UNIVERSITÀ DEGLI STUDI DI PADOVA

DEPARTMENT OF POLITICAL SCIENCE,
LAW AND INTERNATIONAL STUDIES

Master's degree in European and Global Studies

BIG DATA CHALLENGES TO PRIVACY:

MERITS AND LIMITS OF THE GDPR

Supervisor: Prof. GUIDO GORGONI

Candidate: PINAR BOZ

Matriculation No. 2005870

A.Y. 2022 / 2023

Abstract: Big Data, artificial intelligence, and data-driven innovation advancements have an enormous beneficial effect on society as a whole and on various industries. On the contrary, their improper use may cause data processing to violate the purposes of privacy and data protection laws as well as moral obligations. One of the hottest topics in the current stage of study is how to assure big data security and privacy protection. This article begins with big data and its concepts, moves on to its advantages and challenges, and finally emphasizes on the GDPR, which guarantees the protection of personal data as a fundamental human right and gives data subjects more control over how their personal information is processed, shared, and analyzed.

Keywords: big data, privacy, data protection, GDPR

Table of Contents

Chapter 1: Introduction to Big Data in Contemporary Society

Big data promises to produce analytical insights that will expand the body of scientific and social scientific knowledge, dramatically enhance human self-knowledge and comprehension, and significantly improve decision making in the public and private sectors. They have already resulted in the development of totally new categories of products and services, many of which have been enthusiastically accepted by both institutions and people. However, where these data commit to capturing information on human activity, they have been seen as posing a challenge to core principles of autonomy, justice, fairness, due process, property, solidarity, and, perhaps most importantly, privacy.

Given this contrast, some have resorted to asking for complete prohibitions on different big data techniques, while others have found strong reason to finally throw caution and privacy to the wind in the assumption that big data will more than make up for its possible drawbacks. Of course, there are still others who seek a principled approach to privacy that respects the significant ideals that privacy maintains while providing the flexibility required for these promises to be fulfilled (Lane, 2014).

In this chapter, we will be focusing on big data related definitions such as “personal data”, “privacy” and “data protection”, “big data analytics”, “datafication”, “algorithms”, “artificial intelligence” and finally “surveillance capitalism”.

1. The Definition of Personal Data

According to the GDPR (General Data Protection Regulation), “Personal data means any information relating to an identified or identifiable natural person; 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" (Article 4).

Any information that can be used to identify or contact a living human is considered personal data. Personal data also refers to many pieces of information that, when put together, can identify a specific person. According to the GDPR, the definition of personal data is an identifiable natural person who can be identified, directly or indirectly, in particular by reference to an identifier like a name, an identification number, location information, an online identifier, or to one or more factors specific to that natural person's physical, physiological, genetic, mental, economic, cultural, or social identity.

This implies that personal data contain information about an individual. That person must be able to be identified or recognized, either directly or indirectly, from one or more identifiers or from characteristics that are unique to them (*What Is Personal Data?*, 2023).

According to the GDPR, personal data is covered in two ways: Personal information that has been partially or entirely processed automatically (i.e., electronic data) and non-automated processing of personal information that is included in or is meant to be included in a "filing system" (that is, manual information in a filing system).

Some of the personal information you handle might be more delicate in nature and need a higher level of security. Certain data processing activities are referred to as "special categories of personal data" under the GDPR. This refers to a person's personal information, including but also limited to: Political opinions, race, sexual orientation, genetic data, health data, ethnicity, religious or philosophical beliefs, biometric data and even trade union partnership (*What Is Personal Data?*, 2023).

2. Privacy and data protection

Privacy

“Privacy can be seen either as a right to be left alone, the “power to selectively reveal oneself to the world”, or as control over personal information or even as a freedom from judgement by others”. According to Solove, the term “privacy” is an umbrella term, referring to a wide and disparate group of related things and cannot be understood independently from society since privacy, in its core, is a social artefact and without the context of society there would be no need for privacy” (Politou, 2018, page).

Concerns about the use of big data are likely most frequently raised when it comes to privacy and security. Privacy refers to a person's ability to define and restrict access to their personal information. This may have to do with actions, such as secretly entering private areas, or with data derived from people's digital footprints (Richterich, 2018).

Information privacy is the right to some degree of control over the gathering and use of your personal data. Information privacy is the ability of a person or group to prevent information about oneself from being disclosed to individuals or groups other than those to whom the information is disclosed (Jain, 2016). Identification of personal information during Internet transmission is a significant problem for user privacy (Jain, 2016). On the other hand, according to Belanger and Crossler, “Information privacy is a set of the overall concept of privacy, which has been explored and discussed for centuries” (Belanger & Crossler, 2011).

According to Skinner, most interpretations of the term "privacy" pertain to a human right, albeit in various situations. These circumstances inspired Clarke (1999) to name four types of privacy: privacy of an individual, privacy of their personal behavior, privacy of their personal communications, and privacy of their personal data. Personal communication privacy and data privacy can now be combined into the concept of information privacy as the majority of communications are now digital and saved as information (Belanger & Crossler, 2011).

If we take a deeper look, information privacy is defined in a variety of ways, but there is little difference between the definitions' key components, which often include some kind of control over the potential secondary uses of a person's personal information. The process of employing data for reasons other than those for which they were initially acquired is known as secondary use. The four aspects of information privacy identified by Smith. (1996) include

collection, unauthorized secondary use, improper access, and errors. Information gathering, processing, distribution, and invasion are included in another taxonomy. The time, matter, and space dimensions of Skinner's proposed taxonomy of information privacy in collaborative environments reflect the structural view of privacy in information, which includes individual, group, and organizational privacy. The space dimension also reflects the structural view of privacy in information. Clarke clearly stated that "the interest an individual has in controlling, or at least considerably influencing, the processing of data about themselves" is what he meant by information privacy (Belanger & Crossler, 2011).

The right to privacy is regarded as a civic one in democracies. Several national constitutions whether explicitly or implicitly guarantee the right to privacy. The right to privacy is frequently seen as being expanded to include the protection of personal data. However, the European Union's Charter of Fundamental Rights tackles both independently, with Article 8 focused on data protection and Article 7 addressing respect for one's private and family life (The European Union Agency for Fundamental Rights 2007) (Richterich, 2018).

Big data critics have underlined how people lack understanding and control over the personal information that is gathered when utilizing online services, which concerns people's privacy. While supporters of big data, particularly corporate service providers, claim that user information remains anonymous, critics have questioned whether it is even possible to anonymize data with such a wide range of characteristics on such a massive scale (Richterich, 2018).

If we focus on other scholars' definition of privacy, it is the right to some degree of control over the gathering and use of your personal information. Information privacy is the ability of an individual or group to prevent information about themselves from being known to people other than those they disclose the information to. Jain indicates that "One serious user privacy issue is the identification of personal information during transmission over the Internet" (Jain, 2016).

Data Protection

There is a natural link between the right to privacy and the right to data protection. Nonetheless, there is a lot of debate in academia about how the right to privacy and the right

to data protection are related (Kulhari, 2018). Kulhari indicates “the right to privacy is distinct from the right to data protection because the former is rooted in human rights while the latter is treated as an economic matter is a red herring to say the least”.

The foundation of the right to data protection is the idea that data processing happens unplanned. As a result, the GDPR contains specific regulations pertaining to the duties of the data controller and processor. It is suggested that these restrictions on the right to data protection and what qualifies as lawful processing represent a compromise between several legitimate interests (Kulhari, 2018).

Although the right to data protection is closely related to the right to privacy, as well as the rights to freedom of expression, an effective remedy, and a fair trial, it also has some unique characteristics that support its designation as a stand-alone right. As stated in Article 8 of the 2012 Charter of Fundamental Rights of the EU, these components include: the right of access and rectification of collected data; control by an independent authority; and the requirement that data be processed fairly, for specific purposes, and only with the consent of the person concerned or another legal justification (McDermott, 2017).

Let us look at some of the basic principles that guide the right to data protection in the EU legal order (McDermott, 2017).

Privacy. As mentioned earlier, it is obvious that the right to data protection aims to safeguard the value of privacy, which is a basic right in and of itself.

Autonomy. The autonomy of the individual is another crucial value that the right to data protection safeguards, as evidenced by the continuous importance of informed consent in the European data protection policy. Natural persons "should have control of their own personal data," according to Recital 7 of the GDPR. The idea of dignity is undeniably connected to the concept of autonomy and its emphasis on consent.

Transparency. “Today, transparency is a core principle enshrined in Art. 5(1)(a) of the GDPR which states that personal data must be ‘‘processed lawfully, fairly and in a transparent manner in relation to the data subject,’’ thereby illustrating the close connection between transparency, lawfulness, and fairness” (Felzmann et al., 2019, page).

Given the previously mentioned issues with consent and knowledge, several authors have addressed the inequalities that exist in the field of data protection. As a response to this reality, 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."

Non-discrimination. Recognizing that data collection and processing should be done in a way that prevents discrimination against people "on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status, or sexual orientation" is related to the principle of transparency that serves as the foundation of the GDPR (McDermott, 2017).

The right to data protection faces some particularly difficult issues in the current environment of pervasive surveillance techniques and a growing emphasis on the collection and utilization of Big Data (McDermott, 2017). Let us examine these two specific challenges.

The first challenge is the widespread use of "volunteer" data to the right to data protection in the modern day, especially with the increasing number of wearable technology and social media platforms, even though users of these systems might not perceive themselves as providing data to others. Social network data, Internet of Things (IoT) devices, and other methods could lead to the collection of information about user' surroundings in addition to the individual user. Since data might be gathered for one purpose and then used for another, a paradigm that only relies on consent to process data can not completely guarantee that it will be protected (McDermott, 2017).

Secondly, mass surveillance is used to prevent future crimes like terrorist attacks and cyberattacks. As a result, the "watchers" are less noticeable than in the past because a large portion of their surveillance is conducted through "dataveillance," which involves tracking communications and online activity rather than actual movements. According to researchers, any “theory of rights" will face difficulties about the interests it identifies as rights, and the terms in which it identifies them. Those disagreements will in turn be vehicles for

controversies about the proper balance to be struck between some individual interest and some countervailing social considerations" (McDermott, 2017).

We reviewed the guiding principles of the right to data protection and made the case that this distinctive and recently created right represents a few basic features of the European legal system, including privacy, transparency, autonomy, and non-discrimination. The need to consider the values that underpin that right to our society and to advance the recognition of the right to data protection as a basic human right guaranteed to all people is more necessary than ever.

3. Big Data and Data Analytics

Big data explicitly refers to data sets that are so enormous or intricate that conventional data processing software is insufficient to handle them. Every day, a firm is overwhelmed with the vast amount of data, both organized and unstructured. Due to recent technical advancements, the volume of data produced by the internet, social networking sites, sensor networks, healthcare applications, and many other businesses is dramatically growing every day. In other words, big data refers to the vast volume of data created from many sources in many different formats at extremely fast speeds (Jain, 2016).

The term "big data" refers to a new generation of technologies and architectures that are designed to economically separate value from very large volumes of data. These technologies enable high-velocity capture, discovery and analysis. Some of the key properties of big data include volume, velocity and variety. Later studies pointed out that the definition of 3Vs (volume, velocity and variety) is not specific enough to explain the big data we face now. Thus, other terms were added to the definition, such as veracity, validity, value, variability, venue, vocabulary, and vagueness (Jain, 2016).

The diversity of the data, which might include text, audio, image, and video, is a common theme in big data. Variety serves as a metaphor for the various features of the data. Several strategies have been developed in recent years to guarantee the privacy of huge data. The stages of the big data life cycle, including data generation, storage, and processing, can be

used to classify these methods. Access restrictions and data falsification techniques are employed during the data generation phase to safeguard privacy (Jain, 2016).

Large volumes of data are being generated and are constantly being collected by businesses and government organizations. The present growing emphasis on large amounts of data will likely open chances and paths for understanding how such data are processed across many diverse disciplines. Nevertheless, the promise of big data comes with a cost – the customer's privacy is constantly in jeopardy. Current big data analytics and mining techniques have limitations with ensuring compliance with privacy terms and legislation. Independent of modifications to the applications and privacy rules, developers should be able to confirm that their applications comply with privacy agreements and that sensitive data is kept private. The necessity for fresh contributions in the fields of formal techniques and testing procedures should be identified to handle these difficulties (Jain, 2016).

Along with protecting individual privacy, big data has brought "collective privacy" back into the spotlight. This idea suggests that social groups should also have the right to privacy, not just individuals. The need for privacy has frequently been compared with the importance of public safety, as academics note: "Two moral obligations need to be reconciled: promoting human rights and increasing human wellbeing." But he disagrees with the notion that the latter would be a political matter affecting the general public and the former an ethical one affecting people's rights (Richterich, 2018).

4. Datafication, algorithms, artificial intelligence

Datafication

"The datafication of personal information constitutes a new kind of information society. Datafication allows analysis of information in more sophisticated ways and allows analyses across large data sets. It breaks down the traditional understanding of data as numbers and information as texts, movies, music, and so on. A good example is Google Books. When Google digitized books, it scanned them in a way that allowed for full-text searching and

stored the text in a way that allowed people to search for particular words or phrases across millions of books in a few seconds" (Mai, 2016, page).

On the other hand, Mayer-Schoenberger and Cukier define the term "datafication" as "the transformation of social action into online quantified data, thus allowing for real-time tracking and predictive analysis. Businesses and government agencies dig into the exponentially growing piles of metadata collected through social media and communication platforms, such as Facebook, Twitter, LinkedIn, Tumblr, iTunes, Skype, WhatsApp, YouTube, and free email services such as gmail and hotmail, in order to track information on human behavior: We can now collect information that we couldn't before, be it relationships revealed by phone calls or sentiments unveiled through tweets" (Mayer-Schoenberger and Cukier 2013: 30) (Dijck, 2014).

Not only among technological enthusiasts but also among academics who view datafication as a ground-breaking research opportunity to study human behavior, the use of data to access, understand, and monitor people's behavior is increasingly becoming a guiding principle (Dijck, 2014).

The term "datafication" serves as a lens for this analysis because it is increasingly being used in the "ether" to describe how dependent businesses are on data and their data infrastructures, as well as how data is becoming more accessible and, in this case, how it is being transformed into something valuable (Lycett, 2013).

Three inventive ideas, dematerialization, liquification, and density, can be used to conceptualize datafication and allow the logic of value generation to be reexamined. Dematerialization emphasizes the capacity to dissociate the informational component of an item or resource and its use in context from the physical environment. Liquification makes the point that, once dematerialized, information can be easily manipulated and moved around (given a suitable infrastructure), allowing resources and activity sets that were physically linked together to be unbundled and "rebundled" in ways that may have previously been challenging, excessively time-consuming, or expensive. Density is the result of the value creation process and is the best combination of resources that are mobilized for a certain context, at a given time and location (Lycett, 2013).

There isn't much doubt that IT is a key driving force in this logic of value creation because it provides the infrastructure and artifacts that free us from limitations on how, when, where, with whom, and when not to do things (configurations or constellations). Together with information technology, this logic of value creation also offers us some degrees of independence from "frozen knowledge". According to scientists, physical products are effective because they are repeatable and predictable, but they are also tools in which activity and information are frozen at the point of creation, leaving the accumulation of prior knowledge and actions in effect. He further argues that the distinction between goods and services is rather inaccurate and suggests offerings as a richer conception that is a reconfiguration of the entire value-creation process, with the process being optimized in terms of relevant actors, asset availability, and asset costs rather than the physical object (Lycett, 2013).

The value creation process, which is primarily described from the perspective of value-in-use the interaction between the product and the client views offerings as the input rather than the outcome. The dynamics of this interaction are possibly the most fascinating, as the information technology-driven value creation context of today permits a much denser reconfiguration of resources into co-created value patterns as well as a larger (more customized) variety of patterns (Lycett, 2013).

Let us analyze these concepts in the context of Netflix because they are highly abstract.
In many respects, it is reasonable to state that Netflix's streaming business model has undergone a significant shift as IT infrastructure and other technological advancements have completely freed media content from its physical manifestation. With streaming, users may watch a variety of videos at once, sample them before choosing one, and Netflix has far more visibility into audience data. Thus, in the streaming paradigm, a lot more data gets dematerialized. Furthermore, the number and types of data sources have increased, and now include catalog data, search terms, queues, and plays for streams, interactions, and external sources like movie reviews and social media data. Moreover, the streaming model's increasing reliance on recommendations has enhanced the presence of liquefaction. By eliminating time and distance from the business model, more opportunities exist for interaction between service provider and subscriber. These interactions can take the form of dynamic personalization, explanation of content to foster trust, rating, ranking, and reviews,

as well as social influence resulting from what connected friends have viewed or rated (Lycett, 2013).

Datafication has provided the emphasis of discussion, with the ideas of dematerialization, liquification, and density being put up as the cornerstones of comprehending datafication and analytics serving as a crucial method of generating value. We have argued that datafication is a sense-making process powered by information technology while developing this story. Plausibility and enactment, two components of sense-making, stand out in relation to the objectives of this work. Accounts provide perceived realities a tangible form that may be put into action by creating order among groups of entities and giving plausibility to them. It is supposed to be clear from the few instances of conceptualization, algorithmic treatment, and re-representation problems that accounts are closed as datafication stands (Lycett, 2013).

Algorithms

“The dictionary definition of an algorithm is"a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer” (Andrews, 2019, page).

Data must be processed in order to be useful because they are useless on their own. Traditional business intelligence tools started to reveal their limitations when it came to processing data with high velocity, quantities, and types, even though built infrastructures were able to store and retrieve enormous amounts of data. In order to extract knowledge from a massive amount of data, sophisticated "algorithms" with advanced features were required. Machine learning algorithms, a product of artificial intelligence, presented a highly promising approach in this particular field (Adadi, 2021).

While building useful analytic platforms, several machine learning and data mining algorithms have become accessible. Whatever algorithms are employed to separate and process the given information will depend on the goals that have been established. Several algorithms have been created expressly to address business-related issues. Additional algorithms were created to improve on already existing ones or to function differently. Some

algorithms will be more suitable than others, according to data scientists. There are many different algorithms available. They have the ability to do everything from recognizing faces to remind customers they have appointments (Foote, 2016).

The form that algorithms take depends on what they are used for. Comparing data using various methods can yield some unexpected findings about the data being used. A manager will gain a deeper understanding of business problems and solutions by making these comparisons. They may be presented as a set of hypothetical situations, a sophisticated mathematical analysis, or even a decision tree. Certain models only work well with specific types of data and analysis. For instance, problems like a loan applicant with a high likelihood of defaulting can be filtered out using classification algorithms and decision rules.

An organization's dataset can be utilized to uncover relationships using unsupervised clustering methods. These algorithms can be used to identify various consumer groupings or to choose which clients and services can be grouped together (Foote, 2016).

Algorithms play a crucial role in our daily lives. They make it possible for us to do efficient internet searches, discover new works of literature, cinema, and music, and they can be extremely accurate in the early detection of some diseases. In making tax judgments, managing crowds, conducting police investigations, and spotting social security fraud, government agencies use algorithms. Companies apply algorithms for setting prices and selecting employees. In short, algorithms are excellent tools, and we may be lost without them in the modern world (Gerards, 2019).

Algorithms are used for a variety of purposes, from simple suggestions for online search results or social media friends that "you may know" to more serious tasks like assisting doctors in determining your risk of developing cancer, determining if you qualify for a mortgage, or forecasting crimes like gang violence and burglary. You might not have even realized how much our lives are affected by algorithms because it is so difficult to observe. Predictive modeling by algorithms underlies the material we consume on Facebook, Spotify, and Netflix, as well as the music we listen to. The algorithms learn to tailor their "feed" and the marketing for our best possible experience with each click. These examples are employed in the public sector, including in areas like health care, education, criminal justice, and tax administration, even though many of them are motivated by business interests. New methods

of data analysis are increasingly being used by public organizations to enhance public services. We discover the formulas judges use to determine whether a criminal defendant is likely to commit another crime or not. Municipalities utilize algorithms to decide on the best routes for waste pickup. We come upon the algorithms that teachers use to place students in schools. The story continues to help us and algorithms are here to stay. The notion is that algorithms can help with today's biggest problems, including security, government services, healthcare, and protecting the environment (Schuilenburg & Peeters, 2020).

Nevertheless, we have learned to acknowledge that algorithms have some drawbacks over time. They frequently have a dark side, and fundamental rights attorneys in particular are quite interested in them because of this. The three primary, interconnected qualities of algorithms are what make them relevant for the law. In a nutshell, algorithms are non-transparent, non-neutral and human constructs (Gerards, 2019).

The fact that algorithms are designed, coded, learned, and applied by people makes them human creations. That is not a problem because it simply means that we have a lot of control over how they operate and are applied. It also suggests that errors in human reasoning and decision-making may have an impact on algorithms and how they are used. This contributes to algorithms' second characteristic, which is that they are non-neutral. After all, the people in charge of creating and using algorithms all have biases and hold certain values or opinions. Even if this influence is entirely unintentional, it nonetheless affects the decisions humans make while developing an algorithm. As a result, it is almost certain that non-neutral, subjective aspects may appear at different points during the algorithmic decision-making process (Gerards, 2019).

As a consequence, this is directly tied to algorithms' lack of transparency, which is their third characteristic. Some algorithms might be rather straightforward in that they function as a sort of "recipe" that lists several components and decision-making steps and includes options that, given enough effort, we can see and comprehend. Many algorithms, however, are far more complicated, especially if they are built using deep learning or machine learning approaches (Gerards, 2019).

We are going through a major shift in our lives right now, and while it is an outstanding time to live in, we also need to be aware of the side effects that come with it. A stunning quantity of data is generated about our daily activities, and algorithms process and act on this data to monitor, control, and influence our behavior in daily life. In addition to increasing the scope of current control and surveillance, the employment of algorithms brings in a new paradigm characterized by a rise in the rationality of governance, a change in the way power is exercised, and the closure of decision-making processes. We can refer to this as "algorithmic governance," which is the substitution of "black box" algorithms for human, readable, accountable judgments, or as sociologist Aneesh has coined it, "algocracy.". The phrase "algorithmic governance" refers to three different but connected components that have a significant impact on how we behave. While decision-making automation is not especially new, the impact of algorithms on automation, architecture, and anticipatory applications is becoming increasingly systemic (Schuilenburg & Peeters, 2020).

The range of organizational practices based on algorithms has significantly increased in recent years. Perhaps most importantly, algorithms are being employed for more than simply automating internal processes; they are also essential to emerging social governance models. Algorithms are employed to forecast, influence, or constrain human behavior. Via scores, rankings, profiles, and patterns, they accomplish this. The "surveillance capitalism" that supports the digital world and is used by credit card corporations, e-commerce companies, and social media platforms like Facebook and Google is based on the monitoring and forecasting of customer behavior(Schuilenburg & Peeters, 2020).

The operation of both public and private organizations depends on algorithmic governance. Police departments, for instance, employ them to forecast the locations, timing, and suspects of crimes. Algorithms are utilized in criminal justice to foresee a defendant's or convict's potential danger. The consumer audiences are analyzed by marketers using algorithms from online search queries, credit card purchase data, and behavioral data. Algorithms are being used by government agencies to, among other things, spot welfare fraud, provide public services, distribute regulatory oversight resources, and evaluate child safety risk (Schuilenburg & Peeters, 2020).

Many computer scientists have the most ambitious belief that the algorithmic society is merely a stepping stone to the emergence of a new social species, a general AI that would match or surpass human intelligence and maybe challenge human control. According to Moore's law, computer power tends to increase rapidly (Burrell & Fourcade, 2021). If we assume that processing power is the foundation of intelligence and the ability to increase it, then this law indicates a clear trajectory (Burrell & Fourcade, 2021). However, the question of who will gain most from AI's trajectory in society is more important than whether mankind will. A new division in learning pits the knowledgeable against the ignorant; those who make AI function against those who use AI to their own ends. The data capitalists may be able to correct, regulate, or improve how their personal data is portrayed, as well as completely opt out of monitoring systems and enjoy new benefits from AI, in contrast to the vast majority of individuals who are misrepresented, misunderstood, and alienated. Recent talks and actual investments in the development of neural implants, intelligent devices, and data-intensive genetic engineering have revived old cyber imaginations, releasing potential. In such a world, the members of the coding elite would be in the best position to not only expand their material power by claiming ownership over novel data outputs, but also to use new technology to literally enhance their own minds and bodies.

Algorithmic Bias. Commercial big data is gathered from people who have the resources, expertise, and interest to use particular digital platforms and gadgets. Big data may still reflect populations even though it is gathered in enormous volumes. These may be, for example, on average younger or more physically active than ordinary individuals because those included in a huge data sample tend to reflect primarily those using an expensive or innovative technical gadget or service. Selection sampling bias, also known as population bias, results from this. Such bias suggests that generalizing statements based on big data, which are frequently highlighted with reference to the popularity of digital devices or platforms, should be handled with caution: the more exclusive a technology or platform, the higher the likelihood for population bias. However, even though platforms utilized for scientific research are designed with basic sampling biases, they are rarely taken into consideration. (Richterich, 2018).

Since algorithms influence the types of content that users interact with and how data is produced, systematic biases become more common. Software and algorithm studies

academics have long argued that it is important to take into account the agency of such immaterial technological components. These considerations are also applicable to the data generated by interactions between users, algorithms, software, platforms, and their possible corporate providers. These types of bias are particularly problematic since it is frequently challenging to acquire the necessary algorithms, in part because of proprietary claims and because they are constantly changing for commercially motivated reasons (Richterich, 2018).

As was previously stated, these attempts to downplay the importance of informed consent miss the fact that it protects more than just one's bodily integrity as it also protects one's sense of autonomy and dignity. An important discursive function of what Mosco refers to as "digital positivism" in this context is that it takes the form of claims that biases associated with more conventional data collection have been overcome. Big data, however, actually brings about a complicated tangle of new algorithmic biases from humans (Richterich, 2018).

Artificial Intelligence

Although there has been more than 50 years of AI research, interest in the field has grown recently. From the realm of computer science, this extremely complex field was born. The original definition of AI was given by John McCarthy and his colleagues in 1955, who stated that it was the process of "making a machine behave in ways that would be called intelligent if a human were so behaving.". Other researchers define AI as "a growing resource of interactive, autonomous, and often self-learning agency, that can deal with tasks that would otherwise require human intelligence and intervention to be successfully completed," (Kavanagh, 2019).

Natural language processing, machine inference, statistical machine learning, and robotics are just a few of the many subfields that fall under the umbrella of AI. Deep machine learning and machine inference are two examples of certain subfields that are frequently viewed as places along a continuum where progressively fewer humans are needed in sophisticated decision making. Eventually, according to some observers, this will result in artificial general intelligence or super intelligence that equals or surpasses human intelligence. However, as

mentioned earlier, whether AI will ever genuinely match or surpass such cognitive and abstract decision-making powers is a highly debated topic (Kavanagh, 2019).

It is anticipated that developments in the many AI subfields would have a significant positive impact on society and the economy. A few of the fields that have already benefited from advances in AI include communications, healthcare, disease management, education, agriculture, transportation, space exploration, science, and entertainment. However, these advantages could just as readily be eliminated by the dangers associated with the study, development, and application of these technologies (Kavanagh, 2019).

The immediate risks and difficulties include the spread of current cybersecurity threats and vulnerabilities into increasingly crucial AI-dependent systems (like cloud computing); unintended or intended consequences as AI converges with other technologies, including those in the biotech and nuclear domains; algorithmic biases and discrimination; a lack of transparency and accountability in AI decision-making processes; excessively limited approaches to conceptualizing ethical problems; and limitations in the current state of knowledge. In the meanwhile, assumptions about how automation will change various sectors of the economy, the labor force, and the existing structures of social and economic organization are the focus of policymakers. Anxiety has been raised by predictions that enhanced machine learning and automation may worsen economic inequality. Many studies address these worries while also emphasizing and predicting risk on topics like the future of labor, the future of food, and even the future of human society (Kavanagh, 2019).

Large amounts of online data that are gathered, saved, and processed are the source of much of the power used by various AI applications and models. An increasing number of people are worried about data security, privacy, and other principles and values like fairness and equality, autonomy, openness, and responsibility. AI applications' dual-purpose nature makes it challenging to limit their development and control their use (Kavanagh, 2019). Furthermore, AI as a tool for projecting national power in geopolitical, military, economic, and normative terms; this development and use of such technologies may be complicated by increasing strategic competition (Kavanagh, 2019).

In conclusion, additional developments in AI are anticipated to have a considerable impact on how economics, socio-political life, geopolitical competition, and conflict are shaped. Some

analysts claim that technology might possibly cause an existential danger. However, there is still time to consider AI carefully and create a stronger defense against the dangers that lie ahead of us.

5. Surveillance capitalism

According to Shoshana Zuboff, “Invented at Google and elaborated at Facebook in the online milieu of targeted advertising, surveillance capitalism embodies a new logic of accumulation. Like an invasive species with no natural predators, its financial prowess quickly overwhelmed the networked sphere, grossly disfiguring the earlier dream of digital technology as an empowering and emancipatory force. Surveillance capitalism can no longer be identified with individual companies or even with the behemoth information sector. This mutation quickly spread from Silicon Valley to every economic sector, as its success birthed a burgeoning surveillance-based economic order that now extends across a vast and varied range of products and services” (Zuboff, 2019, page).

Simply put, the basis of surveillance capitalism is the exploitation and monetization of personal information as a "new form of information capitalism which aims to predict and modify human behavior as a means to produce revenue and market control”. Scientists indicate that "effective exile from one's own behavior while creating new markets of behavioral prediction and modification" is what surveillance capitalism does. It is supported by organizational usage of behavioral data that results in power and knowledge imbalances. As a result, consumers frequently may not be aware of how much they are reacting to indications that are motivated by commercial goals (Stahl, 2023).

Researchers demonstrate that surveillance capitalists use segmentation (targeting based on behavior, attitudes, and choices), deception (by making false claims and substituting false or inaccurate beliefs for existing ones), and dominance (such as Google's dominance in internet searches and the monopolization of the market through mergers and acquisitions). They list three reasons why businesses keep collecting and keeping personal data: first, they compete

with people's ability to resist persuasion tactics once they become aware of them, and with competitors' abilities to target their customers; second, the organization can acquire an asset for future sales due to the low cost of data aggregation and storage; and third, the penalties for data leaks are minimal (Stahl, 2023).

The emergence of the Internet and ways to access the World Wide Web led to the global ubiquity of computer mediation, both at the institutional interface and in the private spheres of daily experience, moving it beyond constrained places of employment and specialized action. These findings caused high tech companies, led by Google, to see new business potential. Google recognized that these data may significantly impact the value of advertising if it were to collect more of them, store them, and analyze them. Google has built increasingly ambitious procedures that extend the data lens from past virtual conduct to current and future actual behavior as its capabilities in this area have grown and received an unprecedented amount of profit. Thus, a new global architecture of data collection and analysis that generates rewards and penalties intended to change and commercialize behavior for profit is associated with new monetization prospects. (Zuboff, 2015).

Many of the actions done in order to take use of these allegedly new opportunities to profit from them allegedly violated legal and constitutional rights and social norms relating to privacy. As a result, Google and other actors discovered how to cover up their actions. They now choose to encroach upon undefended personal and social space until they are met with resistance, at which point they can make use of their vast resources to reasonably defend what they have already seized. In this approach, surveillance assets are gathered, draw a lot of funding for surveillance, and create unanticipated new social and political dynamics (Zuboff, 2015).

There are several reasons why these new institutional facts have been accepted, including the fact that they have been developed rapidly and designed not to be seen. Few people outside of a small group of professionals knew what they meant. It was impossible for individuals to gain information about these practices due to structural imbalances in knowledge and rights. Leading tech firms were regarded with respect and treated as futuristic messengers. There were few protective defensive barriers because no prior experience had prepared people for these new patterns. People quickly learned to rely on the new communication and informational capabilities as essential tools in the increasingly demanding, competitive, and

stratified desire for an effective life. Thus, the new media, platforms, networks, apps, and technologies became necessary for social interaction. In the end, an overwhelming sense of inevitability was created by the rapid growth of institutionalized facts, including data brokerage, data analytics, data mining, professional specializations, unbelievable cash flows, strong network effects, state collaboration, hyperscale material assets, and unprecedented concentrations of information power (Zuboff, 2015).

These innovations served as the foundation for a newly developed accumulation logic that has been fully institutionalized and is known as "surveillance capitalism." A global architecture of computer mediation in this new system transforms the electronic text of the confined organization into the Big Other, an intelligent organism that spans the entire universe. This inventive institutional logic thrives on unanticipated and unreadable processes of extraction and control that isolate people from their own behavior, creating new opportunities for subjection (Zuboff, 2015).

“Surveillance capitalists... want... your bloodstream and your bed, your breakfast conversation, your commute, your run, your refrigerator, your parking space, your living room, your pancreas” (Zuboff, 2019, page).

Both the public and private sectors are affected by surveillance capitalism, since both see large amounts of personal data being collected and created under a variety of false pretenses. The marketplace, social media, and entertainment industries are some examples of the private sector. Google and Facebook's efforts to strengthen their systems and services are the most frequently cited instances of surveillance capitalism. The healthcare and retail industries in the public sector are especially famous examples of this. Health-related surveillance capitalism has come to light as a result of the COVID-19 epidemic. For instance, it has been argued that the COVID-19 pandemic was a time when telehealth technology was pushed too swiftly (Stahl, 2023).

The development or acquisition of AI solutions has further increased the power of Big Tech, which has been associated with that of nation states. A few large tech companies have a disproportionate amount of power, and because of their influence over political decision-making, market manipulation, and digital lives, they are disrupting economic processes and jeopardizing democracy, individual freedoms, and political and social life.

Privacy and data protection are two other major ethical issues that these instances have brought to the fore. Human autonomy and well-being depend on privacy, which also enables people to defend themselves against intrusion into their life. For instance, employers or health insurance may exploit compromised personal health information against the interests of the affected person. Data must be treated lawfully, fairly, and transparently, have a specific purpose, be accurate, and only be kept for a short period of time in order to be protected. Additionally, it demands that such processing conform to the values of responsibility, confidentiality, and integrity (Stahl, 2023).

Data appropriation, data monetization, and unjust business practices all have a connection to a lack of transparency and explainability. While following transparency requirements by businesses that obtain personal data may appear straightforward from a data protection and societal point of view, this need comes across difficulties. The organization and management of the data business contribute to issues with transparency. The use of transparency principles in public relations campaigns by data brokers to weaken government regulation also poses a challenge to transparency. Additionally, scientists point out that transparency may merely give an image of reform and not address underlying imbalances in power (Stahl, 2023).

The spread of market mechanisms into spheres of life that were previously off-limits to financial transactions appears to be a major factor in the perception of surveillance capitalism as being unethically problematic. This has a slight connection to the idea that the data producers are being taken advantage of. Many consumers of "free" internet services are happy to use social networking or online productivity tools in exchange for the program developers using their data. Although there is a feeling of unfairness, the fact that the service providers have made enormous financial gains without sharing them with the people whose data they utilized to do so also raises concerns. Additionally, it appears that some aspects of social life ought to be exempt from market trade, which is the basis for critique of surveillance capitalism. Use of the word "friend" in social media, where friendships differ greatly from those in the offline world and where the number of followers and friends can result in financial transactions that are considered improper in the offline world, may be an example of this (Stahl, 2023).

The core ethical problem of surveillance capitalism is not one that can be easily identified. The phrase should be interpreted as indicating opposition to technologically assisted societal developments that distribute economic and political power to a small group of well-known institutions.

As Zuboff indicates it "The automated ubiquitous architecture of Big Other, its derivation in surveillance assets, and its function as pervasive surveillance, highlights other surprising new features of this logic of accumulation". As it constructs the company as officially indifferent to and radically detached from its populations, it weakens the historical link between markets and democracies. The conventional reciprocities, in which populations and capitalists depended on one another for employment and consumption, do not apply under surveillance capitalism. Populations are the objects of data extraction in this new model. Another aspect of surveillance capitalism's anti-democratic nature is this severe detachment from the social. Democracy no longer serves as a tool for prosperity under surveillance capitalism; instead, it threatens surveillance profits (Zuboff, 2015).

Confronting Surveillance Capitalism: Responses

Concerns about surveillance capitalism have prompted many different types of solutions, including market-based, societal, and legal or policy-based ones. Antitrust laws, intergovernmental regulations, enhancing consumer or individual data ownership rights, socializing the ownership of developing technologies, requiring large tech companies to use their monopoly profits for governance, mandatory disclosure frameworks, and increased data sharing and access are just a few examples of legal and policy measures (Stahl, 2023).

Market-based responses include valuing the data collected by surveillance capitalists, reducing monopolies, defunding Big Tech, refunding community-oriented services, and users using their market power to reject and avoid businesses that they believe to be acting unethically. Public outrage, personal data spaces or newly developing intermediary services that provide consumers choice over the sharing and use of their data, raising data literacy and understanding of how open a company's data policy is, and enhancing consumer education are some examples of societal responses (Stahl, 2023).

In this part, we will examine three possible solutions. While none of these are answers on its own, they do offer effective methods to reduce the effects of surveillance capitalism and the ethical issues that have been researched in various ways. The impact of these answers and their viability for implementation in the current socioeconomic system have been examined in academic and policy literature. The socio-political context in which AI is produced and employed contributes to the problems of surveillance capitalism, and this context informs the solutions we have highlighted here (Stahl, 2023).

a. Antitrust Regulation. Antitrust refers to practices that limit monopolies, prevent businesses from working together to unfairly set prices, and promote equal commercial competition. Antitrust laws control monopolistic behavior and restrict mergers and illegal corporate practices. Courts examine the validity of mergers case by case. Numerous appeals and suggestions have been made to reduce the influence of Big Tech. There have been several discussions on the use of antitrust laws to break up large tech firms and the appointment of regulators to annul illegal and anti-competitive tech mergers (Stahl, 2023).

Big Tech is viewed negatively because of its concentration of power and control over the economy, society, and democracy, which harms small business competition and innovation. To ensure that we receive the benefits of a data-driven economy while minimizing its dangers, Grunes and Stucke stress the need for competition and antitrust laws. However, using antitrust remedies to regulate dominant companies has drawbacks, including decreasing the incentives for innovation and information exchange, raising privacy issues, and causing platform providers to become hesitant and stagnant (Stahl, 2023).

The employment of antitrust regulations as a tool for minimizing the influence of Big Tech has both benefits and drawbacks. Delaying or irritating acquisitions, improving visibility, openness, and accountability, pressuring Big Tech to reform its operations, and improving prospects for small enterprises are some of the benefits. One drawback is that some Big Tech companies continue to gain power and dominance in spite of the antitrust measures that have already been taken. The difficulties authorities face in enforcing antitrust laws are still another drawback. Antitrust lawsuits are rather costly, they cause economic disruption, and may have a negative impact on innovation (Stahl, 2023).

Nowadays, recent developments in the USA and Europe (legislative measures, acquisition challenges, legal actions, and fines) demonstrate that Big Tech's influence is subject to more scrutiny than ever (Stahl, 2023).

b. Data Sharing and Access. Increased data access and sharing is another answer to concerns about surveillance capitalism. (subject to legal safeguards and restrictions). It is proposed that one way to more effectively handle the restrictions of antitrust law is to make data open and freely available in a highly regulated environment. Similar to this, scientists argue that mandates requiring data sharing (securely enforced through privacy-enhancing technologies) "have become an essential prerequisite for competition and innovation to thrive" and that, in order to challenge the "monopolistic power derived from data, Big Tech should share what they know and make this information widely usable for current and potential competitors" (Stahl, 2023).

c. Strengthening of Data Ownership Claims of Consumers / Individuals. Strengthening consumer and individual data ownership claims is another way to deal with surveillance capitalism. Even though user-held data is intangible, according to some, it satisfies all of the criteria for an "asset" under property laws because it is "specifically defined, has independent economic value to the individual, and can be freely alienated" (Stahl, 2023).

Others believe that although data ownership is still an important notion for defining rights and obligations, it should be reviewed in the light of big data and analytics. They distinguish between three different types of data ownership; data, data platform, and data product ownership which could serve as a guide for defining governance processes and the foundation for deeper data governance roles and frameworks. According to researchers, arguments for data ownership share a commonality that refers to methods of regulating how data is used and the capacity to direct, restrict, and assist the flow of data. Additionally, they contend that ownership has proven to be a double-edged sword in terms of the marketization and commodification of data, and that applying this idea necessitates consideration of the best ways for data subjects to secure their personal information and share it responsibly. Furthermore, they state that "even if legal frameworks forbid true ownership in data, there is still room for debate as to whether they can and should permit such forms of quasi-ownership" (Stahl, 2023).

The vagueness of the ownership idea, the complexity of the data value cycle, the involvement of several stakeholders, as well as the difficulty in defining who may or would be entitled to claim ownership in data, are obstacles that affect this response.

Chapter 2: Privacy in the context of Big Data

1. Big data and their use

As we have mentioned earlier, data currently pours into every sector of the global economy like a flood. A growing amount of transactional data is produced by businesses every day, containing trillions of bytes of information about their clients, suppliers, and daily activities. In the internet of things era, millions of networked sensors are being implanted into the real world in objects like mobile phones, smart energy meters, cars, and industrial machinery that perceive, create, and share data. In fact, as businesses and organizations conduct their operations and engage with people, they produce a vast amount of digital "exhaust data". The availability of social media platforms, mobile devices, and other consumer electronics like PCs and laptops has allowed billions of people around the world to contribute to the amount of Big Data that is available, and the exponential growth of Big Data has been greatly influenced by the volume of multimedia content that is being produced (Ademola, 2015).

Like any other technology, big data has benefits and drawbacks of its own. When it comes to practical applications, there are occasions when big data's drawbacks outweigh some of its benefits. Therefore, sectors and companies must weigh the benefits and drawbacks of big data before using it. Now, let's discuss its potential advantages and disadvantages.

a. Potential benefits of Big Data

Big Data and Healthcare

With the systematic collecting, storage, processing, and analysis of massive amounts of data, data has become an omnipresent concept in our daily lives. This trait is cross-disciplinary, extending from machine learning and engineering to economics and medicine (Pastorino, 2019).

Over the last few decades, there has been increasing excitement about the potential utility of these huge amounts of data, big data, in improving personal care, clinical care, and public health (Pastorino, 2019).

The intricacy of Big Data analysis stems from the combination of several sorts of information that are electronically captured. In recent years, there has been an explosion of new platforms, tools, and approaches for storing and organizing such data, as well as an increase in publications on Big Data and Health. We may currently collect data from electronic healthcare records, social media, patient summaries, genomic and pharmacological data, clinical trials, telemedicine, mobile apps, sensors, and data on well-being, behavior, and socioeconomic indicators. As a result, healthcare practitioners can profit from a massive amount of data (Pastorino, 2019).

Beginning with the collection of individual data elements and progressing to the fusion of heterogeneous data from various sources, can reveal entirely new approaches to improving health by providing insights into disease causes and outcomes, better drug targets for precision medicine, and improved disease prediction and prevention. In light of this, the opportunities and potential for the benefit of patients and, in general, the healthcare system are significant (Pastorino, 2019).

We will be examining big data analytics in healthcare in six applications: disease surveillance, health care management and administration, privacy protection and fraud detection, mental health, public health, and pharmacovigilance.

Disease Surveillance

Disease Surveillance involves understanding the disease's status, origin, and prevention. The information received through EHRs (Electronic Health Record), and the Internet has enormous potential for disease analysis. The numerous surveillance approaches would aid in service planning, treatment evaluation, priority setting, and the development of health policy and practice (Somani, 2022).

Image processing of healthcare data from the big data point of view. Image processing on healthcare data provides vital information about anatomy and organ function, as well as

identifying disease and patient health situations. The approach is being employed for organ delineation, lung tumor identification, spinal deformity diagnosis, arterial stenosis detection, and aneurysm detection. Wavelets are frequently used in image processing techniques such as segmentation, enhancement, and noise reduction. The use of artificial intelligence in image processing is said to improve areas of health care such as screening, diagnosis, and prognosis, and integrating medical images with other forms of data and genomic data will improve accuracy and optimize the illnesses early detection. A better utilization of clinical settings for computer-based healthcare diagnostic and decision-making systems has resulted from the exponential growth in the number of medical institutions and patients (Somani, 2022).

Data from wearable technology. Multinational corporations such as Apple and Google are developing health-related apps and wearable technologies as part of a broader spectrum of electronic sensors, known as the Internet of Things, and toolkits for healthcare-related apps. The ability to collect medical data in real-time (e.g. diet followed, exercise, and sleep cycle patterns), linked to physiological indicators (e.g., heart rate, calories burned, blood glucose level, cortisol levels), is perhaps discrete and omnipresent at low cost, unrelated to traditional health care. "True Colors" is a wearable device meant to collect continuous patient-centric data while also providing the accessibility and acceptability required for accurate longitudinal follow-up. More importantly, this technology is currently being tested as a replacement for daily health monitoring (Somani, 2022).

Medical Signal Analytics. The usage of continuous waveform in health records comprising information provided by statistical disciplines (e.g., statistical, quantitative, contextual, cognitive, predictive, and so on) might influence comprehensive care decision-making. A data collection platform that can control a series of waveforms at varied fidelity rates is required in addition to an ingestion-streaming platform. The combination of this waveform data with the static data from the EHR results in an important component for providing situational and contextual knowledge to the analytics engine. Improving the data obtained by analytics will not only make the process more reliable but will also aid in balancing the sensitivity and specificity of predictive analytics. The signal processing species must rely primarily on the type of disease population being observed (Somani, 2022).

A pre-trained machine-learning model can use a variety of signal-processing techniques to extract several target features, which are then used to deliver actionable insight. These

findings could be analytical, prognosticative, or both. Additionally, such insights can be designed to activate different mechanisms like alerts and doctor notifications. It can be challenging to keep these continuous waveforms-based data and specific data from the other sources in perfect harmony to identify the right patient information to advance diagnosis and therapy for the following generation. For the bedside application of these systems into medical setups, a number of technological criteria and specifications at the framework, analytical, and clinical levels need to be planned and implemented (Somani, 2022).

Healthcare administration

Healthcare practitioners now have access to insights made possible by big data analysis. Data warehousing, cloud computing, patient management, and other areas of healthcare have all benefited from the application of data mining techniques by researchers.

Data storage and cloud computing. In order to improve medical outcomes, data warehousing and cloud storage are largely utilized to safely and affordably store the growing volume of digitized patient-centric data. Data storage is used for research, training, teaching, and quality control in addition to medical applications. In accordance with the set patient privacy policy, users can also extract files using keywords from a repository that contains radiological findings (Somani, 2022).

Patient Data Management. Effective scheduling and patient care delivery throughout a patient's hospital stay are two aspects of patient data management. Electronic medical records are used to design a logistical regression model that estimates the likelihood that patients won't show up for appointments. The model also demonstrates how estimations may be used to create clinical schedules that make the best use of the available clinical capacity with the least amount of waiting time and clinical extra time (Somani, 2022).

More patients can be seen each day, improving clinical performance, if patient no-show models are combined with sophisticated programming techniques. Clinics should take into account the benefits of using planning software, including specific methodologies, when estimating no-show costs (Somani, 2022).

Privacy of medical data and fraudulency detection

It is essential to anonymize patient data, protect medical data privacy, and identify fraud in the healthcare industry. Data scientists must work hard to safeguard big data from hackers as an outcome of this. Researchers brought about the issues with privacy security and presented a special anonymization technique that functions for both distributed and centralized anonymization. The researchers also suggested a model that outperformed the conventional K-anonymization model, in terms of protecting data usefulness without sacrificing any data privacy. Additionally, their system was able to handle large, multi-dimensional datasets. We will be focusing more on data protection techniques in the following chapters (Somani, 2022).

To address the issues with current medical records systems, a mobile-based cloud computing architecture for big data has been developed. EHR data systems are limited by the amount of data, lack of interoperability, and privacy concerns. This innovative cloud-based system intended to store EHR data from various healthcare providers inside the physical space of an internet provider to grant healthcare providers and patients permitted restricted access. To maintain the security of the data, they utilized 2-factor authentication, One Time Password (OTP), and algorithms for encryption (Somani, 2022).

Google's effective technologies, such as MapReduce and big query tools, can be used to analyze big data. Compared to traditional ways for anonymization, this strategy will lower costs, increase efficiency, and ensure data protection. The traditional method typically exposes data to re-identification. In a case study, scientists demonstrated how hacking may connect discrete pieces of data and identify patients. Big data analytics are quite effective for detecting fraud and abuse (i.e., suspicious care behavior, intentional misrepresentation of facts, and unwanted repeated visits) (Somani, 2022).

Yang et al. constructed a method that manually separates characteristics of suspect specimens from a set of treatment regimens that the majority of doctors would choose using data from gynecology-based reports. The method was applied to data from Taiwan's Bureau of National Health Insurance, where it was able to identify 69% of all cases as fraudulent, outperforming the previous model's detection rate of 63%. In conclusion, due to the expanding use of social

media technologies and the inclination of users to post personal information on these platforms, the protection of patient data and the identification of fraud are of major concern. Since a significant amount of everyone's personal information is now accessible through these platforms, the previously successful solutions for anonymizing the data may become less effective if they are not put into practice (Somani, 2022).

Mental Health

52.2% of the US population have been affected by either mental health issues or drug addiction/abuse, according to a national survey on drug use and health. Additionally, some 30 million people experience anxiety disorders and panic attacks (Somani, 2022).

In order to assist medical professionals in managing patients with anxiety problems, scientists created a therapeutic method that focuses on data analysis. The authors created static and dynamic information based on user models using both static and dynamic data. Static data included personal information such as the user's age, sex, body type, and family information. Dynamic data featured stress context, weather, and symptom information. The remaining parameter was mostly used to forecast stress rates under various scenarios for the first three services, where correlations between numerous complex parameters had already been established. Data gathered from 27 participants who were chosen through the anxiety evaluation survey helped to verify this model. As opposed to utilizing analytics to predict cancer or diabetes, applying data analytics to the disease diagnosis, examination, or treatment of patients with mental health is extremely different. The data context (static, dynamic, or non-observable environment) appears to be more significant in this situation than the data volume (Somani, 2022).

Public Health

Data analytics have also been used to track down illness outbreaks. Scientists looked at online records based on media reports of public-affecting elements, consumer behavior patterns, and expert trends in illness outbreaks. They discovered significant elements influencing the search behaviors of experts and laypeople inside public health organizations, along with suggestions for focused communications during emergencies and outbreaks. Body area networks, an IoT-based wireless network of wearable devices, have been proposed as an

emergency response unit. The system is built with "intelligent construction," a model that aids in processing and decision-making using the sensor data. (Somani, 2022).

Online consultations are becoming more popular and could be a solution to the shortage of healthcare resources and ineffective resource delivery. However, many online consultation platforms fail to draw paying clients, and healthcare professionals on these platforms face the additional issue of standing out from a huge number of doctors who can offer comparable services. In a study, scientists used machine learning techniques to analyze vast amounts of service data in order to (1) identify the key characteristics associated with patient payments rather than free trial appointments, (2) examine the relative weight of those features, and (3) comprehend whether or not these attributes affect payment in a linear or nonlinear way (Somani, 2022).

A total of 1,582,564 consultation papers between patient pairs were included in the dataset from 2009 to 2018 from the largest online medical consultation platform in China. The findings demonstrated that, in comparison to features relating to a physician's reputation, service-related features such as service quality (e.g., consultation dialogue intensity and response rate), patient source (e.g., online vs. offline patients), and patient involvement (e.g., social returns and previous treatments revealed), were more important. Promoting numerous immediate responses during patient-provider contacts is crucial for making payments easier (Somani, 2022).

Pharmacovigilance

ADRs (Alternative Dispute Resolution) are rare in traditional pharmacovigilance, however it is possible to identify a link between a medicine and any conceivable ADRs and receive false signals and incorrect results. Because there is already a list of potential ADRs that can be very helpful in potential pharmacovigilance efforts, these false alarms can be avoided (Somani, 2022).

Overcoming the language barrier. The ability to analyze and compare the prevalence of disease and available medical treatments across nations can be enhanced by the global sharing of electronic health records. However, each nation would use its own language to record statistics. Multilingual language models can be used to overcome this language barrier,

opening a variety of options for the growth of data science and helping to create a framework for service customization. The semantics, grammatical structure, and norms of the language, as well as the context and the general meaning of words in various settings, will all be accepted by these models (Somani, 2022).

Big Data and Smart Cities

"The smart city vision rests on the full utilization of information and communication technologies in general, and on data analytics more specifically" (Löfgren & Webster, 2020, page).

Big data is undoubtedly enhancing our understanding of how cities work, providing plenty of new chances for social engagement, and helping us make better decisions based on our understanding of how to interact in cities (Batty, 2013). Many cities are currently competing to become smart cities in an effort to benefit from their advantages in the economic, environmental, and social areas. Many people are therefore keen on the prospects presented by the use of big data analytics in smart city applications (Jaroodi, 2015).

Efficient use of sources

It is crucial for integrating solutions to have a better and more controlled utilization of these resources, as many are becoming either limited or very expensive. It will be helpful to start with technological systems like geographic information systems and enterprise resource planning. With monitoring systems in effect, it will be faster to identify waste spots and distribute resources more effectively while keeping costs under control and consuming less energy and natural resources. The fact that smart city applications are built for interconnection and data collecting, which can help improve collaboration across applications and services, is a further essential aspect of these applications (Jaroodi, 2015).

Improved standard of living

The quality of life for citizens of smart cities may also be higher thanks to improved services, more effective work and living arrangements, and reduced waste (of time and resources). Better planning of living and working settings, more effective transportation systems, better and quicker services, and the availability of sufficient information to make informed decisions will be the result (Jaroodi, 2015).

Increased openness and transparency

Higher interoperability and openness stem from the requirement for better management and control of the various smart city applications and elements. It will become common to share resources and data. The development of further services and apps that further improve the smart city will be aided by this, which will additionally increase cooperation and communication across organizations. In the interest of transparency and openness, the US government is one example that has gathered and made available a wide range of data, publications, and materials. These gave both the general public and the government organizations the possibility to successfully exchange and use data (Jaroodi, 2015).

Opportunities exist for achieving these advantages; but, doing so requires investing in additional technology, greater development efforts, and efficient big data usage. As well as implementing data documentation standards to provide guidance on the content and usage of the datasets, rules must be established to assure data accuracy, high quality, high security, privacy, and control. Technology may also be an excellent tool when managing and protecting natural resources, infrastructure, and environmental resources with the ultimate goal of achieving sustainability (Jaroodi, 2015).

Adopting ICT, Cloud, and big data solutions may assist in resolving several issues in the city, including supplying the tools for storage and analysis. Additionally, this will promote cooperation and communication between the various components of a smart city and assist in advancing to the innovation stage. Building big data communities that function as a single entity to stimulate collaborative and innovative solutions addressing applications for fields like education, health, energy, law, manufacturing, environment, and safety is one way to do

this. Since applications and systems are integrated and information can flow easily between applications and entities, this results in real-time solutions to problems in agriculture, transportation, and crowd management (Jaroodi, 2015).

Smart education

Using education smart services that are flexible to provide better use of information, enhanced control and assessment, and higher support for life-long learning for all people (citizens and stakeholders), ICT provides a solution to enhance the efficiency, effectiveness, and productivity of educational processes. People will be actively involved in their learning through the use of intelligent education applications, which will enable them to quickly adjust to societal and environmental changes. We will also have a good impact on knowledge levels and teaching/learning tools by depending on big data that is collected in the field and properly processed to provide the necessary information (Jaroodi, 2015).

Additionally, technology can make these opportunities accessible everywhere, even in remote or rural areas where people may not be able to commute to school or where their economic situation prevents them from affording other, more expensive models. Big data and ICT use will also contribute to the development of a knowledge-based society, strengthening the competitiveness of the country. Big data in the field of education is primarily created by gathering information on individuals such as students, teachers, parents, administrators, and other support staff, infrastructures as schools, libraries, computing facilities, educational locations, museums, universities, and other related entities, and information like courses, books, exams, grades, economic surveys, assessments, reports, and much more (Jaroodi, 2015).

With the help of this data, analysis and trend and model extraction, it will be possible to provide students with better and more advanced instruction. Big data, for instance, helps educational institutions to "create communities of practice and standardize the presentation of knowledge" and personalize learning. Big data in education can be used to monitor teacher shortages and improve lesson plans (Jaroodi, 2015).

Smart traffic lights

One of the key components of smart cities is effective traffic management, which will improve the city's transportation infrastructure, time spent commuting for residents, and traffic patterns in general. Traffic, pollution, and economic issues all arise as the population grows. As a result, one of the most crucial methods that smart cities employ to deal with heavy traffic and congestion is the installation of intelligent traffic lights and signals. To provide more data regarding traffic patterns, smart traffic lights and signals should be integrated across traffic grids. Each sensor measures a different aspect of the flow of traffic, such as vehicle speeds, vehicle density, time spent waiting at stoplights, and the presence of traffic jams (Jaroodi, 2015).

As a result, this system will be able to make better choices the more data that is available to it. Therefore, it will be best to gather data from all traffic lights throughout the city and create intelligent decision systems using this data in order to provide the best services for smart traffic lights. Big data analytics in real-time are necessary for this.
As an example, this was applied in the US. The Traffic21 project's smart traffic lights and signals in Pittsburgh, Pennsylvania achieved notable results by reducing traffic jams and time spent waiting, which resulted in a 20% reduction in emissions (Jaroodi, 2015).

Smart grid

A smart city's smart grid is a crucial element. It is a modernized electrical grid system that makes use of information and communication technology to automatically gather and add value to accessible data, such as data on the behaviors of suppliers and customers. It enhances the production and delivery of economy, sustainability, efficiency, and dependability. Through system self-monitoring and feedback, a smart grid uses computer-based remote controls with two-way communication technology between power producers and consumers to improve grid efficiency and reliability. In order to obtain accurate near real-time data about the present power production and consumption, this includes installing smart sensors and meters on production, transmission, and distribution systems in addition to consumer access points (Jaroodi, 2015).

Even if numerous smart city applications have shown that the technical obstacles can be resolved, the true difficulties lie in figuring out how to handle organizational diversity, governance, and legal constraints. In addition to local actors in the smart city, this is a concern for regional, national, and occasionally global authorities of government. Although it would take many years to build and establish, formal national and global legislation is expected to emerge in the longer period. Thus, the following basic minimum self-regulatory principles are required: (a) the quality standards of data to be used in the smart city; (b) the ethical standards regarding privacy and data protection; (c) a clearer policy regarding the ownership of unstructured and structured data; and (d) agreed standards regarding safety and protection of storing of data (Löfgren & Webster, 2020).

We examined a number of big data application examples that can serve as models for developing applications for smart cities. Many were successful to varying degrees, and many of them added beneficial components to improve the services and applications of smart cities. Now let us examine a different potential benefit of big data.

Big Data and SMEs

Big Data is regarded as one of the primary drivers of SME growth, according to the Oxford Economics Survey (2013), which highlighted technology and innovation as strategic priority. Big Data analysis and forecasting capabilities represent a paradigm shift for SMEs in terms of market and customer behavior. When properly implemented, it can result in enhanced productivity, responsiveness, anticipation, flexibility, and the capacity to meet consumer needs by identifying blind spots and making wiser decisions (Sen et al., 2016).

Data is considered to be a company's most significant asset in the modern world. Global businesses, from large to small and medium-sized, are looking into innovative ways for utilizing data. Big Data is used by a lot more than global corporations. Small-Medium Enterprises, SMEs can now profit from the vast amount of data to make fast, correct decisions that will enhance the functionality of their businesses. It is essential for SMEs to

give big data adoption some serious thought since, according to several scholars, it represents a paradigm change in how business processes might be improved (Butt, 2018).

Startups in big data and machine learning are frequently purchased by large corporations. However, SMEs can also use big data to more fully comprehend their customers, expand into new markets, and reduce unnecessary business costs in real-time (Butt, 2018).

Enhance Operations

Big data can help SMEs discover new aspects of their system by analyzing data and highlighting correlations, hazards, possibilities, etc. that they weren't previously able to see. The possibility for decision support system improvement arises from this. Additionally, it can aid SMEs in simulating various scenarios. As a result, it can enhance already-existing products and services or aid SMEs in creating new ones. However, decision-makers must initially choose which data to utilize and acquire, how to collect it, how to process it, etc. Additionally, they must be crystal clear about their goals and the questions they need to have resolved (Sen et al., 2016).

Understand the Rivals

Big data may help SMEs analyze their own data in-depth and can also provide competitive intelligence. The strong working correlation of SMEs is mostly to blame for the stated unavoidable truth. Consequently, a sizable collection of datasets can enable a deeper understanding of the current and potential status of the organization within the marketplace and can serve as the foundation for upcoming qualitative and quantitative research (Sen et al., 2016).

Learn about upcoming trends

To achieve a better degree of strategic management and strengthen organizations' positions in the competitive market by creating new opportunities, the deployment and use of big data may very well become the next stage of innovation in firms and SMEs alike (Sen et al., 2016).

Understanding customers comprehensively

It is now possible to track and assess consumer behavior using a variety of communication channels and internal data. For instance, all you need to do is examine their purchasing patterns and predict their future purchases. It's important to consider information published on social media.
When conducting big data analyses, knowing how to ask the right questions is also crucial. The answers to these queries can help you improve your current offerings or create your next top-selling product (*Big Data Benefits for Smes - Dataconomy*, 2022).

According to Thompson et al. (2013), a widely held opinion is that SMEs may grow faster if they are innovatively oriented. The firm's capacity to create information from novel technologies, such as the utilization of big data, is one of the factors determining the competitive advantage. In light of this, innovation can be achieved with the aid of technology leadership, innovative process R&D, and creativity. SMEs must therefore accept business-related risks and should not be afraid of failing in order to succeed in the future (Sen et al., 2016).

Last but not least, SMEs also have a tremendous chance to create novel or creative value propositions for their business models in the industrial ecosystem that is expanding around efforts focused on the use of Big Data for open innovation strategies (Del Vecchio et al., 2018).

Big data and education

In this part, we will go over various data usage patterns in schools. Teachers and administrators in conventional schools utilize student data to make judgments about how to teach, assign grades, provide credit, and produce transcripts. Virtual learning environments are becoming more engaging due to improved speed of the internet and cloud computing. The learner chooses from a variety of modular films, works on problem sets, and examines additional reference material at their own pace. Additionally, schools use these platforms in their curricula to offer "blended" learning experiences that combine in-person and online

learning, or "flip" instruction so that students watch lectures as homework and discuss their material in class (Zeide, 2017).

Technology-aided education

The effectiveness of administrative choices and educational interventions is studied using big data. Big data models can forecast the right times to intervene on behalf of students, such as when they stop participating in online courses. Following that, the detectors served as the foundation for interventions that increased student engagement (Fischer et al., 2020).

As students engage with digital platforms, technologically mediated education technologies produce a constant stream of information. To create student profiles, teachers and students provide details including usernames, emails, and grade levels. Additionally, through emails, online conversations, assignments, and examinations, learners give the standard academic data. Digital education tools gather a huge amount of data on students' performance and behavior throughout the learning process, data that could not be captured or evaluated at scale in physical classrooms. Time stamps, device identifiers, and even geolocation data are examples of metadata that fall into this category. These programs record when students log in, what portions of a video they actually watch, and how long they hesitate before responding to a question (Zeide, 2017).

Data-driven decisions

This vast amount of data is incorporated and analyzed by big data-driven educational technology to support instruction and institutional decision-making. Fewer stressed human teachers can offer a more accurate diagnosis with learning analytics. They use knowledge maps, which divide the relevant subject matter into ideas and "competencies," to monitor student development. Platforms, for instance, can detect that a student's poor chemistry grade was caused by their inability to understand a certain mathematical topic the year before, rather than any issues with the actual scientific concepts. These cognitive models may involve or imply emotional, cognitive, and academic growth (Zeide, 2017).

Most of today's educational technology translates this data for educators and displays it on digital dashboards. These systems involve varying degrees of interpretation and inference. "Skill meters" that graphically portray learners' mastery of particular subjects are displayed on some dashboards. Others summarize a complex variety of data to group students into manageable categories. One early warning technique, for instance, measures students' chances of passing through the use of red, yellow, or green indicators (Zeide, 2017).

In addition, studies on cognitive processes have concentrated on assisting and assessing students' cognitive abilities, knowledge, and skills as well as on assisting instructors (e.g., automated feedback for students, automated assignment grading). The capacity to automate assessments of student learning has grown recently, moving beyond multiple-choice formats to student writing samples. These studies often make use of reading comprehension data sets comprising hundreds of thousands of interactions as well as writing samples from hundreds or thousands of students. Numerous studies show that grading student writing can be automated to greatly minimize the amount of time that must be spent by humans (Fischer et al., 2020).

In order to promote their independent judgment, educators might use data-generated student assessments and estimations. "Stupid" tutoring systems can be designed to inform, rather than replace, "intelligent" human decision-making, as scientists note. A lot of systems in use today are primarily focused on finding and showing these patterns to instructors. For instance, alt schools (alternative schools) continuously observe classrooms to gather digital, audio, and video data regarding student interactions and behavior. In order to make sense of this data and yet determine what children need next, teachers mainly rely on computer analytics (Zeide, 2017).

Individualized platforms

Data-driven education systems, in contrast to data-informed decision-making, do not assist but rather replace human decisions. Instead, by automatically evaluating instructional possibilities in light of student profiles and presenting content accordingly, computers "personalize" learning. They attempt to imitate how teachers respond to the needs of their students in actual classroom environments by doing this (Zeide, 2017).

Computers have a variety of techniques to adapt instruction. Based on the reactions of the students, certain adaptive systems may replace content, such as a review of topics instead of practice problems. Others allow students to progress through the material at their own speed. For instance, in Summit schools, students focus exclusively on one idea until they have mastered it, regardless of how well other students are doing. The most advanced "intelligent tutoring" solutions do more than just guide students down already established paths (Zeide, 2017).

As we already mentioned earlier, big data analytics frequently employ cognitive models to monitor and evaluate student progress. "Smart" learning systems carry out a similar procedure by using "tutoring" models that record many instructive choices. The software determines the option most likely to result in student achievement using past data on how students with similar profiles fared. For instance, let us give an example of a student who attempts the practice problem three times before entering the right response. This data is used by the system to update a student's profile, which results in a data pattern that we will refer to as "ABCD." The tutoring model offers three alternative ways to teach: it can play a video or ask students to study the relevant part of previous lessons. According to statistics analyzed by the platform from previous students with ABCD profile patterns, 70% of students who watch new videos correctly answer the following set of questions, compared to 55% of those who review older content. The data "predicts" how the two choices will impact the student currently utilizing the system, and in this case, plays the new video. Although they haven't yet, intelligent tutoring systems are ready to enter the mainstream (Zeide, 2017).

These effects will have an increasing impact on daily life beyond educational institutions as omnipresent data gathering and mining to feed learning analytics produces ubiquitous surveillance. They alter the information utilized to judge and depict student achievement, academic credit, and intellectual mastery as well as the evaluation criteria and final record-keeping format. These systems' underlying value judgments and commensuration are frequently opaque, may be done unintentionally, but they have significant effects on what "counts" as and toward education and accomplishment in professional and educational contexts (Zeide, 2017).

Supporting and Examining Social Processes

In order to investigate social processes, recent studies investigated dialogue, discussion, and collaboration patterns using video transcripts, intelligent tutoring systems, and online discussion forums. With up to a few million contacts, these studies might involve thousands of students. Interest-based subcommunities arose over time, much like in actual physical settings. Apart from entirely online learning environments, blended learning formats also give students the chance to participate in collaborative learning. As an illustration, speech recognition was used by scientists to review transcripts of classroom recordings. This result has been confirmed by research that analyzes and categorizes discourse sequences in intelligent tutoring systems (Fischer et al., 2020).

It's crucial that the changes brought about by big data in education are not unintentionally and carelessly implemented by using new technology carelessly. Instead, they ought to be the outcome of deliberate decisions that are transparent enough to allow for public review. This may include requiring greater openness, responsibility, or cautiousness in terms of information and privacy procedures in educational settings. Schools must put in place coordinating supervision and governance mechanisms that fit new approaches that are based on technological possibilities.

Even while data mining has a lot of potential advantages for educational research, there are still a lot of challenges to overcome.

To begin with, it can be quite difficult to strike the correct balance between individual privacy and the needs of the public as a whole. This is due, in particular, to how challenging it is to prevent the "reidentification" of de-identified data in enormous data sets, even after all direct identifiers have been eliminated. Therefore, maximizing privacy without minimizing utility is not achievable. Instead, educational institutions and researchers must decide whether to maximize privacy at the expense of the data set's utility or to maximize utility at the expense of potentially leaving the data exposed to re-identification with sufficient effort (Fischer et al., 2020).

The availability of data is further exacerbated by privacy concerns, which is most significant. Concerns about companies' access to vast amounts of private student data and their potential

to take actions that aren't always geared toward improving the futures of particular students are valid concerns among parents, educators, and other stakeholders. There are concerns that student information that is illegally sold or shared could be used to stereotype or profile students, support niche marketing initiatives, or result in identity theft (Strauss, 2019).
Finally, even if we are able to access and analyze huge data, there will still be problems with how these data are used. Education researchers will have to deal with the conflict between explanation and prediction as they use data mining more and more. Some academics go into great length about this tension in relation to psychology. They contend that the field of psychology has been filled with research programs that offer complex theories but have limited ability to reliably forecast future behaviors as a result of psychology's focus on revealing the reasons for behavior. They also contend that a stronger emphasis on prediction through the use of data mining and machine learning methods can eventually result in a better comprehension of behavior (Fischer et al., 2020).

To conclude, big data availability presents intriguing new study directions and the chance to broaden the scope of already established educational themes. We may better understand and respond to individual learner behavior as it manifests in the increasingly widespread digital sphere by recording, retrieving, analyzing, and utilizing various sorts of data. There should also be a strong push toward open science and research structures that value collaborative teams in order to increase the field's capacity for mining big data for education research. We should start working on these issues since big data mining in education has the potential to be beneficial (Fischer et al., 2020).

b. Potential harms of Big Data

Big data analytics applications come across a number of challenges related to data characteristics, methods of analysis, and social issues.

The first difficulty is, which is also the focus of this thesis, privacy. The most delicate topic has consequences for theory, law, and technology. In the context of massive data, this concern becomes of greater significance. The International Telecommunications Union defines privacy as the "right of individuals to control or influence what information related to them

may be disclosed" in the broadest possible sense. For anyone interested in exploring the use of big data for development, privacy is an underlying concern that has a wide variety of consequences for data collecting, storage, retention, usage, and display (Almeida & Calistru, 2013).

The ability to access and share information presents another difficulty that is somewhat connected to the previous one. It is normal to expect resistance from private firms and other organizations when it comes to disclosing information about their users, clients, and internal processes. A culture of secrecy, the desire to safeguard one's competition, legal or reputational concerns, or, more generally, the lack of suitable incentives and information mechanisms, can all be obstacles. When data is stored in locations and formats that make it challenging to access and transfer, there are also institutional and technical obstacles (Almeida & Calistru, 2013).

Another crucial point is security for information sharing in big data use cases. Today, a lot of online services (like Facebook, Linkedin, Instagram, Twitter etc.) ask us to share our private information, but aside from record-level access control, we don't really comprehend what it means to share data, how it can be related, or how to provide consumers fine-grained control over it (Almeida & Calistru, 2013).

The size of massive data structures is another important factor that can limit the system's performance. For many years, managing vast and quickly growing volumes of data has been a difficult problem. In the past, this problem was lessened by faster processors, which gave us the resources required to handle growing data volumes. But now that data volume is growing faster than computer resources, there is a fundamental shift taking place (Almeida & Calistru, 2013).

Let us now employ the famous Mason's methodology to examine these issues and illustrate how privacy, accuracy, property, accessibility, and society relate to the big data analytics system using contemporary situations (Richardson, 2021).

The issue of Privacy

People could share their personal information in the 1980s, but the amount, type, and scope of that sharing was different than it is now. Scholars urged for an ethical balance between risk and better services for residents when thinking about surveillance, which also focused on the expansion of data in government databases. Big data questions the conventional wisdom that says people should be in charge of their data and have the power to decide whether, how, when, and by whom it can be used. The transformation of people into human data generators due to ubiquitous computing, real-time data collecting, facial recognition software, social networks, predictive analytics, and practically infinite bandwidth was unpredicted by researchers (Richardson, 2021).

Inputs to big data have an impact on privacy since businesses see value in gathering data for analysis or to sell to other businesses. Given the widespread distribution of modern technology, people frequently are unaware of the data they generate or how businesses use it to build detailed profiles of their lives. The European Union passed the GDPR, as we will be focusing on the next chapter, became effective in May 2018, in order to hold businesses more accountable and give people more autonomy. According to the GDPR, companies that collect data from EU citizens are required to abide by rules that limit the amount and type of data they can collect, demand that companies clearly state the types of personal data they collect, allow people to request and review the data that companies collect, allow people to ask companies to remove (i.e., delete) their data, and more. This change is an initial regulatory step toward promoting more powerful legal behavior when businesses collect individual data as big data inputs. While some have argued that the European Union legislation fails to provide enough protections, corporations must take GDPR compliance into account when collecting data from consumers (Richardson, 2021).

Privacy as it relates to inputs is a challenging topic because many online businesses need users to contribute data in exchange for service. Google has come under fire for tracking and keeping information about owners of Android smartphones even when those users disable location services. In order to reduce the cost of health insurance, insurance firms encourage policyholders to wear activity trackers and share personal health data. Customers that install

tracking devices in their vehicles are eligible for cheaper premiums from the US automotive insurance market. These methods force people into revealing their personal information without providing them with any warning or protection on how others might utilize it. Those who refuse to provide their data will have to pay more or risk having their service cut off (Richardson, 2021).

The big data procedure increases privacy risks when it makes it possible to identify specific people. One learns more and more about a person as additional characteristics are combined. In one study, researchers discovered that they could recognize over 90% of the persons in the sample with just four data points from an anonymized dataset of credit card records from thousands of stores. People create and present various identities as they engage with the world around them in various contexts. For instance, one might reveal specific aspects of their opinions, personalities, and beliefs with friends in person but not online, or at home but not at work. People lose their autonomy and the right to hold numerous distinct identities when data from their ideas, personalities, thoughts, goals, and behaviors are combined to identify them (Richardson, 2021).

Accuracy

One might mistake objectivity for accuracy since big data eliminates humans from data sources and algorithms. As an illustration, a paper from the US government claims that big data systems "will contribute to removing inappropriate human bias where it has previously existed". The assumptions used in algorithms are made by humans, hence the belief is incorrect. Developers and designers have an impact on the semantics, interpretation, and training data for learning algorithms. Thus, algorithms are objects that were created by humans and contain both human biases and biases from any institutional procedures that had an impact on their construction (Richardson, 2021).

According to Mason (1986) "Misinformation has a way of fouling up people's lives, especially when the part with the inaccurate information has an advantage in power and authority". Therefore, precision is particularly important when big players use big data to make judgments. It can have a detrimental effect on how one analyzes and interprets the

results if one fails to validate and take accuracy into account in datasets (especially when one aggregates or reduces datasets) (Richardson, 2021).

Property

When Mason wrote his essay on ethics in the information age, people shared tangible documents on which they had written their personal information. Data commoditization in the big data era has created the issue of property more challenging. Without being aware of how businesses may use their personal data for other reasons, individuals may provide consent for a third party to acquire and use their data as input for one purpose. For instance, four businesses sold information to Location Smart, an online service that made it possible for anyone to access people's location information in exchange for a small fee. Less tech-savvy people could pay a little amount to obtain any customer's location data. The people who gave the data brokers their personal information did not get paid, were informed when their data was sold, or when someone accessed it. This perspective raises the concerns of who is liable for any problems that result from selling of one's personal data, who is the owner of that data, and how the individuals personal data sold could be compensated or informed when someone buys or sells their data (Richardson, 2021).

Since analytics' algorithms handle data from numerous sources, people might not be able to give their consent when businesses acquire, sell, distribute, analyze, and use their personal data. Although people receive something in return (access to services), they give up ownership or other rights about how businesses utilize their data (Richardson, 2021). To that end, the Facebook / Cambridge Analytica incident, as we will further investigate, serves as an example of a situation in which businesses violated people's property. A Facebook app that collected information from users and their friends via a personality test was used by Cambridge Analytica to acquire information from 50 million Facebook profiles. Following the data collection, Cambridge Analytica examined the correlations between personality factors and predicted voting behavior (Cadwalladr & Graham-Harrison, 2018).

The algorithm used by Cambridge Analytica delivered advertising that was specifically targeted at people who might be influenced in upcoming elections. Particularly targeted ads urged people to support Donald Trump in the US, Brexit in the UK, or other candidates in various elections. Although appropriate authorities in the US and the UK took a look at how

targeted advertising affects elections, more and more individuals see the practice as a characteristic of online environments. One targeted Facebook user has indicated “I’ve come to grips with the fact that you are the product on the internet”. In the era of big data, people have accepted a loss of both personal agency and data property rights. When businesses use individuals' data for financial or competitive advantage, they frequently do not pay them back. Finally, targeted advertising and search results show that big data has increasingly started to shape how people see reality (Richardson, 2021).

Accessibility

The issue of property is intimately related to accessibility. Accessibility in the big data era refers to both the freedom to receive information and the security measures implemented to prevent unlawful or forced access through individuals' or groups' participation in regular daily activities. Accessibility emphasizes big data inputs rather than processes or outputs in this aspect. Since the company followed Facebook's rules when collecting data from 50 million profiles in the Cambridge Analytica case, there was technically no data breach. However, the company at most accessed data dishonestly. Users who responded to Cambridge Analytica survey were tricked into providing the company with access to data belonging to their social network connections as well as their own, even though they had no authorization to do so (Richardson, 2021).

Accessibility relates to people's information technologies, such as mobile phones, data plans, social media, and online services, impacting inputs, processes, and outputs. Opt-in and opt-out procedures favor data collectors over people or groups who might be unwilling or coerced data suppliers. People are frequently required to give permission for their data to be shared or sold by organizations in order to utilize internet services. When giving their assent, customers frequently are unaware that businesses would utilize their information for targeted advertising or as a source of income. When organizations access, utilize, and reuse data without a person's knowledge, informed consent and disclosure become less meaningful. It's still not apparent with the Internet of Things who has the right to access the data that people leave behind as they go about their daily lives. Some of these issues are addressed by the GDPR, but only EU people are subject to these rules, and even then, only to the degree that law enforcement agencies apply them (Richardson, 2021).

Society

Big data is evolving and becoming more integrated into our daily lives at a rate that this discussion cannot keep up with, which poses problems for society (Richardson, 2021).

Threats to society and social institutions include inputs to big data, algorithmic processing, and outputs that have the potential to institutionalize practices that discriminate against and marginalize underprivileged communities. Hoffmann (2019) looked at bias and fairness as the two main concerns in data justice and discovered the ways that big data contributes to unequal distributions of rights, opportunities, and wealth despite actions that are explicitly intended to do the opposite. For instance, firms frequently concentrate on particular features to reduce bias in big data systems. Hoffman counters that this approach "fails to question the normative conditions that produce and promote the qualities or interests of advantaged subjects" since it emphasizes disadvantages while disregarding characteristics linked to advantage (Richardson, 2021).

Focusing on qualities linked to challenges may help eliminate discrimination, but doing so fosters decision-making based on common demographic traits that place people in disadvantageous situations in the first place. Therefore, big data may have greater negative impacts on people who are at the lower end of the socioeconomic range or who are unable to speak up for themselves. For instance, big data is frequently used by banking companies to make lending decisions. However, the institutions consistently charged Latinos and African American borrowers greater interest rates than White and Asian-American borrowers. However, one study indicated that algorithms reduced prejudice based on minority status when it came to loan refusals (Richardson, 2021).

It is challenging to understand how algorithm-driven decisions affect participation in different areas of society because big data is so extensive. Big data frequently classifies people into social groups based on factors including their ethnicity, financial level, age, area of residence, social networks, and daily activities. These factors impact access to opportunities. Big data does this by establishing and maintaining power disparities that limit some people's ability to participate fully in the society in which they are ingrained. For instance, in order to build security among its inhabitants and to advance social welfare, China gathers and aggregates

data to develop an obligatory social credit rating system. Several social credit systems are currently in use in the nation; the government manages and maintains some of them, while private businesses (like Alibaba, one of the biggest technological companies in the world), manage and maintain the others. As a result, as the central repository develops, the inputs the government and organizations use to calculate citizens' scores are probably going to come from sources their creators never intended to measure social credit, which could result in errors in subsequent processes and results and, ultimately, prevent some people from accessing opportunities that enhance life quality (Richardson, 2021).

Organizations frequently create big data procedures to more effectively imitate existing behaviors, which allows them to quickly reproduce the worst of what society has to offer. For instance, Amazon created a tool to screen resumes in 2014 using artificial intelligence approaches. When it realized that the tool had "learned" to bias against female candidates as a result of men predominating the technological sector and the company's workforce, it stopped utilizing the techniques. Another illustration is the 2016 introduction of Tay, a chatbot from Microsoft that used Twitter interactions to demonstrate real-time machine learning. A few hours after its publication, Tay started calling people names, writing rude and racist comments, and denying the Holocaust. These instances show how big data can react to and reinforce biases, power disparities, and even intolerance (Richardson, 2021).

Finally, organizations need to think about how outcomes could be harmful to them and to people as they collect and use data for prediction. Big data collects knowledge from experts and workers and incorporates it into software when combined with artificial intelligence. The future workforce may be significantly impacted by the transmission of knowledge to machines. Moreover, we must take into account important ethical issues with big data given the potential for major and adverse impacts on civil rights and societal participation (Richardson, 2021).

2. Big Data and data analytics privacy exploitation

Data has long been used in marketing to produce an understanding of customer needs. The efficient use of all data sources has grown in importance as an organizational competency in

recent years. This is so because maintaining a competitive advantage depends on the creation and application of new information.

Data exploitation is both a difficulty and an opportunity. The difficulty comes in managing all of the generated data for an organization. It is rather easy to manage transaction data for both incoming and outgoing traffic. However, with the development of the Internet and mobile communications technologies, the complexity of data management has been dramatically enhanced. Huge amounts of data are produced by consumers who shop online, social media accounts, outside parties, emails, blogs, and increasingly, sensors, which give suppliers access to real-time information about the goods their clients are using (Chaston, 2015).

The development of big data and business analytics has significantly changed an organization's ability to gather and analyze external data to better understand the market or fill knowledge gaps regarding customers. When compared to traditional market research, real-time data from internet sources gives synchronization, and this can be gathered for relatively little money, of a size that allows for micro-analysis, and has a significantly lower non-response error rate. As a result, businesses that have implemented big data systems are better able to quickly gather vast amounts of information from the outside world than less progressive rivals who still do not fully understand the advantages of utilizing real-time data and business analytics (Chaston, 2015).

Planning and implementing marketing strategies is a managerial process that used to depend heavily on outside data. Until recently, traditional market research was used to supplement internal organizational data from consumer activities, such as orders, pricing, and delivery, to improve understanding or fill in gaps in understanding. The disadvantage of conventional market research was that it used an inconsistent technique and was only occasionally repeated due to expensive costs. It was based on a tiny sample, which made it difficult to conduct micro-analysis, and the results were subject to be affected by non-response error (Chaston, 2015).

A suitable marketing mix must be developed in order to effectively implement a marketing strategy to keep, modify, or add it. Sathyanarayanan (2012) said that one of the advantages of big data is that, perhaps for the first time, businesses can learn about the entire ecosystem in

which they operate. To better understand the nature of the consumer actions in the past and present and to make more precise predictions about the future, information on behaviors such as product preferences and purchasing habits can be micro-analyzed. The development of more successful marketing campaigns is a result of this outcome (Chaston, 2015).

The 4Ps that are product, promotion, price, and place are the main elements of the marketing mix. Making a good product or service available is ultimately what determines whether marketing efforts are successful or not. Making a superior product or service available to the market is not enough to generate sales. To establish the best marketing mix, data must be acquired and analyzed on the other 3Ps. Marketers have a propensity to view the product or service as the main channel for utilizing innovation. The emergence of the online world has significantly changed reality since marketers can now take advantage of a wide range of distribution, pricing, and promotion options to satisfy customers even more. Big data enables modeling of how creative approaches utilizing one or more of the 3Ps might produce new types of competitive advantage by giving a more in-depth insight of customer behavior (Chaston, 2015).

Competitor disruption eventually affects marketing strategies. Before the development of big data and business analytics, there was frequently a gap between competitors applying new or changed marketing strategies, the organization becoming aware of them, and an evaluation being made of the actual impact of a new competitive threat. By using big data and business analytics, the organization can gain a far earlier insight of changing consumer behavior, considerably reducing risk (Chaston, 2015).

Although known to be significant in the days before big data, it was very challenging to analyze the possible effects of changes in the macro-environment and develop proactive adjustments to the marketing mix. Big data makes it easier for organizations to identify emergent macro-environmental changes and adopt quick modifications to their marketing mix. One instance is seen among major national businesses that have a sizable number of locations in nations where localized differences in weather affect consumer purchasing patterns. Purchasing geographically specialized long-range predictions will help you spot changes in the weather. As a result, if a prediction shows exceptionally cold, rainy weather in one part of the country and hot, sunny weather in another, the retailer might adjust the stock

mix in particular stores to satisfy anticipated consumer purchasing habits in reaction to changing climatic circumstances (Chaston, 2015).

Exploiting Information

In the world we live in today, users can instantly access a great amount of information over the Internet. Customers' resistance to paying a price to use these services is the challenge that organizations seek to profit from the provision of information (Bleyen & Van Hove, 2010). Search engines like Google and Yahoo, which continue to provide Internet users with free access to a plenty of information, have played a significant role in encouraging users in the majority of mass markets to avoid having to pay for information access. By selling ad space, these businesses have maintained a sustainable free access strategy (Chaston, 2015).

According to Koli (2007), there are two aspects to the exploitation of information: using obtained information to find new business prospects and using online information services to improve customer satisfaction. As the convenience of digital technology mediates our lives, information about them is subject to commercial capture, a logic we may refer to as a type of "digital enclosure" of personal data (Chaston, 2015).

Information gathering is simply a portion of the narrative. Participating in the interactive digital economy subjects us to extensive, continuing controlled tests carried out by a new breed of market researchers, turning us into lab rats. Such studies seek to identify patterns of individual characteristics such as personality, location, demography, and historical behavior that increase responsiveness to persuasive advertisements. Every aspect of a person's life is relevant for this new kind of marketing research since they all help to create correlation-based patterns. Marketers are simply not interested in reasoning if someone flossing three times a day means they are more likely to buy a specific brand of car or wine; They simply want to see how well the pattern can be predicted. The likelihood of discovering correlation-based types of prediction increases with the amount of information they can collect. Marketers only need to know that a specific "anonymized" target fits the pattern and will receive a personalized appeal; they are not required to know the names of specific people who meet the pattern (Andrejevic, 2009).

Commercial databases, which are created with the intention of influencing and targeting consumers, are equally prone to "function creep". Law enforcement and national surveillance organizations are just as interested in the data mine as businesses are. The focus of this part, however, will be on the commercial sector because we already have developed the ways of thinking about monitoring, not to mention that the exponential expansion of digital media is built on a surveillance-based business model. Consumer relationship management driven by data is what sponsorship and advertising used to be to the broadcast era (Andrejevic, 2009).

The concept of privacy, when understood narrowly, is inadequate for the task of considering the urgent problems related to information gathering and usage, both legally and in terms of regulations. Personal privacy is a right that people give up in exchange for access to resources, much like labor in the industrial age, and they do so under organized power relations that at best make the idea of free or autonomous consent problematic. For instance, we can give an example for this case about an app called “Appirio”, a business that sells a Facebook program that employers can request their employees to install. Once it has been installed, the application searches the social networks of the users' employees for potential employers, clients, and applicants for jobs (Andrejevic, 2009).

The same information that can lead to new clients by giving information for targeted marketing appeals: Users can decide whether to act on the application's recommendations of friends who might be interested in the offer based on a search of keywords in friend profiles. The tool integrates information from each employee's social media accounts with a private consumer relationship marketing database "to track leads, make follow-up offers, and report on campaign success to see how their viral campaigns stack up to other marketing programs." (Andrejevic, 2009).

A privacy-based approach to an application like Appirio has a number of drawbacks. First of all, ideas of privacy frequently emphasize personal freedom and individual rights. If a worker "voluntarily" decides to use Appirio's software, its defenders will probably claim that it is only a matter of personal preference. Why should regulators or other governmental bodies get in? Additionally, programs like Appirio can work in ways that ostensibly protect employee privacy as defined only broadly by disassociating marketing and consumer appeals from specific employees or by preventing employers from having direct access to information about employees' private lives. However, neither of these assertions seems convincing. Even

if employees have the "choice" to utilize their social networks, doing so seems forced, and this is because the notion of privacy ignores the power dynamics that shape the choice. Even if companies don't have direct access to employee Facebook data, the idea that this information being turned into informational capital for use by other parties is nonetheless troubling. It foreshadows a future in which our social networks will be another resource on the job market, where the effort we put into creating and sustaining these connections will be used as a standing reserve for data mining (Andrejevic, 2009).

Broadly interpreted, the term "user-generated content," which is frequently used in discussions about new media, includes much more than just the rising popularity of social media accounts, personal websites, blogs, and other types of amateur media production. It also includes the vast amounts of data that users of a new generation of networked electronic items create about themselves when they use those devices. Long committed in the study of media audiences, media and cultural studies have tended to concentrate on new manifestations of audience production rather than how these audiences are put to work by these numerous forms of audience monitoring. In dramatic terms, academics and critics have called the new consumer monitoring practices "the end of privacy" and "the destruction of privacy." The loss of privacy, according to other academics, is not the correct description of the current day, in which enormous amounts of personal data are being collected, privatized and exploited. (Andrejevic, 2009).

Industries that rely on this data have a tendency to dismiss privacy issues, arguing that the best way to determine if consumers are prepared to give up control over their personal information is to let the market decide. But this argument assumes two things that aren't supported by the research: first, that consumers are aware of how much control they are giving up over their personal information, and second, that the market has given them a meaningful range of options for control over the gathering and use of that information (Andrejevic, 2009).

According to research, customers' privacy concerns increase rather than decrease when they learn more about methods of monitoring. For example, the 2007 report on Community Attitudes to Privacy commissioned by the Australian Office of the Privacy Commissioner revealed that more than half of the respondents were more hesitant to share information online than they were two years prior, and a similar number indicated that they were hesitant

about where information about them was being collected. A shocking 90% of Australians expressed concern about companies transmitting their personal information abroad, which is exactly what widely recognized social networking sites like Facebook state as one of their terms of use. The commercial sector, on the other hand, has no motivation to carry out or promote studies that would point to public support for restrictions on the power it is given to itself. In the United States, privately financed research has assisted in shaping policy in ways that minimize public concerns and preferences, as stated by Oscar Gandy (Andrejevic, 2009).

Therefore, the current situation seems to create a problem. On the one hand, surveys continue to show that the public is highly concerned about privacy, but on the other, people in this same public seem more and more willing to submit to commercial forms of monitoring on a growing number of platforms and applications, even as they actively engage in public self-disclosure practices like social media, blogging, public journaling, tweeting, and updating their personal profiles for a growing number of Facebook "friends.". The seeming contradiction is advantageous from a business standpoint since the new mass-customized economy depends on accurate descriptions of consumer preferences and behavior to target advertising appeals. According to management experts, it is a total paradox. The amount of personal information put out there is perfect for marketers, and it is an absolute “treasure box” (Andrejevic, 2009).

Therefore, the typical market account claims that the logic driving the commercial online economy is similar to that of free exchange. Similar to the "cost" of watching free-to-air TV, which is giving in to marketing ploys, using commercial internet services comes with a price: giving in to tracking and targeted advertising. If, on this basis, the logic of "free" exchange underlies e-commerce and we voluntarily agree to the terms set by commercial websites, then the notion of exploitation, which involves coercion, is no longer relevant. The goal of critical political economy is to identify the ways that power dynamics and, consequently, forms of coercion shape the parameters of "free" exchange. Any analysis of exploitation within the context of the willing trade-off of privacy for services or the decision to consent to significant types of commercial monitoring must include such a critique (Andrejevic, 2009).

The existence of big data and data analytics then generates concerns about how these technical advancements may affect the privacy of consumers whose data is "exploited" in the use of big data and data analytics. The question that can come up is how much personal

information about customers can be used by businesses employing big data for data analytics. Is the data used for marketing purposes in the form of commercial online advertising only small digits from millions of data points compiled into a single statistical graph for internal research and strategy formulation, or does the data also include information that is extremely individualized to a person's preference for the product, such as products that are frequently searched for, products that were most recently viewed, preferred colors, etc. (Tsaqila, 2019)

As a result, big data users in general face the risk of overlooking consumer privacy issues, which could be financially disastrous for their business. Financial catastrophes are strongly tied to the potential for lawsuits from parties who feel their privacy has been violated or who feel victimized because their personal information is utilized without their consent or in exchange for any form of payment. It is undoubtedly terrifying to think of the possibility of a financial catastrophe if it is connected to big data and data analytics (Altshuler, 2019).

The constraint of the idea of privacy as control is the fictional nature of our control over our personal information. There is a widespread issue with commercial businesses using vast amounts of private information without obtaining genuine agreement to do so. This information can be utilized in a variety of ways, some of which are beneficial and others of which represent serious dangers to the society we live in. Now let us examine in detail one of the most controversial cases of all times: The Cambridge Analytica scandal.

The Cambridge Analytica scandal

People all across the world use social media to maintain their social connections and actively participate in political dialogue. Social media has joined the global public sphere, where information is shared and public discourse takes place, with a spread of news and information (Brown, 2020).

50 million Facebook accounts were harvested for Cambridge Analytica in a huge data breach, according to an article released in March 2018 by the Guardian and the New York Times jointly. This news report revealed the Cambridge Analytica controversy, which is frequently cited as a turning point in the public conversation about businesses' exploitation of personal data. The issue was based on the extensive gathering of 87 million Facebook users' personal

information made possible by the Global Science Research-Cambridge Analytica partnership-developed Facebook app This Is Your Digital Life.

In what started as an academic research project, Cambridge Analytica and Global Science Research offered to pay users to take a personality test within the app in exchange for their consent to the data collection. At the time, unless users adjusted their privacy settings, Facebook's default terms permitted the collection of data from their Facebook friends as well. With the help of this information, Cambridge Analytica and Global Science Research were able to target specific individuals and learn more about their political ideas and personality traits. The 270,000 Facebook users who took the personality test as part of an academic research study also gave access to the data of their Facebook friends. Concerned parties claimed that Cambridge Analytica and Global Science Research violated Facebook's terms of service because none of these individuals or their Facebook friends consented to have their data shared with a third party for marketing purposes (Afriat et al., 2021).

Mark Zuckerberg, the CEO of Facebook, called the Cambridge Analytica situation a "breach of trust". The US authorities criticized Facebook for not taking swift action to stop this privacy breach. Congress demanded Zuckerberg to testify and he was eventually fined (Brown, 2020). This story received extensive publicity in the media worldwide. The "beginning of the end" of Facebook was described by journalists, and studies suggested that once the scandal broke, consumers' trust in Facebook and use of the platform had decreased. Journalists and scholars wondered whether this outrage signaled a turning point that would lead consumers to disconnect.

While several news articles claimed that user disconnections from the network followed the controversy, the real number of Facebook users rose the year the Cambridge Analytica scandal surfaced. Over 1.52 billion individuals used Facebook every day in December 2018, an increase of 9% from the previous year. Additionally, data indicates little proof of user disengagement. Only 9% of users used the new privacy setting Facebook implemented following the Cambridge Analytica controversy, which allowed them to download all the information it collected about them (Afriat et al., 2021).

Institutional and social privacy. Economic surveillance, or the collection of personally identifiable information posted on social media for commercial benefit is a prominent source

of academic concern. However, because of ignorance or a sense of powerlessness, users seem unconcerned with the problem (Afriat et al., 2021).

People often concentrate on social privacy, which is "the control of information flow about how and when personal information is shared with other people," when they are worried about their online privacy (Afriat et al., 2021). In other words, they frequently worry about their social standing and privacy in front of their peers, families, employers, and other people. Social media platforms let users choose which details, such as general biographical data, images, daily activities, thoughts, opinions, and location, to share with online friends. "Any message about the self that an individual communicates to another" is referred to as self-disclosure (Brown, 2020). However, social media platforms that allow users to share information with family, friends, and others also give service providers access to that data (Afriat et al., 2021). The latter is institutional privacy, which concerns "how organizations like banks, businesses, and governments use or abuse personal information" (Afriat et al., 2021).

Because of their ignorance, lack of knowledge, skepticism, and apathy, users are much more concerned about social privacy than institutional privacy, according to previous studies that looked at both categories of privacy. For instance, Stutzman, Gross, and Acquisti (2013) demonstrate that users exposed more information after changes to their privacy settings that increased their level of social privacy and decreased their level of institutional privacy because they were unaware that, in addition to sharing information with their close friends, they were also sharing it with Facebook and other companies. According to Lyon (2017), the explicit study of culture's facilitation of economic surveillance is necessary to comprehend this type of conduct.

Social privacy concerns are made even greater by the fact that, in the case of Facebook, the aggregate acquired data is like the tip of an iceberg, with users only seeing just a little of it, mostly their interactions with other users. Users are also ignorant of the information processing carried out by data collectors. They can only genuinely perceive only a portion of the information that is actually in their social media profiles. Because of this, individuals have a very limited awareness of the kind of profiling that third-party businesses can conduct on the data that social media companies collect on their platforms (Afriat et al., 2021).

The concept of institutional privacy is, overall, much harder for consumers to understand than the concept of social privacy. Additionally, user agreement forms, privacy rules, and copyright enforcement that present harvesting in a way highlighting users' potential benefits, while reducing explanations of its commercial usage, all contribute to economic surveillance being "deeply embedded in and obscured by social media infrastructures". As a result, experts question whether consumers are well-informed enough to comprehend and consent to such tracking (Afriat et al., 2021).

Furthermore, recent research claims that consumers' reactions to institutional privacy include sentiments of feeling powerless, apathy, and cynicism. Users believe they have little choice but to accept the terms and conditions because they understand that "privacy violations are inevitable and opting out is not an option," leaving them with little control over their data. This mindset is known as "resigned pragmatism.". Users' inability to make informed choices regarding their personal data is not due to a lack of knowledge or a concern for their privacy, but rather to the fact that social media platforms have forced them to accept a compromise (Afriat et al., 2021).

In order to reduce risk and safeguard sensitive data, big data must be appropriately managed. Many conventional privacy techniques cannot handle the amount and velocity required because big data consists of enormous and complex data sets. A framework must be developed for privacy protection that can handle the volume, velocity, diversity, and value of big data as it is moved between environments, processed, analyzed, and shared in order to protect it and ensure that it can be used for big data analytics. Let us analyze some of the privacy protecting methods in big data.

3. Privacy protecting methods in Big Data

This will be a brief description of a few traditional techniques for protecting privacy in big data. A certain amount of anonymity is provided by these methods as they have been traditionally employed, but their drawbacks have caused the emergence of alternative approaches.

De-identification techniques

De-identification is a well-known method for protecting individual privacy in data mining (Jain, 2016). Before releasing data for data mining, data should first be cleaned up using generalization (replacing quasi-identifiers with less specific but semantically consistent values) and suppression (not releasing some values at all) (Jain, 2016). To improve traditional privacy-preserving data mining and lessen the risks from re-identification, the concepts of k-anonymity, l-diversity, and t-closeness have been established (Jain, 2016). De-identification is a key privacy protection technology that can be transferred to big data analytics that protect privacy (Jain, 2016). We must be mindful that big data might potentially increase the risk of re-identification because an attacker may be able to obtain additional external information support for de-identification in the big data (Jain, 2016). De-identification is therefore insufficient to guarantee the privacy of big data (Mehmood et al., 2016).

There are three de-identification techniques that protect privacy: K-anonymity, L-diversity, and T-closeness.

K-anonymity. Information security and privacy first experienced the idea of k-anonymity in 1998. It is based on the theory that identifying information about any one of the individuals who contributed to the data can be hidden by combining sets of data with similar features (Devane, 2021). K-Anonymization can be characterized as the ability to "hide in the crowd." Data from different individuals is combined to create a larger group, which means that any

one member could be the subject of any information in the group, hiding the identity of the individual or individuals in problem (Devane, 2021).

A database is a table with n rows and m columns that is used in the context of k-anonymization problems. Each row of the table represents a record pertaining to a specific person from a population, and the entries in the various rows do not necessarily need to be unique. The values in the various columns represent the values of characteristics associated with population members (Jain, 2016).

L-diversity. L-diversity is a type of group-based anonymization that lowers the level of detail in which data is represented to protect privacy in data sets. As a result of this reduction, some privacy-enhancing data management or mining methods may become less viable (Jain, 2016).

The l-diversity model addresses some of the k-anonymity model's flaws, including the fact that protecting identities to the level of k-individuals does not always entail protecting the associated sensitive values that were suppressed or generalized, especially when the sensitive values in a group show homogeneity. The l-diversity model's anonymization process involves a boost of intra-group diversity for sensitive values. This method's dependence on the range of sensitive attributes is a drawback. Although sensitive attributes don't have a lot of different values, fictional data needs to be inserted to make the data L-diverse. This fictional data will increase security, but it could cause issues during analysis (Jain, 2016).

T-closeness. By reducing the granularity of a data representation, it is a further development of l-diversity group-based anonymization, which is used to preserve privacy in data sets. This decrease is a compromise that causes some inadequacy in the data management or mining techniques in exchange for a small increase in privacy. By considering the values of an attribute differently and taking into account the distribution of data values for that attribute, the t-closeness model expands upon the l-diversity model (Jain, 2016).

Analysis of de-identification privacy methods

Powerful data analytics can extract useful information from massive data, but they also represent a significant risk to the privacy of the users (Mehmood et al., 2016). Many different methods have been suggested to protect privacy before, during, and after the big data analytics process. The trade-off between privacy invasion and preservation will become increasingly significant as consumer data continues to expand quickly, and technologies continue to advance (Jain, 2016).

HyberEx. A model for maintaining privacy and confidentiality in cloud computing is the hybrid execution model (Jain, 2016). The model uses public clouds only for non-sensitive data and computation of an organization classified as public, whereas it integrates an organization's private cloud for sensitive, private, data and computation (Mehmood et al., 2016). The model only executes public clouds for safe operations while utilizing an organization's private cloud. It takes data sensitivity into account while offering integration and security (Jain, 2016).

Privacy-preserving aggregation. The foundation of privacy-preserving aggregation is homomorphic encryption, which is a well-liked method of gathering data for event statistics (Jain, 2016). Several sources can utilize the same public key to encrypt their individual data into cipher texts when a homomorphic public key encryption technique is used (Jain, 2016). With the proper private key, the aggregated result of these cipher messages can be decrypted (Mehmood et al., 2016).

Operations over encrypted data. Operations can be done over encrypted data to protect individual privacy in big data analytics, which is motivated by searching over encrypted data. Doing operations over encrypted data might be seen as wasteful in the context of big data analytics since these processes are frequently complex and time-consuming, and big data is large volume and requires us to mine new knowledge in a fair amount of time (Jain, 2016).

Most Recent Techniques of privacy in big data

Differential Privacy. Differential Privacy is a technology that gives researchers and database analysts the ability to access databases that contain personal information about people while maintaining the privacy of those individuals' identities. This is accomplished by minimally interfering with the information the database system provides. The amount of disruption introduced is sufficient to safeguard privacy while still leaving enough room for analysts to make use of the data. Before, several privacy-protection strategies were tried, but they didn't quite work (Jain, 2016).

The degree of distortion that is introduced to the raw data is inversely related to the assessed privacy risk. When there is little privacy danger, distortion added is minimal enough not to degrade the quality of the response but substantial enough to preserve database users' privacy. However extra distortion is created if the privacy risk is significant (Mehmood et al., 2016).

Identity based anonymization. These methods had problems when they were used to analyze usage data while preserving user identities. They successfully merged anonymization, privacy protection, and big data methods. Because of its creative approach and optimistic outlook, cloud computing is a type of large-scale distributed computing paradigm that has recently been a driving force for information and communications technology. It has the potential for enhancing IT system management and is altering how hardware and software are created, acquired, and used. The use of cloud storage services offers significant advantages to data owners, including easing the burden of managing storage and equipment maintenance on users, avoiding the need to invest heavily in hardware and software, enabling access to data regardless of location, and enabling access to data at any time and from anywhere one wants. (Jain, 2016).

To achieve these goals, Intel developed an open architecture for anonymization that made it possible to use a range of methods for both de- and re-identifying web log information (Jain, 2016). Enterprise data differs from the typical examples in the anonymization literature in

terms of attributes, as was discovered during the implementation of the design. This idea demonstrated how big data approaches, especially when used to anonymize data, may be advantageous in the business world. Intel also discovered that the anonymized data was vulnerable to correlation attacks despite masking clear Personal Identifying Information like usernames and IP addresses (Jain, 2016). They looked into the costs and benefits of fixing these flaws and discovered a substantial correlation between user agent information and specific users (Mehmood et al., 2016).

In order to preserve privacy in enterprise data analyzed using big data techniques, this case study on anonymization implementation in a company describes needs, implementation, and experiences observed. K-anonymity based measures were employed in this analysis of the anonymization quality. To analyze the anonymized data and obtain useful information for the Human Factors analysts, Intel employed Hadoop. At the same time it discovered that anonymization requires more than just hiding or reclassifying specific values, and that anonymized datasets must be thoroughly examined to identify their vulnerability to attack (Jain, 2016).

Hiding a needle in a haystack. Current association rule algorithms that protect privacy add noise to the underlying transaction data. As a result, the strategy still leaves open the prospect of the true frequent item set being inferred by an unreliable cloud service provider. It provides sufficient privacy protection despite the possibility of association rule leaking since this privacy-preserving technique is based on the idea of "hiding a needle in a haystack". This approach is founded on the notion that finding a rare class of data, like the needles, in a haystack of data of a vast quantity is difficult. Current methods cannot randomly introduce noise since they must take into account the utility-privacy trade-off. Instead, this method adds noise at an additional computational expense to create a "haystack" to cover the "needle" (Jain, 2016).

Both the dummy and the original objects have their own unique code (Mehmood et al., 2016). After the extraction of frequent items specified by an external cloud platform, the service provider keeps track of the coding information to filter out the dummy items (Mehmood et

al., 2016). The external cloud platform runs the Apriori algorithm on data sent by the service provider. The external cloud platform gives the service provider back the support value and frequently used item set (Mehmood et al., 2016). To extract the relevant association rule using a frequent item set without the dummy item, the service provider filters the frequent item set that is impacted by the dummy item using a code. The extraction association rule method does not place any additional strain on the service provider (Jain, 2016).

Privacy-preserving big data publishing. With a growing number of open platforms, such as social networks and mobile devices, from which data may be acquired, the volume of such data has also grown over time. Publication and dissemination of raw data are essential components in commercial, academic, and medical applications (Jain, 2016). Input and output privacy are two general situations that privacy-preserving models can be used in (Jain, 2016). Publishing anonymized data with models like k-anonymity and l-diversity is the main concern in privacy (Jain, 2016). In terms of output privacy, issues like association rule hiding and query auditing are typically of importance because they require adjusting or auditing the output of various data mining algorithms in order to maintain privacy (Jain, 2016). The quality of privacy preservation (vulnerability quantification) and the value of the published data have received a lot of attention in the privacy field. Just breaking up the data into smaller pieces (fragments) and independently anonymizing each piece is the solution according to scientists (Mehmood et al., 2016). Despite the fact that k-anonymity can defend against identity attacks, due to the lack of variation in the sensitive property inside the equivalence class, it is unable to protect against attacks involving attribute disclosure (Jain, 2016).

We have reviewed some of the privacy protecting methods in big data. Now, let us move on to our last chapter which will be focusing on the GDPR that is considered the world's strongest set of data protection rules, which enhance how people can access information about them and sets limits on what organizations can do with personal data.

Chapter 3: The Legal Environment of Big Data in the EU's GDPR

1. The history of the protection of personal data in the EU

The GDPR (General Data Protection Regulation) started to be applied on May 25, 2018, repealing the EU Data Protection Directive (Directive 95/46/EC known as "Data Protection Directive"), which was first adopted back in 1995 (GDPR, art. 94). The GDPR maintains the traditions of EU data protection law rather than beginning them entirely in a radical direction, while being unquestionably the most important and dominant privacy regulation in the world at the moment. In other words, it is "an evolution rather than a revolution," according to worldwide privacy academics (Richards, 2022).

Many of the GDPR's modifications were made with the intention of addressing identified shortcomings in EU data protection law and modernizing it for digital practices in the 2020s (Richards, 2022).

Internet connectivity and the use of digital technology by our society are having a big positive impact on how effectively we interact, communicate, and collaborate. However, this growing reliance on technology also creates new dangers and exposures for people, economies, and governments, necessitating the need for data protection rights. The GDPR aims to increase data privacy for EU citizens by establishing a unified framework to govern the commercial use of personal data. The GDPR is being implemented to bring together and update data protection rules relating to the internet, social media, and the digital economy as well as to safeguard and empower all EU citizens with regard to personal data privacy (De Carvalho et al., 2020).

"The main data protection principles in the GDPR are revised but are broadly similar to the principles set out in the Data Protection Directive: fairness, lawfulness and transparency (Article 5(1)(a)); purpose limitation (Article 5(1)(b)); data minimization (Article 5(1)(c)); accuracy (Article 5(1)(d)); storage limitation (Article 5(1)(e)); accountability (Article 5(2)); integrity and confidentiality (Article 5(1)(f))." (Politou et al., 2018, 4).

“While the DPD constituted the international standard against which all data protection initiatives, in and out of Europe, were judged, the GDPR brings the novelty of explicitly imposing organizations to enshrine “data protection by design and by default” (Article 25) enforcing measures such as data minimization as a standard approach to data collection and use. Furthermore, the GDPR extends the provision on automated individual decision-making, to include profiling cases as a prime example of enabling individuals to control their personal data in the context of automated decision-making (Article 22) and hence acts as crucial function for mitigating the risks of big data and automated decision making for individual rights and freedoms” (Politou et al., 2018, 4).

The GDPR recognizes that everyone has a right to the protection of personal data that relates to them and has changed the data protection landscape in the EU from that which was outlined under the Directive to a setting that is protected as a fundamental right in Article 8 of the Charter of Fundamental Rights. Personal data itself is at the heart of data protection. As stated in Article 4(1) of GDPR, "any information relating to an identified or identifiable natural person ('data subject')" is included in the definition of personal data, which incorporates many of the fundamental principles of the Data Protection Directive (Clarke et al., 2019).

According to Article 4(2) of the GDPR, almost any use of a person's personal information, from collection and recording to retrieval and dissemination to storage to erasure or destruction, constitutes "processing" and is subject to strict liability. Understanding the obligations of personal data users, including "data controllers" and "data processors," is essential to attaining compliance with these requirements (Clarke et al., 2019).

The GDPR has made major improvements to the rights of the data subject and now contains Articles 16 and 21 that address the right to object to the processing of personal data, the right to rectification or to have personal data erased, and Article 17 that addresses the so-called “right to be forgotten”. Additionally, Article 15 of the GDPR grants data subjects the right to access their own personal information (Clarke et al., 2019).

It is up to the Member States’ Courts to determine how it should be interpreted, The highest authority for interpreting EU legislation remaining the Court of Justice of the European

Union (CJEU), together with the newly established European Data Protection Board's persuasive but non-binding interpretation (Clarke et al., 2019).

To summarize, the GDPR, is a data protection law that establishes rules for handling, storing, and managing personal data from individuals who are currently residing in the European Union (Li et al., 2019). With the advent of new digital technologies, there are new risks to privacy that this new legislation enhances the EU's data protection to address. Although the GDPR only protects EU individuals, its effects are worldwide in nature given its territorial scope affecting any company that serves the European market or offers services there and has access to personally identifiable information about EU citizens (GDPR, art. 3).

2. The right to privacy and to personal data protection as fundamental rights in the EU

“The need for a right to protection of personal data significantly increased with the emergence of information and communication technology in the second half of the twentieth century. Historically, a right to data protection was considered as an aspect of the right to privacy. This has also been referred to as the right to informational privacy” (Custers et al., 2022, 6).

The Charter of Fundamental Rights of the EU was drafted in 2000, entered into force in 2009 via the Treaty of Lisbon. Here for the first time the right to data protection has been separated from the right to privacy and configured as an autonomous fundamental right (Custers et al., 2022, 6).

In essence, the GDPR expands upon the provisions of the EU Directive it replaces while also strengthening several data subject rights, such as the right to data portability and the right to be forgotten, and introducing a number of new ideas, including data protection impact assessments, privacy by design, and data breach notifications (Custers et al., 2022).

According to Article 8 of the Charter “Everyone has the right to the protection of personal data concerning him or her”. Human dignity is closely related to the fundamental right to

privacy protection. Particularly, dignity assumes that a person can freely develop their individuality through self-determination and self-flourishing, two concepts that many authors consider to be the foundation for privacy protection. The relationship between data protection and dignity suggests that this protection is unalienable (Custers et al., 2022, 6).

The EU's Article 8 Fundamental Rights Charter guarantees an outstanding fundamental right to the protection of personal data, since no other country in the world has made protecting personal information a fundamental right. This has a tight connection to the establishment of the high standard of data protection that the EU aimed to attain with its adoption of the GDPR, which is widely regarded as providing the strongest (or at least one of the strongest) legislative instruments for data protection in the entire world (Custers et al., 2022).

If personal data is processed, the GDPR is applicable. Regardless of whether they are publicly or privately funded, universities and research institutions are subject to the GDPR. In addition, the GDPR applies to all people living in the EU, regardless of their citizenship status, when it comes to processing personal data. It soon becomes clear that the GDPR casts a wide net in order to include a variety of situations. Its goal is to change how both public and private organizations, including universities and research facilities, approach data protection (Clarke et al., 2019).

3. The GDPR as a post Big Data regulation: Main aspects

The technology platforms and data architectures that are now used to gather, store, and handle personal data are projected to be significantly impacted by GDPR. Organizations will need to conduct a thorough internal assessment for their technology platforms and data architecture, including different information systems, websites, databases, data warehouses, and data processing platforms, in order to better understand what personal data was collected and how it was used. The GDPR has high requirements for data controllers and processors to handle personal data, including data protection by design and default, and recording all processing activities. Organizations need to make adjustments to their technology platforms and data architecture following the internal review in order to comply with GDPR's standards (Li et al., 2019).

In addition, GDPR mandates that businesses grant EU citizens strong privacy rights, including the Right to Be Forgotten, the Right to Access Data, the Right to Data Portability, and the Right to Information About Automated Decision-Making. A user has the right to request a timely response from an enterprise if they wish to know what personal information a company has acquired about them and why (Right of Access to Data). It is possible that major corporations can daily receive hundreds of inquiries from clients about how they use their personal information. Customers have the right to request that their personal data be deleted if they are unhappy with how the business manages their personal information (Right to be Forgotten). Companies that employ people from the EU or who live in the EU must also manage their employees' personal data, including photos, bank account information, tax and pension information, health and safety reports, sick leave requests, medical records, CVs, job application forms, disciplinary procedures, vacation requests, and salary data. A company may need to improve or reengineer its current platforms and systems to satisfy requests from consumers or workers for efficient access to their personal data and efficient removal of personal data from the system. To be more precise, the company must first locate any personal information about this client or worker in databases, archives, and systems for managing customer and employee relationships. In order to find and retrieve personal data about people, the business must secondly install comprehensive search tools that can search across all technology platforms, systems, archives, and architectures. Without comprehensive search tools, there is no assurance that the business will be able to guarantee that all the personal information associated with a certain client or employee can be handled properly (Li et al., 2019).

The GDPR also makes adjustments intended to improve individual control over personal data by strengthening the rules on transparency and consent, updating the clause on profiling, and announcing new individual rights. Additionally, it expands the duties of controllers and processors through a variety of new accountability provisions.

The types of data generated by big data technology can be used to make nonsensical and improbable assumptions and forecasts about people's preferences, behaviors, and private lives. These inferences are based on extremely varied and feature-rich data that may or may not be valuable, and they open up new possibilities for discriminatory, biased, and intrusive

decision-making, frequently based on delicate aspects of people's private life. The concern over privacy-invading data collection that enables sensitive inferences is a common factor across these concerns. When shared with third parties like insurance companies, financial institutions, or employers, these conclusions can result in prejudice (Watcher, 2019).

4. Aspects of GDPR that provide solutions in respect to Big Data

In contrast to past legislation, the GDPR appears to address the need for individual control more directly and prudently (Van Ooijen & Vrabec, 2019). In fact, as specifically stated in the policy texts that came before the GDPR proposal (Reading 2011) and in the GDPR's actual wording, one of the main objectives of the EU legislator was to increase individual control. Behavioral scientists frequently criticize the GDPR for not being able to adequately handle these threats, despite the fact that it places a lot of emphasis on the control of data subjects (Van Ooijen & Vrabec, 2019, 92).

In view of the changes described in the introduction, GDPR appears to offer the ideal remedy because its authority extends well beyond the EU's borders (Mikkelsen, Soller and Strandell-Jansson, 2017).

Some of the declared goals of data protection under the GDPR are under tension. The GDPR's goal is to defend EU people' rights to privacy, not to halt data transfers within the EU or even across international borders (Richards, 2022). As opposed to this, the GDPR ensures that the fundamental right to data protection is upheld by regulating the inevitable flow of personal data in a reasonable, ethical, and responsible manner. It does this by acknowledging that the flow of personal information is one of the pillars of both the EU economy and the global economy (Richards, 2022).

It is crucial to remember that GDPR's most important end goal is privacy (even though the law as a whole does not use that word explicitly) and how to handle the protection of sensitive data when processing client and employee data. That is not all, though. The issues of "trust" and "risk" are in addition to the "privacy" component. In essence, these three ideas can be regarded as the foundation of GDPR (Larson et al, 2019). We will now provide a closer look to these.

Privacy

The GDPR represents a significant advancement in safeguarding global citizens' privacy interests against data practices that are intriguing, excessive, dangerous, or damaging (Nordstrom, 2020). Depending on the setting, the word "privacy" has been used to indicate many things. Control, use, and disclosure of personal information are at the core of the privacy issue. Since privacy may be thought of as existing on a continuum, a person's level of privacy can either rise or fall. According to the circumstance and the individual's preferences, privacy may also be something that the person chooses to give up, and in varying degrees, frequently in return for benefits they believe to be beneficial (Larson et al, 2019).

Fundamentally, the GDPR attempts to permeate privacy while simultaneously allowing many sectors to contribute to new norms and best practices that may be applicable to the new particular circumstances, which are frequently digitized. Privacy is crucial because breaches of privacy, or "privacy harms," are far more obvious in a digitalized world. Therefore, there are many ways that someone's privacy could be invaded. Six inherent concerns of breaching personal privacy include (Larson et al, 2019):

- **Discrimination.** Actors can leverage the public and commercial sector's use of predictive analytics to make judgments about a person's tendency to fly, get employment, receive clearances, or be approved for a credit card. Predictive analytics' usage of associations may have adverse impacts on some persons, which may result in exclusion from certain services.
- **Breach embarrassment.** This includes data breaches at different companies and organizations that may expose the private information of thousands of clients,

patients, employees, etc. The rate of identity theft and credit card fraud is also at a record high.

- **Elimination of anonymity.** It is possible to integrate data sets, barring regulations for anonymized data files. By integrating different groups of data, this might make it possible to re-identify specific individuals depending on the circumstances.
- **Public sector exemptions.** As an illustration, several government databases will gather personal data. Name, possible aliases, ethnicity, gender, date of birth, social security number, passport and driver's license numbers, home address, phone numbers, pictures, fingerprints, various financial details including bank accounts, employment, and company information, etc. are all included in this.
- **Data exchange.** Consumer profiles that are not expressly protected by the legal frameworks will sometimes be collected and sold by businesses. Data files used for different types of big data analysis may contain inaccurate information about specific people.. They might also use flawed algorithms or data models that are flawed in how they connect to specific people.
- **Data interpretation errors.** While it is possible to find a wide range of political viewpoints on different social media sites, these opinions may not accurately reflect the views of voters. To this end, it has been established that a large portion of tweets and Facebook posts regarding politics throughout the world have actually been generated by computers.

Trust

The concept of "trust" implies that a person has confidence in different data controllers (those in charge of all the personal data an organization holds) to handle personal data fairly and professionally (Buttarelli, 2016; European Commission, 2019). A significant portion of this data consists of vital documents that enable people, businesses, and even governments to continue operating (Larson et al, 2019). This includes, but is not limited to, ownership and land records, financial records, legal papers, contracts, and records of identity and vital statistics. Additionally, Internet of Things related records will be kept on cloud storage platforms (Larson et al, 2019). Even in the absence of malicious intent, there are reasons to doubt about the security of cloud-based record-keeping. This covers concerns including controlling the flow of transnational data, determining who is responsible for data breaches,

and ensuring a fair mechanism for when a cloud service provider stops operating or performing its functions (Larson et al, 2019). Building trust is therefore crucial for service providers that want to draw clients given the potential threats. Building a relationship of trust with the service provider is equally important for people who prefer to use that service over others because they want to maximize their benefits while minimizing any perceived risks (Larson et al, 2019).

Risk

A scenario where there is a chance of losing control over one's personal information is referred to as a privacy risk. Specifically, when someone uses such information about you without your knowledge or consent. As a result, privacy threats arise whenever a party gathers, uses, shares, or manages personal data relevant to their employees, clients, guests, consumers, students, etc. "Likelihood" and "severity" are two distinct values that are frequently used to describe risks. The likelihood that the processing system may cause harm is how the term "likelihood" is defined in the context of privacy. Furthermore, "severity" describes the extent of the effect on the victims (Larson et al, 2019).

The OECD 2016 (The Organisation for Economic Co-operation and Development) states that although the risk can be reduced by using digital risk management, it is difficult to completely eliminate digital security risk when carrying out actions that depend on the digital environment. In light of this, it is the responsibility of Europe's independent data protection authorities to promote risk management through the implementation of GDPR, ensuring accountability and transparency to all people and organizations (Buttarelli, 2016). However, some sorts of businesses may find the cost of GDPR compliance to be excessive. As a result, it's crucial to identify which organizations will be affected by the adoption of the GDPR framework and which would benefit from it (Larson et al, 2019).

Other Solutions

Businesses benefit from GDPR

With the implementation of GDPR, the market has been oversaturated with companies claiming to offer solutions and services that are GDPR compliant. These solutions typically

aim to structure the documentation on the database systems the organization intends to use, as well as assist the organization in keeping track of the types of data that are kept in each database section, the classification the data has, and the justification for storing each specific dataset, along with the procedures for data cleansing (Ashton, 2018).

Companies that offer cloud storage are another industry segment that benefits from the implementation of GDPR. In this usage, "cloud storage" refers to a computer data storage strategy that groups digital information into logical pools. The physical environment is often owned and maintained by a hosting provider, and the physical storage spans numerous servers (perhaps even in various regions). The benefits of choosing a cloud-based solution include greater IT resource utilization because cloud solutions offer nearly unlimited scalability, excellent flexibility, and are typically also cost-effective. Additionally, a lot of GDPR-compliant agreements, such as those between the data controller and the data processor, are stored using cloud services. As a result, it is reasonable to assume that more cloud service providers will emerge in the future, positioning themselves to protect GDPR data (Larson et al, 2019).

Without a doubt, the implementation of GDPR is intended to benefit legal professionals and attorneys' companies. Legal consultants or lawyers will help businesses comprehend the consequences of GDPR for their individual enterprise while assuring full GDPR-compliance. Software and automated digital solutions may guide enterprises through the process effortlessly and make data processing manageable (Larson et al, 2019).

A new occupation known as "data protection officer" (DPO) has emerged as a result of GDPR. A person with this title is responsible for making sure the company manages data in a way that complies with GDPR requirements. That is not to imply that a DPO is a necessary position in every company. Instead, the necessity for a DPO depends on a number of variables, including organizational size, if the organization is an official body, whether its primary function is the extensive or regular monitoring of personal data, etc (Larson et al, 2019).

Due to new opportunities created by GDPR, digital strategy consultants may find themselves performing on new projects. This is especially true given the urgent requirement for complete GDPR compliance across all organizational processes that many businesses are currently

experiencing. Other professions that lean toward structuring or laying the foundations for data management typically stand to gain from GDPR. Professionals in the fields of software architects, solution architects, software developers, and data analysts may fall under this category (Larson et al, 2019).

Customers will unquestionably demand "smarter" goods and services as a result of the digitalization process, which is made possible by businesses' ability to merge various pools of customer data. To that degree, businesses will find themselves walking a fine line between providing perceived value to customers and acting in a way that doesn't make them feel concerned or monitored. That is to say, maintaining honesty and trust will be essential for businesses looking to win over customers. In this regard, the GDPR may contribute to enhancing the relationship of trust between businesses and their clients by giving clients the freedom to decide who they want to share their information with and the ability to revoke their consent whenever they see appropriate (Larson et al, 2019).

Permanent Data Storage

Insights from data have long been used by computing systems. However, with the extensive use of machine learning and big data analytics in system design, this dependence is rising to unprecedented heights, particularly in this decade. Naturally, technology businesses have grown to aggressively capture customer data as well as permanently store it. GDPR stipulates that no data may exist in permanence (Shastri et al, 2019).

Article 17 of the GDPR gives users the unrestricted right to demand that their personal data be deleted promptly from every system component. Along with this, articles 5 and 13 outline additional duties for the data controller: (i) users should be informed at the time of collection of how long their personal data will be kept, and (ii) if the personal data is no longer required for the purpose for which it was collected, it should be erased (Shastri et al, 2019).

Random Data Reuse

When building software systems, programs and models are often given a purpose, but data is seen as a supporting resource that aids these high-level entities in achieving their objectives. Data can be used freely and interchangeably between different systems due to this representation of it as an inert entity. As an illustration, this has made it possible for companies like Google and Amazon to collect user data just once and use it to tailor the user experience across a variety of services (Shastri et al, 2019). GDPR rules, however, forbid this action:

Article 5(1)(b): Purpose limitation. “Personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes”
Article 6: Lawfulness of processing. “(1)(a) Processing shall be lawful only if the data subject has given consent to the processing of his or her personal data for one or more specific purposes”
Article 21: Right to object. “(1) The data subject shall have the right to object at any time to processing of personal data concerning him or her”

The first two articles make it abundantly clear that personal data may only be gathered for those purposes and for no other. Then, article 21 gives users the right to object, at any time, to the use of their personal data for any purpose, including profiling, marketing, and research. Together, these articles mandate that every piece of personal data have its own blacklisted and whitelisted purposes, which are subject to change (Shastri et al, 2019).

The Black Market

The standards for collecting, sharing, and selling personal data are still developing as we are in the early stages of its widespread commoditization. As a result, there are now uncertainties for the general public and conflict among controllers for control of the data. People are worried about lack of visibility once their data is shared or sold in the secondary markets. In response, businesses have created walled gardens and increased the level of obscurity in secondary marketplaces. The GDPR, however, eliminates these procedures (Shastri et al, 2019).

Article 20: Right to data portability. “(1) The data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller the right to have the personal data transmitted directly from one controller to another” (Shastri et al, 2019).

Article 14: Information to be provided where personal data have not been obtained from the data subject. “(1) (c) the purposes of the processing, (e) the recipients, (2) (a) the period for which the personal data will be stored, (f) from which source the personal data originate (3) The controller shall provide the information at the latest within one month” (Shastri et al, 2019).

People have the right to request all personal data that a controller has obtained directly from them under Article 20. Additionally, they could request that the controller send all such personal data immediately to a different controller. Anyone who obtains personal information unintentionally is required to notify the users within a month of (i) how they obtained it, (ii) how long it will be kept, (iii) how it will be used, and (iv) with whom they want to share it (Shastri et al, 2019).

5. Limits of the GDPR in addressing Big Data

Examining a few of the GDPR's provisions will highlight the fact that the regulation has a number of fundamental cross-cutting restrictions. First, the use of personal data is typically made lawful with the consent of the data subject, which is the finest example of how the GDPR is based in part on the idea of individual self-determination. However, there are many reasons to doubt whether people are actually in a position to comprehend modern data ecosystems and the consequences of their consent. Furthermore, the GDPR's emphasis on individual rights falls short of capturing these communal effects, as well as the effect of individual permission on the public interest, given that data analytics also has enormous collective implications through reshaping political, economic, and social landscapes. Second, there are many problems with enforcement. The GDPR is based on the notion that the individual takes a proactive stance in enforcing her rights; however, in practice, individuals rarely do so, and even if they did, it would be difficult, if not impossible, to implement their enforceability without tearing down the entire model that has been used to collect their data.

Thirdly, the GDPR is founded on outdated notions of personal data and associated analysis techniques (Finck, 2021).

The Problems with Personal Autonomy and the "I agree"

According to the GDPR, there are several legal justifications for "processing" personal data, including when it is necessary for a contractual arrangement (such as a contract between a bank and its client). The processing of personal data may also be justified by the consent of the data subject. The most common legal foundation for customization seems to be consent. As personalization cannot typically be assumed to be part of the original objective of data processing (the delivery of the service), it is frequently necessary. In fact, "collection of organizational metrics relating to service or details of user engagement, cannot be regarded as necessary for the provision of the service as the service could be provided without the processing of such personal data," in the majority of situations (Finck, 2021, 2-6). As a result, consent "would almost always be required," especially for digital market research that uses tracking, direct marketing, behavioral advertising, data brokers, and location-based advertising (Finck, 2021, 2-6).

Although it is common for data subjects to give their agreement for the use of their personal information for customization, it is not always obvious if they fully comprehend the implications of that decision. By agreeing to a company's terms of service (without which they cannot use the service), or by selecting "I agree" on cookie alerts that appear while visiting a website, users give their consent to the collection and analysis of their personal data. The e-Privacy Regulation which is yet to be introduced, offers individuals more control over their personal data, reinforces their rights to privacy online. Additionally, it imposes further restrictions on businesses that gather or process this data. It governs cookies, although it adopts the GDPR's requirement of permission. Users acknowledge that their information may be used for customization, such as "providing recommendations, personalized content, and customized search results," as well as personalized advertisements, when they accept Google's terms of service, for instance. Facebook users consent to the usage of their personal information to "personalize features and content and make suggestions for you on and off our products" when they sign up for the service. Therefore, consumers frequently unintentionally consent that their personal data, including details about their online behavior on- and off-site, is used to drive personalization by simply using an online service or by clicking away from a cookie banner to visit a website (Finck, 2021).

A number of requirements set forth by the GDPR for valid consent are difficult to meet in scenarios where customization is taking place (Finck, 2021). According to the Regulation, consent must be "clearly affirmatively given" and must be "a freely given, specific, informed, and unambiguous indication of the data subject's agreement to the processing of personal data.". Recently, the CJEU declared that pre-ticked boxes cannot be used to claim that consent was actively granted (Case C-673/17). However, if we take into account the numerous diverse processes and players engaged in the data ecosystems that support personalization, the practical realization of these needs is extremely unattainable (Finck, 2021).

We highlighted the fact that even when a data subject is unaware that they are truly consenting to such processing or what the ramifications of processing are, permission is frequently claimed to justify personalization. Indeed, in today's intricate data ecosystems, the emphasis on informational self-determination, upon which consent is built, has shown to have major limitations (Finck, 2021).

Mechanisms for Personal Data Control

Prior to being used, recommendation algorithms are trained on training data, which is frequently personal data. This invites the question of how such learning algorithm properties might be harmonized with data subject rights that involve the modification or deletion of personal data. There are in fact a variety of circumstances in which the GDPR demands that data be changed or stopped being processed, including the rights to correction and erasure as well as the data subject's right to refuse permission to the processing of personal data (Finck, 2021).

According to Article 16 of the GDPR, data subjects have the right to ask for the correction of any "inaccurate" personal information relating to them. This can be accomplished by correcting erroneous data while taking the purposes of processing into consideration, for example by offering a supplemental statement. The data controller should ensure that the personal data is corrected when a data subject makes use of her rights under Article 16 and the claim is justified. The data subject might require their information corrected because it is incorrect and because the customized suggestions they receive are unhelpful since they are

based on false assumptions. A data subject might discover, for instance, that incorrect information about them has been processed, such as the fact that they actually live in Sweden rather than Spain. This request could be made out of a simple desire to stop incorrect information about them from spreading or to cease receiving tailored marketing for Spanish events (Finck, 2021).

Additionally, Article 17 GDPR states that data subjects may, under certain conditions, request the deletion of their personal information (the ‘right to be forgotten’). It is unclear whether Article 17 covers observed and inferred data, such as user behavior on a platform, when the data subject wants to exercise their right to erasure. The Article 29 Working Party excluded inferred data from the purview of the right to data portability, regardless of the fact that there is no conclusive answer to that query. However, this was done on the justification that such information is not "provided by the data subject" as stated in Article 20 GDPR, which is why it cannot be applied to Article 17 (Finck, 2021).

In general, data protection law is individual-focused, but it now needs to oversee an interconnected, collective-focused system. As the profile of the data subject is based on information about the behavior or preferences of other people who have been recognized and categorized using training data, the collected data has an undeniably collective dimension. The GDPR's emphasis on the person provides few remedies for collective algorithmic harms, which is once again highlighted by enforcement difficulties (Finck, 2021).

'Proxy' Data: Sensitive Inferences from Non-Sensitive Data

The GDPR provides a more protective regime for what are known as "special categories" or "sensitive data," but it also applies to all personal data. This includes health information, as well as information about a person's sexual life or sexual orientation, as well as personal information that reveals racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership (Article 9). The reasoning behind this distinct category is that since these issues are extremely personal, they should be given more protection. Sensitive information is nevertheless used in the framework of recommender systems in a variety of situations. For instance, research reveals that factors like color, gender, and religion are taken into account by the Facebook algorithm when determining which advertisements

are shown to a particular user (Finck, 2021). As a result and example, more women have been exposed to job listings for secretaries and preschool teachers, while men were more frequently exposed to adverts for taxi drivers (Finck, 2021).

The processing of such data is, in theory, banned under Article 9(1) GDPR unless the data subject "has given explicit consent" or in a number of other circumstances that are less relevant to personalization. If computational intelligence is applied, it can be challenging to determine whether a given data piece qualifies as sensitive data, which determines whether this special regime applies. According to Article 9 of the GDPR, data can be neatly divided into special category data and regular personal data, which are both subject to the general GDPR regime. However, such simplicity is now a thing of the past due to big data analytics. To categorize individuals according to preferences and frequently sell connected profiles for advertising purposes, personalization algorithms may use explicitly sensitive data (such as data regarding sexual orientation if this has been explicitly published, for example on a dating platform). Machine learning, however, is also capable of inferring sensitive information from seemingly unsensitive data. “In this case, ‘ordinary personal data’ becomes proxy data for sensitive data” (Finck, 2021).

Several consequences occur if proxy data is considered to be personal data. First, it is unclear if such data should be classed as sensitive data right away or only after a sensitive attribute has been inferred, and if so, whether this inference needs to be supported by a particular degree of confidence. Once data is classified appropriately, data controllers must comply with the GDPR regime on sensitive data. This, for example, suggests a requirement to do a Data Protection Impact Assessment, which is often necessary if processing personal data puts data subjects at "high risk". Additionally, the legal requirements for obtaining data subject consent and implementing Article 22 GDPR are also subject to change (Finck, 2021).

While the standard for valid consent is high for personalization and frequently not satisfied, it is considerably higher for sensitive data because the GDPR stipulates that "explicit" consent must be granted in this case. Although the term "explicit consent" is not defined in the Regulation, it is generally accepted that it must meet the standard requirements for valid consent (i.e., be precise, informed, and unambiguous) as well as be "affirmed in a clear statement (whether oral or written)". It is reasonable to assume that many data controllers are not GDPR compliant given the requirement for such specific agreement for both the

non-sensitive proxy data that results in a sensitive inference and other sensitive source data that informs a personalized service. Further explanation of these issues is necessary in circumstances where a fair outcome of the processing of special category data is crucial. The inadequate and out-of-date nature of the GDPR's data categories is clearly demonstrated by these circumstances. The examples of sensitive information are not only limited and leave out some of the data that can be used to draw highly sensitive conclusions about an individual, but also the application of the current regime may result in the classification of large amounts of personal data as sensitive data, which was probably not what the lawmakers had in mind (Finck, 2021).

A closer look at the GDPR's consent provisions, regulations for the rectification and deletion of personal data, and special category data in the context of machine learning-driven personalization has exposed some basic flaws in the EU data protection system. The GDPR's emphasis on informational self-determination clashes with people's limited comprehension of today's complex data ecosystems and the multiple barriers to opting out of a customization scheme. Furthermore, while the GDPR is firmly rooted in the idea of individual rights, data analysis in the context of personalization has a strong collective dimension that can reveal highly sensitive information about the person that would not be obvious to a human observer based solely on her own personal data. This then violates the purpose of classifying personal data under GDPR into more sensitive and less sensitive data. The common dimension also affects how data protection rights are enforced, making it challenging to "extract" the individual from the group or personal information from big data (Finck, 2021).

This section briefly discussed some of the GDPR's restrictions and limitations. Beyond this, it may be necessary to take legislative action (either in the form of new legislation or a revision of the GDPR) to address concerns about collective algorithmic harm or other legal issues relating to computational intelligence that are not currently addressed by the GDPR (Finck, 2021).

6. Other legal approaches to the right to privacy and personal data protection

CCPA

Just like the GDPR in the European Union, to protect its citizens, the state of California has taken a number of legislative and regulatory actions as well. CCPA stands for the California Consumer Privacy Act of 2018. It has been effective from January 1, 2020, and is the first law of its kind in the United States.

The Act aims to offer Californians more control over the personal data that businesses gather and process. Transparency, control, and accountability are the main goals of the CCPA. According to the new law, Californians will now have the right to know what personal data is being collected by businesses like Facebook and Google, a way to stop the sharing or selling of personal information, and the power to file lawsuits against companies that violate the CCPA if data breaches, or privacy violations occur (Li, 2020).

The purpose of passing this legislation was to guarantee Californians' belief in the protection of basic human rights. According to the lawmakers, implementing an accountability system for data holders will increase the security of existing privacy. It covers the reason why businesses retain customer information and how they use it for marketing. To guarantee data analysis and system improvement was a further significant feature (Batool, 2023).

Consumers are specifically granted the following rights: the right to know what personal information businesses are collecting about them and how that information is being used and shared; the right to have that information deleted from the business's records; the right to object to the sale of personal information; and the right to privacy rights to be exercised without being subjected to service or price discrimination (Li, 2020).

By taking these actions, the person will have the certainty that his personal information won't be disclosed and that there is a suitable venue where complaints about data breaches can be addressed. A number of the CCPA provisions are connected to GDPR articles. As a result, it

is regarded as the State of California's most concise and reliable piece of legislation in the area of protecting personal data (Batool, 2023).

Businesses are required by the 2021 Amendments to make it simple for customers to opt out, and the opt-out procedure cannot be intended to undermine or restrict a customer's decision to opt out. According to the revised regulations, a consumer must confirm the authorized agent's capacity to act before a business can confirm the agent's right to act on their behalf. Disclosures for customers under the age of sixteen are also included in the 2021 Amendments (Shatz, 2021).

DPA

The amended Data Protection Act, also known as the Data Protection Act 2018 (DPA 2018), was adopted by the UK government on May 24, 2018, one day before the "EU-GDPR" became a legal requirement throughout all of Europe. All of the provisions from the "EU-GDPR" were incorporated into the DPA 2018, which was a complete revision of the previous "Data Protection Act" of 1998. As a result, this legislation became the standard by which "Data Protection" would be evaluated in the United Kingdom (*Uk Gdpr Updated for Brexit | Uk Gdpr*, 2021).

The DPA 2018 governs how private and public businesses, law enforcement organizations, and intelligence agency agencies process the personal data of persons. It accomplishes this by establishing regulations for the processing of general data, data used by law enforcement, data used by intelligence services, and for the regulatory oversight and enforcement of regulations by the national supervisory body, the Information Commissioner's Office (Mc Cullagh et al., 2019).

The UK Government has been praised for seizing the chance to stay in compliance with the GDPR both before and after Brexit, especially given that the EU is promoting it as the "gold standard" in the world for data protection laws. Given that the processing of personal data supports the UK economy, it sends a crucial message to businesses and individuals that the UK intends to uphold high standards of protection regarding such processing. However, adopting the GDPR into UK legislation after the country has left the EU won't be sufficient in and of itself to convince the European Community that a judgment of adequacy should be made with regard to the UK (Mc Cullagh et al., 2019).

The UK will need to make changes to the DPA 2018 and national surveillance laws in order to receive an adequacy judgment, but these adjustments are necessary to ensure regulatory coherence and frictionless European Economic Area and UK trade in personal data (Mc Cullagh et al., 2019).

Conclusion

In this study, we have reviewed the terminologies and ideas around big data, including terms such as privacy, datafication, algorithms, artificial intelligence, and surveillance capitalism. The discussion then focused on the benefits that big data can bring to a society. After having discussed the harms of big data, with reference to the Cambridge Analytica case, we highlighted privacy exploitation. Subsequently, we proceeded to methods that maintain security for big data risks. In the last chapter, we examined the GDPR's impact while debating its advantages and disadvantages.

Big data's emergence has not only opened up significant doors for societal advancement but has also exposed society to numerous information security risks, raising questions about how to protect individuals' privacy.

Big data is prevalent in people's daily lives, but it can improve decision-making in society to a level that has never been possible. Big data has an impact on society for two primary reasons: it makes it possible to make unprecedented predictions about people's private lives and it transfers power from those who gather information to those who keep it (Hijmans, 2016).

Thus, the concept of privacy needs to be reconsidered in the era of big data. Access to information can be restricted, and the flow of personal data can be controlled, according to the concept of privacy. How new personal information is created by businesses and organizations through predictive analytics is an important concern in the era of big data (Mai, 2016).

In the digital age we currently live in, big data is an essential part of most of our daily lives and it has a significant presence across a wide range of industries that vary from monitoring diseases and health care to expanding technology-assisted education. These advantages of course, come with a risk. Since big data has dominated practically every aspect of our lives and businesses, data rights have grown to be a more serious concern.

The GDPR on the other hand, is a significant and ambitious attempt to establish a comprehensive, consistent standard for online privacy and data protection. It is a complicated

enforcement system that solves complex issues, and it will continue to develop over time. Right now, its primary goal is educational, requiring transparency in the name of informing citizens about how their data is used. And since its adoption, it has been rather effective at drawing attention to unethical activities that were previously only well-known to academics and tech specialists. Additionally, the GDPR may prove to be a helpful tool for checking and restraining the most serious violations and exploitations.

Despite all of its benefits, it is crucial to remember that the GDPR offers limited solutions to challenge current structures. The GDPR could undoubtedly increase the influence of companies or reinforce the troubling data usage practices that it was designed to address in the first place. After 5 years of existence, it appears that the GDPR has not been able to lessen the dominance that companies currently hold over the acquisition and use of data. And to be completely honest, additional measures are required if that is what must be done.

To conclude, several ethical, social, and policy issues, risks, and potential barriers are simultaneously raised by big data activities. Concerns about data ownership, the "datafication" of society, privacy concerns, a possible trade between privacy and data analytics advancement, dataveillance, discriminatory practices are just a few of the issues, as mentioned earlier.

In order to create a big data that is fully respectful of human dignity and citizens' rights and capable of further development in an ethically acceptable manner, these and related issues require greater ethical engagement and reflection within the framework of an interdependent ecosystem composed of different and complementary competences.

We have argued that a number of GDPR regulations act as practical initiatives to increase individual control. It is crucial that attorneys, policymakers, and scholars continue to collaborate as these instruments evolve in order to enhance the effectiveness of future data protection law in terms of preserving privacy rights of the individuals.

References

Adadi, A. (2021) A survey on data‐efficient algorithms in big data era. *J Big Data* 8, 24 . https://doi.org/10.1186/s40537-021-00419-9

Afriat, H., Dvir-Gvirsman, S., Tsuriel, K., & Ivan, L. (2021). "This is capitalism. It is not illegal": Users' attitudes toward institutional privacy following the Cambridge Analytica scandal. *The Information Society*, *37*(2), 115–127. https://doi.org/10.1080/01972243.2020.1870596

Almeida, F. L. F., & Calistru, C. (2013). The main challenges and issues of big data management. *International Journal of Research Studies in Computing*, *2*(1). https://doi.org/10.5861/ijrsc.2012.209

Altshuler, T. S. (2019, September 26). Privacy in a digital world. *TechCrunch*. https://techcrunch.com/2019/09/26/privacy-queen-of-human-rights-in-a-digital-world/

Andrejevic, M. (2009). Privacy, exploitation, and the digital enclosure. *Amsterdam Law Forum*, *1*(4), 47. https://doi.org/10.37974/ALF.86

Andrews, L. (2019). Public administration, public leadership and the construction of public value in the age of the algorithm and 'big data.' *Public Administration*, *97*(2), 296–310. https://doi.org/10.1111/padm.12534

Baig, M.I., Shuib, L. & Yadegaridehkordi, E. Big data in education: a state of the art, limitations, and future research directions. *Int J Educ Technol High Educ* 17, 44 (2020). https://doi.org/10.1186/s41239-020-00223-0

Batko, K., Ślęzak, A. The use of Big Data Analytics in healthcare. *J Big Data* 9, 3 (2022). https://doi.org/10.1186/s40537-021-00553-4

Bélanger, F., & Crossler, R. E. (2011). Privacy in the Digital Age: A Review of Information Privacy Research in Information Systems. *MIS Quarterly*, *35*(4), 1017–1041.

Big data benefits for SMEs—Dataconomy. (2022, March 10). https://dataconomy.com/2022/03/10/big-data-benefits-for-smes/

Brown, A. J. (2020). “Should I stay or should I leave? ”: Exploring(Dis) continued facebook use after the cambridge analytica scandal. *Social Media + Society*, *6*(1), 205630512091388. https://doi.org/10.1177/2056305120913884

Burrell, J., & Fourcade, M. (2021). The society of algorithms. *Annual Review of Sociology*, *47*, 213-237

Chaston, I. (2015). *Internet Marketing and big data exploration*. Palgrave Macmillan.

Clarke, N., Vale, G., Reeves, E. P., Kirwan, M., Smith, D., Farrell, M., Hurl, G., & McElvaney, N. G. (2019). GDPR: An impediment to research? *Irish Journal of Medical Science (1971 -)*, *188*(4), 1129–1135. https://doi.org/10.1007/s11845-019-01980-2

Cukier, Kenneth, and Viktor Mayer-Schoenberger. “The Rise of Big Data: How It’s Changing the Way We Think About the World.” Foreign Affairs 92, no. 3 (2013): 28–40.

Custers, B. (2022). New digital rights: Imagining additional fundamental rights for the digital era. *Computer Law & Security Review*, *44*, 105636. https://doi.org/10.1016/j.clsr.2021.105636

Custers, B.H.M., and Malgieri, G. (2022) Priceless data: why the EU fundamental right to data protection makes data ownership unsustainable, Computer Law & Security Review, Vol. 45, p. 1-13, https://doi.org/10.1016/j.clsr.2022.105683.

D. E. O'Leary, "Big Data and Privacy: Emerging Issues," in *IEEE Intelligent Systems*, vol. 30, no. 6, pp. 92-96, Nov.-Dec. 2015, doi: 10.1109/MIS.2015.110

De Carvalho, R. M., Del Prete, C., Martin, Y. S., Araujo Rivero, R. M., Önen, M., Schiavo, F. P., Rumín, Á. C., Mouratidis, H., Yelmo, J. C., & Koukovini, M. N. (2020). Protecting citizens' personal data and privacy: Joint effort from GDPR EU cluster research projects. *SN Computer Science*, *1*(4), 217. https://doi.org/10.1007/s42979-020-00218-8

Del Vecchio, P., Di Minin, A., Petruzzelli, A. M., Panniello, U., & Pirri, S. (2018). Big data for open innovation in SMEs and large corporations: Trends, opportunities, and challenges. *Creativity and Innovation Management*, *27*(1), 6–22. https://doi.org/10.1111/caim.12224

Devane, H. (2021, April 14). *Everything you need to know about k-anonymity*. Immuta. https://www.immuta.com/blog/k-anonymity-everything-you-need-to-know-2021-guide/

Dexe, J., Franke, U., Söderlund, K., Van Berkel, N., Jensen, R. H., Lepinkäinen, N., & Vaiste, J. (2022). Explaining automated decision-making: A multinational study of the GDPR right to meaningful information. *The Geneva Papers on Risk and Insurance - Issues and Practice*, *47*(3), 669–697. https://doi.org/10.1057/s41288-022-00271-9

Dijck, J.V. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *surveillance and society, 12*, 197-208.

Economides, N., & Lianos, I. (2020). Restrictions on privacy and exploitation in the digital economy: A market failure perspective. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3686785

Erdos, David, Accountability and the UK Data Protection Authority: From Cause for Data Subject Complaint to a Model for Europe? (January 17, 2020). University of Cambridge Faculty of Law Research Paper No. 14/2020, Available at SSRN: https://ssrn.com/abstract=3521372 or http://dx.doi.org/10.2139/ssrn.3521372

Felzmann, H., Villaronga, E. F., Lutz, C., & Tamò-Larrieux, A. (2019). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society*, *6*(1), 205395171986054. https://doi.org/10.1177/2053951719860542

Finck, M. (2021). Hidden Personal Insights and Entangled in the Algorithmic Model: The Limits of the GDPR in the Personalisation Context. In U. Kohl & J. Eisler (Eds.), *Data-Driven Personalisation in Markets, Politics and Law* (pp. 95-107). Cambridge: Cambridge University Press. doi:10.1017/9781108891325.008

Finck, M., & Pallas, F. (2019). They who must not be identified—Distinguishing personal from non-personal data under the gdpr. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3462948

Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., Slater, S., Baker, R., & Warschauer, M. (2020). Mining big data in education: Affordances and challenges. *Review of Research in Education*, *44*(1), 130–160. https://doi.org/10.3102/0091732X20903304

Floridi, L. (2018). Soft ethics, the governance of the digital and the general data protection regulation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2133), 20180081. https://doi.org/10.1098/rsta.2018.0081

Floridi, L. Open Data, Data Protection, and Group Privacy. *Philos. Technol.* 27, 1–3 (2014). https://doi.org/10.1007/s13347-014-0157-8

Foote, K. D. (2016, March 22). Techniques and algorithms in data science for big data.*DATAVERSITY*. https://www.dataversity.net/techniques-and-algorithms-in-data-science-for-big-data/

Gerards, J. (2019). The fundamental rights challenges of algorithms. *Netherlands Quarterly of Human Rights*, *37*(3), 205–209. https://doi.org/10.1177/0924051919861773

Gulson, K. N., Sellar, S., & Webb, P. T. (2022). *Algorithms of Education: How Datafication and Artificial Intelligence Shape Policy*. University of Minnesota Press. http://www.jstor.org/stable/10.5749/j.ctv2fzkpxp

Hijmans, H. (2016) “Internet and Loss of Control in an Era of Big Data and Mass Surveillance.” In The European Union as Guardian of Internet Privacy: The Story of Art 16 TFEU, 1st ed., 77–123. Law, Governance and Technology Series 31. Springer International Publishing

Hinds, J., Williams, E. J., & Joinson, A. N. (2020). “It wouldn't happen to me”: Privacy concerns and perspectives following the Cambridge Analytica scandal. *International Journal of Human-Computer Studies*, *143*, 102498. https://doi.org/10.1016/j.ijhcs.2020.102498

Hoofnagle, C. J., Van Der Sloot, B., & Borgesius, F. Z. (2019). The European Union general data protection regulation: What it is and what it means. *Information & Communications Technology Law*, *28*(1), 65–98. https://doi.org/10.1080/13600834.2019.1573501

Iqbal, M., Kazmi, S. H. A., Manzoor, A., Soomrani, A. R., Butt, S. H., & Shaikh, K. A. (2018). A study of big data for business growth in SMEs: Opportunities & challenges. *2018 International Conference on Computing, Mathematics and Engineering Technologies (ICoMET)*, 1–7. https://doi.org/10.1109/ICOMET.2018.8346368

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

Kavanagh, C. (2019). Artificial Intelligence. In *New Tech, New Threats, and New Governance Challenges: An Opportunity to Craft Smarter Responses?* (pp. 13–23). Carnegie Endowment for International Peace. http://www.jstor.org/stable/resrep20978.5

Kulhari, S. (2018). Data Protection, Privacy and Identity: A Complex Triad. In Building-Blocks of a Data Protection Revolution: The Uneasy Case for Blockchain Technology to Secure Privacy and Identity (1st ed., pp. 23–37). Nomos Verlagsgesellschaft mbH.

Kutyłowski, M., Lauks-Dutka, A., & Yung, M. (2020). Gdpr–challenges for reconciling legal rules with technical reality. In Computer Security–ESORICS 2020: 25th European Symposium on Research in Computer Security, ESORICS 2020, Guildford, UK, September 14–18, 2020, Proceedings, Part I 25 (pp. 736-755). Springer International Publishing.

Lane, J. I. (Ed.). (2014). *Privacy, big data, and the public good: Frameworks for engagement*. Cambridge University Press.

Larsson, A., & Teigland, R. (Eds.). (2019). The Digital Transformation of Labor: Automation, the Gig Economy and Welfare (1st ed.). Routledge. https://doi.org/10.4324/9780429317866

Li Y., *The California Consumer Privacy Act of 2018: Toughest U.S. Data Privacy Law with Teeth?*, 32 Loy. Consumer L. Rev. 177 (2020).
Available at: https://lawecommons.luc.edu/lclr/vol32/iss1/6

Li, H., Yu, L., & He, W. (2019). The impact of gdpr on global technology development. *Journal of Global Information Technology Management*, *22*(1), 1–6. https://doi.org/10.1080/1097198X.2019.1569186

Löfgren, K., & Webster, C. W. R. (2020). The value of Big Data in government: The case of 'smart cities.' *Big Data & Society*, *7*(1), 205395172091277. https://doi.org/10.1177/2053951720912775

Lycett, M. (2013). 'Datafication': Making sense of (Big) data in a complex world. *European Journal of Information Systems*, *22*(4), 381–386. https://doi.org/10.1057/ejis.2013.10

Mai, J.-E. (2016). Big data privacy: The datafication of personal information. *The Information Society*, *32*(3), 192–199. https://doi.org/10.1080/01972243.2016.1153010

Marciano, A., Nicita, A., & Ramello, G. B. (2020). Big data and big techs: Understanding the value of information in platform capitalism. *European Journal of Law and Economics*, *50*(3), 345–358. https://doi.org/10.1007/s10657-020-09675-1

Mason, S. (2016). Data protection. In *Electronic Signatures in Law*(pp. 387–396). University of London Press. http://www.jstor.org/stable/j.ctv5137w8.23

Mayer-Schönberger, V., and Cukier K. (2013). *Big data: a revolution that will transform how we live, work and think*. London: John Murray Publishers.

Mc Cullagh, K., Tambou, O., & Bourton, S. (2019). National Adaptations of the GDPR. https://wp.me/p6OBGR-3dP

McDermott, Y. (2017). Conceptualising the right to data protection in an era of Big Data. *Big Data & Society*, *4*(1), 205395171668699. https://doi.org/10.1177/2053951716686994

Mehmood, A., Natgunanathan, I., Xiang, Y., Hua, G., & Guo, S. (2016). Protection of big data privacy. *IEEE Access*, *4*, 1821–1834. https://doi.org/10.1109/ACCESS.2016.2558446

Mooij, A. (2023). Reconciling transparency and privacy through the European Digital Identity. *Computer Law & Security Review*, *48*, 1-9. [105796]. https://doi.org/10.1016/j.clsr.2023.105796

Nair, S. R. (2020). A review on ethical concerns in big data management. *International Journal of Big Data Management*, *1*(1), 8. https://doi.org/10.1504/IJBDM.2020.106886

Narzary, S. (2022). Anthony Larsson and Robin Teigland (Eds.), The Digital Transformation of Labor: Automation, The Gig Economy, and Welfare. NHRD Network Journal, 15(1), 124–126. https://doi.org/10.1177/26314541211064724

Naqvi, S. K. H., & Batool, K. (2023). A comparative analysis between General Data Protection Regulations and California Consumer Privacy Act. Journal of Computer Science, Information Technology and Telecommunication Engineering, 4(1), 326-332.

O'Neil, Cathy. (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* New York: Crown.

Oguntimilehin, Abiodun & Ademola, Ojo. (2014). A Review of Big Data Management, Benefits and Challenges. Journal of Emerging Trends in Computing and Information Sciences. 5. 433-438.

Pastorino, R., De Vito, C., Migliara, G., Glocker, K., Binenbaum, I., Ricciardi, W., & Boccia, S. (2019). Benefits and challenges of Big Data in healthcare: an overview of the European initiatives. *European journal of public health*, *29*(Supplement_3), 23-27. https://doi.org/10.1093/eurpub/ckz168

Politou, E., Alepis, E., & Patsakis, C. (2018). Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions. *Journal of Cybersecurity*, *4*(1). https://doi.org/10.1093/cybsec/tyy001

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 (General Data Protection Regulation). OJ L 119, 4.5.2016, p. 1–88.

Richards N., The GDPR as Privacy Pretext and the Problem of Co-Opting Privacy, 73 Hastings L.J. 1511 (2022).

Richterich, A. (2018). Big Data: Ethical Debates. In *The Big Data Agenda: Data Ethics and Critical Data Studies* (Vol. 6, pp. 33–52). University of Westminster Press.

Rik Crutzen, Gjalt-Jorn Ygram Peters & Christopher Mondschein (2019) Why and how we should care about the General Data Protection Regulation, Psychology & Health, 34:11, 1347-1357, DOI: 10.1080/08870446.2019.1606222

Rubinstein, I. S. (2013). Big data: The end of privacy or a new beginning? *International Data Privacy Law*, *3*(2), 74–87. https://doi.org/10.1093/idpl/ips036

Schuilenburg, M., & Peeters, R. (Eds.). (2020). The Algorithmic Society: Technology, Power, and Knowledge (1st ed.). Routledge. https://doi.org/10.4324/9780429261404

Sen, D., Ozturk, M., & Vayvay, O. (2016). An overview of big data for growth in smes. *Procedia - Social and Behavioral Sciences*, *235*, 159–167. https://doi.org/10.1016/j.sbspro.2016.11.011

Sevignani, S. (2017). Surveillance, Classification, and Social Inequality in Informational Capitalism: The Relevance of Exploitation in the Context of Markets in Information. *Historical Social Research / Historische Sozialforschung*, *42*(1 (159)), 77–102. http://www.jstor.org/stable/44176025

Shastri, S., Wasserman, M., & Chidambaram, V. (2019). The Seven Sins of Personal-Data Processing Systems under GDPR. *arXiv: Computers and Society*.

Stahl, B. C., Schroeder, D., & Rodrigues, R. (2023). *Ethics of Artificial Intelligence: Case Studies and Options for Addressing Ethical Challenges* (p. 116). Springer Nature.

Subrahmanya, S.V.G., Shetty, D.K., Patil, V. *et al.* The role of data science in healthcare advancements: applications, benefits, and future prospects. *Ir J Med Sci* 191, 1473–1483 (2022). https://doi.org/10.1007/s11845-021-02730-z

Tsaqila, Z. Qoulan. (2019, April 29). Big data and customer exploitation, is it occurred? Medium. https://medium.com/@zidnaqoulan/big-data-and-customer-exploitation-is-it-occurred-d5fa8131bfcf

Uk gdpr updated for brexit | uk gdpr. (2021, January 26). https://uk-gdpr.org/

Truong, N.B., Sun, K., Wang, S., Guitton, F., & Guo, Y. (2020). Privacy Preservation in Federated Learning: Insights from the GDPR Perspective. *ArXiv, abs/2011.05411*.

Richardson, S., Petter, S.,, Carter, M.. (2021). Five ethical issues in the big data analytics age. *Communications of the Association for Information Systems*, *49*(1), 430–447. https://doi.org/10.17705/1CAIS.04918

Uricchio, W. (2017). Data, Culture and the Ambivalence of Algorithms. In M. T. Schäfer & K. van Es (Eds.), *The Datafied Society: Studying Culture through Data* (pp. 125–138). Amsterdam University Press. http://www.jstor.org/stable/j.ctt1v2xsqn.13

Van Ooijen, I., & Vrabec, H. U. (2019). Does the gdpr enhance consumers' control over personal data? An analysis from a behavioural perspective. *Journal of Consumer Policy*, *42*(1), 91–107. https://doi.org/10.1007/s10603-018-9399-7

Wachter, Sandra, Data Protection in the Age of Big Data (January 19, 2019). Nature Electronics Vol. 2, 6–7 (2019), DOI: 10.1038/s41928-018-0193-y , Available at SSRN: https://ssrn.com/abstract=3355444 or http://dx.doi.org/10.2139/ssrn.3355444

What is personal data? (2023, May 19). https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/personal-information-what-is-it/what-is-personal-data/what-is-personal-data/

Wulf, A. J., & Seizov, O. (2022). "Please understand we cannot provide further information": Evaluating content and transparency of GDPR-mandated AI disclosures. *AI & SOCIETY*. https://doi.org/10.1007/s00146-022-01424-z

Yeung, K. (2018). Five fears about mass predictive personalisation in an age of surveillance capitalism. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3266800

Zarsky, T. Z. (2016). Incompatible: The GDPR in the age of big data. Seton Hall L. Rev., 47, 995.

Zeide, E. (2017). The structural consequences of big data-driven education. *Big Data*, *5*(2), 164–172. https://doi.org/10.1089/big.2016.0061

Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. *Journal of information technology*, 30(1), 75-89.

Zuboff, S. (2019). Surveillance capitalism and the challenge of collective action. *New Labor Forum*, *28*(1), 10–29. https://doi.org/10.1177/1095796018819461

Zwitter, A., Gstrein, O.J. (2020). Big data, privacy and COVID-19 – learning from humanitarian expertise in data protection. *Int J Humanitarian Action* 5, 4. https://doi.org/10.1186/s41018-020-00072-6