Collaboration in Higher Education for Digital Transformation in European Business

# Big Data: Ethics and Law

Rainer Lenz, rainer.lenz@fh-bielefeld.de

August 2019

This working paper is a result of the EU Erasmus Plus project "Collaboration in Higher Education for Digital Transformation of Corporate Businesses" (CHEDTEB) and in particular the intensive exchanges with IT experts and company managers in the project's working groups on "Big Data" and "Blockchain technology". Further information about Big Data, algorithms, AI and Blockchain technology can be found on our project webpage.

**http://www.chedteb.eu/**

*The author is grateful for the lively discussions within various Big Data workshops and would like to thank in particular the following colleagues for their valuable ideas, and suggestions: Jan Luhan, Jan Budik, Martin Fridrich all of Brno University of Technology/Czech Republic; Viire Täks of Tartu University/Estonia; Margareta Teodorescu and Tahir Lushi, both of the University of Applied Sciences Bielefeld/Germany.*

Co-funded by the Erasmus+ Programme of the European Union

**Abstract**
Personal profiling and predictive behavioural analysis done by Big Data applications pose immense challenges to society and democracy, especially when they violate individuals' constitutionally guaranteed fundamental rights, such as the rights to privacy and data protection, and the right to non-discrimination based on personal attributes. At the same time, Big Data applications also threaten basic ethical principles needed in a democratic society, such as fairness and respect for human autonomy.
An analysis of European data protection laws (GDPR, ePrivacy, Digital Content, Copyright and Trade Secrets) shows far-reaching gaps with regard to the protection of privacy and the non-discrimination of individuals. Governments and legislators have a clear requirement to close these gaps as quickly and comprehensively as possible.
We propose a three-pillared model for the future regulation of Big Data applications, covering three areas of action. Firstly, a fundamental reorientation of the concept of digital identity towards self-sovereignty over private data. According to this, the individual would become the owner of their own personal data and thus be able to decide sovereignly with whom to share which data, for which purposes, and over which time period. Secondly, the empowerment of the individual as a sovereign of their own data must be accompanied by a comprehensive education and training program at all levels of society. Through knowledge and training, individuals must be able to use the opportunity to determine for themselves how their information is used and, at the same time, be able to bear the associated risks. Thirdly, regulators themselves need to use so called "legal-tech" solutions to carry out automated and software-based testing and monitoring of Big Data applications with regard to their compliance with privacy protection regulations. For this purpose, the legislator faces the challenge of implementing the legal principles formulated in written laws into software code.

# Table of Content

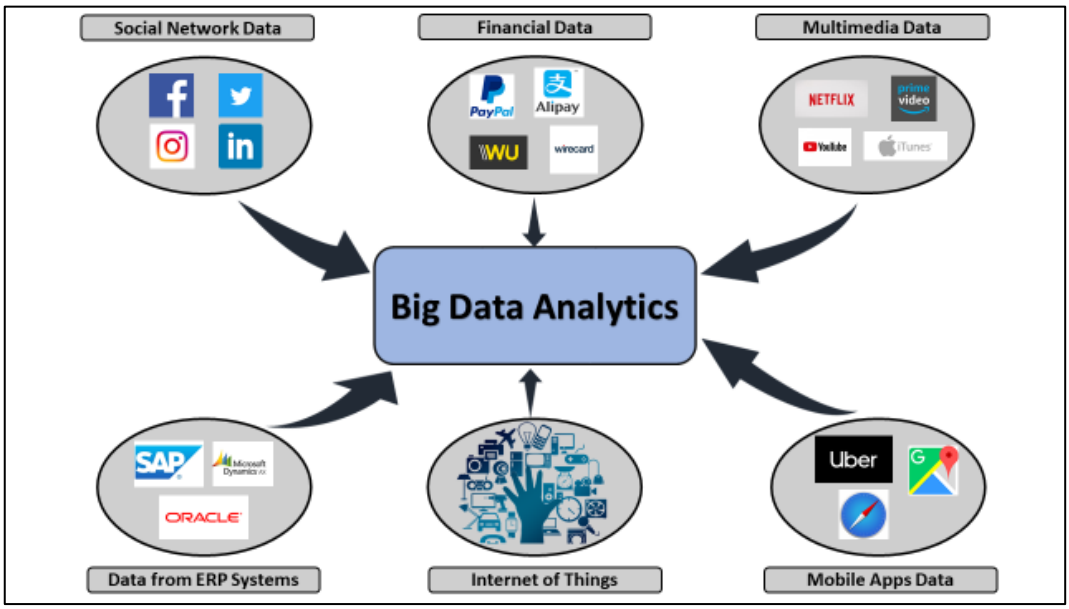
# 1. Understanding Big Data analytics

Almost every minute, each one of us produces new data whenever we use mobile phones, laptops or cars, whenever we are recorded by public cameras, and whenever we stream music, use social media sides or book tickets. Nowadays almost every device collects its user's data via built-in sensors and cameras, and exchanges the data in real-time with other devices via automatic interfaces, or transmits it into a data network. The volume of data generated every minute and its growth rates are gigantic. The pool of data thus generated is of high quality for data analytics, because it contains a wide variety of unstructured data from different sources.

The task of Big Data analytics nowadays is to identify hidden patterns or correlations in the mass of raw data in order to transform the data into information and then into contextualized knowledge to solve a particular problem.

A working definition of Big Data analytics is provided by Treleaven, Barnett, and Koshiyama (2019, p. 35): *"Big data analytics is the process of examining vast and varied data sets to uncover hidden patterns, trends, customer preferences, and so forth. One of the most exciting areas for intelligent algorithms is behavioral and predictive analytics. Behavioral analytics focuses on providing insight into actions of people, whereas predictive analytics extracts information from existing data sets to determine patterns and predict future outcomes and trends."*

Figure 1: Data sources of Big Data

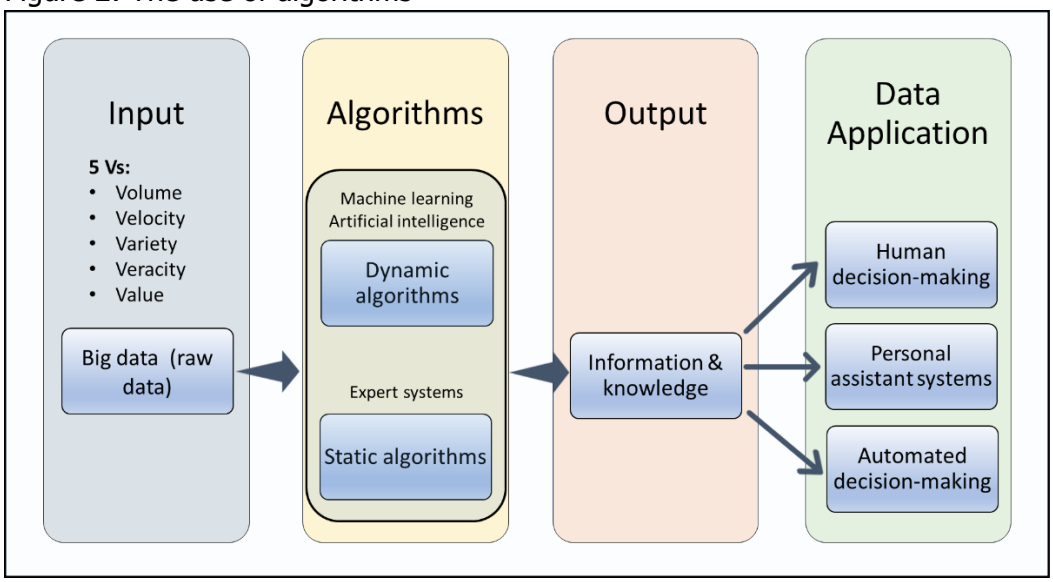



Source: The author

In the literature, the data used as inputs for Big Data analytics is mostly characterized by '5 Vs', which stand for volume, velocity, variety, veracity and value. "Volume" and "variety" describe the amount and types of unstructured data harvested from a variety of sources, regardless of context or analytic purpose. Big

Data technologies also enable comprehensive de-contextualization and re-contextualization of data that is captured, analyzed, and reconnected for different purposes. The term "velocity" refers to the speed at which vast amounts of data are being generated, collected and analyzed. The "veracity" of input data indicates the quality or trustworthiness of the data. Finally, the term "value" indicates the likelihood that the selected mass data - or better the knowledge generated by using the data set - can be turned into business value. The processing of such a mass of data requires the use of extremely powerful computers, storage technologies and fast data networks, but also a variety of qualitative improvements, such as the use of more complex calculation rules (algorithms) in computation-intensive computer simulations and a rationalization, standardization and quality improvement of many work processes. (Demchenko, De Laat, & Membrey, 2014)

Figure 2: The use of algorithms




Source: The author

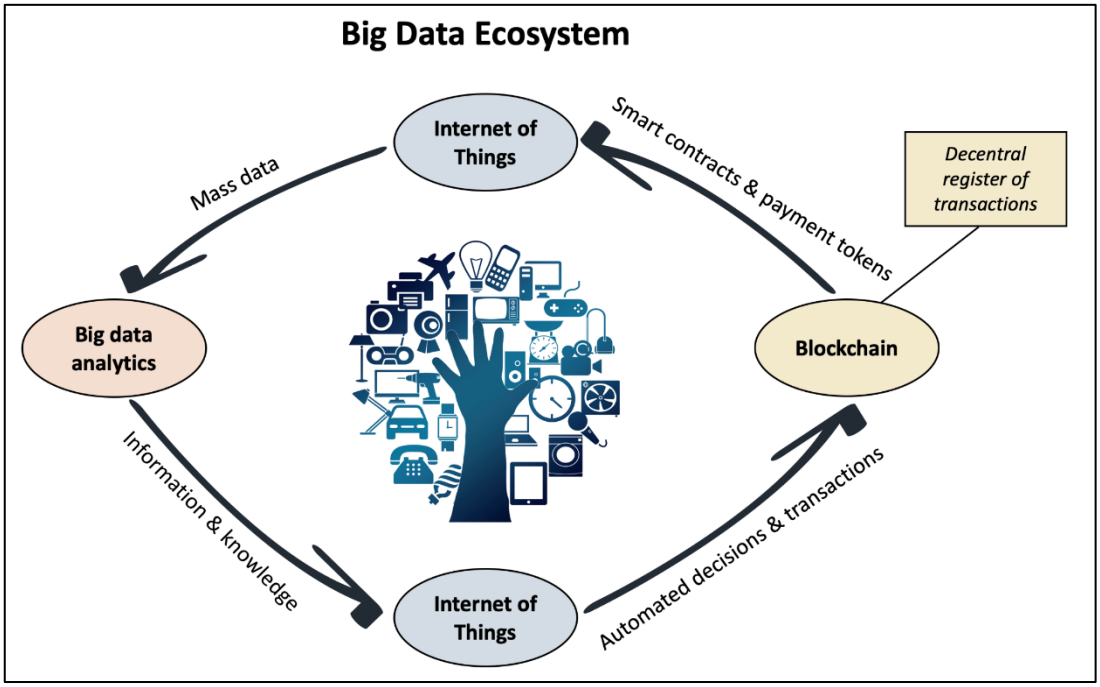
Whereas in the past, the analysis of data was carried out on static algorithms with the rules specified by programming (so called "expert systems"), in recent years the use of dynamic, self-learning algorithms that do not follow pre-programmed learning paths has prevailed. These types of algorithms are termed "machine learning" or "artificial intelligence (AI)" because they are based on data processing by artificial neural networks similar to the neuronal networks that underpin the working of the human brain. Artificial intelligence enables computers to make decisions almost in real time and to learn without explicit programming.

A shortcoming of artificial intelligence is the "black box" effect, whereby outputs can be difficult to explain due to the complexity of near-autonomous AI algorithms choosing their own paths for decision-making and selecting their own input data and weightings. Even the developer of artificial intelligence algorithms may be unable to explain the reasoning and logic behind a certain decision.

The strategic importance of having unlimited access to unstructured mass data and the ability to analyse the data by using AI becomes apparent when observing the general trend towards disintermediation and decentralisation towards peer-to-peer business (P2P). Peer-to-peer platforms create value by facilitating the exchange of information or values between their users. In order to make these exchanges happen, platforms harness and create large, scalable networks of users and resources that can be accessed on demand. They are subcontracting every client order by setting up a platform, where all the interfaces with external stakeholders (users, subcontractors, payment systems etc.) are fully standardized and automated. Big Data provides the analytics and intelligence for the automated decision-making of platform businesses. Big Data analytics is thus the nucleus of the platform economy and of digital business models.

The emergence of Blockchain technology facilitates the overall trend of automated and speedy decision-making by providing a shared database for the recording and registration of decentralised transactions between P2Ps, P2Ms and machine-to-machines (M2M). Additionally, Blockchain technology enables the storage of software code with 'if-then' relationships within the database, which facilitates the use of so-called smart contracts in which Blockchain users store automated transactions, ready to be executed given a certain external event as trigger.

Figure 3: Big Data Ecosystem




Source: The author

But for the execution of autonomous decisions in the physical real world, machines with a variety of built-in sensors and a fast connection to the internet are needed. The development of the Internet of Things (IoT) provides the necessary link between the digital and real worlds, without which Big Data analytics would lack both the mass of data generated by the sensors for the analysis and the executive power

of machines for automated decision-making. Devices located in the IoT both feed algorithms with data and are to some extent controlled by those same algorithms.

Treleaven et al. (2019, p. 34) describe the close connection between Big Data, AI, IoT and Blockchain as four core algorithm technologies which "…*are intimately linked, i.e. AI provides the algorithms, Blockchain provides data storage and processing infrastructure, the IoT provides the data and Big Data (behavioural/predictive) provides the analysis."*

Given the interconnectedness between Big Data Mining, the use of the Internet of Things and Blockchain Technology, it is clear that any regulation of Big Data applications will have an impact on the benefits of IoT and Blockchain Technology.

# 2. Opportunities and risks need regulation

This section looks at the opportunities and risks associated with Big Data. From an analysis of these opportunities and risks, we conclude that Big Data applications require an appropriate regulatory and legal framework.

## 2.1. Opportunities

Automated data analysis means that decision-making can become more data driven as patterns become easier to recognize and the future behaviours of stakeholders become easier to predict. Decisions supported by real-time data analysis can be made faster and could be taken by software in an automated way. This enables IoT-applications such as automatic driving of cars or robotic applications.

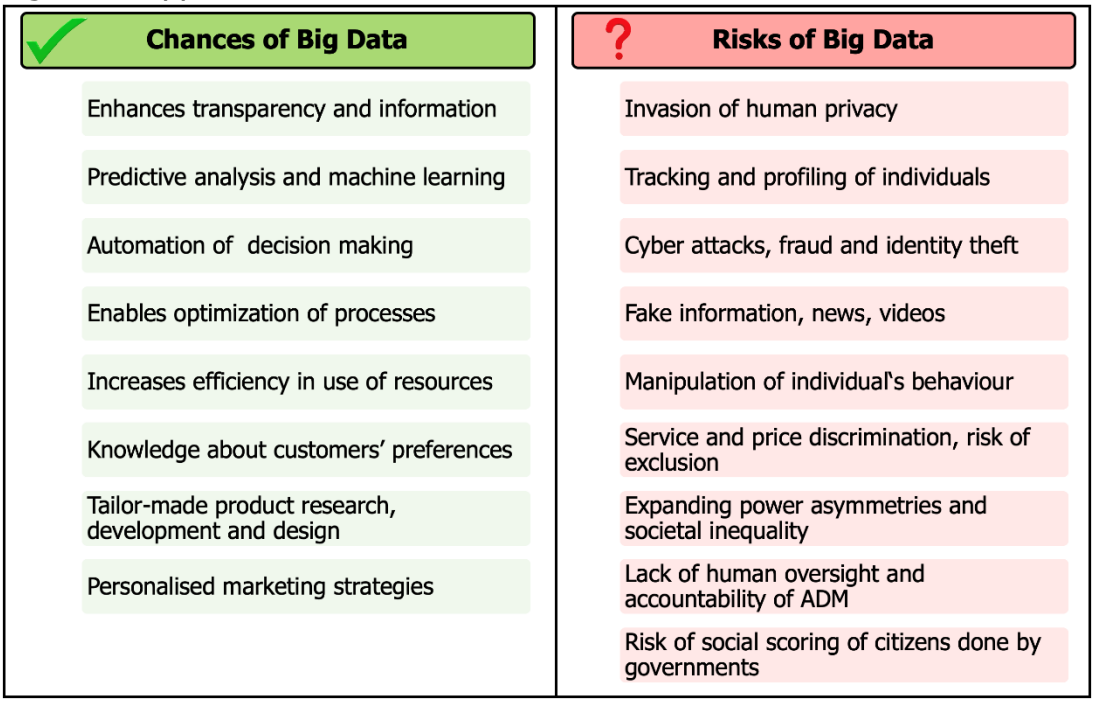
The analysis of large data pools is particularly useful for optimizing processes in production and logistics. Almost every machine in these industries today is equipped with some combination of sensors, cameras and GPS transmitters and creates a mass of data when used. The analysis of these data pools in combination with the communication between the machines allows a completely different production structure: processes can be more agile and scalable, saving considerable resources and costs.

Most beneficial from a company perspective is the collection and analysis of customer data, which allows companies to generate a near-perfect image of customer preferences and behaviour. This allows product and service offerings to be individually tailored to the customer. Marketing strategies can be personalized in communications, advertising, pricing and payment conditions. Customers can even be profiled to assess their credit risk or the likelihood of them returning goods they have purchased online.

The large number of apps that can be used on mobile phones have been joined by personal assistance systems in households, through which the user can communicate directly with machines via devices in the homes. This starts with smart home applications from Amazon's Echo or mobile phone voice assistants such as

Apple's Siri or Google's Assistant. The better the knowledge of the user's habits, the more helpful and pleasant such assistance systems can be.

Figure 4: Opportunities and Risks




Source: The author

## 2.2. Risks

However, the benefits are offset by significant risks for individuals and for society, which are briefly described below.

Tracking the behaviour and activities of customers can constitute a violation of their right to privacy. In public life, in one's own home and especially in digital life, the individual hardly has any chance of privacy in communication and behaviour. Everywhere, cameras or sensors or GPS transmitters are installed that track individuals continuously. But privacy is a fundamental human right recognized in the UN Declaration of Human Rights, the International Covenant on Civil and Political Rights and in many other international and regional treaties.

The use of customer data for profiling and scoring ultimately leads to discrimination between citizens through different prices, payment terms and even access to digital services, such as online shopping. Weaker and more vulnerable social groups are most likely to be affected, meaning that data-based discrimination tends to aggravate already existing societal inequalities.

As Barocas and Selbst (2016, p. 674) wrote: "*Approached without care, data mining can reproduce existing patterns of discrimination, inherit the prejudice of prior decision makers, or simply reflect the widespread biases that persist in society. It*

*can even have the perverse result of exacerbating existing inequalities by suggesting that historically disadvantaged groups actually deserve less favorable treatment. Algorithms could exhibit these tendencies even if they have not been manually programmed to do so, whether on purpose or by accident."*

Our economic lives are increasingly being determined by automatic decisions made solely by software agents without human involvement. These decisions can sometimes have far-reaching consequences for the economic and social well-being of the individual, especially if the decisions are wrong. But it is hard to know how a person on the receiving end of an automated decision could contest it, especially when even the programmer is unable to explain how the decision was made or what goes on inside the 'black box' behind it. Furthermore, the accountability and liability for incorrect decisions and program failures need to be clarified. Finally, it needs to be discussed if automated decision-making should be allowed for every sector of the economy. For decisions with a high social impact, it might be appropriate to retain some human oversight and powers of intervention.

The knowledge of the personality and behaviour of the individual can easily lead to a loss of human autonomy in decision-making and open the door to manipulation, heteronomy in individual decisions and exploitation of dependencies. The danger of the manipulation of the individual is manifold and begins with the commercial interests of companies seeking to change consumers' purchasing decisions. But it can go much further, even to the level of manipulating citizens in their voting decisions in democratic elections of states as the recent data scandal around Cambridge Analytica shows (Scott, 2018).

Moreover, governments have an interest in positively influencing citizens by nudging them into socially better behaviour. In more authoritarian regimes, this well-intentioned state paternalism can easily become a complete surveillance of the social behavior and social scoring of citizens, see for instance China (Marr, 2019).

Wachter and Mittelstadt (2019, p. 4) wrote: "*Inferential analytics methods are used to infer user preferences, sensitive attributes (e.g., race, gender, sexual orientation), and opinions (e.g., political stances), or to predict behaviors (e.g., to serve advertisements). These methods can be used to nudge or manipulate us, or to make important decisions (e.g., loan or employment decisions) about us. The intuitive link between actions and perceptions is being eroded, leading to a loss of control over identity and how individuals are perceived by others. Concerns about algorithmic accountability are often actually concerns about the way in which these technologies draw privacy-invasive and non-verifiable inferences that cannot be predicted, understood, or refuted."*

Knowledge of the behaviour of individuals always constitutes power, which can all too often be exploited, especially when the power structures are asymmetrical, as between a state and its citizens, an employer and its employees, or a (health)

insurance company and an insured individual. Employers as well as health insurance companies have an interest in the health data of their contractual partners (employee and insured person), since in the event of illness high costs can arise. The monitoring of patients' or employees' behaviour provides the basis to incentivise a healthy lifestyle or sanction an unhealthy lifestyle (Ethikrat, 2017, p. 18).

With regard to power structures and the resulting danger for democracy, it should be questioned whether a few global technology companies such as Google, Facebook, Amazon, Alibaba etc. should be allowed to establish data monopolies over the information that users of their technology generate. The transfer of power previously reserved for the state to private companies continues in the management of personal identities and can easily be extended to the issuance of private cryptographic currencies.

Meanwhile, one can use one's Google ID for a variety of non-Google products and services, i.e. Google takes over in the digital world the function of the providers of digital identities of persons and in this way collects more and more user data. In the analogue world, the state has a monopoly on the issuance of official identity cards. Facebook plans to issue its own cryptographic currency "Libra" (a stable coin); this is another step towards breaking the state monopoly of the Central Bank on the issuance of means of payment. The question remains whether the state should replicate its powers in the digital world and follow as quickly as possible with the issuance of a digital state identity for its citizens and with the issuance of a state cryptographic currency as legal tender.

### 2.3. The need for regulation

The comparison of risks and opportunities above clearly shows the need for appropriate regulation of Big Data to protect individuals' rights to privacy and non-discrimination. This especially applies to economically weaker groups in society.

The need for regulation arises especially from the fact that the opportunities and risks of using Big Data in society are asymmetrically distributed. The ability to analyse large data pools primarily benefits large private companies and organisations, while smaller companies and individuals are more exposed to the risks. Here the legislator has a duty to empower the rights of individuals and users in the digital world. Regulation must prevent the individual from being degraded to a mere object of Big Data applications and ensure that the individual always retains their autonomy to act and to access all social and democratic rights.

# 3. The Ethics of Algorithms and AI

It is of overall importance for society to agree on common values of data protection, the protection of the human right to privacy, and protection against the danger of discrimination of individuals. Finding a compromise in the conflict between the legitimate commercial interests of companies and the protection of individuals' liberties is extremely difficult. Ethics or moral philosophy could serve as a yardstick

for judging what is right and wrong or good and bad. A social consensus on the common values needed, based on an open dialog and discussion process within society, would be an important step towards suitable regulation and legislation. Regulation and legislation can be seen as a social contract that aims to achieve a fair balance between the interests of certain stakeholder groups within society.

However, it is an illusion to believe that with legislation, state regulation and supervisory authorities it is possible to determine, control and monitor all Big Data activities. On the contrary, the more detail-based regulation is, the more likely it is that it will become a box-ticking exercise without having the desired effect. In this respect, ethical principles based on a broad societal consensus are of particular importance for legislation, which should reflect such principles. This applies in particular to the European Civil Law (called "code law"), which is driven by principles. Ethical principles are also important for the self-regulation of developers and designers of Big Data applications, and should inform codes of conduct, certifications and industry standards.

The ethics of Big Data applications, artificial intelligence and machine learning is a hotly disputed topic. Hence industry associations, trade unions and other civil society institutions have already published a number of guidelines of ethical principles in Big Data. AlgorithmWatch, an NGO started to map the landscape of these frameworks, has set up an AI Ethics Guidelines Global Inventory.[1]

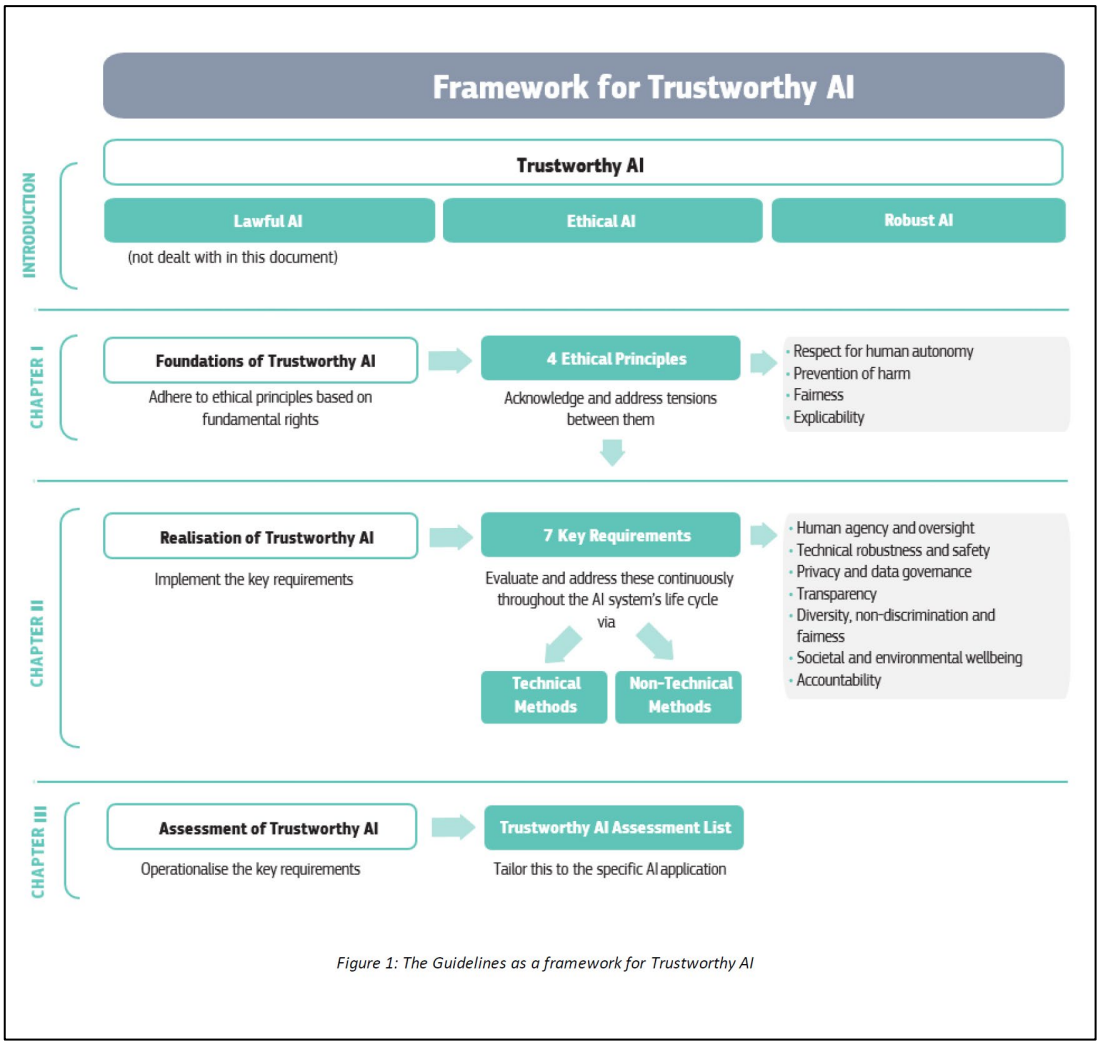
Particular attention is paid to the initiative of the EU Commission, which in June 2018 set up a high level expert group on artificial intelligence (HLEG AI), consisting of 56 experts from industry, science and civil society. This working group recently published a comprehensive "Ethics Guidelines for trustworthy AI" and "Policy and investment recommendations for trustworthy Artificial Intelligence" to the EU Commission

The HLEG AI developed a framework for trustworthy AI, which contains three main elements:

- Four ethical principles, which build the foundation of trustworthy AI
- Seven key requirements derived from those four principles, which a trustworthy AI needs to fulfil
- Finally, the HLEG AI developed a concrete and non-exhaustive assessment list for Trustworthy AI aimed for operationalising the key requirements set out beforehand.

[1] For further details see webpage of AlgorithmWatch: https://algorithmwatch.org/en/project/ai-ethics-guidelines-global-inventory/

Figure 5: Framework for Trustworthy AI




Source: European Commission High Level Expert Group on Artificial Intelligence (April 2019, p. 7)

According to the HLEG, a trustworthy AI needs to comply with given laws and regulations ("Lawful AI"), follow ethical rules, and be robust. The term "robust" is used to express the request that the AI should do what it is intended to do, both from a technical and social perspective, and should not cause any unintentional harm.

Nearly all guides put the well-being of the individual at the centre of these ethical considerations and emphasize the human-centric approach of Big Data ethics.
As the HLEG (p. 4) notes, *"AI systems need to be human-centric, resting on a commitment to their use in the service of humanity and the common good, with the goal of improving human welfare and freedom."*

Consequently, the use of Big Data, algorithms and machine learning should not in any way restrict or curtail the human being's fundamental rights as defined in Articles 2 and 3 of the Treaty of the European Union and in the Charter of Fundamental Rights of the EU as well as in most constitutions of EU member states.

Based on these general considerations, the HLEG derived four major ethical principles for AI**:**

1. **Respect for human autonomy**
    - *Humans interacting with AI systems must be able to keep full and effective self-determination over themselves, and be able to partake in the democratic process.*
    - *AI systems should not unjustifiably subordinate, coerce, deceive, manipulate, condition or herd humans. Instead, they should be designed to augment, complement and empower human cognitive, social and cultural skills.*
    - *The allocation of functions between humans and AI systems should follow human-centric design principles and leave meaningful opportunity for human choice. This means securing human oversight over work processes in AI systems.*
2. **Prevention of harm**
    - *This entails the protection of human dignity as well as mental and physical integrity.*
    - *AI systems and the environments in which they operate must be safe and secure. They must be technically robust and it should be ensured that they are not open to malicious use.*
    - *Vulnerable persons should receive greater attention and be included in the development, deployment and use of AI systems.*
    - *Particular attention must also be paid to situations where AI systems can cause or exacerbate adverse impacts due to asymmetries of power or information, such as between employers and employees, businesses and consumers or governments and citizens.*
    - *Preventing harm also entails consideration of the natural environment and all living beings.*
3. **Fairness**

    There are both substantive and procedural dimensions of fairness:
    - Substantive dimension of fairness: *Ensuring equal and just distribution of both benefits and costs, and ensuring that individuals and groups are free from unfair bias, discrimination and stigmatisation. If unfair biases can be avoided, AI systems could even increase societal fairness. Equal opportunity in terms of access to education, goods, services and technology should also be fostered. Moreover, the use of AI systems should never lead to people being deceived or unjustifiably impaired in their freedom of choice.*
    - Procedural dimension of fairness: *The ability to contest and seek effective redress against decisions made by AI systems and by the humans operating them. In order to do so, the entity accountable for the decision must be identifiable, and the decision-making processes should be explicable.*

4. **Explicability**
    - *Explicability is crucial for building and maintaining users' trust in AI systems. This means that processes need to be transparent, the capabilities and purpose of AI systems openly communicated, and decisions – to the extent possible – explainable to those directly and indirectly affected. Without such information, a decision cannot be duly contested.*

These ethical principles are translated into seven key requirements to achieve a trustworthy AI. Designers and developers of AI systems should implement key requirements. Buyers or deployers of AI-system, whether from the public or private sector, have to prove that the AI product being deployed meets ethical principles and the derived key requirements. End-users and the broader public should be made aware of ethical requirements and should be informed about their rights if the AI-system violates ethical principles.

Figure 6: Seven key requirements – HLEG

1 **Human agency and oversight**
*Including fundamental rights, human agency and human oversight*

2 **Technical robustness and safety**
*Including resilience to attack and security, fall back plan and general safety, accuracy, reliability and reproducibility*

3 **Privacy and data governance**
*Including respect for privacy, quality and integrity of data, and access to data*

4 **Transparency**
*Including traceability, explainability and communication*

5 **Diversity, non-discrimination and fairness**
*Including the avoidance of unfair bias, accessibility and universal design, and stakeholder participation*

6 **Societal and environmental wellbeing**
*Including sustainability and environmental friendliness, social impact, society and democracy*

7 **Accountability**
*Including auditability, minimisation and reporting of negative impact, trade-offs and redress.*

Source: European Commission High Level Expert Group on Artificial Intelligence (April 2019, p. 14)

Finally, the HLEG has developed a self-assessment test to help managers of private and public organisations assess whether the AI system deployed fulfils the ethical key requirements (European Commission High Level Expert Group on Artificial Intelligence, April 2019, p. 26).

# 4. Big Data legislation and regulation

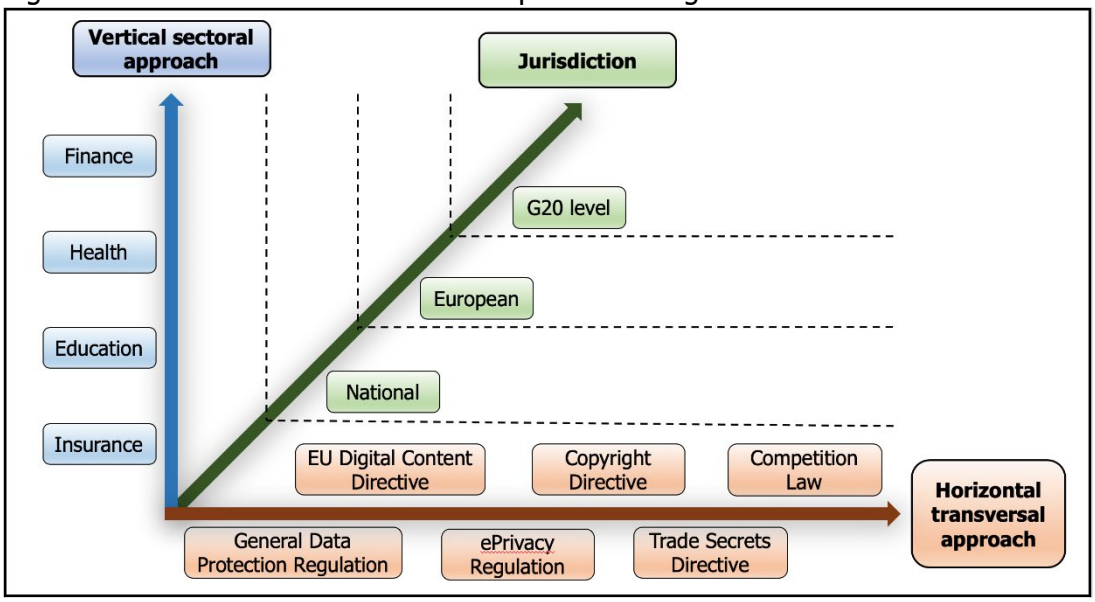
Ideally, data protection legislation should have three dimensions:

- Jurisdiction: the internet and the exchange of data have no geographic boundaries, so a borderless regulation is needed. National regulations, whether they are horizontal or vertical in their approaches, are not particularly effective because of geographic limitations. In this sense, the ideal approach would be a global (multilateral) agreement on common data protection legislation and

the creation of a multilateral supervisory authority, which is part of the network of national supervisory authorities.

- Horizontal transversal approach: Data protection and the associated human right to privacy is a transversal task or a cross-sectional task that affects all areas of our lives. In this respect, independent legislation is required for the collection, use, storage, transfer and protection of data that applies to all areas of our lives.
- Vertical sectoral approach: in addition to this horizontal regulation, which is universally applicable to all sectors, more sector-specific regulation should be applied to certain sectors in which the protection of private data is of particular importance, for example providing specific rules for the finance, health, education and other sectors. Big data applications will contain sector-specific features and should therefore also be reviewed by an expert regulator with sector-specific regulation as an add-on to horizontal regulation.

Figure 7: The three dimensions of data protection regulation




Source: The author

The following analysis of Big Data regulation focuses solely on international and European legislation at the horizontal level and thus includes only two of the three dimensions mentioned above. Sectoral legislation on data protection, such as in the health sector or the financial sector, is not the subject of this study.

## 4.1. International legal framework

At the international level, there is no legally binding multilateral agreement on data protection, the use of Big Data, the use of personal data, etc., nor any international data management supervisory authority. But there are several multilateral commitments, most of which include respect for human rights, and in particular the right to privacy, and the prohibition of discrimination on the basis of personal attributes, which have been signed by a large number of states.

### 4.1.1. United Nations level

Data Protection is not explicitly mentioned in the Universal Declaration of Human Rights (UDHR), but a human right to protection of personal data could be derived from the human right to privacy as written in Article 12:

Article 12 UDHR: "*No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks." United Nations (UN) (1948)*

Article 12 UDHR was the first time that the right of individuals to protection of their privacy against unjustified interference and external surveillance had been recognized as a human right at international level by all signatory states. Despite the fact that the UDHR is a non-binding declaration, it had considerable impact on international politics and the development of further conventions, declarations, resolutions and laws on human rights.

The UN General Assembly agreed in 1966 on The International Covenant on Civil and Political Rights (ICCPR) which entered into force in 1976. The ICCPR is an international treaty that commits its 169 signatory states to respecting and ensuring the exercise of individuals' civil rights, including privacy. In the past years the UN agreed on further resolutions the right to privacy in the digital age. (Council of Europe (CoE), 2018, p. 22)

### 4.1.2. G20 level

A global data law with supranational supervisory authority does not yet exist. But at the last G20 summit in Osaka in June 2019, 24 countries signed a declaration of intent to conclude a global agreement on data exchange, the so-called "Osaka Track" Declaration (The Japan Times, 2019). However, the non-committal wording of this declaration shows that data protection is a very controversial topic and leaders of powerful G20 countries show little interest in signing a multilateral treaty about it in the near future. In particular, the United States is concerned with commercial interests and preserving competitive advantages, and in addition, in some G20 states, the protection of human rights is not such a high priority on the political agenda.

It is advantageous that in the final G20 statement in Osaka world leaders explicitly emphasized the human-centric approach, which is the core element of Big Data ethics: "*We share the notion of a human-centred future society, which is being promoted by Japan as Society 5.0. As digitalization is transforming every aspect of our economies and societies, we recognize the critical role played by effective use of data, as an enabler of economic growth, development and social well-being. We aim to promote international policy discussions to harness the full potential of data." (G20, 2019)*

## 4.2. Pan-European Conventions

At the European level, a distinction must be made between the European Union of 28 member states and the European Council on Human Rights with 47 member states, which includes the 28 EU member states. After the Second World War, European countries formed the European Council on Human Rights and committed themselves to respecting human rights within the European Convention on Human Rights (ECHR), which entered into force in 1957.

### 4.2.1. The European Convention on Human Rights

The European Convention on Human Rights makes explicit reference to the United Nations' Universal Declaration of Human Rights and contains similar articles. The right to personal data protection is implicitly covered by Article 8 "Right to respect for private and family life"(European Court on Human Rights, 1950).

To ensure compliance with the ECHR, the European Council established the European Court of Human Rights (ECtHR) in Strasbourg in 1959, The ECtHR decides on complaints by individual persons as well as groups of persons and states that refer to violations of the rights recognized in the European Convention on Human Rights. Since 1998, the ECtHR has been a permanent court of law. Citizens can turn to it directly only after their domestic remedies have been exhausted.

The European Court on Human Rights has examined many situations involving data protection issues. These include interception of communications, various forms of surveillance by both the private and public sectors, and protection against storage of personal data by public authorities.

However, as the Council of Europe (CoE) (2018, p. 23) wrote in the Handbook on European data protection law: "*The respect for private life is not an absolute right, as the exercise of the right to privacy could compromise other rights, such as freedom of expression and access to information and vice versa. Hence, the Court strives to find a balance between the different rights at stake.*"

### 4.2.2. Council of Europe Convention 108

In 1981, the Council of Europe set up a Convention for the protection of individuals with regard to automatic processing of personal data (called "Convention 108") and opened it for signature by European council member states and non-member states. To date (August 2019), some 51 states have signed and 44 of them have ratified the convention. It is worth mentioning that eight non-European states ratified Convention 108: Argentina, Cabo Verde, Mauritius, Mexico, Morocco, Senegal, Tunisia and Uruguay (a chart of signatures and ratifications can be found on the Council of Europe Portal). For those states that have ratified the convention it is legally binding. However, it is not subject to the judicial supervision of the ECtHR and lacks a supranational regulatory authority for monitoring and enforcement.

According to the Council of Europe (CoE) (2018, p. 24), *"Convention 108 applies to all data processing carried out by both the private and public sectors, including data processing by the judiciary and law enforcement authorities. It protects individuals against abuses that may accompany the processing of personal data, and seeks, at the same time, to regulate the transborder flows of personal data. As regards the processing of personal data, the principles laid down in the convention concern, in particular, fair and lawful collection and automatic processing of data, for specified legitimate purposes."*

It is no coincidence that the latter principles of data protection are very similar to the principles of the EU General Data Protection Regulation. Because based on the Convention 108, the EU adopted in 1995 the General Data Protection Directive, which was converted into an EU regulation in 2016. In parallel, the European Council modified the "old" Convention 108, renaming it Convention 108+, and declared it open in 2018 for signing and ratification by states.

But unlike the European Council's conventions, EU legislation is legally binding on every EU member state, there are national and European regulatory bodies overseeing its implementation, and EU laws are enforceable. Consequently, the emphasis will be on the description of EU data protection legislation below.

## 4.3. European Union Legislation

For an understanding of the EU legislation and its legal implications or the EU member states, it is important to distinguish between different types of EU legislation: The starting point for any EU legislation is always the EU Treaty, which is also referred to as primary law in the EU. The Treaty of Lisbon amending the founding Treaty on European Union was adopted in 2007 and entered into force in December 2009. The Lisbon Treaty lays down the objectives of the European Union, the rules for EU institutions, how decisions are made, and the relationship between the EU and its member countries.

The Lisbon Treaty's principles and objectives underpin the body of EU law, which is known as secondary law and includes regulations, directives, decisions, recommendations and opinions. EU-Regulations are binding in their entirety on all EU countries. They are legal acts that apply automatically and uniformly to all EU countries as soon as they enter into force, without needing to be transposed into national law. EU-Directives leave EU countries a higher degree of liberty, how to incorporate them into national law in order to achieve the objective set by the directive. For further details and explanation on different types of EU-law see EU-webpage.

### 4.3.1. Charter of Fundamental Rights of the European Union

The Charter of Fundamental Rights of the European Union (CFR) guaranties EU citizens and residents certain political, social, and economic rights by EU law.
Referring to the European Parliament (2019): "*The Charter was solemnly proclaimed by EU Parliament, the Council and the Commission in Nice in 2000. After*

*being amended, it was proclaimed again in 2007. However, only with the adoption of the Treaty of Lisbon on 1 December 2009 did the Charter come into direct effect, as provided by Article 6(1) TEU, thereby becoming a binding source of primary law."*

The Charter is of particular importance for Big Data regulation as Article 7 and 8 enshrine the rights of EU citizens and residents to privacy and data protection into EU primary law. Furthermore, it states that legitimate grounds for data processing and a consent of the person are needed, everyone has the right of access to his/her data, has the right to rectify it and that the monitoring and control of compliance with these rules has to be done by an independent authority:

*"Article 7 Respect for private and family life*
*Everyone has the right to respect for his or her private and family life, home and communications.*

*"Article 8 Protection of personal data*

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."*

Additionally, it might be worth mentioning Article 21 on non-discrimination:
*"Article 21 Non-discrimination*

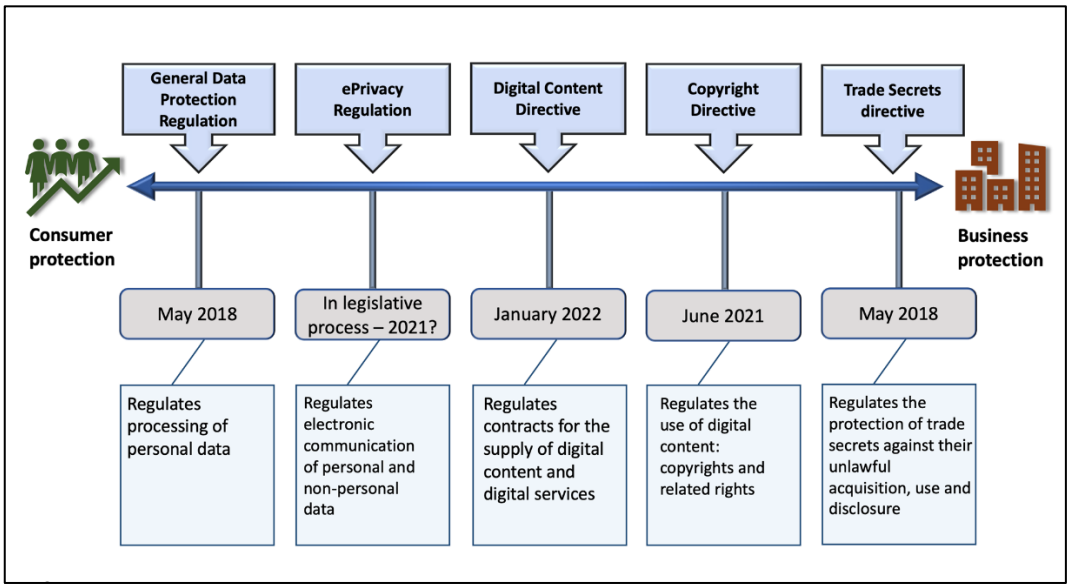
1. *Any discrimination based on any ground such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation shall be prohibited."*

### 4.3.2. EU Regulations and Directives

At the EU level (horizontal transversal approach) there are currently five pieces of legislation that directly or indirectly relate to data protection, the use and management of data and the transfer of data. These five EU laws serve as the minimum standard for any Big Data application in various sectors of the economy and society. Sector-specific legislation on data protection and data usage may go beyond these minimum standards, for example, when dealing with particularly sensitive personal health data.

Roughly, the five EU pieces of legislation can be differentiated by their objectives, whether they are more for the protection of the consumer, such as the General Data Protection Regulation (GDPR), or more for the protection of the commercial interests of companies, such as the Trade Secrets Directive (see figure 8).

Figure 8: European legislation related to Big Data




Source: The author

Strikingly, all five pieces of horizontal legislation are quite new. They have either just been implemented, or adopted and implementation is still pending, or, as with ePrivacy regulation, are still in the legislative process. This shows that legislation and regulation are lagging behind and react to societal demands with considerable delays. As legislative processes are lengthy and accompanied by intensive trilateral discussions between the EU Parliament, Commission and European Council it seems to be difficult to find the right balance between commercial and consumer interests. However, it is worth noting that consumer data protection laws such as GDPR and ePrivacy are EU regulations while the business related ones are EU directives.

The European Union seeks to compensate for this lack of a data protection agreement at the G20 level by declaring an extra territorial scope for its EU General Data Protection Regulation (GDPR). Article 3 of the GDPR ("Territorial Scope") includes data relating to products or services (paid for or free) that are offered to people in the EU regardless of where the data processor is located. Hereby the EU tries to extend the scope of the GDPR far beyond its own jurisdiction towards a worldwide level. Because of its power as a major economic bloc, it may be possible to enforce its own regulation in some non-EU countries.

#### 4.3.2.1. General Data Protection Regulation

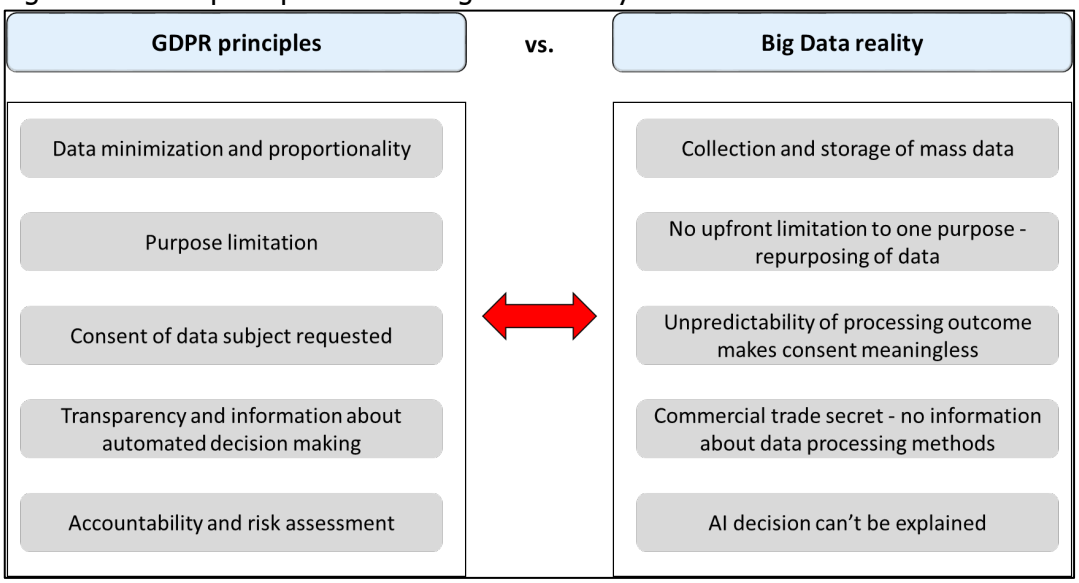
The EU General Data Protection Regulation is based on seven fundamental principles that are intended to strengthen the user's rights as a data subject and to give him or her sovereignty and autonomy over the use of their data. The principles of GDPR are lawfulness, fairness and transparency, purpose limitation, data minimization, accuracy, storage limitation, and integrity and confidentiality. As a seventh principle the data controller is kept accountable to comply with all these principles.

In addition, the GDPR obliges the EU member states to establish an independent regulatory authority for data protection, which monitors and enforces GDPR compliance (Walden & Christou, 2018).

It is obvious that the Big Data concept contradicts the basic principles of the GDPR. According to the GDPR, personal data may be collected only for a specific purpose and with the consent of the data subject. The best protection is data minimization - avoiding the collection of data that is not needed for the given purpose. But for Big Data applications it is important to get a mass of unstructured data from different sources and the outcome of data analysis needs to be open and not predetermined by a specific purpose of data collection. The data can be subsequently analyzed for multiple purposes, which implies repurposing of data and deviation from the purpose for which the consent of the data subject was given.

As Mitrou (2018, p. 37) wrote: "*It remains very questionable if the so called "notice and consent" model is suitable or practical in a Big Data context. For giving the consent the person needs all relevant information in easy language: the data subject, the identity of the controller, categories of data to be processes, purpose of processing, if multiple purposes than multiple consents are needed."*

Figure 9: GDPR principles versus Big Data reality




Source: The author

Although Big Data applications are obviously in contradiction with the principles of the GDPR, they are nevertheless classified as GDPR compliant, according to the standing jurisprudence of the European Court of Justice (ECJ) and the General Advocat (GA). The problem lies in the definition of personal data and thus in the limitation of the scope of the GDPR.

*"Article 4 GDPR Definition:*

*(1) personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is 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;"*

According to Wachter and Mittelstadt (2019, p. 30), the ECJ noted that only the *"name, date of birth, nationality, gender, ethnicity, religion and language of the applicant,"* or only data that is *"about"* the data subject are personal data. In this regard the judgement followed the opinion of the Advocate General (AG). She argues that *"only information relating to facts about an individual can be personal data,…"*

Based on the standing jurisprudence, the process of combining, restructuring, analysing different sources of sensitive data for commercial interests does not fall under GDPR regulation. Even if the interferences drawn from this data processing endanger the privacy of data subjects, Big Data applications are assessed as GDPR compliant.

The GDPR definition of personal data relies on the idea that personal sensitive data is static and could be exactly defined. But the reality is different, since with skilful combination of different data sets, which have a sufficient correlation with the sensitive personal data as defined in the GDPR, a person can be easily identified and tracked at any time and the data becomes de-anonymized.

According to van der Sloot and van Schendel (2016)… "*the nature of the data is also becoming less and less static; rather, data increasingly goes through a lifecycle in which its nature might change constantly. While the current legal system is focused on relatively static stages of data, and linked to them specific forms of protection (e.g. for personal data, sensitive data, private data, statistical data, anonymous data, non-identifying information, metadata, etc.), in reality, data go through a circular process: data is linked, aggregated and anonymized and then again de-anonymized, enriched with other data and profiles, so that it becomes personally identifying information again, and potentially even sensitive data, and is then once again pseudonymised, used for statistical analysis and group profiles, etc."*

In this respect, the distinction between sensitive and non-sensitive personal data, as defined in the GDR definition in Article 4, is not relevant for Big Data applications. Instead of putting the focus on the input data, legislatures should focus on the outcome of data processing, understood here as inferences or decisions, regardless of the type of data informing them. Under a human-centered approach, data protection is needed to protect the privacy of individuals regardless of the technology being used for data processing.

Wachter and Mittelstadt (2019, p. 81) have recommended to *"reconfigure privacy as a holistic concept and be more in line with the ECHR, the Council of Europe's "Modernised Convention for the Protection of Individuals with Regard to the Processing of Personal Data" and their guidelines on AI."*

#### 4.3.2.2. ePrivacy Regulation

The Regulation on Privacy and Electronic Communications (short "ePrivacy" Regulation) will apply to electronic communications in a broad sense covering any content exchanged by electronic means including text, images, speech, videos, and metadata. The regulation is still in the legislative process but could be adopted around the end of 2019 and enter into force in 2021 or later.

While the GDPR's focus is to protect the individual's personal data (Article 8 Charter of Fundamental Rights of the European Union), the ePrivacy Regulation aims to protect a person's private and family life, home and communications (Article 7 Charter of Fundamental Rights of the European Union). The ePrivacy Regulation is complementary to the GDPR and has a wider outreach as it targets the protection of an individual's privacy at every stage of every online interaction. But, unlike the GDPR, the ePrivacy regulation protects the privacy of natural persons as well as of legal entities. The rules of the GDPR are always relevant and are an overall part of the legislative aspects of the ePrivacy regulation.[2]

According to the proposal of the EU Commission (2017) *"Confidentiality of electronic communications ensures that information exchanged between parties and the external elements of such communication, including when the information has been sent, from where, to whom, is not to be revealed to anyone other than to the parties involved in a communication….. the principle of confidentiality enshrined in this Regulation should also apply to the transmission of machine-to-machine communications (IoT)."*

For Big Data applications, the most relevant aspect within the ePrivacy regulation – beside the content of digital communication - is the use of metadata for data analysis and interferences:

"*… metadata derived from electronic communications may also reveal very sensitive and personal information. These metadata includes the numbers called, the websites visited, geographical location, the time, date and duration when an individual made a call etc., allowing precise conclusions to be drawn regarding the private lives of the persons involved in the electronic communication, such as their social relationships, their habits and activities of everyday life, their interests, tastes etc."*

Metadata needing extra protection are described as data that "*(20) …reveal details of an individual's emotional, political, social complexities, including the content of*

[2]For further details on complementarity of ePrivacy to GDPR see webpage of Usercentris

*communications, pictures, the location of individuals by accessing the device's GPS capabilities, contact lists, and other information already stored in the device, the information related to such equipment requires enhanced privacy protection."*

The EU Commission's proposal strongly emphasizes the importance of consent and appears to be driven by the consideration that metadata in electronic communications should be treated as a special category of personal data which is inherently sensitive. The enhanced privacy protection implies that such interferences are only allowed with the end user's consent and for specific and transparent purposes:

*"(22) The methods used for providing information and obtaining end-user's consent should be as user-friendly as possible."*

The need for user's consent in particular applies to the storage of third party tracking cookies. "*(24) …they should, among others, require a clear affirmative action from the end-user of terminal equipment to signify his or her freely given, specific informed, and unambiguous agreement to the storage and access of such cookies in and from the terminal equipment."*

Wachter and Mittelstadt (2019, p. 65) see a loophole for the use of Big Data in Article 7 "Storage and erasure of electronic communications data" of the ePrivacy. According Article 7 (2), the *"provider of the electronic communications service shall erase electronic communications metadata or make that data anonymous when it is no longer needed for the purpose of the transmission of a communication."* The loophole is that data could be easily anonymised but then de-anonymised by enrichment with data obtained by other sources.

The EU Commision's proposal is from 2017 and since then several amendments have been proposed. At this stage it is not clear how the regulation will turn out at the end of the legislative process.

#### 4.3.2.3. Digital Content Directive

In January 2019, the EU adopted two new Directives: The directive on contracts for the supply of digital content and services (Digital content directive - DCD), and a directive on contracts for the sale of goods (Sales of goods directive - SGD). These two EU directives are linked because goods for sale could be either pure digital content and digital services or a digital service included in "smart" physical goods. Physical goods, such as for instance smart watches containing digital components, would be covered by the Sales of goods directive. Both directives come into force January 2022. The GDPR and any other EU data protection fully apply to the processing of personal data in connection with any contract falling within the scope of both Directives.

From a Big Data perspective, the most relevant are those implicit contracts where the user does not pay with a price but with their personal data for the supply of digital service or content. In this case, according to DCD, the consumer is entitled

to terminate the contract for a minor lack of conformity. This is different to paid service contracts, where termination requires a non-minor lack of conformity.

The question is what happens to the right to use the content or service and the right to use the personal data after the contract is terminated?

The Council of European Union (2019 (38)) wrote: "*Where personal data are provided by the consumer to the trader, the trader should comply with Regulation (EU) 2016/679* (GDPR)*. These obligations must also be complied with in cases where the consumer pays a price and provides personal data. Upon termination, the trader should also refrain from using any content other than personal data that was provided or created by the consumer when using the digital content or digital service supplied by the trader."*

In other words, there is no need for the contracting supplier of digital content or services to delete the personal data if the consumer does not explicitly request it by executing their right of deletion under the GDPR. Allowing Big Data applications in this way to retain personal data after the contract is terminated might incentivise traders to create more non-paid contracts in exchange for personal data and facilitate the Big Data business.

#### 4.3.2.4. Copyright Directive

The Directive on copyright and related rights in the Digital Single Market (short "Copyright Directive) was adopted in March 2019 and will come into force in 2021. The Copyright Directive gained prominence with the public discussion about the use of "up-load filters" on social media platforms. The Directive establishes a legal framework for regulating the rights and interests of authors and other rights holders on the one hand, and of users on the other. The GDPR and any other EU data protection law fully apply to the Copyright Directive.

From a Big Data perspective, most relevant is Article 3 which grants research organisations and cultural heritage institutions an exception to do text and data mining for the purposes of scientific research without needing permission from rights holders or authors.

Article 2 defines text- and data mining as follows: "'*text and data mining' means any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations;"*

This wide-reaching exception enables research institutions such as universities to use any digital content and the related metadata published on any public social media platform (YouTube, Instagram etc.) to be used for Big Data analysis for scientific purposes without having the consent or permission of the rights holder. With this exception, the data subject's right over personal data will be severely restricted for the sake of knowledge discovery and public research.

In reality, it might be difficult for regulators to distinguish where the scientific purpose of data mining ends and the commercial interest starts. Many research institutions are obliged to partner with commercial entities and to document the commercial use of results when applying for public project funding at national and European level. Furthermore, many research institutions of public universities are nowadays very much aligned with commercial partners. Researchers are quite often on the payrolls of companies and universities at the same time, or switch between the scientific and commercial sectors during their careers.

The press release of European Council (25/05/2018) on the adoption of the Copyright regulation confirms these concerns: "*In line with existing European research policy, which encourages universities and research institutes to develop collaboration with the private sector, research organisations should also benefit from the exception when their research activities are carried out in the framework of public-private partnerships. A mandatory exception for uses of text and data mining technologies in the fields of scientific research is included in the directive.*"

#### 4.3.2.5. Trade Secrets Directive

In 2016, the EU adopted the Directive on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure. It came into force in May 2018.

The Directive aims to "*harmonise the definition of trade secrets in accordance with existing internationally binding standards. It also defines the relevant forms of misappropriation and clarifies that reverse engineering and parallel innovation must be guaranteed, given that trade secrets are not a form of exclusive intellectual property right." European Commission (2016)*

From a Big Data perspective, it is crucial whether information about customers, suppliers, business plans, market research and strategies are also classified as trade secrets and as such benefit from special protection and confidentiality. The definition of a trade secret (Article 2) is very broad and allows any processing of data by algorithms or AI handled by a commercial entity to be classified as a trade secret.

However, the GDPR regulation puts some limits on the protection of trade secrets when using sensitive personal data: "*this Directive should not affect the rights and obligations laid down in Directive 95/46/EC (GDPR), in particular the rights of the data subject to access his or her personal data being processed and to obtain the rectification, erasure or blocking of the data where it is incomplete or inaccurate and, where appropriate, the obligation to process sensitive data in accordance with Article 8(5) of Directive 95/46/EC. European Union (2016 (35))*"

Since Big Data applications mostly do not use data that is explicitly classified as sensitive under the GDPR, they do not fall under the GDPR and hence can be

classified without further reason as a trade secret, enjoying full protection and confidentiality. Trade secrets might include customer habits, personalized marketing plans, credit risk assessments, predictive analysis about health risks etc.. Once data is classified as a trade secret, the public - and in particular affected users - have no right to transparency, disclosure of underlying algorithms or explanations of automated decision-making.

## 4.4. Summary of findings on Big Data legislation and regulation

At the international level, there is no legally binding multilateral agreement on data protection, the use of Big Data, the use of personal data, etc., nor any international data management supervisory authority. At the last G20 meeting in Osaka, only a Memorandum of Understanding was signed for the development of a multilateral data exchange agreement.

The right to privacy has been included as a human right in a number of international conventions (UN Convention on Human Rights, European Convention of Human Rights), Treaties (Lisbon Treaty), and national constitutions. In this respect, the signatory states of these conventions have committed themselves to the observance of human rights and in particular to the protection of privacy.

Figure 10: EU data protection laws and short comings for regulation of Big Data

| EU data protection law | Relevance for Big Data |
|---|---|
| GDPR | compliant as long as not using data classified as personally sensitive |
| ePrivacy Directive (draft) | collection of metadata needs end-user's consent, no right on deletion of data |
| Digital Content Directive | service provider could keep data of non-paid contracts after termination |
| Copyright Directive | facilitates data mining of research institutions for scientific purposes |
| Trade Secrets Directive | data processing of personal data of customers, suppliers etc. is classified as a trade secret |

Source: The author

But the above analysis of EU data protection legislation shows that the legislation does not adequately protect EU citizens in their right to privacy and non-discrimination against Big Data applications. Big Data applications are classified as GDPR compliant as long as they do not use any data directly classified as personally sensitive. This is the case, even though the Big Data concept contradicts all the principles of data protection formulated in the GDPR. The ePrivacy Regulation extends the data protection of individuals to the metadata of digital communication and requires the explicit consent of the data subject to have their metadata captured. This will not significantly restrict personal profiling and predictive behavioural analysis via Big Data. This is especially the case as the individual has no right

of deletion over the metadata. The three other EU data laws (Digital Content, Copyright, Trade Secrets) are formulated to be very business-friendly and are rather supportive for the use of Big Data. The Digital Content Directive provides digital service providers with the opportunity to keep the personal client data of non-paid contracts after termination. The Copyright Directive facilitates text- and data mining for scientific purposes done by research institutions. This opens the backdoor for commercial data mining as borders between commercial and scientific purposes are very fluid. Finally, the Trade Secrets Directive enables companies to classify the processing of clients' or suppliers' personal data as a trade secret and thus deprives them of any transparency and control over their data.

De facto, the privacy of EU citizens is currently not adequately protected against Big Data applications. The approach of defining which input data is considered to be personally sensitive and therefore cannot be used for data mining is proving to be ineffective. Ultimately, it depends on whether personal profiling and predictive behavioural analysis by Big Data applications constitute an encroachment on the privacy of the individual. Under a human-centered approach to data protection, it is necessary to protect the privacy of the individual regardless of the technology used for data processing.

With the GDPR, ePrivacy Regulation, the Digital Content Directive, the Copyright Directive and the Trade Secret Directive, the current EU data protection legislation seeks to find a balanced compromise between the commercial interests of technology providers and the protection of the private sphere of the individual. But this is also about setting priorities and properly weighing civil rights against business interests. The protection of commercial interests may be a legitimate concern, but the protection of privacy is a human right superior to commercial interests and should be a top priority, even if it limits the commercial opportunities of some companies.

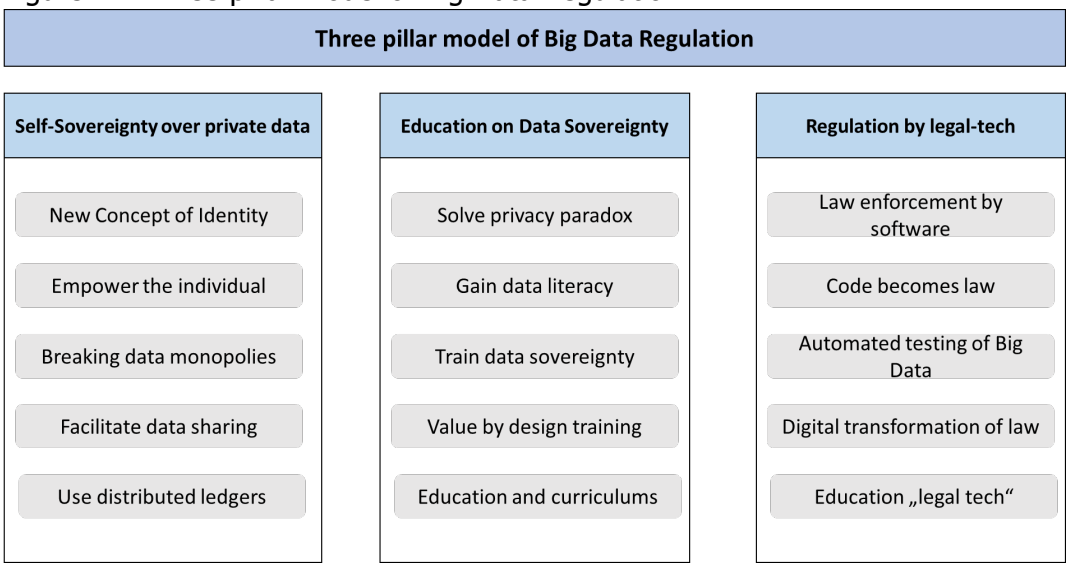
There is an urgent mandate for policy makers and public legislators to close this gap as quickly and comprehensively as possible.

# 5. Closing the gap of Big Data regulation

The gap in Big Data regulation cannot be filled by closing loopholes in the existing legal framework. Rather, it requires a fundamental system change that addresses several areas: the individual's right of self-determination over the use of his or her own personal data, a changed concept of the individual's digital identity, a broad education and training program for all citizens, as well as a fundamental reorientation of legislation and regulatory authorities in the use of intelligent software to enforce data protection laws.

As a vision for an effective protection of citizens' privacy in the future, we propose the following three-pillar model for the regulation of Big Data Applications: (1) self-sovereignty over private data, (2) education regarding data sovereignty, and (3) Big Data regulation by legal tech.

Figure 11: Three-pillar model of Big Data Regulation




Source: The author

## 5.1. Self-sovereignty over private data

The basic concept of GDPR and ePrivacy is based on the empowerment of the individual by strengthening their rights to information, transparency, deletion of personal data and the need for consensus. This strengthening of citizen rights is very beneficial and a step in the right direction, but unfortunately does not go far enough. What GDPR and ePrivacy regulations neglect is the existing unequal distribution of power and the information asymmetry between bigtech companies and their users. In the current model, the company – and not the user - owns the personal data. Furthermore, all the data is centrally owned by only a few bigtech companies, which became data monopolists. This is neither beneficial for users of digital services nor for competitors nor society.

The Bank for International Settlement (2019, p. 20) has recognised this problem, albeit from a competition point of view alone, and proposes to assign property rights on private data to customers: *"The issue, therefore, is how to promote data-sharing. Currently, data ownership is rarely clearly assigned. For practical purposes, the default outcome is that big techs have de facto ownership of customer data, and customers cannot (easily) grant competitors access to their relevant information. This uneven playing field between customers and service providers can be remedied somewhat by assigning data property rights to the customers. Customers could then decide with which providers to share or sell data. In effect, this attempts to resolve inefficiencies through the allocation of property rights and the creation of a competitive market for data – the decentralised or "Coasian" solution."*

Assigning property rights over personal data to the individual implies the creation of the individual's self-determined digital identity. Personal attributes, such as colour of skin, DNA, name, age, fingerprint etc. constitute the digital identity of the

individual. If the individual has full control of his or her private data, they become the sovereign of their own data and should decide for themselves when to collect, disclose and share the data with others. Therefore, the human right to privacy is contingent on the right to generate one's own identity.

As Agre and Rotenberg (1998) have stated "*the right to privacy is the freedom from unreasonable constraints on the construction of one's own identity".*

In the current economic system, the identity of individuals is provided by organisations, public administration and corporates. Nowadays, each individual has a barely manageable number of passwords for the use of online services, processing of online transactions, payment services, credit card transactions, etc. because every online merchant, every bank, and every platform registers and collects personal user data and assigns passwords to determine the user's identity when logging in. This means that the same private data of a single user is stored redundantly in a large number of corporate data silos. The private corporate provides the user of a service with his or her identity and not the other way around. Each administrative identity system is proprietary and owned by the organisation that provides it. (Lenz, 2019, p. 21)

When the management of digital identities falls to business organisations, the individual has no control over their personal data and has no possibility to monitor the external use of their personal data for business purposes. It is also highly inefficient. If, for example, the personal address or credit card number changes, hundreds of data records from different organizations have to be updated because there is no automatic data reconciliation between private organizations. When relocating from one place to another, an individual will probably change medical doctors but their health data does not move to the new doctor or hospital and remains in the disconnected data silo of the former practitioner. And even from a company perspective, the handling of private data becomes costly and burdensome as the new European General Data Protection Regulation strengthens user rights by imposing clear legal duties on the corporate's handling of the user data. (Lenz, 2019, p. 22)

Applying the new Distributed Ledger Technology would provide an opportunity to establish a decentralised concept for identification and data. Under this technology, every network participant is the sovereign of their own private data and hence their digital identification. The private data and its attributes are owned and controlled by the individual, are stored by them in a digital safe and could be shared partly or fully, temporally or permanently, with restricted or unrestricted access to network peers via a public key. Every access by a third party to the private data base is registered, recorded and time-stamped. As the GDPR grants every user the right to data portability, individuals could request businesses to transfer their personal data to a private storage place.

To sum up, the concept of self-sovereignty over private data has the following advantages:

- It grants the individual the basic right of informational self-determination and is therefore strongly aligned with the ethical principles of dignity and human autonomy.
- It changes the power relationship between individuals and platform businesses with regard to the use of private data for data analytics and algorithms.
- It breaks the existing data monopolies as it is a decentralised solution of data sharing between peers.
- It facilitates data sharing between network peers and could lead to efficiency gains within the economy.
- Decentralised data bases based on distributed ledger technology provide a higher resiliency against hacker attacks than centrally organised data storage.

The implementation of the self-sovereignty concept would require the following measures:

- Establishing an EU-wide accepted and publicly governed Distributed Ledger Infrastructure for storing and sharing digital identities and personal data between EU citizens and legal entities.
- Creating an EU legal framework for sharing digital identities: With eIDAS regulation (electronic IDentification, Authentication and trust Services regulation), adopted in 2014, the EU has created already a legal framework including electronic signatures, electronic seals, time stamp, electronic delivery service and website authentication for sharing national digital identities and personal data. The eIDAS regulation needs to be extended towards a distributed ledger solution. (European Union Blockchain Observatory and Forum, 2019, p. 19)
- Providing citizens of EU member states with national digital identities testifying to certain personal attributes, as is currently done with physical passport documents. These digital national identities need to follow EU standards to enable secure and seamless electronic interactions between businesses, citizens and public authorities.
- Amending the GDPR regulation towards the use of distributed ledger technology. GDPR regulation was designed for the current business world, which collects, stores and processes data in a centralised way and not for the decentralised distributed ledger world. Both worlds are so different that the GDPR seems hardly applicable to DLT in its current form and needs to be adapted.

## 5.2. Education on data sovereignty

The use of Distributed Ledger Technology combined with a concept of self-sovereignty on private data would strengthen the power of the individual vis-à-vis institutions, but at the same time shift the risk to the individual. Intermediaries absorb risks for their users and, should they cease to exist, individuals would have to bear

the risks. For some individuals it will be a blessing but others will find it difficult to cope with this new responsibility to take care of their own data.

The second pillar of this reform proposal is therefore a broad-based teaching, training and awareness program to help citizens understand their new social role as owners and managers of their personal data. A change in awareness of the risks from disclosing their personal data and creating a change in their behaviour is probably the most difficult part of the three-pillar reform program to implement. Taking people with you on a safe journey to the new digital world of algorithms and AI, so that it leads finally to welfare gains for all, is far more difficult, lengthy and costly than "just" developing new technologies to protect privacy. But without the societal change in behaviour, the other reform proposals will come to nothing.

In the past decade, several studies have investigated users' awareness of privacy issues and reported the existence of a so-called "privacy paradox" (Norberg, Horne, & Horne, 2007). Users are very much aware and concerned about the disclosure risks but this attitude is not reflected in their actual behaviour.

Pötzsch (2008) describe this paradox as an *"existing dichotomy between the intentions of people towards disclosure of personal data and their behaviour."*

The "privacy calculus theory" explains the paradox, stating that users are willing to disclose private information if the perceived benefits outweigh the perceived risks (Li, Sarathy, & Xu, 2010). Referring to the findings of behavioural economics, users' decisions to disclose data are taken under restrictions of bounded rationality. Mostly decisions are made within a second by pushing the button and without giving a second thought about potential risks. Quite often users are nudged by digital service providers - in a bad way - by offering digital services for free or for a discount in exchange for private information. Besides monetary incentives, the social reason to stay connected or being part of group are strong influencers of information disclosure (Vervier, Zeissig, Lidynia, & Ziefle, 2017, p. 88).

The self-sovereignty concept for private data should lead to a reversal of this decision-making situation: Disclosure of personal data is supposed to be rewarded with a benefit e.g. in a monetary kind. At least the user is compensated for taking the risk. Nevertheless, this does not solve the bounded rationality problem.

The question is what is needed to change decision-making so that users give more priority to non-disclosure? How can they be empowered to better calculate the perceived risk against the perceived benefits? A comprehensive education and training program for all layers of society for the handling of private data has to be set up as an integrated part of all curriculums. This should start in schools and continue in universities and continuing education institutions. However, it is not enough just to impart knowledge; in order to achieve changes in behaviour, it is necessary to integrate digital education into all contexts of learning. The focus should therefore be on real simulation and training programs for exercising user

rights given by the GDPR, in storing private data in digital safes, in deciding which data are to be shared partly or fully, temporally or permanently and restricted for a specific purpose or unrestricted with which business, and in the handling of public and private keys for encrypting or decrypting data.

As Vervier et al. (2017, p. 89) emphasise, "*….practical, demonstrative and concrete training programs should be developed which allow persons to see and feel consequences of their digital traces in the Internet, thereby possibly influencing their digital behaviors to the better. Many learners refuse to respond to dictating tutor systems with a superior attitude in the sense "you should" or "you must". Therefore, privacy behaviors need to be mediated by quite seamless assistant which let the users know about their current digital traces and how valuable the data might be for external or illegal access."*

The hands-on approach should apply in particular to courses of study that train data analysts and data miners. Here the students should learn and train in "value by design" courses how Big Data ethics can be implemented in the development of data models, algorithms and software. The accreditation of degree programmes should essentially depend on the extent to which value by design, ethical consideration of Big Data, and knowledge about data protection laws are integrated into curriculums.

## 5.3. Big Data regulation by legal-tech

The two reforms proposed above are not sufficient to protect the individual from interference with his or her privacy. Once the personal data have been passed on to third parties with or without the consent of the data subject, they are beyond the control of their use. While the principle of data minimization should be taken into account in any decision-making when buying new "smart" devices, in the Big Data world where mass data is collected, stored, and analyzed for patterns and formations anywhere, at any time, with or without consensus, and without a specific purpose, the principle of data minimization no longer plays a role.

Data protection legislators and regulatory authorities have to face the reality and challenges of the dynamic development in the Big Data sector and this unfortunately cannot be done with static instruments of the "old legal world" such as written law of books, paper documents and simple bans on the use of sensitive data. Here, the lawyers themselves must climb up to the next digital development stage by developing intelligent algorithms in cooperation with IT experts that monitor and test Big Data applications with regard to privacy protection and non-discrimination. In the financial sector, many banks and FinTechs are already using so-called "RegTech" software to comply with the regulator's legal static requirements. Now, it is time for legislators and regulators to use "legal-tech" that will use software to enforce the legal principles laid down in written legal codes.

Hildebrandt (2010, p. 428) and Koops were among the first to call for law to be embedded in the technological environment, an idea they called "Ambient Law":

"*Current law's articulation in the technology of the printed script is inadequate in the face of the new type of knowledge generation. A possible solution is to articulate legal protections within the socio-technical infrastructure. In particular, both privacy-enhancing and transparency-enhancing technologies must be developed that embed legal rules in ambient technologies themselves. This vision of 'Ambient Law' requires a novel approach to law making which addresses the challenges of technology, legitimacy, and political-legal theory. Only a constructive and collaborative effort to migrate law from books to other technologies can ensure that Ambient Law becomes reality, safeguarding the fundamental values underlying privacy, identity, and democracy in tomorrow's ambient intelligent world.*"

They noted two requirements for this (Hildebrandt (2010, p. 445): *"code cannot become law unless it fits two requirements: first it must be enacted by the democratic legislature and second, it must provide the possibility of contestation in a court of law."*

Beside these two basic requirements there will be more hurdles to overcome: Natural human language and in particular the language used by lawyers has its own interpretations and natural subjectivity which does not really fit with binary code. Frequently, lawyers use terms in legal contracts that are not clearly defined such as "may" or "in good faith" or "by mutual agreement" or "commercially reasonable manner" which are highly contextual and open to interpretation. These unspecified terms do not translate into a discrete yes-or-no interpretation of a binary code. A conceivable solution might be an approved translator or something like a compiler of computer language in the natural language of legal drafting and vice versa. It remains to be seen how this problem of the missing link between software code and the legal code of the lawyers will be solved.

## Bibliography


Agre, P. E., & Rotenberg, M. (1998). *Technology and privacy: The new landscape*: Mit Press.

Bank for International Settlement. (2019). *III. Big tech in finance: opportunities and risks*. Retrieved from https://www.bis.org/publ/arpdf/ar2019e.htm

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *Calif. L. Rev., 104*, 671.

Council of Europe (CoE). (2018). *Handbook on European data protection law - 2018 edition*. Retrieved from https://www.echr.coe.int/Documents/Handbook_data_protection_02ENG.pdf

Proposal for a Directive of the European Parliament and of the Council on certain aspects concerning contracts for the supply of digital content - Confirmation of the final compromise text with a view to agreement, (2019).

Demchenko, Y., De Laat, C., & Membrey, P. (2014). *Defining architecture components of the Big Data Ecosystem.* Paper presented at the 2014 International Conference on Collaboration Technologies and Systems (CTS).

Ethikrat, D. (2017). Big Data und Gesundheit–Datensouveränität als informationelle Freiheitsgestaltung. *Vorabfassung vom, 30*, 2017.

Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications), (2017).

European Commission. (2016). Trade Secrets. Retrieved from https://ec.europa.eu/growth/industry/intellectual-property/trade-secrets/

European Commission High Level Expert Group on Artificial Intelligence. (April 2019). *Ethics Guideline for Trustworthy AI*. Retrieved from https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

European Council. (25/05/2018). Copyright rules for the digital environment: Council agrees its position [Press release]. Retrieved from https://www.consilium.europa.eu/en/press/press-releases/2018/05/25/copyright-rules-for-the-digital-environment-council-agrees-its-position/

European Court on Human Rights. (1950). *European Convention on Human Rights*. Retrieved from https://www.echr.coe.int/Documents/Convention_ENG.pdf

European Parliament. (2019). *The Protection of Fundamental Rights in the EU*. Retrieved from

DIRECTIVE (EU) 2016/943 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure, (2016).

European Union Blockchain Observatory and Forum. (2019). *Blockchain and Digital Identity*. Retrieved from

https://www.eublockchainforum.eu/sites/default/files/report_identity_v0.9.4.pdf
G20. (2019). G20 Osaka Leader's Declaration [Press release]. Retrieved from https://g20.org/en/documents/final_g20_osaka_leaders_declaration.html
Hildebrandt, M., Koops, E-J.,. (2010). The challenges of ambient law and legal protection in the profiling era. *Modern Law Review, 73(3)*, 428-460. Retrieved from https://pure.uvt.nl/ws/portalfiles/portal/1248058/Koops_The_Challenges_of_Ambient_Law_100712.pdf
Lenz, R. (2019). Managing Distributed Ledgers: Blockchain and Beyond. *Available at SSRN 3360655*.
Li, H., Sarathy, R., & Xu, H. (2010). Understanding situational online information disclosure as a privacy calculus. *Journal of Computer Information Systems, 51*(1), 62-71.
Marr, B. (2019). Chinese Social Credit Score: Utopian Big Data Bliss Or Black Mirror On Steroids? *Forbes*. Retrieved from https://www.forbes.com/sites/bernardmarr/2019/01/21/chinese-social-credit-score-utopian-big-data-bliss-or-black-mirror-on-steroids/#3e93ac9048b8
Mitrou, L. (2018). Data Protection, Artificial Intelligence and Cognitive Services: Is the General Data Protection Regulation (GDPR)'Artificial Intelligence-Proof'? *Artificial Intelligence and Cognitive Services: Is the General Data Protection Regulation (GDPR)'Artificial Intelligence-Proof*.
Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of consumer affairs, 41*(1), 100-126.
Pötzsch, S. (2008). *Privacy awareness: A means to solve the privacy paradox?* Paper presented at the IFIP Summer School on the Future of Identity in the Information Society.
Scott, M. (2018). Cambridge Analytica helped 'cheat' Brexit vote and US election, claims whistleblower. *POLITICO*. Retrieved from https://www.politico.eu/article/cambridge-analytica-chris-wylie-brexit-trump-britain-data-protection-privacy-facebook/
The Japan Times. (2019). Abe heralds launch of 'Osaka Track' framework for free cross-border data flow at G20. Retrieved from https://www.japantimes.co.jp/news/2019/06/28/national/abe-heralds-launch-osaka-track-framework-free-cross-border-data-flow-g20/#.XSW9m4_grD5
Treleaven, P., Barnett, J., & Koshiyama, A. (2019). Algorithms: Law and Regulation. *Computer, 52*(2), 32-40.
United Nations (UN). (1948). *Universal Declaration of Human Rights (UDHR)*. Retrieved from https://www.ohchr.org/EN/UDHR/Documents/UDHR_Translations/eng.pdf
van der Sloot, B., & van Schendel, S. (2016). Ten questions for future regulation of big data: A comparative and empirical legal study. *J. Intell. Prop. Info. Tech. & Elec. Com. L., 7*, 110.
Vervier, L., Zeissig, E.-M., Lidynia, C., & Ziefle, M. (2017). *Perceptions of Digital Footprints and the Value of Privacy.* Paper presented at the IoTBDS.

Wachter, S., & Mittelstadt, B. (2019). A right to reasonable inferences: re-thinking data protection law in the age of big data and AI. *Columbia Business Law Review*.

Walden, I., & Christou, T. A. (2018). *Implications of Disruptive Technologies in Emerging Market Economies*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3230674