 'Property and (Intellectual) Ownership of Consumers' Information: A New Taxonomy for Personal Data' (2016) 4 PinG 133; Jacob M. Victor, 'The EU General Data Protection Regulation: Toward a Property Regime for Protecting Data Privacy' (2013) 123(2) Yale Law Journal 266

[162] Nadezhda Purtova, 'The Illusion of Personal Data as No One's Property' (2015) 7(1) Law, Innovation and Technology 83

[163] Wolfgang Kilian, 'Informationelle Selbstbestimmung und Marktprozesse' (CR 2002) 921 (926) cited in Martin Sebastian Haase, *Datenschutzrechtliche Fragen des Personenbezugs* (Mohr Siebeck 2015) 109; Benedikt Buchner, *Informationelle Selbstbestimmung im Privatrecht* (Mohr Siebeck 2006) 203 ff, 223 ff

[164] E.g., Michael Dorner, *Big Data und "Dateneigentum"* (2014) 9 CR 617, 619 ff

[165] République Française, 'Explanatory Memorandum' (*République numérique*) <https://www.republique-numerique.fr/pages/digital-republic-bill-rationale> accessed 23 August 2018

[166] Willy Mikalef, 'New Data Protection Provisions in France: Are You Ready?' (2016) 16(45) WDPR

3.5.2 Description of the challenges in relation to personal data ownership

The rights and obligations under data protection law emanate from the fundamental right to privacy. Accordingly, such rights relate to a personality right granted to individuals. The fact that the GDPR, and some EU Member States' laws, grant important rights to data subjects does not as such regulate the question of data ownership and therefore does not recognise a "property" right of individuals in their data. In our view, the GDPR only regulates the relationship between the data subject and the data controller(s)/processor(s), without creating and regulating the issues of commercially exploitable rights in personal data.[167]

This view is supported by the manner in which the right to property is recognised in the EU Charter of Fundamental Rights; i.e. *the right to own […] his or her lawfully acquired possessions*. Personal data is not a possession that can be acquired by the data subject, be it lawfully or not. It is information that attaches to an individual because of his/her persona. Consequently, personal data protection is not conditional upon an act of acquisition on behalf of the data subject. To claim otherwise would go against the data protection principles of the GDPR and the rights to respect for private and family life and to protection of personal data enshrined in the EU Charter of Fundamental Rights.

Whereas personal data is something inherent to and indivisible from the individual, it may be lawfully – i.e. in compliance with the data protection rules – acquired by third parties, either directly from the data subject or through other sources. Such interpretation would fit within the definition of the right to property under the EU Charter of Fundamental Rights. This being said, any such "ownership" right subsisting in personal data to the benefit of third-party natural or legal persons, would be restricted by the application of the GDPR and notably by the rights of data subjects.[168]

In a big data ecosystem, this tension between data subjects wanting to "own" their personal data and third parties (be it private sector companies or public organisations) claiming ownership over entire datasets could stifle innovation. Indeed, as long as data subjects do not volunteer their personal data, they retain some type of *de facto* ownership or at least control (see Section 3.3.2.3 above). Therefore, data subjects may refrain from providing their personal data as soon as they realise this would entail forsaking "ownership" or control over such data.

In addition, even if data subjects willingly provide their personal data, big data analytics imply the use of datasets containing data from various types and originating from various sources. It consequently proves highly difficult, if not impossible, to establish ownership over the separate (personal) data components in those datasets.

[167] European Data Protection Supervisor,' Ethics Advisory Group Report 2018' (EDPS 2018) 25 <https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf> 23 August 2018

[168] See in the same vein, in the context of disclosure of chemical data, the Court Order of the EU General Court in case T-189/14 wherein the President examines in obiter dictum the question of privacy (Case T-189/14 R *Deza a.s v Agence européenne des produits chimiques* [2014] ECLI:EU:T:2014:686). More particularly, the President acknowledges the relevance of the question of privacy of legal entities but nevertheless reminds, on the basis of the decision of the Court of Justice of the EU in case C-450/06, that it may be necessary to prohibit the disclosure of information qualified as confidential in order to preserve the fundamental right to privacy of an undertaking (Case C-450/06 *Varec SA v État belge* [2008] ECLI:EU:C:2008:91).

Furthermore, taking into account the various actors involved in the big data ecosystem, many different entities may try to claim ownership in (parts of) the dataset, including in the personal data components.

An additional complicating factor is that the definition of personal data is constantly evolving.[169] Certain types of information (e.g. IP addresses) that would as such not have been qualified as personal data under the previous Data Protection Directive, are now recognised to be personal data under the GDPR. This is not only due to the fact that the legal definition of personal data has been broadened, but also because of continuous technological developments facilitating the identification or linking back to an individual.

3.5.3 Description of the opportunities in relation to personal data ownership

As mentioned above, the concept of "personal data ownership" is a fiction that cannot have legal consequences. The use of big data – be it in the transport sector or not – is unlikely to provide any opportunities in this respect, even from an ethical perspective. We believe however that this should not mean end users should bid farewell to the "sensation" of ownership. If we look at the reasons why data subjects would want to claim ownership over their personal data, it mainly boils down to the wish to maintain "control" over said personal data. Therefore, the opportunities identified in relation to control (see Section 3.3.3.3 above) may prove useful in this context.

The opportunities in relation to data ownership in general will be discussed in Deliverable D2.2 "Report on legal issues."

3.5.4 Examples of personal data ownership in the transport sector

Personal data ownership in the transport sector – Example 1
The developments in relation to connected and autonomous vehicles have also raised questions with respect to personal data ownership.[170] The on-board computing systems present in connected and autonomous vehicles will allow for the transfer of substantial amounts of information, including about the driver and its location. At the current stage, it is still unclear who will "own" this information among the many different actors involved; i.e. the driver who the personal data relates to, the owner of the vehicle (if different from the driver), the manufacturer of the vehicle, the national department of transportation, or any other third party. Any data ownership claim may have a far-reaching impact on the further implementation of the technology concerned. In any event, the personal data protection rules will need to be respected.

[169] Václav Janecek, 'Ownership of Personal Data in the Internet of Things' (2018) forthcoming in the Computer Law & Security Review

[170] Caitlin A. Surakitbanharn and others, 'Preliminary Ethical, Legal and Social Implications of Connected and Autonomous Transportation Vehicles' <https://www.purdue.edu/discoverypark/ppri/docs/Literature%20Review_CATV.pdf> accessed 23 August 2018

Personal data ownership in the transport sector – Example 2
An accelerated development and implementation of intelligent transportation systems (ITS) can be observed in the transport sector.[171] By relying on data regarding the movement of vehicles, ITS can help improve road transport aspects such as infrastructure, traffic circulation, traffic offense enforcement, and road safety. An example of ITS technology currently already widely deployed is red-light cameras. Red-light cameras collect information on a vehicle's location and its owner by linking the licence plate to the vehicle registration data. Generally, the cameras also take a picture of the vehicle's driver (who may be different from the owner). The use of red-light cameras allows the collecting and processing personal data on a large scale and in an automated manner. The data subjects (i.e. the owner of the vehicle and, when applicable, the driver) may not expect this type of processing activity (e.g. because there is a lack of transparency) and may therefore lose "ownership" or control over the data concerned.

3.5.5 Summary

As discussed in this Section, a claim of ownership by a data subject in its personal data would be hard to sustain. This however does not mean that data subjects have to give up all control over their personal data. The advent of the GDPR, with its novel and/or strengthened data subject rights, has increased the means of data subjects to exercise control over the processing of their personal data.

Nevertheless, in a big data context, different third-party entities may try to claim ownership in (parts of) a dataset, including in the personal data components thereof, which may hinder the production of, access to, linking and re-use of big data, including in the transport sector. This challenge will be further elaborated in Deliverable D2.2 "Report on legal issues."

[171] Jaimee Lederman, Brian D. Taylor and Mark Garrett, 'A Private Matter: The Implication of Privacy Regulations for Intelligent Transportation Systems' (2016) 39(2) Transportation Planning and Technology 115

Opportunities in relation to personal data ownership in the context of big data in the Transport sector	Challenges in relation to personal data ownership in the context of big data in the Transport sector
Given the nature of the issue, no opportunity in relation to personal data ownership has been identified. We believe however that the opportunities related to control, discussed above, may (at least partly) address the data subjects' concerns with respect to the "ownership" of their personal data.	As the definition of 'personal data' continues to evolve, information that is qualified as non-personal data today may be classified as personal data in the future.
	End users may be reticent to provide their personal data for big data analytics in transport as this would entail forsaking "ownership."
	Difficulty to establish ownership of different data components within a set of data of various types and coming from various sources.
	Multiple actors involved in big data analytics may try to claim ownership of the data concerned, which may lead to a gridlock.

Table 5 Summary table of opportunities and challenges in relation to personal data ownership in the context of big data in the transport sector

3.6 Data-driven Social Discrimination

3.6.1 Introduction

According to the Oxford English Dictionary, the term 'discrimination' is defined as "*treating a person or particular group of people differently, especially in a worse way from the way in which you treat other people, because of their skin colour, sex, sexuality, etc*." or in more general terms: "*the unjust or prejudicial treatment of different categories of people, especially on the grounds of race, age, or sex*."

The principles of non-discrimination and equality are to a great extent covered in Title III of the EU Charter of Fundamental Rights. Thus, the EU Charter recognises the following fundamental rights, freedoms and principles in relation to discrimination: (i) equality before the law; (ii) non-discrimination; (iii) cultural, religious and linguistic diversity; (iv) equality between women and men; (v) the rights of the child; (vi) the rights of the elderly; and (vii) the integration of persons with disabilities.[172]

Elements discriminatory treatment can be based on are, as mentioned above, skin colour, race, sex, but also for example income or education level, gender, residential area, and others. Using big data analytics to improve business processes or to provide personalised services may lead to discrimination of certain groups of people. At any step of the big data analytics pipeline[173], unintended data biases may be created due to wrong statistical treatment or poor data quality. Big data poses certain challenges requiring expert knowledge to estimate the accuracy of conclusions drawn from it.[174]

A great interest exists in personalised services, individually targeting advertisements, and customised services and product offers. Personalising services means nothing else than to exclude people from or include them into certain target groups on the basis of personal data such as gender, income, education, consumption preferences, etc. Big data analytics relies on the categorisation of information and conclusions drawn from it. In that sense, the definition of discrimination in contrast to personalisation does not seem to be straightforward and discrimination might therefore be an inherent part of the analytics process.[175]

Another important aspect of data-driven discrimination concerns the access to and knowledge of technology needed to use digital services or gather valuable information from online

---

[172] EU Charter of Fundamental Rights, Art. 20-26

[173] Kim Hee and others, 'Big Data Methodologies, Tools and Infrastructures' (LeMO 2018) <https://static1.squarespace.com/static/59f9cdc2692ebebde4c43010/t/5b6d4674032be489a442fa8b/1533888127770/20180716_D1.3_Big+data+methodologies%2C+tools+and+infrastructures_LeMO.pdf> accessed 24 August 2018

[174] D. R. Cox, Christiana Kartsonaki and Ruth H. Keogh, 'Big Data: Some Statistical Issues' (2018) 136 Statistics & Probability Letters 111

[175] Rena Coen and others, 'A User Centered Perspective on Algorithmic Personalization' (UC Berkeley 2016)

platforms or applications. The social differences in the corresponding access to technology and education or skills to use it, is often referred to as the “Digital Divide”.[176]

3.6.2 Description of the challenges in relation to discrimination

The challenges related to data-driven social discrimination and equity discussed in the framework of this Deliverable are (i) unintended data bias; (ii) intended data bias, i.e. personalised services, offers and advertisements; and (iii) the “Digital Divide”.

3.6.2.1 Unintended data bias

Biases in datasets or in statements or predictions based on the analysis of datasets can originate from various errors, shortcomings or misinterpretations along the analytics pipeline. The data collection process might be biased by design because of a biased formulation of a survey, a biased selection of data sources, an insufficient length of the surveyed time-period, or the neglect of relevant parameters or circumstances. Along the analytics process, correct statistical treatment and accuracy estimation require expert knowledge. The procedure is therefore prone to methodical and technical errors. Some of the main reasons for unintended data biases are mentioned hereafter.

The sample size of a dataset directly influences the validity of a statement or the conclusions drawn from the data sample. The accuracy of a statistical analysis depends on the nature of the sample and its estimation. In the context of big data, the available data often consist of many subsets, demanding careful statistical treatment to normalise the estimating procedure to the single subsets in order to avoid overfitting or wrong conclusions. This heterogeneity of the data calls for clean and careful aggregation of data from different data sources corresponding to different subsets where some unique features are not shared by all sets.

Dealing with a huge amount of data generated from a large amount of individuals or sensors makes analysis prone to errors resulting from bad data quality. Data derived from device measurements or automated studies must be carefully checked for various types of errors that may arise during the collection process. Since the cleaning and checking procedures are usually automated processes themselves, even more attention is required. In some sectors, well-established quality control and assurance procedures exist and should be standardised in order to ensure reliable conclusions and predictions.[177]

Due to the usually high dimensionality, the analysis of big data requires the estimation of simultaneously different variables. Every estimation relates to a corresponding error leading to accumulated errors if a conclusion or algorithm-based prediction is based on many variables. This effect is referred to as noise accumulation and can make it difficult to refer to

[176] Massimo Ragnedda and Glenn Muschert, *The Digital Divide: The Internet and Social Inequality in International Perspective* (Routeledge 2013)

[177] Jules J. Berman, ‘15 - Big Data Failures and How to Avoid (Some of) Them’ in Jules J. Berman (ed) *Principles and Practice of Big Data* (Second edition, Academic Press 2018); D. R. Cox, Christiana Kartsonaki and Ruth H. Keogh, ‘Big Data: Some Statistical Issues’ (2018) 136 Statistics & Probability Letters 111; Pierre-André G. Maugis, ‘Big Data Uncertainties’ (2018) 57 Journal of Forensic and Legal Medicine 7

the original signal. Statistical techniques dealing with this issue require special expertise. Parameter selection and reduction of dimensionality is also crucial to overcome noise accumulation in classification and prediction analytics. This can be challenging due to other big data challenges like spurious correlation, incidental endogeneity, heterogeneity and data quality.

Spurious correlation and incidental endogeneity are two other effects that may lead to wrong conclusions and predictions. Variables or instances might “spuriously” correlate if the correlation is caused by an unseen third variable or event and not by the original variables. High dimensionality makes this effect more likely to occur. It may also be that variables are actually correlated but without any meaning or cause. Incidental endogeneity occurs as a result of selection biases, measurement errors and omitted variables. These phenomena arise frequently in the analysis of big data. The possibility of collecting many different parameters with available measurement techniques increases the risk to create incidental correlation. Big data aggregated from multiple sources with potentially different data generation procedures increases the risk of selection biases and measurement errors causing potential incidental endogeneity.[178]

Learning algorithms are often highly complex. This complexity combined with a lack of transparency or comprehensibility for a broader community (see Section 3.3.2.1 on transparency) increases the probability of uncovered errors. Often algorithms are black boxes within a company with limited reproducibility. Open communication, in particular about accuracy levels, uncertainties within the algorithms, or implicit assumptions may often be insufficient.

The causes for data bias discussed above are all relevant in the transport sector. They may differ in importance in a specific domain, e.g. for freight and passenger transport. In route optimisation using big data, a huge amount of various sensor data of freight transport-related items might be aggregated with data from other sources (e.g. weather data), which calls for an accurate data merging and cleaning process to ensure good data quality.

3.6.2.2 Intended data bias

Increasing knowledge on customer or user behaviour and access to personal data creates, besides new business opportunities and the possibility of growth, also strong power. Power is referred to in the sense that personal data of individuals or groups such as their gender, race, income, residential area and even patterns of their behaviour (e.g. movement profiles) can be aggregated to detailed profiles. Power inherited from such profiles may be unintentionally or intentionally used to discriminate people. The distinction between value-added personalisation and segmentation and discrimination is not well defined therefore depending largely on the experience and perception of the affected individuals.

[178] Ian L. Dryden and David J. Hodge, ‘Journeys in Big Data Statistics’ (2018) 136 Statistics & Probability Letters 121; David D. Dunson, ‘Statistics in the Big Data Era: Failures of the Machine’ (2018) 136 Statistics and Probability Letters 4

Some personalised services or advertisements might be discriminatory because it excludes certain groups or it is only offered to the people who communicated their personal data. This also includes the selective visibility of a service due to personalised online search results: different groups are not provided with the same information or are offered the same product or service with different pricing or availability options.[179]

Personalisation may also lead to discriminatory treatment if it is based on statistical analysis assuming wrong segmentation criteria that are not really representing the target groups or are addressing it in a prejudicial way. Since the underlying algorithms are typically not accessible to the target groups themselves, their ability to object is limited and it may lead to the manifestation of existing prejudice. In other words, data-based predictions or conclusions are more likely perceived to be objectively true since they rely on “objective” data. This might lead to even worse discrimination of a social group since prejudicial data can serve as evidence for the confirmation of prejudice.

By way of example, personalised job offers may limit the possibility of individuals to explore new opportunities if the algorithms based on educational backgrounds, professional experience and other underlying factors do not make them aware of possibilities not fitting their profiles.

Lange, Coen and Berkeley confirmed in their study[180] that users negatively perceive personalisation based on race or household income level. Their study surveyed the opinion of 748 participants. Information on the income level, residence area and gender were considered as very private information, and negative responses to the use of it for individualised services were recorded. The use of race as a parameter for personalisation was also seen as unfair across all researched domains, which were chosen to be targeted advertising, filtered search results, and differential pricing.

One might consider that these forms of information offering and service platforms are often operated by corporations. Accordingly, the online communication environment is to a large degree dictated by commercial actors who aim to maximise profits. Discrimination might emerge from the fact that people with e.g. lower income or other traits that do not correlate to the business models of those corporations are of less interest.[181]

Conferred to the domain of passenger transport, this could mean that a segregation of services based on specific characteristics of individuals, such as income or residential area and implicitly race or gender, might take place. The possibility to create new mobility offers according to individualised needs, e.g. private shuttle services combining different modes and optimising routes, might lead to a graduated system of offers dedicated to different social groups with low permeability.

[179] Michael Schrage, ‘Big Data’s Dangerous New Era of Discrimination’ *Harvard Business Review* (2014) <https://hbr.org/2014/01/big-datas-dangerous-new-era-of-discrimination> accessed 24 August 2018; Rena Coen and others, ‘A User Centered Perspective on Algorithmic Personalization’ (UC Berkeley 2016)

[180] Rena Coen and others, ‘A User Centered Perspective on Algorithmic Personalization’ (UC Berkeley 2016)

[181] Wendy Arianne Günther and others, ‘Debating Big Data: A Literature Review on Realizing Value from Big Data’ (2017) 26(3) Journal of Strategic Information Systems 191

3.6.2.3 Digital Divide

Discrimination based on social factors and the "Digital Divide" are interconnected: different levels of access and skill in technology use influence and are influenced by individuals' social position, which includes characteristics like age, gender, race, income and education level amongst others.

Increasing Internet diffusion may decrease social inequality in certain fields if people of lower social status profit relatively more from the use of the Internet than the ones with higher social status. Accordingly, the opposite situation would manifest and increase social inequity. The term "Digital Divide" was first referred to as the diffusion of Internet access throughout population but is nowadays extended to a "second-level Digital Divide", which includes the different degrees of skill, time, knowledge and usage possibilities. It turns out that social status directly influences the online usage behaviour, as higher education for example correlates with a higher online user experience in the fields of information retrieval and transactional purposes. Certain user groups are more likely to become more disconnected from the benefits of Internet usage, which might lead to reinforcement of existing social inequities.[182]

In countries with high diffusion rates of Internet access (see comparison for Europe[183]), the ability and skill to use online services or platforms becomes a substantial part of social life and individuals depend on it in various fields of their professional and private life.[184]

In the transportation sector, this is for example the case in route planning. Route planning is increasingly managed by applications or navigation programmes ensuring, among others, the online availability of public transport schedules, the purchase of tickets, and access to real-time information about the route. This however requires a certain level of skills, access to technology in the form of appropriate devices and some financial contributions.

3.6.3 Description of the opportunities in relation to discrimination

Personalisation and segmentation for customised services or targeting may resolve biases. Big data analytics might indeed also be utilised to decrease social inequity and to improve existing discriminatory situations or services. Discriminatory situations can be made visible using big data analysis, which is the basis to resolve biases. Personalised or individualised services could in a second stage offer the possibility to people with special needs, who are not fitting the

---

[182] Massimo Ragnedda and Glenn Muschert, *The Digital Divide: The Internet and Social Inequality in International Perspective* (Routeledge 2013); Monica Răileanu Szeles, 'New Insights from a Multilevel Approach to the Regional Digital Divide in the European Union' (2018) 42(6) Telecommunications Policy 452

[183] Eurostat, 'Internet Access of Households, 2017' (Eurostat) <http://ec.europa.eu/eurostat/statistics-explained/index.php/Digital_economy_and_society_statistics_-_households_and_individuals> accessed 24 August 2018

[184] Petter Bae Brandtzæg, Jan Heim and Amela Karahasanović, 'Understanding the New Digital Divide—A Typology of Internet Users in Europe' (2011) 69(3) International Journal of Human-Computer Studies 123; Petya Chipeva and others, 'Digital Divide at Individual Level: Evidence for Eastern and Western European Countries' (2018) 35(3) Government Information Quarterly 460

majority, to improve their inclusion into society. This could be seen as “positive discrimination.”[185]

Several ongoing projects aim to improve existing discrimination situations in the transport sector.

Mobility services to rural and periphery areas are a big challenge. This coincides with the rapidly changing age structure in these areas, where people are getting much older on average. The MobiDig project in the region of Northern Bavaria in Germany aims to tackle these issues by improving mobility services in rural areas in order to increase social inclusion. The project led by five partner institutions (including the Technical University of Munich and the Fraunhofer Group for Supply Chain Services) intends to evaluate and promote new mobility concepts in order to provide efficient and sufficient transport services.[186]

Gender equality in the transport sector seems to be another issue. The systematic analysis of the situation based on big data allows identifying discriminatory practices and the reasons therefor. This is what several EU projects are aiming to do. They seek to make recommendations in order to improve the situation, such as implementing, as a starting point, the information about the gender of workers in the transport sector in existing databases.[187]

[185] ‘Positive discrimination’ is defined by the Oxford English Dictionary as "*the practice or policy of favouring individuals belonging to groups known to have been discriminated against previously*".

[186] Bundesministerium für Verkehr und digitale Infrastruktur, ‘Mobilität digital Hochfranken – MobiDig’ (BMVI) <https://www.bmvi.de/SharedDocs/DE/Artikel/DG/mfund-projekte/mobilitaet-digital-hochfranken-mobidig.html?nn=326002> accessed 24 August 2018

[187] WISE ‘Project Wise - Project Report: Women Employment in Urban Public Transprot Sector’ (WISE) <http://www.wise-project.net/downl/final_wise_project_report.pdf> accessed 23 August 2018; Peter Turnbull, Julia Lear and Huw Thomas, ‘Women in the Transport Sector - Promoting Employment by Preventing Violence against Women Transport Workers’ (International Labour Organization 2013) <http://www.ilo.org/wcmsp5/groups/public/---ed_dialogue/---sector/documents/briefingnote/wcms_234882.pdf> accessed 24 August 2018; Anne Loehr, ‘Big Data for HR: Can Predictive Analytics Help Decrease Discrimination in the Workplace’ (The Blog Huffpost 2015) <https://www.huffingtonpost.com/anne-loehr/big-data-for-hr-can-predi_b_6905754.html> accessed 24 August 2018

3.6.4 Examples of social discrimination in the transport sector

Social discrimination in the transport sector – Example 1
Uber, the ride-sharing company, has allowed making discrimination visible thanks to its online platform technologies. Several forms of discrimination have been observed in the Uber environment. Uber rating system used by passengers to give feedback about drivers at the end of a ride has allowed highlighting discrimination against drivers from racial minority groups. This is problematic as the data collected via the tool are used to evaluate drivers, and eventually dismiss them if their ratings do not meet Uber's expected standards. Another form of discrimination concerns passengers. It has been observed that drivers are sometimes less keen to offer their services to black riders or riders willing to go to poor neighbourhoods. Besides highlighting those discriminatory situations, the Uber platform could also be used to deter or prevent discrimination by for example configuring the level of passengers' information available to the drivers in order to decrease discrimination against them.[188]

3.6.5 Summary

Using big data analytics to improve business processes or provide personalised services may lead to discrimination of certain groups of people. Big data analytics relies on the categorisation of information and conclusions drawn from it. In that sense, discrimination might therefore be an inherent part of the analytics process. As another important aspect of data-driven discrimination, the social differences in access to technology and education or skills to use it, are often referred to as the "Digital Divide". In this regard, this Section discussed three specific challenges related to data-driven discrimination: (i) unintended data bias; (ii) intended data bias; and (iii) the "Digital Divide".

Big data analytics can be a tool to make existing discriminatory decisions visible, hence this social issue may be resolved by personalised services (as "positive discrimination") based on big data analytics. In spite of this opportunity, there are still biases because of big data's characteristics (e.g., heterogeneity, data size and quality, noise, etc.). Furthermore, also personalised services may cause discriminatory treatment by excluding certain groups. Finally, big data creates new visibilities and makes it possible to discern between people on a whole range of behaviour-related and other personal aspects. This also provides fertile ground for 'new discriminations'.

These issues are of course highly relevant for the use of big data in the transport sector, for instance, for the planning of different routes on the basis of quality data or technologies used.

---

[188] Alex Rosenblat and others, 'Discriminating Tastes: Uber's Customer Ratings as Vehicles for Workplace Discrimination' (2017) 9(3) Policy and Internet 256; Brishen Rogers, 'The Social Costs of Uber' (2015) (University of Chicago Law Review Dialogue, Forthcoming, Temple University Legal Studies Research Paper No. 2015-28) DOI: 10.2139/ssrn.2608017; Yanbo Ge and others, 'Racial and Gender Discrimination in Transportation Network Companies' (2016) NBER Working Paper No. w22776 <http://www.nber.org/papers/w22776.pdf> accessed 24 August 2018

Therefore, it is essential and important to reduce the likelihood of discrimination in the processing of big data and its analytics.

Opportunities in relation to social discrimination in the context of big data in the transport sector	Challenges in relation to social discrimination in the context of big data in the transport sector
Big data analytics can be a tool to make existing discriminatory decisions visible.	Big data analytics are vulnerable to technical and systematic biases which can lead to discriminatory conclusions. These biases may be caused by heterogeneity of data, the size of the data sets, data quality, noise accumulation, spurious correlation, incidental endogeneity, and algorithms complexity.
Big data analytics can be used to tailor customised services that meet the needs of certain social groups in order to improve their inclusion into the society.	Personalised services can exclude certain social groups or lead to discriminatory treatment.
	Big data analytics requires digital accessible services, which makes it difficult for people with limited access to or knowledge of technology ("Digital Divide").

Table 6 Summary table of opportunities and challenges in relation to social discrimination in the context of big data in the transport sector

3.7 Environmental

Big data is regarded as today's "digital oil" because of the valuable information hiding inside.[189] Many question importance and relevance of big data to society, but when evaluated and analysed, we can see how it has made advancements in technology and in social developments, affecting each of us on a daily basis.[190] Suitable data processing and management are appropriate to expose new knowledge and facilitate response to emerging opportunities and challenges in a timely manner.[191]

While one of the biggest transportation planning challenges of the twentieth century was to provide efficient systems to accommodate the automobile revolution, the twenty-first century has seen a shift towards a focus on sustainability.[192] One possible definition was proposed at the 2004 European Conference of Ministers of Transport, stating, "*a sustainable transport system is one that is accessible, safe, environmentally-friendly, and affordable*;" another by Transport Canada states "*the goal of sustainable transportation is to ensure that environmental, social, and economic considerations are factored into decisions affecting transportation activity*." While many transport organisations would like to achieve environmental sustainability, they have the lack of knowledge to design appropriate eco-tactics that can accomplish environmental sustainability.[193]

In this regard, there are trade-off or rebound effects limiting the effect of big data exploitation or creating unintended consequences. Therefore, trade-off or rebound effects from the use of big data in transport should be assessed. Within the LeMO project, a separate Task 2.4 is dedicated to the environmental aspects of big data in the transport sector in terms of the trade-off and rebound effect. The outcome of this Task will be provided in Deliverable D2.4.

---

[189] Xiaomeng Yi and others, 'Building a Network Highway for Big Data: Architecture and Challenges' (2014) 28(4) IEEE Network 5

[190] Alonzo L. Foster Jr and Dewayne R. Brown, 'Big Data and its Impact on Society and Industry and the Growing Need for Big Data' (2015) 12(2) International Advanced Research Journal in Science, Engineering and Technology 6

[191] Uthayasankar Sivarajah and others, 'Critical Analysis of Big Data Challenges and Analytical Methods' (2017) 70 Journal of Business Research 263

[192] Caitlin D. Cottrill and Sybil Derrible, 'Leveraging Big Data for the Development of Transport Sustainability Indicators' (2015) 22(1) Journal of Urban Technology 45

[193] Pei-Ju Wu and Yi-Ching Chen, 'Big Data Analytics for Transport Systems to Achieve Environmental Sustainability' in Teen-Hang Meen, Stephen D. Prior and Artde Donald Kin-Tak Lam (eds) *2017 International Conference on Applied System Innovation* (IEEE, 2017) 264-267, DOI: 10.1109/ICASI.2017.7988401

4 Conclusion: Moving Forward on Ethical and Social Issues in relation to the Use of Big Data in the Transport Sector

This Deliverable has looked into the challenges and opportunities in relation to six pre-defined ethical and social issues that may arise in the context of big data usage generally and in the transport sector specifically. A full list of the identified challenges and opportunities is available in Appendix A.

As one can observe from these lists, for most ethical and social issues big data may constitute both a challenge and an opportunity, depending on the exact use made thereof. This observation applies to the use of big data generally, but also to the use of big data in the transport sector specifically. Whether big data is used in the transport sector or within another sector does not seem to have a particular impact on how the ethical and social issues play out.

Nevertheless, it is apparent from Appendix A that at this stage the list of challenges is still longer than the list of opportunities. Therefore, in order to enable and/or facilitate the production of, access to, linking and re-use of big data, including in the transport sector, certain remedial actions should be taken.

In the context of ethical and social issues, we may broadly distinguish two types of remedial actions: (i) regulatory intervention, i.e. by means of legislation, standards or soft law; and (ii) design, i.e. by ensuring that systems or applications are designed in such a way that the decisions they take are ethical.[194]

4.1 Regulatory intervention

Although governments in most countries intervened in some kind of way in relation to (big) data and the transportation sector, regulating ethical and social issues – be it through "hard" legislation or soft law – proves to be particularly challenging.

Any regulatory intervention requires compromises and striking a balance between the different societal objectives it may pursue, such as economic efficiency, social justice, or environmental sustainability.[195] Indeed, "*regardless of the philosophical standings used to judge the desirability of public interventions, one should consider that policy makers are not a neutral apparatus applying policies to achieve the maximum social good.*"[196]

In those events where ethical or social justice is retained as one of the main driving factors of regulation, ethical principles "*are not necessarily universal, as each society has different*

[194] Virginia Dignum and others, 'Ethics by Design: necessity or curse?' in *AAAI/ACM Conference on Artificial Intelligence, Ethics and Society* (AAI/ACM, 2018) <http://www.aies-conference.com/wp-content/papers/main/AIES_2018_paper_68.pdf> accessed 24 August 2018

[195] Paulo Rui Anciaes and Nikolas Thomopoulos, 'Ethical Issues in Transportation' in Mark Garrett (ed.), *Encyclopedia of Transportation: Social Science and Policy* (SAGE Publications: Thousand Oaks, California, USA 2014)

[196] Ibid

concerns at each moment in time" and even within the same society competing ethical principles may exist as a consequence of the differences between various users and non-users of big data analytics.[197]

Taking into account this inherent contradiction between regulation, which requires hard and fast choices to be made, and ethics, which is variable between and within societies and may evolve over time; regulation of ethical and social issues in itself will not provide an adequate solution to encourage the production of, access to, linking and re-use of big data, including in the transport sector.

4.2 Ethics-by-design

According to Dignum et al., ethics-by-design concerns "*the methods, algorithms and tools needed to endow autonomous agents with the capability to reason about the ethical aspects of their decisions, and the methods, tools and formalisms to guarantee that an agent's behavior remains within given moral bounds.*"[198]

The idea to achieve a certain higher purpose through technology design is not novel.

In cybersecurity, for instance, the concept of 'security-by-design' is well established and functional. Also privacy or data protection-by-design is a widely discussed concept, even more since its explicit codification as a legal principle in the GDPR.[199] Its first roots in the EU legal framework could however already be observed in Recital 46 of the old Data Protection Directive[200], according to which "*protection of the rights and freedoms of data subjects with regard to the processing of personal data requires that appropriate technical and organizational measures be taken, both at the time of the design of the processing system and at the time of the processing itself*." Neither the Directive nor the GDPR provide much guidance on how to comply with the data protection-by-design principle, but the GDPR mentions the example of pseudonymisation.[201]

Therefore, in theory, ethics-by-design seems an ideal approach to ensure that systems or applications are inherently ethical. However, several issues may arise in this respect.

From the perspective of the software developer or engineer, we may envisage *inter alia* the following concerns:

---

[197] Ibid

[198] Virginia Dignum and others, 'Ethics by Design: necessity or curse?' in *AAAI/ACM Conference on Artificial Intelligence, Ethics and Society* (AAI/ACM, 2018) <http://www.aies-conference.com/wp-content/papers/main/AIES_2018_paper_68.pdf> accessed 24 August 2018

[199] GDPR, Art. 25

[200] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data [1995] OJ L 281/0031-0050

[201] Pseudonymisation is defined in the GDPR as the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person (GDPR, Art. 4(5)); GDPR, Art. 25(1)

1. An ethics-by-design approach requires ethical and societal values to be taken into account in the development of big data applications. This requires a mental shift from developers and programmers, in the sense that they can no longer solely focus on performance of the application but also have to consider the ethical consequences of actions and/or decisions taken by the application.[202]
2. As discussed in Section 3.6, it should be ensured that the input data is collected, created and used in a fair and non-discriminatory manner.[203]
3. Provided that the two previous points are fulfilled, would it be safe to assume that the big data system or application is able to take ethical decisions in an autonomous way or should there be a "human in the loop", i.e. a human supervisor, at all times?[204]
4. Should a system that is claimed to be ethical-by-design be accompanied by a software debugger, i.e. a tool that would allow software developers to explore the code and understand the decision being made and the learning behind it, which would allow for algorithmic transparency (which would also benefit the users – cf. infra)?[205]

From the user perspective, we may envisage the following expectations of a system that is claimed to be ethical-by-design:

1. In light of the importance of transparency discussed in Section 3.3.1.1 above, the big data system or application should ideally be able to explain and justify to the user why it performed a certain action and if and how it used the user's personal data.[206]
2. Contrary to humans, machines are generally perceived to be incapable of ethical reasoning. Users may therefore wish to claim some type of seal of approval or certification guaranteeing that a system is able to make ethical decisions.[207]
3. Similarly to a data subject's right not to be subject to a decision solely based on automated processing and therefore to obtain human intervention from the

[202] Virginia Dignum and others, 'Ethics by Design: necessity or curse?' in *AAAI/ACM Conference on Artificial Intelligence, Ethics and Society* (AAI/ACM, 2018) <http://www.aies-conference.com/wp-content/papers/main/AIES_2018_paper_68.pdf> accessed 24 August 2018

[203] Ibid

[204] Wenchao Li and others, 'Synthesis for Human-in-the-Loop Control Systems' in Erika Ábrahám and Klaus Havelund (eds) *Tools and Algorithms for the Construction and Analysis of Systems* (TACAS 2014, Lecture Notes in Computer Science, vol 8413, Springer, Berlin, Heidelberg, 2014)

[205] Interview with Pedro Domingos, Professor of computer science, University of Washington (18 June 2018) http://www.odbms.org/blog/2018/06/on-artificial-intelligence-machine-learning-and-deep-learning-interview-with-pedro-domingos/

[206] The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, 'Ethically Aligned Design: A Vision For Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems, Version 1' (IEEE, 2016) 20 <http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html> accessed 24 August 2018; Virginia Dignum and others, 'Ethics by Design: necessity or curse?' in *AAAI/ACM Conference on Artificial Intelligence, Ethics and Society* (AAI/ACM, 2018) <http://www.aies-conference.com/wp-content/papers/main/AIES_2018_paper_68.pdf> accessed 24 August 2018

[207] Virginia Dignum and others, 'Ethics by Design: necessity or curse?' in *AAAI/ACM Conference on Artificial Intelligence, Ethics and Society* (AAI/ACM, 2018) <http://www.aies-conference.com/wp-content/papers/main/AIES_2018_paper_68.pdf> accessed 24 August 2018

controller, as enshrined in the GDPR[208], a user may want to have the possibility to request human intervention in the big data analytics process.

4. In the event a system or application does take an unethical decision, there should be means to trace back the system's decision to the different actors involved (including the user itself) in order to share liability between those actors so that the user receives adequate compensation.[209]

Thus, looking at the achievements of design concepts such as security-by-design and privacy-by-design, ethics-by-design seems a promising solution to tackle the ethical and social challenges that today hinder the production of, access to, linking and re-use of big data, including in the transport sector. However, taking into account the concerns that may arise, it cannot be a standalone solution. Regulatory intervention to a certain extent and to a certain degree will most likely be necessary to define the core principles to be respected by all actors involved in the big data lifecycle.

4.3 Ethics-by-design supplemented by (self-)regulation

In light of the foregoing, we would advocate a combined approach whereby, on the one hand, ethics-by-design, like privacy-by-design, is recognised as a principle of EU law and, on the other hand, this principle is further elaborated by (self-)regulation.

Considering the challenges raised with regard to regulatory intervention in Section 4.1, and notably the fact that ethical issues are variable between and within societies and evolutionary, self-regulation may be more appropriate to address ethical and social issues as it provides more flexibility and can be more easily adapted to advances in technology and unforeseen circumstances. Such self-regulation could be achieved through the creation of ethical codes of conduct. Therefore, stakeholders in the big data lifecycle should be encouraged to create and implement ethical codes of conduct. Against the background of its efforts to establish a Digital Single Market and EU data economy, the EU seems best placed to create this incentive at EU level. Further EU guidance on a certain minimum content can be envisaged.

Also, in order to avoid a too great disparity between different organisations' ethical codes of conduct, which would defeat the purpose of creating an ethics-by-design legal principle at EU level, agreement should be reached on the general implementation principles of ethics-by-design and further guidance on how to interpret these should be provided.

In this respect, we believe it is the EU's responsibility to oversee the specification of the ethical framework that should be complied with in the design of the systems concerned.[210] To this end, working groups should be established at EU level to define the ethics-by-design core implementation principles. These working groups may also be helpful in providing further

---

[208] GDPR, Art. 22

[209] Virginia Dignum and others, 'Ethics by Design: necessity or curse?' in *AAAI/ACM Conference on Artificial Intelligence, Ethics and Society* (AAI/ACM, 2018) <http://www.aies-conference.com/wp-content/papers/main/AIES_2018_paper_68.pdf> accessed 24 August 2018

[210] Ibid

guidance on the content of ethical codes of conduct. In order to ensure that the working group findings are universally acceptable throughout the EU, participation of all different actors involved in the big data analytics lifecycle should be secured, including that of end users either directly or through a representative organisation.

4.3.1 Ethics-by-design core implementation principles

By way of inspiration for the ethics-by-design core implementation principles, we list below some of the most interesting approaches uncovered during our research:

- **Addressing asymmetries in information collection**: Stakeholders involved in the big data value cycle should be required to engage in full disclosure regarding the amount and type of data collected about users. All users should be able to receive a copy of all data stored about them. Public authorities and law enforcement should be held to the same standards, with exceptions or delays to protect the public interest or ongoing investigations.[211]

- **Limits on repurposing of data**: One of the practices to which people object most strenuously is the re-use of their personal data for a range of purposes other than those for which it was originally collected. Restrictions on repurposing would likely entail the deletion of certain types of transactional data at regular intervals. All forms of data not needed for archival and record-keeping purposes might be provided with a built-in expiration date.[212]

- **Ability to opt out of tracking**: All commercial interactive services and platforms should provide a no-track option – including an option that covers third-party ad servers and data collectors.[213]

- **Accountability**: When decisions are made about access to goods and services based on data mining and predictive analytics, consumers should have the right to have those decisions explained and the relevant data verified.[214]

- **Privacy-enhancing technologies (PETs)**: Innovative solutions should be put in place to tackle privacy issues in big data analytics. Such solutions should be easy to understand, and simple to the users.[215] Anonymisation and pseudonymisation techniques are examples of PET that are already commonly deployed.

[211] Mark Andrejevic, 'Surveillance in the Big Data Era' in Kenneth D. Pimple (ed) *Emerging Pervasive Information and Communication Technologies (PICT)* (Law, Governance and Technology Series, vol 11, Springer, Dordrecht, 2014)

[212] Ibid

[213] Ibid

[214] Ibid

[215] European Union Agency for Network and Information Security, 'Privacy by Design in Big Data – An Overview of Privacy Enhancing Technologies in the Era of Big Data Analytics' (ENISA 2015) <https://www.enisa.europa.eu/publications/big-data-protection> accessed 23 August 2018

- **Standardised transparency**: The use of standardised icons and pictograms should be considered, as this would allow users' instant understanding of any privacy policy with no further reading needed. A possible disadvantage of this solution however relates to the potential unwillingness of organisations to use negative icons identifying intrusive practices.[216]

- **Sticky policies**: "Sticky policies" allow organisations processing personal data as well as individuals to specify their privacy policies and their requirements in terms of privacy. The process of sticky policies can also be described as the sticking of "*machine-readable policies […] to data to define allowed usage and obligations as it travels across multiple parties*."[217] This would be a manner for organisations and individuals to align and agree on the way personal data are processed, giving individuals more control over their personal data. Encryption techniques might be useful in this field to improve the tool and should be further explored. Although this solution cannot replace consent, it can help increase control of individuals over their personal data.[218]

- **Tracking consent**: Where consent is the adequate personal data processing ground, the use of automated mechanisms to track consent should be considered in order to ensure that processing in a big data context is lawful at any time. Also, automated means for data subjects to collect and withdraw consent should be explored.[219]

- **Personal data management systems for users ("personal data stores")**: The use of systems through which individuals can control their personal data and digital identity, so-called personal data stores, should be explored. The European Commission has named such systems "Personal Information Management Systems" (PIMS) and defines them as "*a control function (e.g. console or dashboard provided via a web page or app) allowing the individual to define access to and usage of the data at a highly granular level in terms of which data (or sub-sets of data) can be accessed, by whom and for what purpose and foresee also the option to withdraw consent at any moment (concept of 'dynamic consent').*"[220] In the context of PIMS, third parties willing to use the data can access and process it in line with the preferences determined by the individual through secure exchange protocols. In addition, rather than accessing the data, third parties could apply their computing algorithm to the data on the platform without actually seeing the data or perform analytics in an "analytics-as-a-service"

[216] Ibid

[217] Privacypatterns.org, 'Sticky Policies' (Privacypatterns.org) <https://privacypatterns.org/patterns/sticky-policy> accessed 23 August 2018

[218] European Union Agency for Network and Information Security, 'Privacy by Design in Big Data – An Overview of Privacy Enhancing Technologies in the Era of Big Data Analytics' (ENISA 2015) <https://www.enisa.europa.eu/publications/big-data-protection> accessed 23 August 2018

[219] Ibid

[220] European Commission Directorate-General for Communications Networks, Content and Technology, Media and Data, Unit G.1. – Data Policy and Innovation, 'An Emerging Offer of "Personal Information Management Services". Current state of service offers and challenges' (European Commission 2016) <https://ec.europa.eu/digital-single-market/en/news/emerging-offer-personal-information-management-services-current-state-service-offers-and> accessed 23 August 2018

format.[221] Personal data stores would allow giving back control to data subjects over their data and allow them to retain "ownership" also after the data has been released (although there would be no legal qualification of property or ownership).[222]

- **Awareness and education**: Investing in education and raising awareness among citizens are paths that should be considered by governments but also industry to promote big data analytics and clarify personal data processing activities, including in the transport sector.

- **Anti-discrimination-by-design**: Effective methods to deal with bias have recently been developed and are a topic of further research. These methods can be integrated in auditing tools and in an anti-discrimination-by-design approach, similar as in privacy-by-design. Systems like DCUBE[223] can be used to detect discriminatory classification rules from historical data in order to intervene in these practices. Further pre- and post-processing techniques can be used to eliminate or recompense for discrimination within training datasets, including massaging the data, reweighting particular variables, resampling or applying model correction methods. Each of these methods can combat discrimination as well as ensure responsible and ethical data practices moving forward within the big data landscape.

- **Cooperation between data protection authorities and other dedicated bodies**: Data protection authorities should work together with other bodies dedicated to certain ethical and/or social issues (e.g. equality or anti-discrimination bodies) in order to share best practices and mainstream big data policies. Such cooperation could, for instance, play a role in addressing anti-discrimination in big data; first, by promoting the development of an anti-discrimination-by-design approach, and second, by developing a transparency and accountability framework based on both the anti-discrimination principles and the data protection framework.

In any event, in order for the implementation principles to be meaningful, they should be defined in such a manner that they are able to keep pace with developing technological capacities.[224]

---

[221] Ibid

[222] EDPS, Opinion 9/2016

[223] Salvatore Ruggieri, Dino Pedreschi and Franco Turini, 'DCUBE: Discrimination Discovery in Databases' in Ahmed Elmagarmid (ed) *Proceedings of the 2010 International Conference on Management of Data* (ACM Press, 2010) 1127 DOI: 10.1145/1807167.1807298; Dino Pedreschi, Salvatore Ruggieri and Franco Turini, 'The Discovery of Discrimination' in Bart Custers and others (eds) *Discrimination and Privacy in the Information Society* (Studies in Applied Philosophy, Epistemology and Rational Ethics, vol 3, Springer Berlin Heidelberg 2013)

[224] Virginia Dignum and others, 'Ethics by Design: necessity or curse?' in *AAAI/ACM Conference on Artificial Intelligence, Ethics and Society* (AAI/ACM, 2018) <http://www.aies-conference.com/wp-content/papers/main/AIES_2018_paper_68.pdf> accessed 24 August 2018

References

Books:

Baudrillard J, *Simulacra and simulation* (University of Michigan Press 1994)

Bogard W, *The Simulation of Surveillance: Hyper-control in Telematic Societies* (Cambridge University Press 1996)

Buchner B, *Informationelle Selbstbestimmung im Privatrecht* (Mohr Siebeck 2006)

Hanneman R A and Riddle M, *Introduction to Social Network Methods* (University of California 2005)

Howard A, *Data for the Public Good* (O'Reilly Media Inc. 2012)

MacIntyre A, *A Short History of Ethics: A History of Moral Philosophy from the Homeric Age to the 20th Century* (Routledge 2003)

Ohlhorst F J, *Big Data Analytics: Turning Big Data into Big Money* (John Wiley & Sons 2012)

Vallor S, *Technology and the virtues: A philosophical guide to a future worth wanting* (Oxford University Press 2016)

Journal articles:

Bertot J C and Choi H, 'Big Data and e-Government: Issues, Policies, and Recommendations' (2014) 19(12) Information Polity 5

Brandimarte L, Acquisti A and Loewenstein G, 'Misplaced Confidences: Privacy and the Control Paradox' (2013) 4(3) Social Psychological and Personality Science 340

Brandtzæg P B, Heim J and Karahasanović A, 'Understanding the New Digital Divide—A Typology of Internet Users in Europe' (2011) 69(3) International Journal of Human-Computer Studies 123

Chipeva P and others, 'Digital Divide at Individual Level: Evidence for Eastern and Western European Countries' (2018) 35(3) Government Information Quarterly 460

Cottrill C D and Derrible S, 'Leveraging Big Data for the Development of Transport Sustainability Indicators' (2015) 22(1) Journal of Urban Technology 45

Cox D R, Kartsonaki C and Keogh R H, 'Big Data: Some Statistical Issues' (2018) 136 Statistics & Probability Letters 111

Cruz-Jesus F and others, 'The Education-related Digital Divide: An Analysis for the EU-28' (2016) 56 Computers in Human Behavior 72

Dorner M, 'Big Data und "Dateneigentum"' (2014) 9 CR 617

Dryden I L and Hodge D J, 'Journeys in Big Data Statistics' (2018) 136 Statistics & Probability Letters 121

Dunson D D, 'Statistics in the Big Data Era: Failures of the Machine' (2018) 136 Statistics and Probability Letters 4

EL Jamiy F and others, 'The Potential and Challenges of Big Data - Recommendation Systems Next Level Application' (2015) abs/1501.03424 CoRR arXiv:1501.03424v1

Ferdowsi A, Challita U and Saad W, 'Deep Learning for Reliable Mobile Edge Analytics in Intelligent Transportation Systems' (2017) abs/1712.04135 CoRR arXiv:1712.04135

Finn R L and Wright D, 'Privacy, Data Protection and Ethics for Civil Drone Practice: A Survey of Industry, Regulators and Civil Society' (2016) 32(4) Computer Law & Security Review 577

Foster A L Jr and Brown D R, 'Big Data and its Impact on Society and Industry and the Growing Need for Big Data' (2015) 12(2) International Advanced Research Journal in Science, Engineering and Technology 6

Ganesh B R and Appavu S, 'An Intelligent Video Surveillance Framework with Big Data Management for Indian Road Traffic System' (2015) 123(10) IJCA 12

Giber L and Kazantsev N, 'The Ethics of Big Data: Analytical Survey' (2015) 2(3) Cloud of science 400

Günther W A and others, 'Debating Big Data: A Literature Review on Realizing Value from Big Data' (2017) 26(3) Journal of Strategic Information Systems 191

Holstein T, Dodig-Crnkovic G and Pelliccione P, 'Ethical and Social Aspects of Self-Driving Cars' (2018) abs/1802.04103 CoRR arXiv:1802.04103

Hosmer L T, 'The Call for Transportation Ethics' (1996) 50(1) Transportation Quarterly 22

Janecek V, 'Ownership of Personal Data in the Internet of Things' (2018) forthcoming in the Computer Law & Security Review

Kilian W, 'Informationelle Selbstbestimmung und Marktprozesse' (CR 2002) 921 (926) cited in Martin Sebastian Haase, *Datenschutzrechtliche Fragen des Personenbezugs* (Mohr Siebeck 2015) 109

Kumfer W and Burgess R, 'Investigation into the Role of Rational Ethics in Crashes of Automated Vehicles' (2015) 2489 Transportation Research Record 130

Lederman J, Taylor B D and Garrett M, 'A Private Matter: The Implication of Privacy Regulations for Intelligent Transportation Systems' (2016) 39(2) Transportation Planning and Technology 115

Malgieri G and Custers B, 'Pricing Privacy – The Right to Know the Value of Your Personal Data' (2017) CLSR 289

Malgieri G, 'Property and (Intellectual) Ownership of Consumers' Information: A New Taxonomy for Personal Data' (2016) 4 PinG 133

Mallikarjuna C, Phanindra A, and Rao K R, 'Traffic Data Collection under Mixed Traffic Conditions Using Video Image Processing' (2009) 135(4) Journal of Transportation Engineering 174

Mantelero A and Vaciago G, 'The "Dark Side" of Big Data: Private and Public Interaction in Social Surveillance' (2013) 14(6) Computer Law Review International 161

Maugis P-A G, 'Big Data Uncertainties' (2018) 57 Journal of Forensic and Legal Medicine 7

Mikalef W, 'New Data Protection Provisions in France: Are You Ready?' (2016) 16(45) WDPR

Mittal A, 'Trustworthiness of Big Data' (2013) 80(9) International Journal of Computer Applications 35

Murphy P R, Smith J E and Daley J M, 'Executive Attitudes, Organizational Size and Ethical Issues: Perspectives on a Service Industry' (1992) 11(1) Journal of Business Ethics 11

Ohm P, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2009) 57 UCLA Law Review 1701

Pagallo U, 'The Legal Challenges of Big Data: Putting Secondary Rules First in the Field of EU Data Protection' (2017) 3(1) EDPL 36

Priscoli F D and others, 'Ensuring Cyber-security in Smart Railway Surveillance with SHIELD' (2017) 7(2) International Journal of Critical Computer-Based Systems 138

Purtova N, 'The Illusion of Personal Data as No One's Property' (2015) 7(1) Law, Innovation and Technology 83

Rogers B, 'The Social Costs of Uber' (2015) 82(1) University of Chicago Law Review Online 85

Rosenblat A and others, 'Discriminating Tastes: Uber's Customer Ratings as Vehicles for Workplace Discrimination' (2017) 9(3) Policy and Internet 256

Sivarajah U and others, 'Critical Analysis of Big Data Challenges and Analytical Methods' (2017) 70 Journal of Business Research 263

Snijders C, Matzat U and Reips U-D, '"Big Data": Big Gaps of Knowledge in the Field of Internet Science' (2012) 7(1) International Journal of Internet Science 1

Szeles M R, 'New Insights from a Multilevel Approach to the Regional Digital Divide in the European Union' (2018) 42(6) Telecommunications Policy 452

Tam W P, Chan S and Chan S, 'Smart Light Rail: Integrated Speed and Position Supervision System' (2017) 24(4) HKIE Transactions 237

Tene O, and Jules Polonetsky, 'Big Data for All: Privacy and User Control in the Age of Analytics' (2013) 11(5) Northwestern Journal of Technology and Intellectual Property 239

Tien J M, 'Big Data: Unleashing Information' (2013) 22(2) Journal of Systems Science and Systems Engineering 127

Turow J and others, 'The Federal Trade Commission and Consumer Privacy in the Coming Decade' (2007) 3 ISJLP 723

Victor J M, 'The EU General Data Protection Regulation: Toward a Property Regime for Protecting Data Privacy' (2013) 123(2) Yale Law Journal 266

Wielki J, 'The Social and Ethical Challenges Connected with the Big Data Phenomenon' (2015) 11 Polish Journal of Management Studies 192

Wigan M R, and Roger Clarke, ‘Big Data’s Big Unintended Consequences’ (2013) 46(6) Computer 46

Wu X and others, ‘Data Mining with Big Data’ (2014) 26(1) IEEE Transactions on Knowledge and Data Engineering 97

Yi X and others, ‘Building a Network Highway for Big Data: Architecture and Challenges’ (2014) 28(4) IEEE Network 5

Zhang J and others, ‘Data-driven Intelligent Transportation Systems: A Survey’ (2011) 12(4) IEEE Transactions on Intelligent Transportation Systems 1624

Zwitter A, ‘Big Data Ethics’ (2014) 1(2) Big Data & Society 1

Working papers:

Ge Y and others, ‘Racial and Gender Discrimination in Transportation Network Companies’ (2016) NBER Working Paper No. w22776 <http://www.nber.org/papers/w22776.pdf> accessed 24 August 2018

Sweeney L, ‘Simple Demographics often Identify People Uniquely’ (2000) Carnegie Mellon University, Data Privacy Working Paper 3 <https://dataprivacylab.org/projects/identifiability/paper1.pdf> accessed 24 August 2018

Publications in press:

Anderson C, ‘The End of Theory: The Data Deluge Makes the Scientific Method Obsolete’ *Wired Magazine* (16.07, 2008) <https://www.wired.com/2008/06/pb-theory/> accessed 23 August 2018

Duhigg C, ‘How Companies Learn your Secrets’ *The New York Times* (New York, 19 February 2012) 30

Crocket M, ‘The Trolley Problem: Would you Kill one Person to Save many Others?’ *The Guardian* (12 December 2016) <https://www.theguardian.com/science/head-quarters/2016/dec/12/the-trolley-problem-would-you-kill-one-person-to-save-many-others> accessed 23 August 2018

Smith O, ‘A Huge Global Study on Driverless Car Ethics Found the Elderly are Expendable’ *Forbes* (21 March 2018) <https://www.forbes.com/sites/oliversmith/2018/03/21/the-results-of-the-biggest-global-study-on-driverless-car-ethics-are-in/#7b297b254a9f> accessed 23 August 2018

Schrage M, ‘Big Data’s Dangerous New Era of Discrimination’ *Harvard Business Review* (2014) <https://hbr.org/2014/01/big-datas-dangerous-new-era-of-discrimination> accessed 24 August 2018

Loehr A, ‘Big Data for HR: Can Predictive Analytics Help Decrease Discrimination in the Workplace’ *The Blog Huffpost* (2015) <https://www.huffingtonpost.com/anne-loehr/big-data-for-hr-can-predi_b_6905754.html> accessed 24 August 2018

Edited books:

Kalyvas J R and Overly M R, *Big Data: A Business and Legal Guide* (Auerbach Publications 2014)

Ragnedda M and Muschert G, *The Digital Divide: The Internet and Social Inequality in International Perspective* (Routeledge 2013)

Chapters in edited books:

Anciaes P R and Thomopoulos N, 'Ethical Issues in Transportation' in Mark Garrett (ed), *Encyclopedia of Transportation: Social Science and Policy* (SAGE Publications: Thousand Oaks, California, USA 2014)

Andrejevic M, 'Surveillance in the Big Data Era' in Pimple K D (ed) *Emerging Pervasive Information and Communication Technologies (PICT)* (Law, Governance and Technology Series, vol 11, Springer, Dordrecht, 2014)

Berman J J, '15 - Big Data Failures and How to Avoid (Some of) Them' in Berman J J (ed) *Principles and Practice of Big Data* (Second edition, Academic Press 2018)

Cukier K and Mayer-Schoenberger V, 'Rise of Big Data: How it's Changing the Way We Think about the World' in Pitici M (ed) *The Best Writing on Mathematics 2014* (Princeton University Press 2015)

Larsson A O 'Studying Big Data – Ethical and Methodological Considerations' in Fossheim H and Ingierd H (eds) *Internet Research Ethics* (Cappelen Damm Akademisk 2015)

Melis A, 'Public Transportation, IoT, Trust and Urban Habits' in Bagnoli F and others (eds) *Internet Science* (INSCI 2016, Lecture Notes in Computer Science, vol 9934, Springer, Cham, 2016)

Ming Z and others, 'BDGS: A Scalable Big Data Generator Suite in Big Data Benchmarking' in Rabl T and others (eds) *Advancing Big Data Benchmarks* (WBDB 2013, Lecture Notes in Computer Science, vol 8585, Springer, Cham)

Pedreschi D, Ruggieri S and Turini F, 'The Discovery of Discrimination' in Custers B and others (eds) *Discrimination and Privacy in the Information Society* (Studies in Applied Philosophy, Epistemology and Rational Ethics, vol 3, Springer Berlin Heidelberg 2013)

Pingel F and Steinbrecher S, 'Multilateral Secure Cross-community Reputation Systems for Internet Communities' in Furnell S, Katsikas S K, Lioy A (eds), *Trust, Privacy and Security in Digital Business* (TrustBus 2008, Lecture Notes in Computer Science, vol 5185, Springer, Berlin, Heidelberg)

Simon J, 'Distributed Epistemic Responsibility in a Hyperconnected Era' in Floridi L (ed) *The Onlife Manifesto* (Springer, Cham, 2015)

Steinbrecher S, 'The Need for Interoperable Reputation Systems' in Camenisch J, Kisimov V and Dubovitskaya M (eds) in *Open Research Problems in Network Security* (Lecture Notes in Computer Science, vol 6555, Springer, Berlin, Heidelberg 2011)

Published proceedings:

Albanese M 'Measuring Trust in Big Data' in Aversa R and others (eds) *Algorithms and Architectures for Parallel Processing* (ICA3PP, Lecture Notes in Computer Science, vol 8286, Springer, Cham 2013) https://doi.org/10.1007/978-3-319-03889-6_28

Belov N and others, 'Computational Trust Assessment of Open Media Data' in *2013 IEEE International Conference on Multimedia and Expo Workshops* (IEEE, 2013) DOI**:** 10.1109/ICMEW.2013.6618392

Dignum V and others, 'Ethics by Design: necessity or curse?' in *AAAI/ACM Conference on Artificial Intelligence, Ethics and Society* (AAI/ACM, 2018) <http://www.aies-conference.com/wp-content/papers/main/AIES_2018_paper_68.pdf> accessed 24 August 2018

Gal-Oz N, Grinshpoun T and Gudes E, 'Privacy Issues with Sharing Reputation across Virtual Communities' in Truta T M and others (eds) *Proceedings of the 4th International Workshop on Privacy and Anonymity in the Information Society* (ACM, 2011) DOI: 10.1145/1971690.1971693

Golle P, 'Revisiting the Uniqueness of Simple Demographics in the US Population' in Juels A (ed) *Proceedings of the 5th ACM Workshop on Privacy in Electronic Society* (ACM, 2006) DOI: 10.1145/1179601.1179615

Josang A, 'Robustness of Trust and Reputation Systems' in *2010 Fourth IEEE International Conference on Self-Adaptive and Self-Organizing Systems Workshop* (IEEE, 2010) DOI**:** 10.1109/SASOW.2010.33

Lee T and Tso M, 'A Universal Sensor Data Platform Modelled for Realtime Asset Condition Surveillance and Big Data Analytics for Railway Systems: Developing a "Smart Railway" Mastermind for the Betterment of Reliability, Availability, Maintainability and Safety of Railway Systems and Passenger Service' in *2016 IEEE SENSORS* (IEEE, 2016) 1-3, DOI: 10.1109/ICSENS.2016.7808734

Li W and others, 'Synthesis for Human-in-the-Loop Control Systems' in Ábrahám E and Havelund K (eds) *Tools and Algorithms for the Construction and Analysis of Systems* (TACAS 2014, Lecture Notes in Computer Science, vol 8413, Springer, Berlin, Heidelberg, 2014)

Lukoianova T and Rubin V, 'Veracity Roadmap: Is Big Data Objective, Truthful and Credible?' (2013) 24(1) Advances In Classification Research Online 4 DOI: 10.7152/acro.v24i1.14671

Plale B, 'Big Data Opportunities and Challenges for IR, Text Mining and NLP' in Liu X and others (eds) *Proceedings of the 2013 International Workshop on Mining Unstructured Big Data Using Natural Language Processing* (ACM, 2013) DOI: 10.1145/2513549.2514739

Qian J and others, 'A Passive UHF Tag for RFID-based Train Axle Temperature Measurement System' in *2011 IEEE Custom Integrated Circuits Conference* (IEEE, 2011) 1-4, DOI: 10.1109/CICC.2011.6055420

Ruggieri S, Pedreschi D and Turini F, 'DCUBE: Discrimination Discovery in Databases' in Elmagarmid A (ed) *Proceedings of the 2010 International Conference on Management of Data* (ACM Press, 2010) 1127 DOI: 10.1145/1807167.1807298

Sänger J and others, 'Trust and Big Data: A Roadmap for Research' in Morvan F, Wagner R R and Tjoa A M (eds) *2014 25th International Workshop on Database and Expert Systems Applications* (IEEE, 2014) 278-282, DOI: 10.1109/DEXA.2014.63

Someh I A and others, 'Ethical Implications of Big Data Analytics' in *Proceedings of the Twenty Fourth European Conference on Information Systems* (2016)

Swan M, 'Philosophy of Big Data: Expanding the Human-data Relation with Big Data Science Services' in *'15 Proceedings of the 2015 IEEE First International Conference on Big Data Computing Service and Applications* (IEEE, 2015) 468-477 DOI: 10.1109/BigDataService.2015.29

Wenlong P and others, 'Access Point Research in Rail Train Safety Monitoring Sensor Network' in *2012 Third International Conference on Digital Manufacturing and Automation* (IEEE, 2012) 157-160, DOI: 10.1109/ICDMA.2012.38

Wu P-J and Chen Y-C, 'Big Data Analytics for Transport Systems to Achieve Environmental Sustainability' in Meen T-H, Prior S D and Lam A (eds) *2017 International Conference on Applied System Innovation* (IEEE, 2017) 264-267, DOI: 10.1109/ICASI.2017.7988401

Zang H and Bolot J, 'Anonymization of Location Data Does not Work: A Large-scale Measurement Study' in Ramanathan P (ed) *Proceedings of the 17th Annual International Conference on Mobile Computing and Networking* (ACM, 2011) 145-156 DOI: 10.1145/2030613.2030630

Thesis:

Awad E, 'Moral Machine – Perception of Moral Judgment Made by Machines' (Master's Thesis, Massachusetts Institute of Technology 2017)

Büchi M, 'Digital Inequalities: Differentiated Internet Use and Social Implications' (Thesis, University of Zurich 2017)

Coen R and others, 'A User Centered Perspective on Algorithmic Personalization' (UC Berkeley 2016)

Conference papers:

Schroeder R and Cowls J, 'Big Data, Ethics, and the Social Implications of Knowledge Production' (Data Ethics Workshop, KDD@Bloomberg, New York, 2014) <https://pdfs.semanticscholar.org/5010/f3927ca8133a432ac1d12a8e57ac11cb3688.pdf> accessed 23 August 2018

Sweeney L, 'Foundations of Privacy Protection from a Computer Science Perspective' (Proceedings Joint Statistical Meeting, AAAS, Indianapolis, 2000) <https://dataprivacylab.org/projects/disclosurecontrol/paper1.pdf> accessed 23 August 2018

Websites:

Bundesministerium für Verkehr und digitale Infrastruktur, 'Mobilität digital Hochfranken – MobiDig' (BMVI) <https://www.bmvi.de/SharedDocs/DE/Artikel/DG/mfund-projekte/mobilitaet-digital-hochfranken-mobidig.html?nn=326002> accessed 24 August 2018

Denham E, 'Blog: Consent is not the "Silver Bullet" for GDPR Compliance' (Information Commissioner's Office, 2017) <https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2017/08/blog-consent-is-not-the-silver-bullet-for-gdpr-compliance/> accessed 24 August 2018

Eurostat, 'Internet Access of Households, 2017' (Eurostat) <http://ec.europa.eu/eurostat/statistics-explained/index.php/Digital_economy_and_society_statistics_-_households_and_individuals> accessed 24 August 2018

Cukier K, 'Big Data is Better Data' (TED 2014) <https://www.youtube.com/watch?v=8pHzROP1D-w> accessed 23 August 2018

Marr B, 'Why only one of the 5 Vs of Big Data really Matters' (IBM Big Data & Analytics Hub, 19 March 2015) <http://www.ibmbigdatahub.com/blog/why-only-one-5-vs-big-data-really-matters> accessed 21 August 2018

Noorman M, 'Computing and Moral Responsibility' (The Stanford Encyclopedia of Philosophy 2012) <https://stanford.library.sydney.edu.au/entries/computing-responsibility/> accessed 23 August 2018

Privacypatterns.org, 'Sticky Policies' (Privacypatterns.org) <https://privacypatterns.org/patterns/sticky-policy> accessed 23 August 2018

République Française, 'Explanatory Memorandum' (*République numérique*) <https://www.republique-numerique.fr/pages/digital-republic-bill-rationale> accessed 23 August 2018

Smith R, '5 Core Principles to Keep AI Ethical' (World Economic Forum, 19 April 2018) <https://www.weforum.org/agenda/2018/04/keep-calm-and-make-ai-ethical/> accessed 22 August 2018

Reports and studies:

Custers B and others, 'Deliverable 2.2 Lists of Ethical, Legal, Societal and Economic Issues of Big Data Technologies. Ethical and Societal Implications of Data Sciences' (e-SIDES, 2017) <https://e-sides.eu/resources/deliverable-22-lists-of-ethical-legal-societal-and-economic-issues-of-big-data-technologies> accessed 22 August 2018

European Commission Directorate-General for Communications Networks, Content and Technology, Media and Data, Unit G.1. – Data Policy and Innovation, 'An Emerging Offer of "Personal Information Management Services". Current state of service offers and challenges' (European Commission 2016) <https://ec.europa.eu/digital-single-market/en/news/emerging-offer-personal-information-management-services-current-state-service-offers-and> accessed 23 August 2018

European Data Protection Supervisor,' Ethics Advisory Group Report 2018' (EDPS 2018) <https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf> 23 August 2018

European Data Protection Supervisor, 'Opinion 4/2015 Towards a New Digital Ethics' (EDPS 2015) <https://edps.europa.eu/sites/edp/files/publication/15-09-11_data_ethics_en.pdf> accessed 23 August 2018

European Data Protection Supervisor, 'Opinion 7/2015 Meeting the Challenges of Big Data: A Call for Transparency, User Control, Data Protection by Design and Accountability' (EDPS 2015) <https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf> accessed 23 August 2018

European Data Protection Supervisor, 'Opinion 9/2016 EDPS Opinion on Personal Information Management Systems. Towards more User Empowerment in Managing and Processing Personal Data' (EDPS 2016) <https://edps.europa.eu/sites/edp/files/publication/16-10-20_pims_opinion_en.pdf> accessed 23 August 2018

European Union Agency for Network and Information Security, 'Privacy by Design in Big Data – An Overview of Privacy Enhancing Technologies in the Era of Big Data Analytics' (ENISA 2015) <https://www.enisa.europa.eu/publications/big-data-protection> accessed 23 August 2018

Evodevo, 'The Ethics of Big Data: Balancing Economic Benefits and Ethical Questions of Big Data in the EU Policy Context' (European Economic and Social Committee 2017) <https://www.eesc.europa.eu/resources/docs/qe-02-17-159-en-n.pdf> accessed 23 August 2018

Federal Ministry of Transport and Digital Infrastructure, 'Ethics Commission, Automated and Connected Driving' (BMVI 2017) <https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile> accessed 22 August 2018

Hee K and others, 'Big Data Methodologies, Tools and Infrastructures' (LeMO 2018) <https://static1.squarespace.com/static/59f9cdc2692ebebde4c43010/t/5b6d4674032be489a442fa8b/1533888127770/20180716_D1.3_Big+data+methodologies%2C+tools+and+infrastructures_LeMO.pdf> accessed 24 August 2018

Norwegian Data Protection Authority, 'Big Data Report' (Datatilsynet 2013) <https://www.datatilsynet.no/globalassets/global/english/big-data-engelsk-web.pdf> 23 August 2018

Sciencewise Expert Resource Centre 'Big Data: Public Views on the Collection, Sharing and Use of Personal Data by Government and Companies' (Sciencewise 2014)

Surakitbanharn C A and others, 'Preliminary Ethical, Legal and Social Implications of Connected and Autonomous Transportation Vehicles' <https://www.purdue.edu/discoverypark/ppri/docs/Literature%20Review_CATV.pdf> accessed 23 August 2018

The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, 'Ethically Aligned Design: A Vision For Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems, Version 1' (IEEE, 2016) <http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html> accessed 24 August 2018

Turnbull P, Lear J and Thomas H, 'Women in the Transport Sector - Promoting Employment by Preventing Violence against Women Transport Workers' (International Labour Organization 2013) <http://www.ilo.org/wcmsp5/groups/public/---ed_dialogue/---sector/documents/briefingnote/wcms_234882.pdf> accessed 24 August 2018

US General Accounting Office, 'Record Linkage and Privacy: Issues in Creating New Federal Research and Statistical Information' (GAO, 2001) <https://www.gao.gov/assets/210/201699.pdf> accessed 23 August 2018

WISE 'Project Wise - Project Report: Women Employment in Urban Public Transprot Sector' (WISE) <http://www.wise-project.net/downl/final_wise_project_report.pdf> accessed 23 August 2018

World Economic Forum in collaboration with Bain & Company, Inc., 'Personal Data – The Emergence of a New Asset Class' (World Economic Forum 2011) <http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf> accessed 23 August 2018

Young L and others, 'Personal Data in Transport: Exploring a Framework for the Future' (Open Data Institute 2018) <https://theodi.org/wp-content/uploads/2018/06/OPEN-Personal-data-in-transport-.pdf> accessed 23 August 2018

EU legislation and cases and European Commission documents:

Case C-450/06 *Varec SA v État belge* [2008] ECLI:EU:C:2008:91

Case T-189/14 R *Deza a.s v Agence européenne des produits chimiques* [2014] ECLI:EU:T:2014:686

Charter of Fundamental Rights of the European Union [2012] OJ C 326/391

Commission, 'Proposal for a Directive of the European Parliament and of the Council on certain aspects concerning contracts for the supply of digital content' COM (2015) 634 final

Commission, 'Towards a Thriving Data-Driven Economy' (Communication) COM(2014) 442 final, 4

Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data [1995] OJ L 281/0031-0050

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC [2016] OJ L 119/1

Interviews:

Interview with Gerhard Kress, Responsible for Data services, Siemens (31 July 2018) http://www.odbms.org/blog/2018/07/on-ai-and-data-technology-innovation-in-the-rail-industry-interview-with-gerhard-kress/

Interview with Steve Lohr, Reporter, New York Times (19 December 2016) http://www.odbms.org/blog/2016/12/big-data-and-the-great-a-i-awakening-interview-with-steve-lohr/

Interview with Pedro Domingos, Professor of computer science, University of Washington (18 June 2018) http://www.odbms.org/blog/2018/06/on-artificial-intelligence-machine-learning-and-deep-learning-interview-with-pedro-domingos/

Interview with Oren Etzioni, CEO, Allen Institute for Artificial Intelligence (15 January 2016) http://www.odbms.org/blog/2016/01/on-artificial-intelligence-and-society-interview-with-oren-etzioni/

Interview with John Markoff, Reporter, The New York Times (11 August 2016) http://www.odbms.org/blog/2016/08/machines-of-loving-grace-interview-with-john-markoff/

White papers:

Cloud Security Alliance Big Data Working Group, *Expanded Top Ten Big Data Security and Privacy Challenges* (White Paper, 2013)

Appendix A Opportunities and challenges

Opportunities

1	**TRUST** Higher volume in big data leads to trustworthiness, such as reputation.
2	**TRUST** A solid partnership between organisations can lead to trust in information sharing.
3	**SURVEILLANCE** As the degree of sophistication increases, the centralised management of data, such as for instance traffic data, needs to be enabled, combined with an optimisation of massive data usage. If this is successful, a high degree of precious data can be gathered to support decision-making. This way, big data can facilitate smart transport. In other words, it can encourage: • Reliability; • Availability; • Maintainability; • Safety; but also • Efficiency.
4	**PRIVACY** If organisations involved in big data analytics are able to address the users' privacy needs, this could open the door to more people engaging in big data.
5	**PRIVACY** In certain situations, individuals may be willing to forsake part of their privacy in return for the benefits big data applications may bring.
6	**TRANSPARENCY** Providing transparent information to individuals whose personal data is involved in big data analytics may increase trust in the processing activities and the technology used. When people feel they can trust a technology, they tend to be more willing to engage in it.
7	**PRIVACY / CONSENT / TRANSPARENCY** Both industry and government should take up responsibility to eliminate the misconceptions that exist regarding personal data protection. Data subjects should be educated, notably through transparent notices from industry, about the grounds for processing and the possible impacts on privacy.
8	**CONTROL** Giving control to data subjects should not necessarily stifle the use of big data. Instead, a bigger involvement of data subjects may lead to improved analytics given that the data subjects can correct mistakes and detect unfair decisions.

9	**FREE WILL** Social and political phenomena can be extended into new domains by big data, it achieves greater accuracy in pinpointing individual behaviour to use as supporter for free will.
10	**FREE WILL** Increasing accessibility and personalisation for people may provide proper choices through a better understanding and serving people's needs.
11	**DATA OWNERSHIP** Given the nature of the issue, no opportunity in relation to personal data ownership has been identified. We believe however that the opportunities related to control, discussed above, may (at least partly) address the data subjects' concerns with respect to the "ownership" of their personal data.
12	**DISCRIMINATION** Big data analytics can be a tool to make existing discriminatory decisions visible.
13	**DISCRIMINATION** Big data analytics can be used to tailor customised services that meet the needs of certain social groups in order to improve their inclusion into the society.

Challenges

1	**TRUST** The huge amount and diversity of data and data sources reduce the average quality of overall data.
2	**TRUST** Trust related to big data can be easily polluted by various kinds of people who act against a moral requirement.
3	**SURVEILLANCE** There are serious ethical issues posed by the emerging regimes of population-level monitoring, whereas recent privacy-protection initiatives fall short of addressing the challenge to democracy posed by big data surveillance.
4	**SURVEILLANCE** If predictive analytics succeed in altering behaviours, surveillance cannot be understood as a passive technique but as a dynamic shaping of behaviours (impacting for instance free will).
5	**SURVEILLANCE** Various commentators consider that the privacy risks related to big data analytics are low, pointing out the large amount of data processed by analytics and the de-identified nature of most of this data. This conclusion is likely to be wrong in practice, including from a legal perspective. This is notably due to the fact that anonymity by de-identification is a difficult goal to achieve, as demonstrated by different studies.
6	**PRIVACY** Difficulty of striking a balance between the right to privacy and the use of individuals' personal data in the context of big data applications used in the transport sector.
7	**PRIVACY** As needs for privacy may vary between individuals or between situations (e.g. depending on the benefits the individual gets in return), it will be difficult for companies and developers to adopt a one-size-fits-all approach.
8	**TRANSPARENCY** The lack of transparency of personal data processing activities in a big data context negatively affects data subjects' trust in such activities and the related technology. Data subjects may be reluctant to use big data applications in the transport sector.
9	**PRIVACY / CONSENT** The misconceptions regarding data protection concepts, such as consent, cause confusion both among data subjects and organisations. This general trend will also affect the use of big data in the transport domain.

10	**CONTROL** There is an asymmetry of control of personal data between data subjects and the organisations processing the data. Data subjects may fear losing control over their digital identity by engaging in big data analytics.
11	**FREE WILL** The rapidly increasing size and scope of information of big data technologies could lead to unfair use, in terms of the free will and autonomy of humans.
12	**DATA OWNERSHIP** As the definition of 'personal data' continues to evolve, information that is qualified as non-personal data today may be classified as personal data in the future
13	**DATA OWNERSHIP** End users may be reticent to provide their personal data for big data analytics in transport as this would entail forsaking "ownership."
14	**DATA OWNERSHIP** Difficulty to establish ownership of different data components within a set of data of various types and coming from various sources.
15	**DATA OWNERSHIP** Multiple actors involved in big data analytics may try to claim ownership of the data concerned, which may lead to a gridlock.
16	**DISCRIMINATION** Big data analytics are vulnerable to technical and systematic biases which can lead to discriminatory conclusions. These biases may be caused by heterogeneity of data, the size of the data sets, data quality, noise accumulation, spurious correlation, incidental endogeneity, and algorithms complexity.
17	**DISCRIMINATION** Personalised services can exclude certain social groups or lead to discriminatory treatment.
18	**DISCRIMINATION** Big data analytics requires digital accessible services, which makes it difficult for people with limited access to or knowledge of technology ("Digital Divide").