LUISS Guido Carli

LIBERA UNIVERSITA' INTERNAZIONALE DEGLI STUDI SOCIALI

Department of Business & Management

MARKETS, REGULATIONS AND LAW

WHAT DOES THE GDPR ACTUALLY MEAN FOR AI STARTUPS

SUPERVISOR:

Prof. Giuseppe Colangelo

CANDIDATE:

Alessia Treta

ID Nr. 722071

CO-SUPERVISOR:

Prof. Luca Arnaudo

ACADEMIC YEAR 2020/2021

Table of Contents

Introduction

The right to privacy is part of the 1950 European Convention on Human Rights, which states that everyone has the right to respect for his/her private and family life, home and correspondence. On this basis, the European Union has attempted to ensure the protection and respect of this right through legislation. The EU recognized the need for modern types of protection, as technology was becoming highly developed with the invention and spreading use of Internet. Thus, in 1995 the European Data Protection Directive came into force and established a minimum data privacy and security standards, upon which each Member State based its own regulation. Then, as the amount of personal data held by organizations continues to rise, so does the responsibility for its secure storage and ethical use. Regulators recognized that the existing laws were insufficient and, in response, the GDPR replaces the Data Protection Initiative of 95/46/EC.

Due to the rapid development of technology and the Internet, the European Union felt the need to protect citizens' rights and to harmonize the data protection laws of the 28 EU member states. To accomplish this goal, the European General Data Protection Regulation (GDPR) was adopted. The EU decided that one major way to enhance harmonization through the new law was to enact it in the form of a regulation, rather than a directive.

The regulation aims to harmonize data privacy laws across Europe, to protect and empower data privacy for all EU residents, and to reshape the way organizations across the region approach data privacy for EU residents wherever they work in the world. The GDPR is also designed to address technological and societal changes that have taken place over the last 20 years by adopting a technology-neutral approach to regulation. This law applies to both EU and non-EU companies that collect, process, or store information on EU citizens as well as on non-citizens while they reside in the EU. The GDPR has brought significant changes for both EU and non-EU organizations, such as the following ones: the increased scope, as it will impact almost every organization that is based in the EU, as well as every organization that does business in the EU, even if based outside the EU. It substantially increases the maximum penalties for non-compliance to the greater of €20 million, or 4% of an organization's worldwide turnover. It raises the bar for compliance by requiring greater openness and transparency about how organizations process personal data and it imposes tighter limits on the use of personal data. It also gives individuals more powerful rights towards their personal data.

One of the hopes is that by slim-lining data legislation with GDPR, it can bring benefits to business. Indeed, the European Commission claims that GDPR will save €2.3 billion per year across Europe. Moreover, having a Single Supervisory Authority (SSA) for the entire EU will make it simpler and cheaper for businesses to operate within the region.

However, regulations have costs, which are meant to be recovered by the expected benefits. In terms of additional spending on consulting services and technological solutions, 74% of small- and medium-sized organizations spent more than $100.000. Notably, 20% spent more than $1 million. More than one-third (34%) of these organizations spent anywhere from $100,000 to $499,999 on GDPR compliance. Only 6% of all organizations spent less than $50,000. But, the full GDPR cost was actually much higher than that, counting in this lost time. Indeed, the average company spent 2.100 hours in meetings. And, for larger enterprises, that figure is much closer to 9.000 hours. All of those hours increased the total GDPR cost notably. However, there is nothing that demonstrates the effort was a waste of resources. The value of personal data being safe, or at least handled according to this regulation, could be worth this sum plus whatever other costs there are. Organizations need to balance the cost of implementation with the benefits of becoming compliant.

Small- and medium-sized companies and startup companies have struggled to comply with the GDPR. This is because the data in most cases represent a fundamental asset for startups. It is through data collection that a business or management has the quality information they need to make informed decisions based on further analysis, study, and research. Data collection allows them to stay on top of trends, provide answers to problems, and analyze new insights to great effect. Indeed, a database, if well constituted and GDPR compliant, could represent a form of gain and competitive advantage for a startup. However, in order to become compliant, startups have to comply with the three basic principles of the GDPR: accountability, privacy design, and privacy by default. This causes an increase in the costs to be incurred. But, despite the initial effort, the GDPR urges the startup firms to manage their data in a transparent, responsible, and accountable way. This is fundamental in order to obtain a permanent position of advantage on the market by mitigating risk and saving money in the long run.

Nevertheless, irrespective of the expenses that GDPR compliance calls for, following the regulations has become a mandate, owing to the growing public concerns over data collection, storage and dissipation. Indeed, it creates a harmonized duty to secure personal data for controllers and processors and guarantees the protection of EU citizens’ fundamental rights to data protection.

The work is structured as follows. Chapter 1 illustrates the new emerging phenomenon of Big Data and the synergic relation with Artificial Intelligence, from which the GDPR came into force. Then, Chapter 2 illustrates the intended and unintended consequences of the GDPR, by carrying out an analysis of the impact of the GDPR on vendor market concentration and vendor usage. Finally, Chapter 3 illustrates how the GDPR has impacted on startup firms, particularly on those using AI, through the analysis of Elaisian case study.

Chapter one

Why Big Data matters

1.1 How Big Data and AI work together

1.1.1 Big Data: 3V's or 7V's

Since the mid 1980s, the world has experienced an unprecedented explosion in the capacity to produce, store, and communicate data, primarily in digital formats (Hilbert and Lopez, 2011, 60–65). Indeed, as far back as 2001, a new phenomenon "Big Data" was beginning to emerge. Big Data is a term applied to data sets whose size is beyond the ability of commonly used software tools to capture, manage and process within a tolerable elapsed time (The McKinsey Global Institute, 2012).

The emerging technological development of big data is recognized as one of the most important areas of future information technology and is evolving at a rapid speed, driven in part by social media and the Internet of Things (IoT) phenomenon. The technological developments in big data infrastructure, analytics, and services allow firms to transform themselves into data-driven organizations. Due to the potential of big data becoming a game changer, every firm needs to build capabilities to leverage big data in order to stay competitive (Lee, 2017).

Gartner analyst Doug Laney defined the three main dimensions of data management, affirming that they refer to Volume, Velocity and Variety (Laney, 2001). According to Laney, "Big data" is a high-volume, velocity, and variety information asset that demands cost-effective and innovative forms of information processing for enhanced insight and decision making. The first dimension refers to the massive amount of data which are being generated, gathered and processed. The second dimension refers to the speed at which data are generated, processed and moved between different systems and devices. Finally, the third one refers to the different types of data that can be used for achieving the desired information or results (Younas, 2019, 105-107).

Though the three V's are the most widely accepted core of attributes, there are several extensions that can be considered and that lead to the extensions of the V's of Big Data, including the 4V's and, more recently, the 7V's approach.

IBM added Veracity as a fourth dimension, which represents the unreliability and uncertainty latent in data sources. Uncertainty and unreliability arise due to incompleteness, inaccuracy, latency, inconsistency, subjectivity, and deception in data. Indeed, managers do not trust data when veracity issues are prevalent. Customer sentiments are unreliable and uncertain due to the subjectivity of human opinions. Statistical tools and techniques have been developed to deal with these issues of big data with specified confidence levels of intervals (Lee, 2017).

In addition to the Vs mentioned above, three dimensions more have been added, leading to the elaboration of the so called 7V's theory: Variability, Visualization and Value. The latest attribute, Value, sits at the top of the Big Data pyramids and refers to the ability of transforming a myriad of data into business, and - through it - a model that answers sophisticated queries, delivers counterintuitive insights, and creates unique learnings.

Currently, the dimensions of big data have been proposed separately in the computing industry, but so far there is a lack of an integrated view of big data.

Lee (2017) proposed an integrated view, which shows how these dimensions are interrelated with each other. The three edges of the integrated view of big data represent three dimensions of big data: volume, velocity and variety. Inside the triangle there are the five dimensions of big data that are affected by the growth of the three triangular dimensions: veracity, variability, complexity, decay and value. The growth of the three-edged dimensions is negatively related to veracity, but positively related to complexity, variability, decay and value.

The integrated view shows that traditional data is a subset of big data with the same three dimensions, but the scope of each dimension is much smaller than that of big data. Traditional data consist mostly of structured data, and relational database management systems have been widely used to collect, store and process the traditional data. As the scope of the three dimensions continues to expand, the proportion of unstructured data increases. The magnitude of big data expands with the growth of velocity, volume and variety. The arrows represent the expansion of each of the three dimensions. The expansion of velocity, volume and variety are intertwined with each other. The expansion of each dimension affects the other seven dimensions.

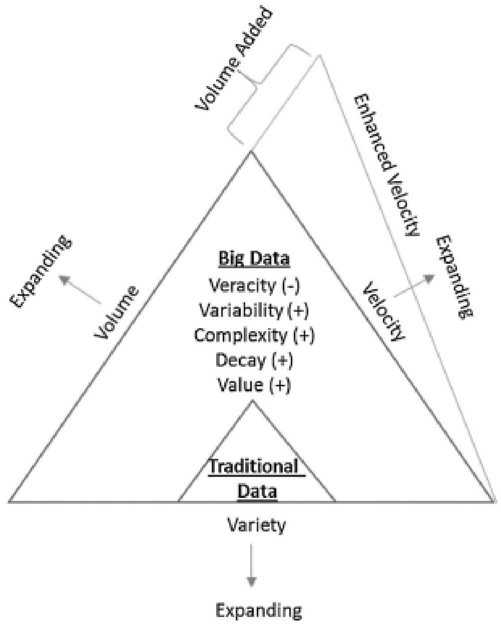


Figure 1: Integrated view of Big Data (Lee, 2017)

1.1.2 How AI fuels better insights

The term Artificial Intelligence (AI) appeared for the first time in 1956 in a conference at Dartmouth College, New Hampshire. It is used to define a technical discipline that researches and develops theories, methods, technologies and application systems for simulating the extension and expansion of human intelligence. Its main goal is to let machines perform some complex tasks that require intelligent humans to complete (Yeung, 2020).

As described in the AI Index 2018 Annual Report (Shomhan et al., 2018), Artificial Intelligence has advanced rapidly in recent years, measured both in terms of the amount of resources devoted to it and also in terms of output. Many scholars believe that AI has the potential to boost human productivity and economic growth (Furman and Seamans, 2019). Scholars also worry that these gains may come at a cost, potentially including labor displacement, income inequality and loss of privacy.

Aggregate statistics provide ample evidence that the deployment and use of AI and other advanced technologies have increased over the past decade. For example, academic papers focused on AI have increased nine times since 1996; in comparison, computer science papers have increased six times since 1996. Moreover, the number of students enrolled in artificial intelligence and machine learning courses at Stanford has increased eleven times since 1996; similar trends are observed at other universities including UC Berkeley, University of Illinois, Georgia Tech, and others.

It is useful for companies to look at AI through the lens of business capabilities rather than technologies (Davenport and Ronanki, 2018).

Broadly speaking, AI can support three important business needs: automated business processes, gaining insight through data analysis, and engaging with customers and employees.

For what concerns process automation, robot process automation (RPA) technologies are becoming more and more advanced as compared to earlier business-process automation tools, because the robots act similarly to human who is in the process of imputing and consuming information from multiple IT systems. RPA technologies perform different tasks such as transferring data from e-mail and call center systems into systems of record, replacing lost credit cards or ATM cards, and reading legal and contractual documents to extract provisions using natural language processing.

Secondly, for what concerns cognitive insight, AI has introduced the usage of algorithms to detect patterns in vast volumes of data and interpret their meaning. These machine-learning applications are being used to automate personalized targeting of digital ads, predict what a particular customer is likely to buy, identify credit fraud in real time and detect insurance claims fraud. Cognitive insights provided by machine learning differ from those available through traditional analytics. They differ from these because they are usually much more data-intensive and detailed. There are versions of machine learning which can perform feats such as recognizing images and speech; and they can also make new data available for better analytics.

Finally, cognitive engagement technologies refer to technologies that engage employees and customers using natural language processing chatbots, intelligent agents, and machine learning. They are being used to interact with employees rather than with customers.

However, Perry and Uuk (Perry et al., 2019) state that there are two categories of risk associated with AI systems: AI technical safety and AI governance. The Future of Life Institute has identified two scenarios where they believe AI will pose a risk to society. The first concerns autonomous AI systems that are programmed to kill and destroy human life, a risk highlighted by many international organizations such as the IEEE, the EU, the UK Office for Artificial Intelligence and smaller companies. The second scenario is based on an AI system which benefits society but does so through developing a destructive method for achieving its goal, generally when human and machine goals are misaligned.

Cheatham et al. (2019) identify five areas that can lead to AI risks. “Data difficulties” refer to the correct usage of data including compliance with the European General Data Protection Regulation (GDPR, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016) and other regulatory bodies. “Technology troubles” refer to when an AI system fails to do the job as expected, for example missing a fundamental outlier. “Security snags” are concerned with risks associated with AI-driven cybersecurity and the implications to the data and hence the data model. “Misbehaving models” refer to models developed from biased or unrepresentative data. Finally, the nature of human-machine AI system interactions causes risks. An example of this type of risk can be when human misinterpretation is caused by inappropriate training.

Bias is one of the biggest risks in using AI systems and this includes bias that is embedded into organizational or industrial cultures, personal and unconscious bias and data bias. Data bias must be considered and addressed in the selection of training data for AI systems. Data which has been labelled by humans for training may be subjective. Where training, validation and testing are dynamic and models continually evolve and learn, it should be monitored to ensure that there is no bias creep. It is also important to recognize that applications such as human profiling, and a one size solution to fit all humans may not be appropriate. Different models may need to be developed for different genders, cultures etc. as it may not be possible for the models to generalize on the human population.

Arnold et al. (2019) propose the use of an AI factsheet which provides information on statement of purpose, basic performance, safety, security, and lineage which are aligned with AI trust principles. The aim is that suppliers of AI systems and services voluntarily populate these factsheets. Customers who purchase from these suppliers can therefore review the factsheet and decide if that product meets their ethical and data governance standards.

Explainable AI (XAI) is necessary in building trust so that users and subjects can understand how the AI made a decision. However, the big question is to whom? Solutions such as Google Cloud's AI Explanations product include end users as stakeholders “who want to understand a model prediction to incorporate it into their decision-making process”.

This is not the same, for example, as providing an explanation to a user who applies for health insurance and is rejected based upon the automatic profile of an AI system based on facial micro-expressions (Neil, 2019).

Crockett et al. (2020) propose a new Hierarchy of Explainability and Empowerment that allows information and decision-making complexity to be explained at different levels depending on the personal perception of the knowledge level of knowledge of each individual.

Accountability of the AI system also affects trust and is difficult to determine due to the limited legislation and the lack of substantive case law (Doshi-Velez et al., 2017).

In order to be accountable, decisions need be explainable so that errors can be identified.

In a lecture given by Lord Sales, Justice of the UK Supreme Court, in 2019, the clear need for direction within the legal system was noted, "we need to build a structure of legal obligations on those who design and operate algorithmic and AI systems which requires them to have regard to and protect the interests of those who are subject to those systems" (Lord Sales, 2019). Similarly to all software products, usability and reliability of the AI system will also factor in how much people trust the system. Amershi et al. (2019) propose 18 human-AI interaction design guidelines to produce more usable, AI-centric systems.

Data governance and data privacy also play a significant role in perceived trust, especially following the widely published Facebook Cambridge Analytica scandal which continues to reveal leaked documents in 2020 (Cadwalladr, 2020).

1.1.3 The birth of a new intelligence: AI and Big Data

Big Data and Artificial Intelligence are merging into a synergic relationship, where AI is useless without data and data is unsurmountable without AI (Marville University, 2020). The latter depends heavily on the former for success, while also helping organizations unlock the potential in their data stores in way that were previously cumbersome or impossible (Casey, 2019).

According to Glenn Gruber, a senior Digital Strategist at Anexinet, there is a virtuous cycle in evidence: the more data are put through the machine learning models, the better they get (Gruber, 2019). There are three key ways that AI can deliver better insights with Big Data (Chan, 2020).

Firstly, AI is creating new and enhanced methods for analyzing data. Previously, determining insight from data required much manual effort by staff (i.e. usage of a SQL query). With the advent of AI, an array of enhanced methods to obtain data insights have become available. Secondly, data analytics is becoming less labor-intensive. The value of Big Data sets is related to data quality and data sets that are of inferior quality are of little or no worth for the organizational decision-making process. In many big data projects, eighty percent of the effort is spent towards cleansing and preparing the data for analytics. While, AI applications can discover outlier and missing values, duplicate records and standardize data for Big Data analytics. Finally, analytics become more predictive and prescriptive. In the past, data analytics were primarily backward-looking with a post-analysis of what happened. Predictions and forecasts were essentially historical analyses. Big Data decisions were therefore based on past and present data points with a linear ROI. AI is now creating new opportunities for enhanced predictions and forecasts. Indeed, an AI algorithm can be set up to decide or take an action based upon forward-looking insights.

Moreover, in the age of Big Data, companies have more consumer data, and more kinds of data, available to them than ever before. As companies fall further and further behind, they miss opportunities to learn from data and apply what they learn to how they connect. AI closes the gap by moving far past human limitations to consume and analyze data on a scale that previously was only imagined. In fact, the “intelligence” in Artificial Intelligence is exactly what it sounds like: the ability to think independently, to become more knowledgeable from being exposed to more information and to adapt and adjust when things change.

All of this is becoming possible now because there are more data than ever for AI to work with, and because AI is better than ever at working with it.

Thanks to the latest advancements in the field of molecular and computational techniques and information and communication technologies (ICTs), AI and Big Data can help in handling the huge, unprecedented amount of data derived from public health surveillance, real-time COVID-19 epidemic outbreaks monitoring, trend now-casting/forecasting, regular situation briefing and updating from governmental institutions and organisms, and health facility utilization information (Bragazzi et al., 2020).

The short-term applications of AI and Big Data could enable a quick and effective pandemic alert. Indeed, Big Data can enable monitoring of the disease outbreak in real time. With respect to previous epidemics and pandemics outbreaks, COVID-19 is unprecedented in that open-access datasets containing daily numbers of new infections broken down by country, and, in some cases, even cities, are widely available. Combined with the information available on the movement of people, it represents the perfect dataset to combine mathematical methods modelling and AI.

Secondly, these short-term applications are useful for tracking and diagnosing COVID-19 cases. In fact, having a reliable, sensitive and specific diagnostic test is of paramount importance in the prevention and control of infectious outbreak. Researchers have succeeded in establishing a molecular kit able to quickly and accurately capture the proper diagnosis and distinguish between COVID-19 and SARS-CoV. Salivary diagnostics seems to hold great promise in effectively detecting the virus. Together with molecular assays and tests, either multiplex nucleic acid amplification or microarray-based, high-resolution CT of the chest is fundamental for monitoring the disease course and its evolution in terms of severity and response to treatment. Currently ongoing research is also trying to identify early radiological predictors of prognosis, which would be extremely helpful in stratifying patients with COVID-

19 and in their clinical management (Wong et al., 2019, 44–48; McCall, 2020, 166–167; Ai et al., 2020).

AI can facilitate the diagnosis of COVID-19 cases. For instance, Infervision is a startup that employs deep learning medical imaging platforms for facilitating quick diagnosis of COVID-19 cases via the recognition of specific lung features (Wong et al., 2019, 44–48; McCall, 2020, 166–167). Furthermore, block-chain technology is a unique decentralized system of recording, verifying and approving data and carrying out a series of transactions. It is characterized by a high level of security and enables the delivery of patient-centered healthcare services, enhanced public health surveillance, management of outbreaks and a quick and effective decision-making process (Mashamba-Thompson and Crayton, 2020, 198; Chattu et al., 2019, 25).

A low-cost block-chain and AI-coupled self-testing and tracking system has been proposed for managing the COVID-19 pandemics, in developed settings so as to avoid overwhelming and straining public health capacity and healthcare/laboratory infrastructures, and in developing, resource-limited contexts (Mashamba-Thompson and Crayton, 2020, 198).

From a medium-term view, AI and Big Data can facilitate the implementation of public health interventions. During an outbreak, resources are limited and can be quickly consumed. As such, to avoid wasting resources and to better allocate those available, the Chinese government has supplemented classical data collection methods with sophisticated computational systems and advanced techniques that have helped identify subjects at risk. A startup company, called Megvii, has announced the development of sophisticated body and face detection dual sensing via infrared cameras and visible light, as thermal scanners for rapidly screening subjects transiting in a crowded place and identifying fever and high temperature, potentially related to COVID-19. Thus, AI and Big Data appear to have enormous potential for the management of COVID-19 and other emergencies, and their role is anticipated to increase in the future. AI and Big Data can be used to track the spread of the virus in real time, plan and lift public health interventions accordingly, monitor their effectiveness, repurpose old compounds and discover new drugs, as well as identify potential vaccine candidates and enhance the response of communities and territories to the ongoing pandemic. These emerging approaches can be exploited together with classical surveillance: while the latter enables data analysis and interpretation, the former uncovers hidden trends and patterns, which can be used to build predictive models (Bragazzi et al., 2020).

1.2 Data markets

1.2.1 The emerging data economy

Data are expected to become the fuel of the digital economy as they can be used to reduce information asymmetries, improve resource management, and identify casual relationships using artificial intelligence and statistical analyses.

Because of its increasing volume and perceived importance, the data economy is a central element of the Digital Single Market (DSM) policy framework as envisioned by the European Commission.

Data are rarely valuable alone and are usually inputs into analytics - embedded in a software program - to generate insights that can become expressed as content-based information goods. Data are thus primarily intermediate goods, produced with the intent of being combined and transformed to create other information goods (Koutroumpis et al., 2019).

Data have long been shared and traded. In recent years, the lower cost of data collection and the adoption of digital communication networks have dramatically increased volumes of data collected (Reinsel et al., 2017).

Much of the collected data are “exhausted data”, created as a by-product of other activities such as online shopping or socializing, rather than specifically for analytical purposes (Manyika et al., 2011; Mayer-Schonberger and Cukier, 2013). Indeed, the purchasing patterns of consumers have become the first data market segment that has experienced a significant commercial activity and raised privacy concerns regarding reading practices. Digital platforms such as Apple, Amazon, Facebook and Google enable trackers that collect and aggregate data from online sources, including mobile phones, and provide access or sell the data to third parties.

Furthermore, in the shadow of a digital economy, there have always been thriving marketplaces for stolen data (Holt and Lampke, 2010), such as credit card numbers or user profile data (Shulman, 2010). The growing amount of data has thus enabled highly controversial commercial practices.

The organizations and institutions in data markets are rapidly evolving. New regulations such as the GDPR of the European Union and the California Consumer Privacy Act have been implemented, as privacy has become a heightened concern. New types of data intermediaries have been envisioned that would either carry out data trading as their core activity, or trade data that arise from their core operations (Parmar et al., 2014; Thomas and Leiponen, 2016). Such entities would allow third parties to upload and maintain datasets, with access, manipulation and use of the data by others, and regulated through varying licensing models (Schomm et al., 2013).

In principle, data marketplaces could resemble multi-sided platforms, where a digital intermediary connects data providers, data purchasers, and other complementary technology providers (Parker and Van Alstyne, 2005; Eisenmann et al., 2006). Such platforms could generate value for both data buyers and sellers through lower transactional friction, resource allocation efficiency, and improved matching between supply and demand (Bakos, 1991; Soh et al., 2006). However, in practice, data are rarely traded on a large scale through multilateral platforms (Borgman, 2012). There are large-scale open data repositories such as the London Datastore set up by the Greater London Authority that do not actually sell data. Commercial data "platforms" such as Acxiom (consumer data), Bloomberg (financial data), or LexisNexis (insurance data) operate as intermediaries that buy and sell data via bilateral and negotiated contractual relationships. Moreover, there are abundant examples of failed data platforms (Markl, 2014; Carnelley et al., 2016): for instance, the Microsoft Azure DataMarket closed down in March 2017 after seven years of poor performance.

It thus appears challenging to set up large-scale systems to trade data through open markets in the same way many other goods are traded, including intangible goods such as content and inventions.

Markets for data exhibit similar characteristics to those for ideas and patents. Ideas, patents, and data are intangible goods and therefore mostly not used for competitive purposes. An idea or a data point, if digitized, may be usable by many individuals and replicated at low marginal costs (Romer, 1990; Koutroumpis et al., 2019).

Even though the – strategic – value of an idea or data may diminish due to wide dissemination, this will not prevent its application and use by many parties. Furthermore, ideas and patents need to be combined with complementary inputs for their commercialization (Teece, 1986; Bresnahan and Trajtenberg, 1995; Gans and Stern, 2010).

Much the same as inventions, data are intermediate goods and need to be further processed and combined with complementary input such as analytic technologies to become final goods and contribute to utility or productivity (Chebli et al., 2015; Koutroumpis et al., 2019).

Gans and Stern (2010) suggest that markets for ideas may exist in settings where intellectual property protection is sufficiently strong, which increases the likelihood that sellers appropriate enough of the value of an idea to justify the investment by excluding illegitimate trades and uses (Arrow, 1962; Teece, 1986). However, Hagiu and Yoffie (2013) have argued that multilateral digital marketplaces for patents are not viable due to the burdensome arrangements that would be required to ensure that high-quality patents are offered for sale. When the quality of the good is imperfectly observable, markets tend to be flooded with low-quality goods (Akerlof, 1970), and electronic markets for such goods may function particularly poorly (Overby et al., 2009). Nevertheless, companies such as Ocean Tomo orchestrate both public and private auctions for intellectual property portfolios. Thus, scholarship into markets for ideas and patents implies that specific governance mechanisms may be needed for a data market to launch.

1.2.2 Three assumptions regarding data markets

Data markets consist of three main links along the data value chain: collection, processing and the use of data-generated information and knowledge. Collection is related to the extraction of the data and its datafication, where processing is related to optimizing, cleaning, parsing or combining different datasets in order to organize the data for future extractions and to find correlations. Finally, use is defined as employing data-based information or knowledge for prediction and decision-making in relevant markets (Strandburg, 2014, 10-12; Nissen-Baum, 2019, 8-9).

The increased importance of data in our information economy as an input for innovation, economic growth, and societal interactions, along with the associated surveillance and security concerns, has brought data collection, processing and use issues to the foreground. Such issues are affected by a combination of private incentives and public measures. Understanding the interaction between private incentives and regulatory measures is thus essential for designing data governance models that can enable a well-functioning and social-welfare-enhancing digital economy.

The analysis below is based on three assumptions regarding data markets, which depend on the current understanding of the information economy.

The first assumption concerns the importance of data (Gal and Rubinfeld, 2019). As numerous studies have emphasized, with data “rapidly becoming the lifeblood of the global economy”, the efficiency of its use significantly affects both social and private welfare. Predictions based on patterns and correlations identify in data numerous aspects of daily life, including health, education, transportation and sustainability.

The second assumption regards the quality of knowledge that can be extracted from data, which is correlated with the volume of data, its variety, its veracity and its velocity (Stucke and Grunes, 2016). This is due to three reasons (Gal and Rubinfeld, 2019). First, data analysis is often characterized by economies of scale and scope (Stucke and Grunes, 2016, 352–55). This implies that until such economies are reached, the more data is available and the more varied the data, the better the knowledge that can be mined from it. Second, the 4V's of the data may affect the quality of the algorithm used for its analysis, due to the algorithm's feedback loop, with the algorithm evolving from learning based on an analysis of past predictions (Stucke and Grunes, 2016, 170; Farboodi et al., 2019).

Accordingly, the better the data, the better the algorithm and the better its predictions. Finally, the qualities of a dataset can also create positive externalities as compared to other datasets. This is because of "transfer learning", which is an algorithm that can learn from a high-value dataset to perform tasks that can then be performed on other datasets.

Lastly, the third assumption is related to the important role of data sharing in realizing potential data-based benefits (Gal and Rubinfeld, 2016; De Streel and Tombal, 2019; Mayer-Schonberger and Ramge, 2018). This is because much data is collected in a system that is largely modular and distributed (Allen and Chan, 2017). Broadening and improving the use of data through data sharing is likely to increase the competitive advantage of firms and nations.

At the same time, the sharing of personal data can be welfare-reducing due to price increases, price discrimination and intangible harms such as psychological discomfort and harm to freedom of speech. Sharing of personal data therefore creates complex, often ambiguous, tradeoffs that require a careful and conscious balancing among the competing considerations.

Yet, at least in some situations, data sharing has a significant potential for increasing both private and public welfare.

1.2.3 Five potential business models for obtaining relevant data

When data constitutes an important input toward the operations of a firm, it must choose its strategy for collecting and processing such data. This requires a switch from the current linear model of economy to a circular one that has recently attracted increased attention from major global companies such as Google, Unilever, Renault and policymakers attending the World Economic Forum. The reasons for this are huge financial, social and environmental benefits (Lewandowski, 2015).

Thus, five main strategies employed by market players for amassing relevant data have been identified (Gal and Aviv, 2020, 9-11):

1. "First-party data" refers to when the firm gathers the data directly from the Data Subject;
2. Merging with an entity and using its data in one's own operations;

3. “Third-party data” consists in buying or receiving the data from an external supplier. This strategy also includes the sharing of data by using an application programming interface (API). An API is an interface or communication protocol between a client and a server so that the server will initiate a defined action, including providing data, in response to a recognized request by the client for data in a specific format;
4. Becoming part of a joint venture in which firms pool their data for specific defined purposes;
5. Buying or receiving data-based knowledge, or aggregate data, from an external provider.

Firms choose between these options based on their relative cost-effectiveness, as well as their feasibility and scalability, which, in turn, are affected by a combination of technological, financial, strategic and legal barriers.

Technological barriers are those factors which impede the collection or sharing of data, for example barriers to interoperability between databases which were organized by different entities in accordance with different internal logics. Other technological barriers may be prohibitive, such as the inability to directly collect historical data ex post. Some such barriers may be overcome, but at a cost. Financial barriers are those factors which prevent data from being amassed in a cost-effective way. For example, internal data collection is financially feasible only when its benefits outweigh the costs of putting in place and operating such a system. Strategic barriers are those erected by data owners in order to retain their market power. For example, data controllers may set highly restrictive terms, or may be reluctant to share their data at all, in order to preserve a comparative advantage. Finally, legal barriers are those imposed by laws and regulations, relating to the collection, processing, and use of data, the GDPR being a prime example.

However, the GDPR imposes obligations regarding the collection, processing, storage and use of data. The choice between the business models for collecting data is affected by this legal regime. The relative costs of meeting these legal obligations under different business models may therefore lead firms to make choices they would otherwise have not made.

The stronger the contractual, legal and reputational sanctions for non-compliance with the GDPR, the more important it becomes for firms which need data for their operations to adopt business models that ensure compliance.

While the extent of costs imposed by the GDPR might differ from one factual scenario to another, general observations can be made. Assuming that Firm A requires a certain type of data for its operations, which can be lawfully collected by it or by an external firm (Firm E), absent technological, strategic and legal barriers to data collection and sharing, the choice of business models will be based on the relative costs of internal versus external data collection.

If the costs of internal collection by Firm A are lower than those of Firm E, Firm A will collect the data internally, and vice versa. If Firm E can collect the relevant data more cheaply and efficiently, or if it has already collected the data and has incurred the sunk costs involved, Firm A will buy the data from firm E, thereby saving on data collection costs and limiting duplication in collection. Considering the implication of the GDPR, buying data from Firm E adds several types of costs and obstacles. These include, inter alia, the costs incurred by Firm A for ensuring it has a lawful basis for the use of data. As noted above, a major obstacle may involve obtaining the consent of data subjects to an additional use of their data. The costs are incurred by Firm A for vetting the compliance of Firm E with the GDPR in collection and processing actions which pertain to the shared data. Furthermore, acquiring an external dataset may require Firm A to re-evaluate its own data life cycle to ensure compliance. In addition, the GDPR imposes costs on Firm E, to ensure that Firm A does not use the shared data in a way which was not agreed upon and which infringes its obligations to its data subjects. These costs will be reflected in the price charged by Firm E for the data. The greater the original data controller's loss of control over the data, and the higher the risk of non-compliance, the higher such costs.

1.3 From Big Data to GDPR

1.3.1 Background to the GDPR

Europe has long recognized privacy explicitly as a human right (Fuster, 2014, 164–166). The commitment of the Europeans foes beyond the home, similar to the focus of much of U.S. law, which is to include protection for family life, communications and reputation, with the rise of the information age, for privacy in the context of data processing (Koops et al., 2017).

While U.S. lawyers may refer broadly to 'privacy' or to 'information privacy', European law discusses information privacy as 'data protection' (Schwartz and Solove, 2009). Data privacy and information privacy refer to roughly the same concept. In Europe, data protection is increasingly seen as separated from the right to privacy. Data protection focuses on whether data is used fairly and with due process while privacy preserves the Athenian ideal of private life.

There was a key point of privacy divergence in the 1970s between the U.S. and Europe. The U.S. articulated Fair Information Practices ('FIPs') (Gellman, 2017; Hoofnagle, 2014; Bennett, 1992), the building blocks of all information privacy laws, but applied them in a serious sense only to the government in the form of the Privacy Act of 1974, and to the private sector only in the credit reporting sector. Europe embraced the FIPs, deepened and expanded them, and applied them to all information processing – both ‘vertically,’ that is government to citizen, and ‘horizontally,’ that is, business to citizen. The U.S. followed a sectoral model, leaving many forms of information practices regulated only by the Federal Trade Commission’s general consumer protection authority.

By 1990, the European Commission feared that diverging national data protection laws would hinder the internal market in the EU. That year, it published a proposal for a Data Protection Directive. After five years of negotiations, the final Data Protection Directive was adopted in 1995. The Directive laid down an omnibus regime based on the FIPs, which applied to most of the private and public sector - with exceptions to the latter -. The Directive required member states to enact implementing legislation. Problems quickly emerged with the Directive. The Directive did not fully harmonize national privacy laws, and even within Europe, countries behaved opportunistically to court big tech with signals of weak enforcement and advantageous tax schemes (Albrecht, 2013).

Even among countries committed to privacy, enforcement was weak, with the French fining Facebook a mere 150,000 Euros in 2017. This enforcement gap left Europe with a reputation of a region with rules with no real policing. The U.S. was seen as not being rule bound, yet the Federal Trade Commission applied enforcement (Bamberger and Mulligan, 2015).

The GDPR is the EU attempt to address these and other shortfalls. It did so in a process completely unlike U.S. legislative efforts. European policy makers started a process that involved a multitude of expert consultation and deep sophistication about how information practices can be manipulated to evade regulatory goals. Consultation began in 2009 and the European Commission published a proposal text in 2012.

Two years later, the European Parliament adopted a compromise text, based on almost 4,000 proposed amendments. The Council of the European Union published its proposal for the GDPR in 2015, to start negotiations with the European Parliament. In December 2015, the Parliament and Council reached an agreement on the text of the GDPR.

The GDPR was officially adopted in May 2016, and it has been applied since May 2018. Now that the GDPR is enforceable, its interpretation is entrusted to the courts, combined with persuasive, albeit non-binding, interpretation by the newly created European Data Protection Board. The Court of Justice of the European Union (CJEU) is the highest authority on the interpretation of EU law. The CJEU has delivered a number of dramatic, pro-privacy decisions, including striking down the data retention mandates, striking down the US–EU Safe Harbor agreement (Schrems, 2015), and granting people, under certain conditions, a ‘right to be forgotten’. Thus information industries face a more hostile judicial environment than in the U.S.

As in the 1995 Directive, the first article of the GDPR stresses a dual goal of promoting the free flow of personal data within the EU – to help businesses – as well as protect people and their personal data (GDPR, Article 1). Yet, the GDPR emphasizes the latter goal. The GDPR sets normative preferences in conflict with information-intensive industry practices, particularly those performed by third parties. The GDPR puts pressure on big data and machine learning business models, at least in their current form.

1.3.2 The principles enshrined in the GDPR

Article 2 of GDPR states that the Regulation is applicable to the “processing of personal data wholly or partly by automated means” (GDPR, Art. 2). Therefore, the material scope of data protection depends on personal data that must be processed. The definition of personal data is provided by Article 4(1), according to which four elements can be deduced: it is (i) any information, (ii) relating to (iii) an identified or identifiable (iv) natural person.

The first element indicates that broad categories of data are covered, regardless of the data’s content or format. The second one emphasizes the personal nature of data protection and excludes certain types of data from the Regulation’s scope of application, such as data that are solely about companies. Then, the third element means that the correlation to criterion is met when data reveal information about an individual, therefore considering data related to an individual when the data are about the individual, which can be in the guise of contents, purpose or resulting element. The final element is referred to as the identifiability element, because it is the lower threshold for application of data protection as compared to identified individuals (Oostveen, 2016).

Article 5 summarizes the inspiring criteria of the entire personal data protection principle, listing all the general principles. These are as follows: lawfulness, fairness and transparency; purpose limitation; data minimization; accuracy; storage limitation; integrity and confidentiality; accountability.

Any data collection and processing must comply with the above principles.

Operationalizing and enshrining these principles (Goddard, 2017) in the research cycle require proactive design and conceptualization of privacy as the default for any data collection exercise. It also needs to be embedded both in the design systems of any IT architecture and general organizational business practices of research agencies and clients.

Accountability requires organizations to put in place appropriate technical organizational measures, and to be able to demonstrate what they did and its effectiveness when requested. This may also include the use of privacy impact assessment for high-risk processing. Moreover, GDPR introduced a mandatory data breach notification regime.

A key change to note in lawful processing is the standard required for consent. Indeed, GDPR consent must be freely given, specific, informed and evidenced by clear affirmative action. It must also be verifiable, with a higher standard of explicit consent required to process sensitive data.

Processing of data is fair only if it is transparent and this means there must be openness in data processing through effective communication with individuals including in the use of information notices. GDPR is user-centric, so transparency in a GDPR context means a move away from legal tick-box compliance to a tailored, reflective and dynamic approach. Extensive information must be provided to individuals including details about recipients, retention periods and the range of their individual rights such as access and portability. All of this needs to be provided in an accessible language to ensure that it can be easily understood.

1.3.3 The One-Stop-Shop principle

The GDPR provides for a decentralized implementation system, whereby each Member State "shall provide for one or more independent public authorities to be responsible for monitoring the application of the GDPR" (GDPR, Article 51(1)). Each supervisory authority "shall be competent for the performance of the tasks assigned to and the exercise of the powers conferred on it in accordance with the GDPR" (GDPR, Article 55(1)).

When cross-border processing takes place, the GDPR established a One-Stop-Shop system, according to which the authority of the main establishment of the controller or processor will have the primary responsibility for dealing with and investigating any complaints from data subjects across the EU regarding the processing of personal data (Article 29 WP, 2017). This allows businesses operating in different countries to deal with one DPA – the DPA of their main establishment being the "sole interlocutor of the controller or processor for the cross border possessing carried out by that controller or processor" (GDPR, Article 56(6)).

While, under certain circumstances, a supervisory authority other than that of the main establishment of the controller or processor may be competent to handle a complaint or to investigate a possible infringement of the GDPR – notably when "the subject matter relates only to an establishment in its Member State or substantially affects data subjects only in its

Member State” (GDPR, Article 56(2)) – this can only be done if the lead supervisory authority decides not to handle this case (GDPR, Article 56(3) and (5)).

As a result, the GDPR has granted disproportionate enforcement power to certain DPAs, namely the DPAs of EU Member States, in which the large digital platforms, such as Apple, Google, Facebook and Amazon are established. In the ad tech sector, Ireland is at the forefront of GDPR enforcement. Its ability and willingness to investigate and sanction these companies therefore determine whether these companies will be able to get away with questionable data processing activities or will be held accountable. At the same time, controllers and processors that fall under the supervision of zealous DPAs are more likely to be subject to lengthy investigations and hefty fines of up to 20 million Euros or 4% of the total worldwide annual turnover of the preceding financial year, whichever is higher (GDPR, Article 83).

While for large digital platforms the imposition of such financial penalties would not significantly impact their business, for small players it can drive them out of the market.

When an alleged breach of the GDPR involves cross-border data processing investigations, it will be led by a Lead Supervisory Authority (LSA).

A supervisory authority is an independent public authority in each European Union member state, tasked with protecting the fundamental rights and freedoms of individuals in relation to any processing of their personal data and with monitoring and ensuring a consistent application of applicable data protection laws in the country in which they are situated.

This mechanism is attractive to organizations facing data privacy issues, as they can liaise with just one regulator for one decision, despite the issue impacting data subjects in multiple Member States. Once an enforcement issue arises, the LSA must cooperate with any concerned supervisory authorities in other Member States to attempt to achieve consensus.

Both controllers and processors involved in cross-border processing of personal data may be able to benefit from the One-Stop-Shop principle under the GDPR by identifying a Lead Supervisory Authority. The authority has the primary responsibility for coordinating investigations involving multiple member states, meaning businesses only have to deal with one lead regulator.

The LSA mechanism is only applicable in the context of a company’s cross-border processing activities. Consequently, companies must assess whether they meet one of the following criteria where either (i) processing takes place in the context of the activities of businesses or organizations in more than one member state where the business or organization is established in more than one member state; or (ii) processing takes place in the context of the activities of a single establishment but substantially affects or is likely to substantially affect individuals across more than one member state.

The GDPR does not permit “forum shopping”, which means that it is not possible for an organization to appoint a particular supervisory authority to be its Lead Supervisory Authority, on the basis that is may reputedly be more lenient on enforcement, as compared to another authority. For instance, if an organization claims to have its main establishment in one EU state, but no effective and real exercise of management activity or decision making over the processing of personal data takes place there, the relevant supervisory authorities will decide which supervisory authority is the “lead”, using objective criteria and looking at the evidence.

Article 29 Working Party adopts draft guidelines on issuing administrative fines. In the draft guidelines, the degree of cooperation with supervisory authorities, the previous contacts with these on previous infringements, and account taken of guidance are all considerations in the complex matrix of factors when determining fines.

While all supervisory authorities are equal under the GDPR, there will be organizational, technical, financing, structural, and cultural difference among them. While it is also difficult to assess and compare these differences today, as national laws creating and empowering GDPR regulators are still emerging, undoubtedly these differences will have an impact on the enforcement landscape.

1.4 Data portability

The right to data portability is one of the most important novelties within the EU General Data Protection Regulation, since in the Data Protection Directive (94/46/EC) text no relevant references may be found (Beslay and Sanchez, 2018, 193-203). Indeed, no field of law has been experimented before with anything resembling personal data portability (Custers and Hursic, 2016).

Data Portability is the ability and capacity to export data collected or stored digitally concerning a data subject and the ability to receive data concerning the data subject and to allow another controller to receive portable data.

The Data Portability requirement entails both a technical design requirement and a data subject rights requirement. From a technical perspective, data controllers will need to ensure their systems, connected products, applications and devices that collect and store information on data subject also have the added functionality of porting and transmitting data. In some cases, this will require controllers to tweak or redesign some systems, products, applications and devices.

Furthermore, the new porting functionality must export data in a structured, commonly used and machine-readable format so that reuse of the data is possible.

From the perspective of a data subject, the right to data portability creates a new right for individuals to exercise more control over their own data. It enables individuals to receive personal data concerning him or her, which he or she has provided to a controller. Thus, data controllers will need to establish and implement processes, in addition to added systems and digital propositions/products functionality, that aid in processing data subject requests whether manually or in automated fashion. After receiving the data the individual must be able to transmit this data to another controller without creating additional burden or hindrance to the previous data controller. The right to port data is also entailed where technically feasible, the personal data will be transmitted directly from one controller to another.

This is both in terms of warranting control rights to data subjects and in terms of being found at the intersection between data protection and other fields of law (competition law, intellectual property, consumer protection, etc.). The first example of portability of users' data referred to telephone numbers. Then, in the GDPR text it was extended to all digital services. It constitutes, thus, a valuable case of development and diffusion of effective user-centric privacy enhancing technologies and a first tool to allow individuals to enjoy the immaterial wealth of their personal data in the data economy. Indeed, a free data portability of personal data from one controller to another can be a strong tool for data subjects to foster competition of digital services and interoperability of platforms and as well as to enhance controllership of individuals on their own data (Article 29 WP, 2017).

The right of data portability is regulated by the article 20 of the GDPR. Firstly, it states that data portability is a right of the data subject to receive a subset of the personal data processed by a data controller concerning him or her, and to store those data for further personal use. Such storage can be on a private device or on a private cloud, without necessarily transmitting the data to another data controller. In this regard, data portability complements the right of access. One specificity of data portability lies in the fact that it offers an easy way for data subjects to manage and reuse personal data themselves. The data should be received "in a structured, commonly used and machine-readable format".

Secondly, Article 20(1) provides data subject with the right to transmit personal data from one data controller to another data controller "without hindrance". In accordance with Article 20(2), data can also be transmitted directly from one data controller to another on request of the data subject and where it is technically feasible. In this respect, recital 68 encourages data controllers to develop interoperable formats that enable data portability but without creating an obligation for controllers to adopt or maintain processing systems which are technically compatible. The GDPR does, however, prohibit controllers from establishing barriers to the transmission.

The impact of the right to data portability is relevant both for businesses, in particular for e-business involved in the digital market, and for individual users (data subjects).

From the business perspective, this impact is tangible in several fields, it is both a challenge to the traditional system of competition law (Graef, Verschakelen and Valcke, 2013, 53-63) and a problematic opportunity in terms of interoperability of systems. From the user perspective, this impact is evident both in terms of control of personal data and in terms of a more user-centric interrelation between services. Thus, the right to data portability is one the most remarkable novelties of the GDPR.

On the one hand, it can be the opportunity to foster interoperability and develop more and more user-centric platforms for the management of personal data (Article 29 WP, 2013, 47). On the other hand, it represents the first theoretical step towards a default ownership of personal data to data subjects.

However, there are several points that remain a challenge. The role of European Commission to give incentive for interoperability has been removed from the first proposal of GDPR. As a result, the efforts needed for the development of interoperable formats and interfaces to port data are minimum. Instead, a very prudent balancing structure was chosen.

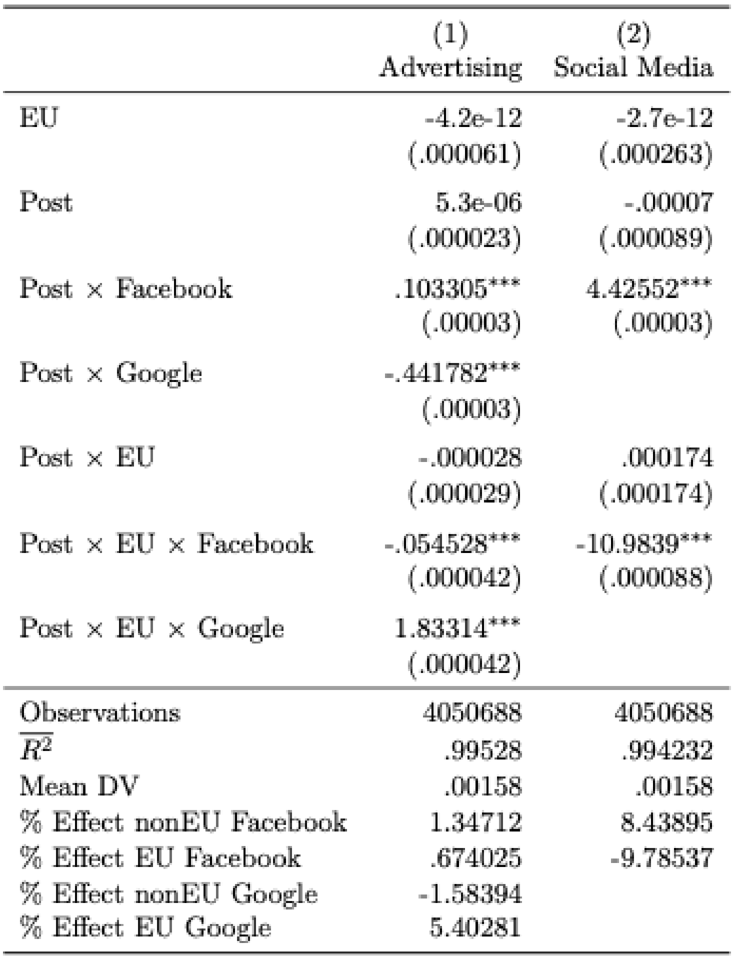
Chapter two

GDPR’s intended and unintended consequences

2.1 Privacy law and antitrust policy

An increasing concentration and an increasing market share of the dominant firm are most likely not what the European legislature had in mind when designing GDPR. Indeed, the European Commission (EC) stressed in 2012 how the pro-competitive effects of the future GDPR would increase the attractiveness of Europe as a location to do business (European Commission, 2012, 148-149).

GDPR implemented and enforced the consent requirement for websites on a large scale, which disproportionately benefits larger firms offering a broader range of services (Campbell et al., 2015). The concentration in the markets for web technologies may additionally be explained by economies of scale. Why would only the firm with largest pre-GDPR market share increases its market share?



	(1) Advertising	(2) Social Media
EU	-4.2e-12	-2.7e-12
	(.000061)	(.000263)
Post	5.3e-06	-.00007
	(.000023)	(.000089)
Post × Facebook	.103305***	4.42552***
	(.00003)	(.00003)
Post × Google	-.441782***	
	(.00003)	
Post × EU	-.000028	.000174
	(.000029)	(.000174)
Post × EU × Facebook	-.054528***	-10.9839***
	(.000042)	(.000088)
Post × EU × Google	1.83314***	
	(.000042)	
Observations	4050688	4050688
$\overline{R^2}$	.99528	.994232
Mean DV	.00158	.00158
% Effect nonEU Facebook	1.34712	8.43895
% Effect EU Facebook	.674025	-9.78537
% Effect nonEU Google	-1.58394	
% Effect EU Google	5.40281	

Table 1: Change in market shares, by submarket, Google and Facebook vs. all other firms (Garret et al., 2021)

In Table 1, looking at the advertising network submarket, Google and Facebook both increased market share in EU-targeted websites, but Google's increase is four times larger. This may be due to network effects that make it more appealing for websites to use large advertising networks since Google's pre-GDPR market share in the advertising technology market was already 3.5 times larger than Facebook's. Also, Facebook's reputation concerning privacy issues has suffered over the last years. Facebook has lost 9.8% of the market share with websites catering to EU audiences in their strongest market before GDPR (market share of 59.5%): the market for like/share/login buttons, which additionally allows Facebook to track users across websites (Roosendaal, 2012). This is consistent with the issue of joint responsibility, and recent European case law that involved websites and their usage of Facebook's like buttons.

The increased concentration in web technology markets may have been an unintended but unavoidable consequence of the GDPR. This raises the question of how privacy law and antitrust policy are related.

Under European law, antitrust and privacy laws have traditionally been distinct. The enforcement of European antitrust laws was well-developed and done by both public antitrust authorities at the EU and member state levels. Enforcement of privacy laws was traditionally weak and left to member state authorities, which sometimes delegated this task further in their federal structure.

As recently as 2014, the EC noted in its approval of merger between Facebook and WhatsApp that "any privacy-related concerns flowing from the increased concentration of data within the control of Facebook as a result of the transaction do not fall within the scope of the EU competition law rules but within the scope of the EU data protection rules" (European Commission, Decision in Case COMP/M. 7217 (Facebook/WhatsApp), Oct 3., 2014, C(2014) 7239 final, note 164).

It indicates that it is increasingly difficult to conceptualize antitrust and privacy law as two distinct areas of the law with different goals, remedies, and enforcement mechanism (Economides and Lianos, 2019). On the one hand, network effects, lacking competition in terms of service and privacy policies as well as the limited effectiveness of user consent in privacy law (Acquisti and Grossklags, 2005; Bart and de Jong, 2017), may enable firms to increase their dominant position by violating privacy laws. On the other hand, laws aimed at increasing privacy protection may, at the same time, decrease competition in related technology markets.

In a world where processing personal data, analyzing user-profiles and predicting consumer behavior are cornerstones of highly concentrated Internet markets, designing privacy laws that do not have immediate implications for antitrust policy – or vice versa – is nearly impossible.

2.2 GDPR and market concentration

While there is a growing public concern about the unparalleled amounts of data accumulated by a few digital platforms (Gonzalez, 2018), one of the paradoxes of the GDPR is that it may strengthen these large platforms to the detriment of smaller market actors. As pointed out by Gal and Aviv in a recent paper, the GDPR could therefore lead to further market concentration (Gal and Aviv, 2020). Gal and Aviv's concern that the GDPR could increase market concentration applies with full force in the ad tech sector for the reasons discussed hereafter.

First, the implementation of the GDPR and the related compliance costs may create barriers to entry or may cause exit.

Indeed, compliance with the GDPR is a particularly onerous task for small and medium-size ad tech providers, as it places a particularly heavy burden on their resources. They must – inter alia – put in place consent gathering mechanism (e.g. by having a Consent Management Platform), provide detailed information regarding their data processing activities, implement technical and organizational measures to ensure compliance with the GDPR, monitor and document GDPR compliance (e.g. by keeping detailed records of their processing activities), carry out Data Protection Impact Assessments and have a designed DPO. Companies with data presence in the EU have been required to spend millions to comply with the GDPR (Chivot and Castro, 2019).

The human resources and capital costs involved in ensuring compliance with the GDPR disproportionately burden small and medium-size vendors – which are limited on both financial resources and personnel. While a big company has dozens if not hundreds of experts working on GDPR compliance (e.g. Microsoft has 1,600 engineers working on GDPR compliance since its enactment in 2016) (Brill, 2018), most ad tech companies do not have lawyers, data experts and programmers necessary to make compliance with the GDPR a smooth and effective process (Kottasova, 2018). Additionally, compliance with the burdensome requirements of the GDPR, such as adopting technical and organizational measures and monitoring and documenting data flows, exhibits economies of scale and scope, which tend to create a competitive advantage for large organizations.

In a sector with a high concentration in some segments, that has already suffered from a large drop in investments in recent years (Murgia, 2017; Shields, 2018) and has seen several ad tech players struggling or even exiting the market, the additional costs generated by the GDPR could further precipitate market concentration. While a large company has no difficulty to absorb the compliance costs of the GDPR, the situation may be different for market players that are already struggling to make money and attract investment (Kuchler, 2018). The anticipated compliance costs of the GDPR could also delay or discourage market entry by making it more costly and risky, hence depriving advertisers and publishers of new and innovative tools, as well as the competitive pressure they would bring on Google.

This imbalance is further aggravated by the uncertainty surrounding the GDPR. For instance, immediately after the entry into force of the GDPR, numerous independent ad exchanges and other vendors in the ad tech ecosystem saw their demand volumes shrink dramatically from 20 to 40 percent (Davies, 2018). Indeed, the text of the GDPR left open a number of questions and authorities had not clarified fundamental issues, such as what legal basis is appropriate in the context of online advertising, how to obtain user consent or who can be considered as a controller or processor in the complex online advertising ecosystem, which ultimately was a benefit for Google as advertisers decided to fly to safety. Regulatory uncertainty may further penalize small and medium-size ad tech vendors when DPAs are not responsive to the need for clarity on some opaque areas of the GDPR.

Second, large online platforms, such as Google, benefit from the advertisers' trust, which therefore tend to concentrate their ad expenditure on them. Thus, the fear of liability and the large fines that can be imposed because of the GDPR have led advertisers to concentrate their ad spending on the largest players, as they trust that they are compliant with their regulatory requirements (Kostov and Schechner, 2019). Since the entry into force of the GDPR, Google and Facebook's market dominance in online advertising has been further strengthened (Scott et al., 2018; Weissman, 2019).

Trust in these players follows from three assumptions. First, that such companies have the resources to comply with the GDPR. Second, that companies holding vast amounts of data will be closely monitored by regulatory authorities and thus will be compliant. Third, that such companies will be more careful with users' personal data as they have more to lose in case of non-compliance.

Third, it is easier for a company like Google to obtain end-user content through its numerous customer-facing products. Indeed, the strengthened data protection framework set out by the GDPR has made it harder for small and medium-size players to collect and use data. Privacy-aware consumers might be hesitant to give consent to the processing of their data to market players they do not necessarily know – especially when seeing that their data might be used for a variety of purposes, such as advertising and measurement.

On the contrary, Google's market dominance allows it to easily obtain users' consent to the collection of personal data (Markman, 2018). Billions of people are dependent on its service, such as Gmail, Search, YouTube and such, which they consider an indispensable part of their personal or even professional life – a part that they do not consider letting go. Evidence shows that about half of all Internet users and about two thirds of users aged 14-29 classify Google Search as "absolutely essential" (Niedermann, 2019). While some users are unhappy about the way consent is sought – for example, that it is often a take-it-or-leave-it approach or that it is too cumbersome to read all relevant notices – the perceived essential nature of the services provided by Google, combined with the lack of credible alternatives, outweigh their concerns about their data being collected.

Moreover, in the case of logged-in environments (e.g. Android), user consent needs to be obtained only once, when the user is required to accept the terms and conditions in order to use the service. In contrast, outside the "walled gardens", user consent must be obtained each time a user visits a publisher or advertiser's website. Users faced with repetitive requests to consent to the collection and processing of their data are more likely to refuse granting the required consent. Additionally, the fragmentation of consent in the open web hampers the business of ad tech vendors which must ensure that user consent has been obtained in various touchpoints, covering the entire ad tech ecosystem, in order to be able to provide their services without infringing data protection laws.

Fourth, the GDPR's restrictions in data sharing give a competitive advantage to ad tech players that can acquire large troves of data through their consumer-facing products. Indeed, the GDPR has considerably limited data sharing, by requiring free, specific, informed and unambiguous consent for data transfers, as well as by requiring the data suppliers to monitor and follow the data transferred – as the data collector must ensure that data are only used in accordance with the data subject's consent and that the data subject can exercise its rights – and by imposing liability in case of violation of the GDPR.

Data sharing is, therefore, risky and many data holders may decide to take the extreme measure of refusing to share their data with smaller ad tech players as they may not trust their ability to comply with the GDPR. This is problematic as these smaller actors are generally the ones that would benefit the most from accessing third-data parties.

In contrast, there is limited incremental value from data transfer for large market actors holding massive amounts of data, as they already capture the data they need within their ecosystem. At the same time, when entities decide to engage in data sharing, they prefer to deal with large market actors, whom they trust to comply with the GDPR. This affects competition and creates a competitive advantage for large, well-known players, as smaller suppliers or new entrants will often be overlooked. The limitations to data sharing have therefore widened the gap between Google and Facebook and small players, making the former much more attractive to advertisers who value the amount of data and the identification possibilities they allow.

Finally, one challenging issue is to quantify the impact of the GDPR on market concentration in the ad tech sector. A couple of recent empirical studies suggest that the short-run impact of the GDPR was indeed increased by market concentration. For instance, in their paper analyzing the impact of the GDPR on web technology vendors, Johnson and Shiver show that the highest impact on such vendors both in terms of market shares and concentration ratio was felt in the advertising category (Johnson and Shiver, 2019).

Similarly, in their paper analyzing the impact of the GDPR on web technology services, Peukert et al. (2019) show that “with the introduction of GDPR, the dominant firm in many markets for web technologies, Google, increases its market share whereas all other firms that supply web technology either do not see a change in market share or suffer losses”. They also find among the service categories analyzed in their paper, Google’s largest market share increases were in the analytics market (7.2%) and the advertising market (4.5%). Thus, the GDPR does influence market concentration and competition in the advertising field.

2.3 A decline in competition in Data markets

The importance of the GDPR cannot be overstated. It seeks to protect consumers and users from harm resulting from unauthorized and excessive use of their personal data, in ways that might negatively affect human dignity and well-being, including but not limited to price discrimination, other forms of discrimination, black-mail, intangible nuisance, identity theft and harm to autonomy (Acquisti et al., 2016; Elkin-Koren and Gal, 2019). The GDPR also seeks to change the balance of power between data subjects and data controllers, potentially enabling data subjects to enjoy a larger portion of the fruits from sharing their data.

Additionally, it seeks to ensure the free flow of data between EU member states – inter alia – by eliminating differences among such states regarding data processing.

Furthermore, it seeks to strengthen the trust users have that their personal data will not be used in ways that do not conform to their reasonable expectations, which is necessary for the efficient working on the market and for society to realize the value of technology.

At the same time, the GDPR creates inherent tradeoffs between data protection and other dimensions of welfare, including competition and innovation (Price et al., 2019). While some of these effects were acknowledged when constructing the legal data regime, many were disregarded. Furthermore, the magnitude and breadth of such effects may well constitute an unintended and unheeded welfare-reducing consequence. Indeed, the GDPR limits competition and increases concentration in data and data-related markets, and potentially strengthens large data controllers. It also further reinforces the already existing barriers to data sharing in the EU, thereby potentially reducing data synergies that might result from combining different datasets controlled by separate entities (OECD, 2015).

Seven main parallel and cumulative market dynamics that may limit competition and increase market concentration have been identified so far.

First, as some commentators have already observed in the IAPP-EY Annual Governance Report 2018 (2019), the costs of organizing a dataset in a way which complies with the GDPR may be high and are characterized by economies of scale. Accordingly, some small entrants might find it unprofitable to collect data.

Second, also as previously observed, the GDPR prohibits or makes it more difficult to engage in some methods of data collection, creating comparative advantages to some data controllers. For instance, in their seminal article Campbell, Goldfarb and Tucker showed that the need to receive a user’s consent to use his/her data imposes transaction costs for internal data collection, whose effects fall disproportionately on less diversified or new firms. Both dynamics reduce the number of potential competitors in data collection (Campbell et al., 2015).

Third, the GDPR reduces the economic incentives of firms to share any data collected. This is because those sharing data are still liable for monitoring its use by anyone with whom the data is shared. This, in turn, further reduces the number of data suppliers.

Fourth, even where data is shared, the GDPR may limit its use. To illustrate, it is often costly, and sometimes impossible, to obtain informed consent from data subjects to have their data shared with the data receiver, as may be required by the GDPR. This effect is strengthened in a multi-product and/or multi-service environment, in which consent is required for each different use of data. The stronger the legal limitations on using data collected by an external entity, the stronger the motivation to collect it internally.

Fifth, the costs of non-compliance are high. Indeed, Article 83 points to the size of fines that can be imposed on firms which fail to comply with the GDPR (GDPR, Article 83). In addition to that, the effect of the virality of non-compliant data should be analyzed. The GDPR imposes a duty on the data receiver to ensure that any data received from an external entity is GDPR-compliant. Accordingly, should non-compliant data be transferred from an external data controller and combined with the receiver's data, the whole dataset could be polluted and considered non-compliant. Virality may affect all types of data included in the dataset, including non-personal data, so long as it is combined with – and cannot be easily separated from – the non-compliant personal data. Furthermore, and potentially more troubling, even if the dataset cannot be separated ex post, any learning by an algorithm based on the combined dataset cannot be easily reversed, especially if such learning was already translated into products or services. Undoing such effects could significantly disrupt business operations. To avoid such consequences, data receivers must engage in ongoing monitoring of their data suppliers' collection and processing practices. This, in turn, might further reduce incentives to use externally collected data, and strengthen incentives for internal data collection.

Sixth, the GDPR creates uncertainty, which may impose higher costs on smaller players, and might also enable large firms to use such uncertainty strategically, limiting the sharing of their data based on broad interpretations of the GDPR.

Finally, the GDPR, and especially the discussion surrounding it, could have an indirect effect on data subjects, who might be more willing to provide their data to larger, more reputable firms, or to firms with which they must interact, at least until the trust of data subjects in the actual enforcement of data protection obligations is increased.

The cumulative effect of such dynamics causes a decline in competition in data and in data-based markets. More often than in the pre-GDPR period (Data Protection Directive 95/46EC), firms may now prefer to collect data internally. Where internal collection is costly or impossible, firms will prefer to purchase data from external data suppliers.

Yet, the GDPR reduces the number of potential data suppliers and increases the costs of and barriers to data-sharing transactions. Accordingly, it is now substantially more difficult for firms to realize data synergies through data sharing. Furthermore, where data-based analysis requires a combination of personal and non-personal data, or where difficulties arise in separating these two types of data, the effects of such obligations may carry over to non-personal data (Graef et al., 2018). Thus, the GDPR might also indirectly affect the free flow of non-personal data.

The dynamics identified offer partial explanations for some of the troubling empirical evidence regarding investment in data-driven markets following the adoption of the GDPR (Chivot and Castro, 2019). Indeed, a study conducted by Merrill Corporation, for example, found that 58% of mergers and acquisitions professionals surveyed reported having worked on transactions that did not go through due to concerns about the parties' compliance with the GDPR (Merrill Corporation, 2018).

By having identified the markets dynamics that affect data collection and competition, the short-term and long-term competitive effects of the GDPR could be analyzed.

2.4 Vendor concentration analysis

Analyzing vendor market concentration introduces two definitional requirements. First, markets in terms of vendor membership must be defined. Second, vendor market shares, from which concentration measures are derived, must be defined. The U.S. Department of Justice and Federal Trade Commission (DoJ & FTC, 2010) suggest best practices for such choices in their guidance on the analysis of horizontal mergers. With regard to demand metrics, the FTC guidelines note: "In cases where one unit of a low-priced product can substitute for one unit of a higher-priced product, unit sales may measure competitive significance better than revenues. For example, a new, much less expensive product may have great comparative significance if it substantially erodes the revenues earned by older, higher-priced product, even if it earns relatively few revenues".

The web technology industry generally conforms to this description, due to high rates of service innovation and cost reduction. Vendor "unit sales" could be considered as its reach. For example, in the case of the advertising technology market, vendor "unit sales" are measured by the number of websites that interact with a domain owned by the vendor, and market shares represent the fraction of website-vendor interactions attributable to the vendor.

With regard to market definition, the industry in aggregate as well as category-level markets have been considered.

For a robust quantification of a vendor concentration in a – N vendor – market, three concentration metrics have been identified. These are as follows:

1. Herfindahl–Hirschman Index (HHI): HHI summarizes market concentration as the sum of the squared market shares. Market shares are on a 0 to 100 scale, so that HHI varies from 0 (perfectly competitive) to 10,000 points (monopoly).

$$HHI = \sum_{j=1}^{N} s_j^2$$

2. Concentration ratio (CR): CR is calculated as the sum of the market share percentage held by the largest specified number of firms in an industry. It indicates the size of firms in relation to their industry as a whole. Low concentration ratio in an industry would indicate greater competition among the firms in that industry, compared to one with ratio nearing 100%, which would be evident in an industry characterized by a true monopoly (Kenton, 2020).

$$CR(M) = \sum_{j=1}^{M} s_j$$

3. Head-to-head win rate: A metric that quantifies which vendor sites are more likely to drop. In particular, it seeks to quantify how often the sites drop each vendor, prior to and after the GDPR.

First, the web technology industry has been considered as a single market by including all vendors. Figure 2 plots the evolution in relative market concentration over 2018, as measured by aggregate HHI. Aggregate HHI is 146 points before the GDPR and HHI reaches its maximum of 171 points one week post-GDPR, a 17.1% increase. It can be noticed that the Figure 2 provides empirical support for part (b) of Proposition 3: market concentration increases in response to GDPR's implementation.

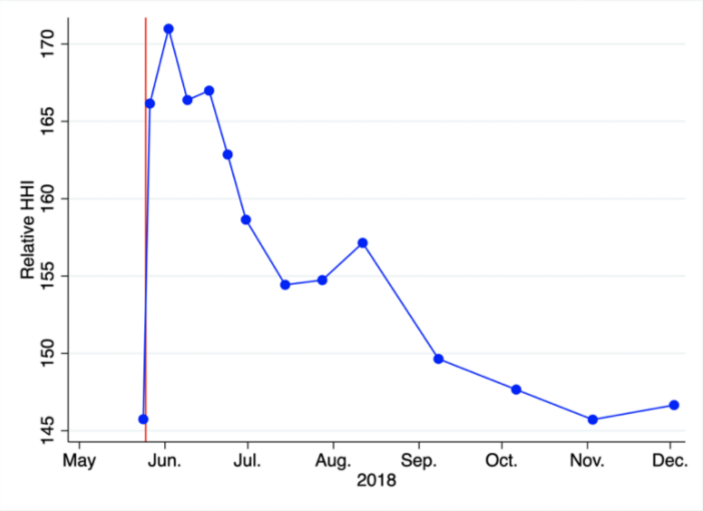
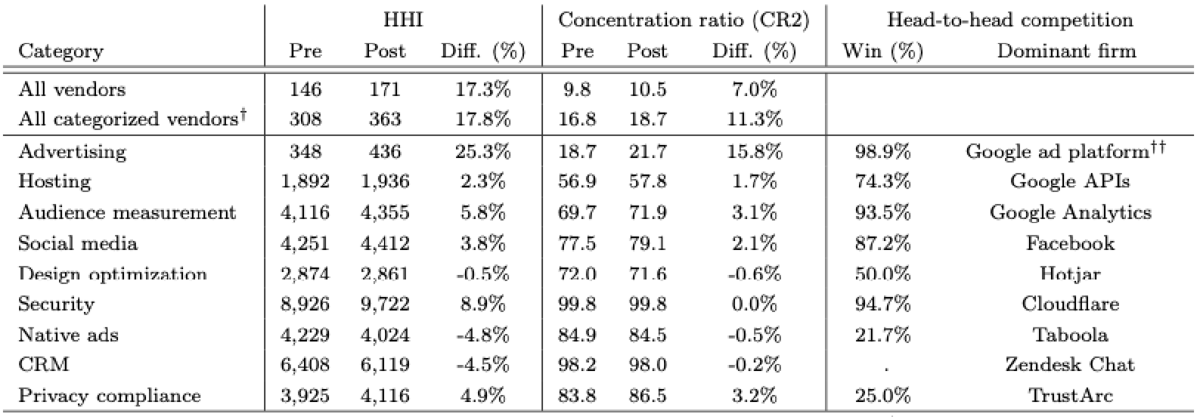


Figure 2: Evolution of web technology vendor concentration (HHI) (Garrett et al., 2021)

Then, the effect of the GDPR on market concentration by vendor categories has been analyzed.

2.4.1 GDPR short-run concentration impact



	HHI			Concentration ratio (CR2)			Head-to-head competition	
Category	Pre	Post	Diff. (%)	Pre	Post	Diff. (%)	Win (%)	Dominant firm
All vendors	146	171	17.3%	9.8	10.5	7.0%		
All categorized vendors[†]	308	363	17.8%	16.8	18.7	11.3%		
Advertising	348	436	25.3%	18.7	21.7	15.8%	98.9%	Google ad platform[††]
Hosting	1,892	1,936	2.3%	56.9	57.8	1.7%	74.3%	Google APIs
Audience measurement	4,116	4,355	5.8%	69.7	71.9	3.1%	93.5%	Google Analytics
Social media	4,251	4,412	3.8%	77.5	79.1	2.1%	87.2%	Facebook
Design optimization	2,874	2,861	-0.5%	72.0	71.6	-0.6%	50.0%	Hotjar
Security	8,926	9,722	8.9%	99.8	99.8	0.0%	94.7%	Cloudflare
Native ads	4,229	4,024	-4.8%	84.9	84.5	-0.5%	21.7%	Taboola
CRM	6,408	6,119	-4.5%	98.2	98.0	-0.2%	.	Zendesk Chat
Privacy compliance	3,925	4,116	4.9%	83.8	86.5	3.2%	25.0%	TrustArc

Table 2: Short-term GDPR impact on concentration (1 week) (Garrett et al., 2021)

Table 2 reports short-run changes in market concentration by category. The columns labeled “Pre” show the baseline HHI and concentration ratios for the top two vendors (CR2) in every category. It can be seen that all categories, except advertising and hosting, have HHI’s above the 2,500-point threshold that American regulators define as a “highly concentrated market” (US DoJ & FTC, 2010).

Advertising has the lowest HHI (348 points) and CR2 (18.7), as ad-supported websites often employ several vendors to boost ad revenue. Advertising contains 165 vendors and site use 4.35 vendors on average, so that even the dominant vendor – Google Marketing Platform – has a relative share of only 14.5 although it reaches 78.3% of sites.

Turning to the GDPR’s short run impact on market structure, Table 2 shows that aggregate HHI increases 17.3% among all vendors and 17.8% among all classified vendors. The top four vendors categories represent 94.3% of categorized vendors pre-GDPR, and HHI increases post-GDPR in each of these categories. The advertising category sees the largest increase in HHI, growing 25.3% from 348 to 436 points. The increases in HHI among the next three top categories are more moderate: 2.3% in hosting, 5.8% in audience measurement and 3.8% in social media.

Beyond the top 4 categories, also mixed results have been analyzed. Design optimization changes little (-0.5%), whereas HHI in security increases 8.9%. the native ads and CRM categories become less concentrated: HHI falls -4.8% and -4.5% respectively. Both categories are highly concentrated and so small that they represent only 1.1% of total categorized vendor reach. The increase in HHI in the advertising category (25.3%) is proportional to the decrease in the average number of vendors (24.1%), through this relationship is less than proportional in the remaining categories.

Several categories see HHI increases near or above the 100-point threshold that American regulators use to scrutinize mergers: advertising gains 88 points, audience measurement gains 239 points, social media gains 161 points and security gains 796 points.

As the total share of the top two firms, CR2 can be a more intuitive metric than HHI. In Table 2, the sign of the short-run change in CR2 reflects the change in HHI in all categories except security, where the baseline CR2 of 99.8 creates a ceiling effect.

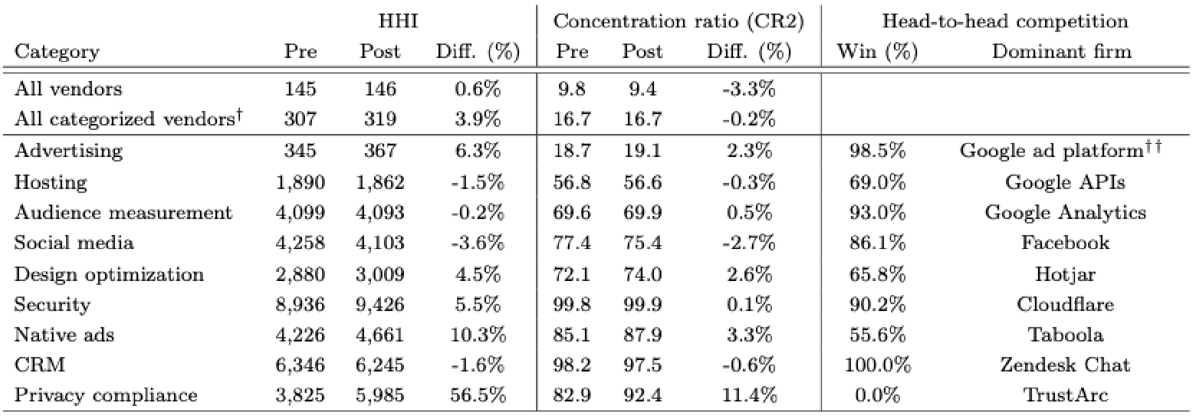
As with HHI, the largest increase in CR2 is in the advertising category, with a relative increase of 15.8% from a CR2 of 18.7 to 21.7; for the remaining top 4 categories, the relative increase in CR2 lies between 1.7% and 3.1%. The decreases in CR2 for design optimization native ads and CRM are small at -0.6%, -0.5% and -0.2% respectively.

Finally, Table 2 shows the head-to-head win rate of the dominant firm in each category. This metric reflects the probability that a website keeps the dominant category vendor and drops a competitor post-GDPR, conditional on employing both vendors pre-GDPR. The top 4 categories suggest that the increase in concentration is in part a story of Google and Facebook's dominance. In advertising, Google Ad Manager wins an exceptional 98.9% of these head-to-head battles. Google also wins in hosting (Google APIs) 74.3% of the time and in audience measurement (Google Analytics) 93.5% of the time. For its part, Facebook wins 87.2% of its head-to-head battles in social media.

Below the top 4 categories, sites tend to use a single category vendor and less than 75 head-to-head battles per category. In addition, the dominant firm's win rate also helps to explain the change in HHI for smaller categories. For instance, Hotjar wins only half of its head-ho-head battles in the design optimization category, which helps explain why the category's HHI is flat. In the security sector, Cloudfare wins 94.7% of the time, which helps to explain why that category sees the second largest increase in HHI. Taboola wins only 21.7% of the 23 head-to-head battles in the native ads category, which helps to explain why that category sees a 4.8% reduction in HHI.

Thus, the concentration ratios and win rate results suggest that sites prefer to keep the dominant firm over alternatives.

2.4.2 GDPR long-run concentration impact



	HHI			Concentration ratio (CR2)			Head-to-head competition	
Category	Pre	Post	Diff. (%)	Pre	Post	Diff. (%)	Win (%)	Dominant firm
All vendors	145	146	0.6%	9.8	9.4	-3.3%		
All categorized vendors†	307	319	3.9%	16.7	16.7	-0.2%		
Advertising	345	367	6.3%	18.7	19.1	2.3%	98.5%	Google ad platform††
Hosting	1,890	1,862	-1.5%	56.8	56.6	-0.3%	69.0%	Google APIs
Audience measurement	4,099	4,093	-0.2%	69.6	69.9	0.5%	93.0%	Google Analytics
Social media	4,258	4,103	-3.6%	77.4	75.4	-2.7%	86.1%	Facebook
Design optimization	2,880	3,009	4.5%	72.1	74.0	2.6%	65.8%	Hotjar
Security	8,936	9,426	5.5%	99.8	99.9	0.1%	90.2%	Cloudflare
Native ads	4,226	4,661	10.3%	85.1	87.9	3.3%	55.6%	Taboola
CRM	6,346	6,245	-1.6%	98.2	97.5	-0.6%	100.0%	Zendesk Chat
Privacy compliance	3,825	5,985	56.5%	82.9	92.4	11.4%	0.0%	TrustArc

Table 3: Long-run GDPR impact on concentration (27 weeks post) (Garrett et al., 2021)

Table 3 explores the change in concentration by category 27 weeks after the GDPR by replicating the calculation in Table 2 for the later time period. While aggregate HHI returns to baseline levels (0.006% higher), the aggregate HHI among vendors, whose purpose is classified, is still 3.9% higher than the baseline.

The largest category – advertising – remains 6.3% more concentrated than the pre-GDPR baseline, while the next three categories see small decreases. Average vendors in the native ads category (0.07 vendors) remain lower than the pre-GDPR baseline (0.08 vendors), though the sign of the long-run HHI impact has reserved to +10.3% from -4.8% one week post-GDPR.

In sum, the GDPR coincided with a short-run increase in aggregate web technology concentration. While market concentration does not always follow a reduction in vendor use, the largest web technology categories become more concentrated. Many categories are highly concentrated initially and several categories exhibit significant increases in concentration, relative to both the underlying change in category use and the 100-point threshold that regulators use to scrutinize mergers. The three different concentration metrics paint a consistent picture of these results. In aggregate, short-run concentration effects appear to dissipate over the long run, though increased concentration in the advertising technology market persists to some extent. This model again suggests that this dissipation is consistent with publishers' declining beliefs about regulatory enforcement in the absence of enforcement actions.

2.5 Vendor usage analysis

2.5.1 GDPR impact on vendor usage

This empirical strategy relies on before-after comparisons to measure the effect of the GDPR. Identifying a control group poses a key challenge for studying the GDPR.

First, non-EU websites may not be representative and are still subject to the GDPR if they target EU users. Given confusion of the interpretation of "targeting" – which the European Data Protection Board clarified in November 2018 – non-EU sites may treat EU traffic with caution. GDPR therefore can affect how non-EU websites treat EU users. Second, non-EU users may experience some spillover effects of the GDPR: that is, websites may implement GDPR measures for non-EU users to reduce administrative costs or enforcement scrutiny.

The analysis showed that non-EU websites implement GDPR measures for EU users and that websites expose EU users to fewer vendors than non-EU users. Thus, neither non-EU sites nor non-EU users represent clean controls.Then, it is favored pre- vs. post-GDPR comparisons for a set of sites to quantify the effect of the GDPR on web technology vendors. Specifically, it has been analyzed how the number of vendors engaged by a website evolves after enforcement of the GDPR, by using the following fixed effects regression:

$$y_{it} = \mu + \lambda_t \cdot GDPR_t + \theta_i + \varepsilon_{it}$$

where y_{it} is website *i*'s number of technology vendors at time *t*, $GDPR_t$ is an indicator for the post-GDPR enforcement date *t*, θ_i is a site fixed effect, and ε_{it} is the error term. The coefficient λ_t therefore captures the difference in the average number of web technology vendors relative to the pre-GDPR baseline (μ), after conditioning on website fixed effects.

It can be noticed that the econometric specification attributes all common temporal changes in outcomes to the GDPR.

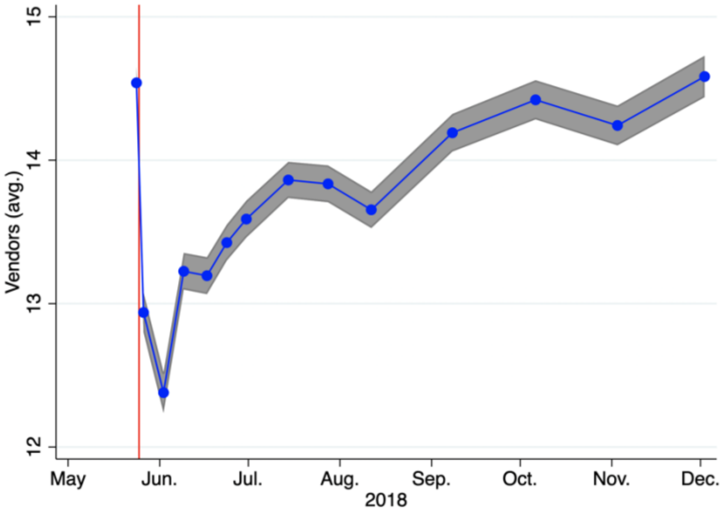


Figure 3: Evolution of average web technology vendor usage per website (Garrett et al., 2021)

Figure 3 shows how website use of web technology vendors evolves over 2018, by plotting the regression estimates (λ_t) as a function of time. It can be seen that average web technology vendor use drops sharply after GDPR enforcement on May 25, 2018 – denoted by the vertical red line. Vendor use reaches its minimum one week later. The comparison between the pre-GDPR baseline and one week post-GDPR is used as short-run GDPR effect estimate.

The short-run estimate shows that sites reduce web technology vendors 14.9%, from an average of 14.5 to 12.4 vendors. Three quarter of this reduction happens right after the enforcement deadline as the number of vendors falls 11% between the initial scan on May 23-24 and the second scan on May 25-28. This finding suggests that most publishers waited until the last minute to adjust the vendors on their site.

One of the starkest findings is that the short-run GDPR effect appears to erode over time: by the end of 2018, the average number web technology effectively returns to its pre-GDPR level. This result is still consistent with Zhuo et al. (Zhuo et al., 2019), who do not find that GDPR affected the Internet’s network level connectivity: due to the modest contribution of vendors to aggregate data flows as well as the magnitude and duration of the GDPR effect.

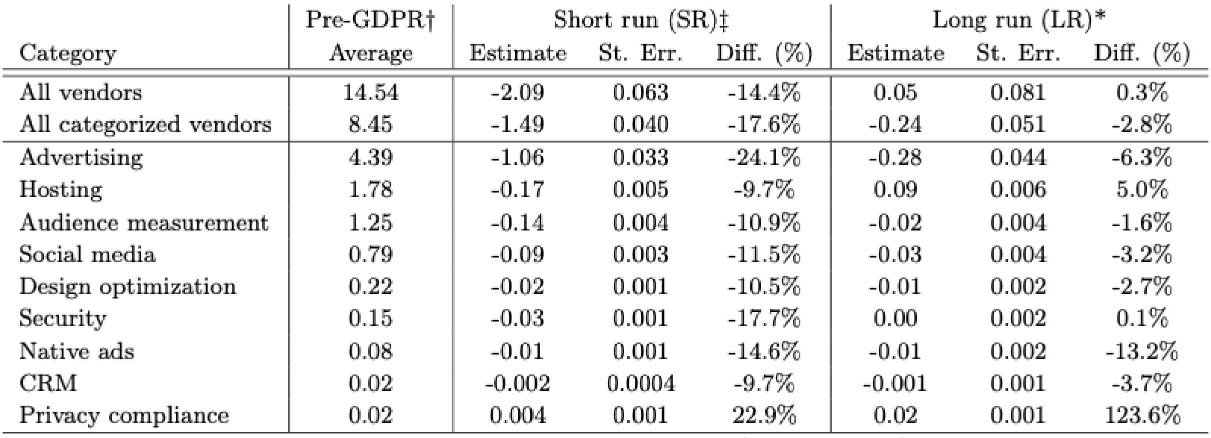
In interpreting the pattern of Figure 3, the descriptive nature of the results has been emphasized. Thus, the post-GDPR growth in vendor use arises from a combination of: (1) declining site beliefs about GDPR enforcement, and (2) natural industry growth.

These two explanations may be understood within the context of this theory model. Explanation (1) corresponds to enforcement beliefs declining over time following the GDPR implementation (α → 0 as t → ∞), a corollary to Proposition 3. These beliefs about GDPR enforcement risk were most heightened around the May 25 enforcement deadline and subsequently declined as regulators did not levy fines and signaled, they would first study the industry. The enforcement expectations explanation is also consistent with the European Commission review of GDPR, which cited lack of enforcement as an obstacle to the full realization of the regulation (European Commission, 2019).

Explanation (2) posits that the GDPR induced a downward shift in vendor usage after May 25, which eroded over time due to innovation among web technology vendors. Proposition 5 implies that vendor innovation could materialize in the form of declining data processing costs (δ) or reduced service substitutability (γ). Growth and innovation are typical for the vendor industry: Lerner et al. (2016) show a steady increase in website use of third-party domains between 1966 and 2016. Moreover, Peukert et al. (2020) document an increasing trend in vendor use through 2017 and 2018, excepting the enforcement deadline.

Explanations (1) and (2) highlight the rationale for using short-run effect estimate as the relevant benchmark for the GDPR's impact on the web technology industry. To the extent that GDPR effects diminish over time due to flagging enforcement expectations, the short-run estimate captures the regulation's effect when enforcement beliefs were highest in 2018. The short-run estimate is also preferred on the economic grounds, because the long-run estimates become increasingly confounded by industry growth trends.

Then, the analysis shifts to vendor usage by service category. Category-level analysis lends insight into the GDPR's effect on different types of web technology vendors and links directly to the discussion of market concentration, which relies upon the same categorization scheme. With nine categories, Table 4 is a comparison of the short-run GDPR effect (1 week pre- vs. 1 week post-GDPR) and a single long-run effect (1 week pre- vs. 27 weeks post-GDPR) for each category, as reported below.



	Pre-GDPR†	Short run (SR)‡			Long run (LR)*		
Category	Average	Estimate	St. Err.	Diff. (%)	Estimate	St. Err.	Diff. (%)
All vendors	14.54	-2.09	0.063	-14.4%	0.05	0.081	0.3%
All categorized vendors	8.45	-1.49	0.040	-17.6%	-0.24	0.051	-2.8%
Advertising	4.39	-1.06	0.033	-24.1%	-0.28	0.044	-6.3%
Hosting	1.78	-0.17	0.005	-9.7%	0.09	0.006	5.0%
Audience measurement	1.25	-0.14	0.004	-10.9%	-0.02	0.004	-1.6%
Social media	0.79	-0.09	0.003	-11.5%	-0.03	0.004	-3.2%
Design optimization	0.22	-0.02	0.001	-10.5%	-0.01	0.002	-2.7%
Security	0.15	-0.03	0.001	-17.7%	0.00	0.002	0.1%
Native ads	0.08	-0.01	0.001	-14.6%	-0.01	0.002	-13.2%
CRM	0.02	-0.002	0.0004	-9.7%	-0.001	0.001	-3.7%
Privacy compliance	0.02	0.004	0.001	22.9%	0.02	0.001	123.6%

Table 4: GDPR impact on average vendor use by category (Garrett et al., 2021)

The first column of Table 4 reports the pre-GDPR mean and the next three columns report the short-run GDPR coefficient estimate, standard error and percentage difference relative to the pre-GDPR mean. Each coefficient represents a separate fixed effect regression corresponding to each vendor category outcome. It can be seen that web technology vendors overall fall 14.5% and the subset of categorized vendors falls 17.7% from 8.4 to 6.9.

The category-level results in Table 4 reveal that the average number of vendors falls for all but one category in the short run. The exception is the "privacy compliance" category, which is expected to benefit from the GDPR. However, few sites use vendors in the privacy compliance category, as these increase from only 0.017 to 0.021 vendors on average.

Advertising is both the largest category and the category that falls the most (24.3%) from 4.35 to 3.29 average vendors. Hosting, audience measurement and social media are the next largest categories and these categories fall by 9.7%, 10.9% and 11.5% respectively.

The remaining categories appear infrequently with means of at most 0.22 vendors per site.

The last three columns of Table 4 report the long-run change in web technology vendors by category. For all categories, vendor usage increases in the long run, compared to the measured short-run effects. The advertising category shows the largest attenuation of the short-run effect, which is consistent with either significantly revised GDPR enforcement expectations or greater innovation in this category.

In sum, the short-run estimates demonstrate clear evidence of the GDPR effect. This drop is sudden with 74% of the short-run reduction in vendors arising within a couple days of the enforcement deadline. There is no comparable change in vendors usage for the rest of 2018, and other research suggests no such change in the year prior to the GDPR either (Peukert et al., 2020). Vendor usage falls in all categories but privacy compliance, where it increases. This pattern is consistent with a GDPR effect rather than technology change or some other transitory shock.

2.5.2 Usage effect heterogeneity and GDPR mechanisms

The GDPR's Recital 9 expresses the need for the GDPR to standardize EU Data Protection Law, though important differences between countries persist. The central EU privacy regulator recognizes that EU country-level regulators vary both in their resources and their enforcement intensity (European Data Protection Board, 2020).

This cross-country variation has been exploited to gain insight into GDPR's effects across regions with different expectations of regulatory strictness. Specifically, the analysis is based on a European Commission survey of 4,835 data controllers that asks if the country's regulator is more or less strict than other countries in the EU (European Commission, 2008). According to this survey, the strictest data regulators are Germany and Sweden whereas the laxest regulators are Bulgaria and Greece. In addition, a regulatory strictness measure for each website as the weighted average of the four-point country strictness survey measure, with weights determined by the proportion of website users from each EU country – sourced from Alexa –, has been introduced.

The role of enforcement expectations has been analyzed by comparing vendor usage among websites in high regulatory strictness regimes to those in low regulatory strictness regimes. To infer both short and long run effects, the same model of Paragraph 2.5.1 has been replicated, by estimating equation on two data subsamples and plotting the model-predicted average usage over time.

There is a focus on majority-EU sites – site receiving more than 50% of their traffic from EU users – to ensure a tighter link between the website and a given EU country's regulator and to avoid any confounding relationship between foreign sites and traffic from certain EU countries. Then, the sample has been divided into quartiles of regulatory strictness.

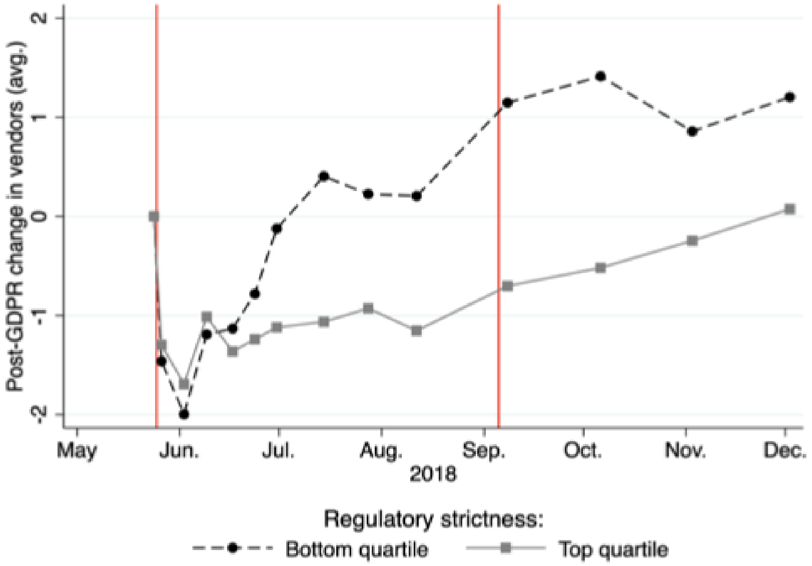


Figure 4: GDPR effects moderated by expectations of regulatory enforcement (Garrett et al., 2021)

Figure 4 summarizes the analysis by plotting average vendor usage among websites in the top quartile of regulatory strictness and websites in the bottom quartile. It can be noticed that sites in stricter regimes return to pre-GDPR levels at the end of 2018, whereas sites in laxer regimes do so in July. These results strongly suggest that enforcement expectations materially contribute to the pattern of GDPR usage effects captured in Figure 3.

Other GDPR studies also find that regulatory strictness moderates the impact of the GDPR technology venture capital (Jia et al., 2019) and recorded site outcomes (Goldberg et al., 2020). Then the analysis focuses on website characteristics that potentially explain variation in vendor choice. The theory model suggests an important role for exogeneous website revenue shifters, denoted by β. Specifically, Proposition 4 predicts that higher revenue websites will reduce vendor usage more than their lower revenue counterparts.

Three observed variables shift website revenue can be observed. First, the traffic rank of the site – which is inversely related to the volume of site traffic – as measured by Alexa.

Second, the average user income, which is expected to correlate with advertising and e-commerce revenue. To construct this measure, the share of users of each site by country – from Alexa – was used to weigh income per capita at the country level.

Third, the number of ads per page, which is expected to shift site revenue from advertising. This variable was collected in August 2018 by visiting each website homepage while using an ad blocker to count the number of blocked ads, following Shiller et al. (Shiller et al., 2018).

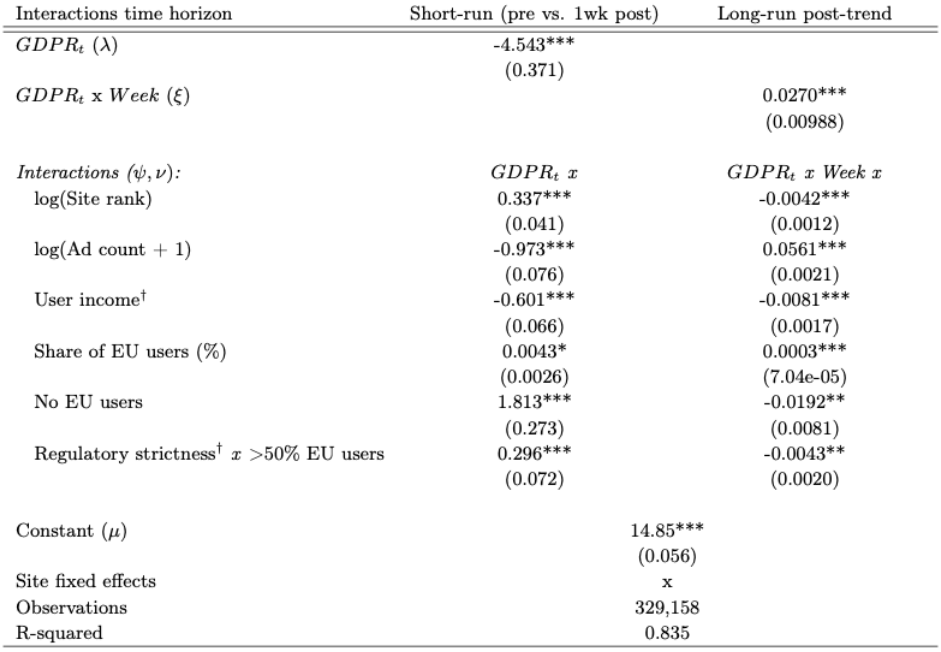
The analysis further includes three website characteristics that are not related to revenue but are of interest. First, the normalized measure of regulatory strictness – for majority-EU sites – as a co-variate in the analysis. Second, the website's share of traffic from the EU, again using data from Alexa. The share of users from the EU potentially affects the relative benefit of sharing personal data given that GDPR fines are a share of global revenue. Third, an indicator for sites without traffic from the EU as sites that do not target EU users are not subject to the GDPR. No potentially endogenous changes in website characteristics over time are included, by fixing the associated variables using pre-GDPR levels whether possible.

The analysis seeks to analyze differences in both the short-run and long-run effects of the GDPR – on vendor usage – by website characteristics. To do this, a panel dataset has been used to estimate a regression equation that interacts the six website variables discussed above with a post-GDPR indicator and a post-GDPR time trend:

$$
\begin{aligned}
y_{it} \quad = \quad & \mu + \lambda \cdot GDPR_t + \sum_k \psi_k \cdot X_{ik} \cdot GDPR_t + \\
& \xi \cdot GDPR_t \cdot t + \sum_k \nu_k \cdot X_{ik} \cdot GDPR_t \cdot t + \theta_i + \epsilon_{it}
\end{aligned}
$$

where y_{it} indicates the number of vendors used by website *i* at time *t* and $K \in \{1, ..., 6\}$ indicates website characteristics to that X_{ik} is the K^{th} (time-stationary) characteristic of website *i*. The model includes site fixed effects θ_i, so only site characteristics X_{ik} as an interaction are included. This regression is designed to test whether website characteristics explain post-GDPR differences in usage levels – short-run effects – and usage trends – long-run effects. As such, the full panel has been used except the May 25th scan, so that the short-run GDPR interactions better capture the 1 week post-GDPR effects.

The first column of Table 5 estimates the short-run GDPR interaction effects and illuminates the role of website incentives. The short-run effect of the GDPR is to increase the chance of a penalty from 0 to some $\alpha_{SR} > 0$. This theory model thus predicts how different sites will cut vendors as α increases. From Proposition 4, sites with greater revenue shifters (β) engage more marginal vendors ex-ante and thus will cut more vendors post-GDPR. It can be noticed that the sign on site rank is opposite those on ad count and user income because site rank is inversely related to site traffic, the corresponding coefficients in column (1) have the predicted sign and are reach statistically significant with $p < 0.01$, lending empirical support to Proposition 4.



Interactions time horizon	Short-run (pre vs. 1wk post)	Long-run post-trend
$GDPR_t$ (λ)	-4.543***	
	(0.371)	
$GDPR_t$ x $Week$ (ξ)		0.0270***
		(0.00988)
Interactions (ψ, ν):	$GDPR_t$ *x*	$GDPR_t$ *x Week x*
log(Site rank)	0.337***	-0.0042***
	(0.041)	(0.0012)
log(Ad count + 1)	-0.973***	0.0561***
	(0.076)	(0.0021)
User income†	-0.601***	-0.0081***
	(0.066)	(0.0017)
Share of EU users (%)	0.0043*	0.0003***
	(0.0026)	(7.04e-05)
No EU users	1.813***	-0.0192**
	(0.273)	(0.0081)
Regulatory strictness† *x* >50% EU users	0.296***	-0.0043**
	(0.072)	(0.0020)
Constant (μ)	14.85***	
	(0.056)	
Site fixed effects	x	
Observations	329,158	
R-squared	0.835	

Table 5: Heterogeneity in GDPR effect by website characteristics (Garrett et al., 2021)

The interactions with the share of EU users is positive and marginally significant ($p < 0.1$). This result suggests that sites with the smallest share of EU users make the deepest cuts to their vendors. This finding is expected to relate to the design of the GDPR penalties. Since GDPR penalties are 4% of global revenue, site with a small share of EU users have comparatively little to gain but more to lose from violating GDPR provisions, leading them to cut more vendors. There is an expectation for a positive and significant discontinuity in the interaction between the GDPR and an indicator for sites without EU users. This discontinuity is expected since sites that do not serve EU users are excluded from the GDPR and should therefore respond little to the GDPR, if at all. The positive interaction between share of EU users and the GDPR indicator illuminated how the design of the GDPR's penalties shapes website's choice of vendors. This represents an unintended consequence of the GDPR penalty design: sites that serve more EU users are less privacy-protective in this sense.

Finally, the coefficient on regulatory strictness is positive and highly significant ($p < 0.01$), so that sites facing stricter data regulators cut vendors less in the short run. However, the magnitude of this effect is small, as suggested by Figure 4. A one standard deviation increase in regulatory strictness implies 2.0% (0.296/14.54) more vendors one week after the GDPR.

There are three explanations. First, website facing stricter regulators may have been more careful with their vendors use prior to the GDPR in response to existing laws and the EU e-Privacy Directive (Goldfarb, 2015). Second, this could be interpreted as a weak evidence of anticipatory behavior by sites that face a strict regulator. Third, if sites are risk averse and low-strictness regimes are also more uncertain, it can therefore be observed that sites in low-strictness regimes are cutting vendors more.

The third column of Table 5 reports the estimates of the long-run GDPR interaction effect, which capture heterogeneity in post-GDPR usage trends. As expected from Figure 4, the baseline trend after the GDPR in column (3) is a positive: the number of vendors increases by 0.0027 per week. The post-GDPR trend linked to regulatory strictness is negative (-0.0043) and significant ($p < 0.05$). This result provides evidence that the central insight from Figure 4 is robust to the inclusion of covariates: website beliefs about the enforcement probability play an important role in determining vendor usage, with sites in stricter regimes being slower to add vendors post-GDPR.

The estimated signs of the revenue shifters and post-GDPR time trend interactions are largely as expected. Suppose that, in the absence of enforcement, website beliefs about the likelihood of enforcement fall – that is, α falls from its short run value α_{SR} to its long run value $\alpha_{LR} < \alpha_{SR}$. Then, Proposition 4 predicts that the sign on interactions with the revenue shifters should flip.

The results validate this prediction for the interactions with site rank and ad count: sites with more traffic and more ads lead the post-GDPR growth in vendor use. The sign on income is negative and highly significant, which represent the sole finding contrary to the theory-driven hypothesis. Thus, a correlation between income and user preferences for privacy might discipline sites with high-income user traffic.

Finally, it can be noted that the interaction of the post-GDPR time trend with the share of EU users is positive and significant while the interaction with the indicator for sites without EU users is negative and significant. Then, GDPR-exempt sites, which experienced smaller short-run usage declines, actually tend to reduce more rapidly post-GDPR than sites with small shares of EU traffic. This is consistent with the GDPR's incentive structure, which penalizes firms based on their global revenues.

2.6 Is GDPR worth the cost?

The GDPR revolutionizes Europe's data protection regime and significantly affects how organizations worldwide collect, use, manage, protect and share personal data that come into their possession.

As personal data increasingly represents an important class of economic asset for organizations, the regulatory environment across European member states is undoubtedly shifting and regulators have greater powers of enforcement. GDPR replaces a regime under which fines for a data breach were limited and enforcement actions infrequent. By contrast, the scale of the fines under GDPR – which can reach as high as €20 million or 4% of an enterprise's annual global turnover – has understandably generated concern in boardrooms, and many are keen to know whether these fines can be insured (Leemans and Molony, 2018). Indeed, there have been several high-profile fines since the GDPR came into force, with GDPR Fines Quarterly Report finding that organizations were fined more than €182 million in 2020 alone (Irwin, 2021).

A recent discussion suggests that the compliance costs of the GDPR may be substantial. Indeed, during the first year of implementation of the GDPR, approximately 500,000 European organizations hired a data protection officer (a DPO) – whose average annual salary is approximately 80,000 euros in Europe – (International Association of Privacy Professionals, 2019). This means an additional cost of 40 billion euros solely for the employment of the DPOs. In addition, organizations have needed to use their resources to adapt their information systems and practices to comply with the requirements of the GDPR. During the first nine months of the GDPR, the total fines issued amounted to approximately 56 million euros. Fifty million euros of this total amount is due to a single fine that the French Data Protection Agency (CNIL) issued to Google (Koch, 2019).

However, given that European data protection agencies have handled approximately 100,000 self-reported breaches and user complaints under the GDPR, it becomes clear that the amount of fines issued does not give a complete picture of the extent to which European firms have achieved and maintained compliance with the GDPR.

Achieving and maintaining compliance with the GDPR is likely to increase costs and, hence, deteriorate the financial performance of all EU companies collecting and handling personal data. However, it is not clear what the actual magnitude of these effects is or whether the effects vary across different types of firms (e.g., small vs. large) and across different industries (consumer data intensive vs. non-intensive).

Currently, as the GDPR came into force only recently, the empirical literature on the economic impacts of the GDPR is scarce. Yan and Li (2019, 1-15) look at the early effects of GDPR compliance investments on the group of organizations dealing with highly sensitive personal data: they estimate the pre-2018 financial impacts of the GDPR on EU hospitals that provide digital health services as their primary business. They use the difference-in-difference method and employ hospitals with at most a small share of digital services in their operating business as their control group. Yan and Li (2019, 1-15) find a negative effect on the financial performance of the hospitals in the treatment group. Another group of firms that is heavily affected by the GDPR is online firms collecting and managing vast quantities of individual-level user data. Goldberg, Johnson and Shriver's (2019) study among approximately 1,500 online firms shows that the enforcement of the GDPR further adversely affected European web traffic and e-commerce sales. Their weekly data from the 2nd through 38th weeks of 2017 and 2018 indicate approximately a 10% post-GDPR drop in recorded pageviews and revenues in Europe.

The GDPR may affect the performance of not only incumbent companies but also, via investors' future profit expectations, the funding available for newly established companies. Jia et al. (2019) compare venture capital investments in new and emerging technology firms in Europe and in the U.S. before and after the enforcement of GDPR. Their estimations using monthly investment data from January 2014 to April 2019 suggest that the short-term negative impacts of the GDPR on EU technology venture investments were substantial.

Thus, compliance may incur costs in various ways, as discussed below.

First, achieving GDPR compliance requires that decision makers in firms understand what the GDPR implies and what kind of actions it requires. Second, training of employees may be necessary for the skills needed for achieving and maintaining compliance. The actual responsibilities depend on whether the firm is what the GDPR defines as a data controller or a data processor. A firm that owns data is called data controller. It is responsible for determining the purposes for which the data are used and the means of data processing. The data controller is also responsible for responding to consumers' requests concerning their personal data. On the other hand, a firm that processes data for a third party is called a data processor. It processes the data only by following instructions from the data controller and keeps records of how the data are processed. Regardless of whether a firm is a data controller or a data processor, it is responsible for securing the data and keeping records of the actions taken for data security. A firm may have the responsibilities of both a data controller and a data processor. Under certain conditions, the firm needs to designate a data protection officer who is responsible for designing and implementing the firm's data protection plan as well as monitoring the implementation. In short, compliance calls for a considerable allocation of human resources. A firm may allocate the tasks that GDPR compliance brings about either to the current employees or to new hires. Reallocating current employees to take care of GDPR compliance reduces output, while hiring new employees raises salary costs. Either way, GDPR compliance decreases the firm's financial performance.

In addition to human resources, achieving and complying with the GDPR imposes needs on IT infrastructures and software. If the current IT infrastructure is insufficient or the current software does not cater to the requirements imposed by the GDPR, the firms needs to invest in IT capital. Similar to human resources, IT investments made to comply with the GDPR incur costs without increasing output. In other words, the obligation to invest in this type of non-productive capital has a negative effect on firm profitability.

Moreover, implementing new systems and practices may temporarily disrupt or slow down the normal operations of a firm. If a firm uses consumer data intensively, GDPR compliance may require the firm to make substantial changes to its procedures. Thus, GDPR compliance may cause a negative productivity shock or even a lasting productivity drop. A decrease in productivity reduces the firm's output and affects the firm's performance negatively.

Many firms also lack legal and technical expertise to implement the changes that GDPR compliance necessitates. These firms are likely to purchase legal and technical services from a third party. Even in a firm which has expertise, it may be more cost-effective to outsource the design and implementation of the changes that GDPR compliance requires. An increase in service expenses naturally has a negative effect on firm profitability.

In the longer term, the GDPR may affect firms' performance by changing the efficiency of resource allocation among firms and sectors. The capability of the firm to cost-effectively meet the requirements of the GDPR affects its financial performance and, in the end, whether or not the firm can operate in the given market or industry. The GDPR may thus affect market structure and competition. Campbell, Goldfarb and Tucker (2015) propose a theoretical model to examine how regulation regarding the privacy of consumers' data affects the competitive structure of data-intensive industries. They assume that consumers incur a cost when prompted to give consent to use their data. They make this assumption following the EU's Data Protection Directive (95/46/EC) and Privacy and Electronic Communications Directive (2002/58/EC) and its amended (2009/135/EC), which were active before the GDPR came into force. Their model implies that privacy regulation imposes transaction costs that fall disproportionately on small and new firms, which may deter the entry of new firms. This conclusion is in line with the empirical finding of Jia et al. (2019) that the adverse effects of the GDPR on venture capital investments in emerging technology companies in Europe may block the entry of innovative companies.

The GDPR concerns any firm that owns or processes personal data of EU consumers. In other words, an organization, irrespective of its location, needs to comply with the GDPR if it tracks, collects, stores, uses or analyses the data of citizens and residents of the EU. Consequently, the implementation of the GDPR has generated compliance costs not only to European firms but also to companies outside of the EU dealing with data of individuals located in the EU area. Hence, the largest U.S. companies, also active in the EU, have been hit hard by the compliance requirements. The International Association of Privacy Professionals and Ernst & Young suggest that the combined GDPR compliance costs of Fortune 500 firms amount to 7.8 billion U.S. dollars. It seems that it has primarily been the largest U.S. multinational companies and data-centric firms dealing with EU citizens data that have swiftly acted on the GDPR in the United States. Consequently, it seems possible that large U.S.-based firms, particularly those that are data-intensive high-technology companies, have not had a change in their financial performance different from that experienced by their European counterparts.

Moreover, the compliance costs of the GDPR are likely to differ across industries. The required actions and costs of the GDPR compliance are presumably higher the more intensively the firm collects, processes and/or manages consumer data. The financial implications of GDPR compliance are therefore likely to differ across industries. Information and communication as well as finance are examples of sectors where consumer data are used intensively and where the financial impacts of GDPR are expected to be significant. It can also be noted that the GDPR brings about an increase in the demand for legal and technical counselling – perhaps even gives birth to new kind of services – these counselling industries can, in fact, gain from the GDPR.

In addition, if there are returns to scale in adopting GDPR compliant systems and processes, the financial effects of GDPR compliance are likely to differ across firms of different sizes. Irrespective of whether becoming GDPR compliant requires the firm to reallocate human resources or hire new employees, invest in IT capital, undergo a drop in total factor productivity, or purchase technical or legal services, the financial impacts can be seen in the firm's profit margin.

2.7 The way forward

The GDPR is the Magna Carta of data protection, the importance of which cannot be overstated. Data protection creates an inherent clash with competing values, most importantly the potential loss of benefits from better data-based knowledge (Acquisti et al., 2016). To balance these tradeoffs, the GDPR does not prohibit data collection or sharing. Rather, it provides a mechanism for control, accountability and liability over data collection, processing and use, by combining individual rights with systemic governance (Kaminski, 2019).

Yet, the price of data protection through the GDPR is much higher than previously recognized (Campbell et al., 2017). As elaborated, the GDPR has two main harmful effects: it limits competition in data markets, creating more concentrated market structures and entrenching the market power of those who are already strong, and it limits data synergies, thereby preventing the creation of some data-based knowledge. Such effects belie the confidence expressed by the European Commissioner for Justice, Consumers and Gender Equality Vera Jourovà, according to whom "the big guys increasing market share? I don't believe the GDPR will have such a consequence" (Schechner and Kostov, 2018).

In a world where "those that know how to use data have a decisive competitive advantage… through raising performance, offering more user-centric products and services, fostering innovation," (EU Commission, 2017) and where the battlefield over data-based advantages has become global, the GDPR's effect on data markets cannot be disregarded. First, the GDPR has implications well beyond the geographic borders of the EU. This is because many international firms which operate in the EU, or trade with it, must comply with its rules (Batikas et al., 2020). Once such firms adopt the internal mechanism necessary for GDPR compliance, these may be used for non-EU data as well. In addition, some jurisdictions are following in the footsteps of the EU and adopting laws which resemble the GDPR (e.g. California Consumer Protection Act

Jan 2020). Widespread adoption of such laws could lead to even greater concentration of firms in international markets.

Second, most of the effects analyzed are long-term ones, which will not disappear once the market adjusts to the existence of the GDPR. It is thus worth reevaluating the overall welfare effects of the legal data regime chosen. But it is not to say the overall welfare effects of the GDPR are necessarily negative. The GDPR may negatively affect competition but still be welfare-enhancing. This will be the case if the harm to data subjects reduced by the GDPR is sufficient to compensate for its competitive effects, including its potential to increase user participation in the market based on increased trust (Campbell et al., 2017). The overall balance depends on the relative magnitude of these effects.

Several suggestions could be made for creating a more welfare-enhancing equilibrium.

Most importantly, competition law should give more weight to factors which might balance the negative effects of the GDPR on competition and innovation. For example, when evaluating the competitive effects of a merger or a joint venture, more weight should be given to considerations such as the ability of firms to engage in welfare-enhancing data sharing which may facilitate reductions in market concentration, or the potential for significant data synergies that could not be realized otherwise. This implies, for example, a more lenient policy towards mergers or joint ventures between small or medium-sized data controllers, which would enable them to reach economies of scale and scope in data analysis and compete more effectively with those who already enjoy such economies. It also implies that when at least one data controller in a proposed merger or joint venture already possesses strong comparative advantages in data analysis, a careful balance is required between the benefits of increased data synergies and the need to ensure the ability of other firms to effectively complete, in light of the increased hurdles to data collection and processing resulting from the GDPR. The conditions for applying the essential facilities doctrine and granting access to data might also need to be redefined in light of the effects of the GDPR. The interface between the GDPR and competition law, in cases where harm to privacy is minimal and benefits to competition and innovation are large, may also need to be reevaluated (Zingales, 2018).

In addition, assessments of market power and potential competition should take into account the actual competitive effects of the GDPR. No longer can it be assumed that new players seeking to accumulate large volumes of data face only low barriers, as was the case in several earlier Commission decisions, especially where separate entities collect different parts of the dataset. Considering, for example, the Commission’s reasoning when it approved a joint venture between Google and the global biopharmaceutical firm Sanofi aimed at using big data analysis to improve the management and treatment of diabetes (COMP/M.7813 SANOFI/GOOGLE/DMI JV, rec. 36 et seq. 4). The competitive analysis at the time addressed concerns that the venture would allow the parties to lock in patients by restricting their ability to direct their data towards alternative services. The Commission dismissed such claims on the grounds that data subjects had the right to data portability. Yet, such a right is generally not sufficient to address competitive concern. This is because the comparative advantages of the joint venture are partially based on existing large datasets owned by the parties. It may be difficult for competitors to overcome such comparative advantages unless they can convince a sufficiently large number of users to sign up to their services, or unless they can combine different datasets. The first option is limited by user "stickiness", by the fact that the data potentially arrive in a fragmented form and at different points in time, and by the fact that economies of scale and scope and network effects in data analysis create significant first-mover advantages that are difficult for new competitors to overcome. The second option, data sharing, must overcome high hurdles.

Thus, several other suggestions, that might go some way toward making the GDPR more welfare-enhancing, could be made.

First, where uncertainty regarding how to meet the GDPR's legal obligations contributes to concentration, it may be useful to consider ways of limiting such uncertainty. For instance, the GDPR establishes a right to data portability, but does not specify technical requirements for meeting this commitment. The development – by regulators together with industry – of technological standards for data portability and interoperability might help reduce the consequent uncertainty as to what standards might ultimately be applied (Gal and Rubinfeld, 2019).

Second, governments might support investment in the development of better and privacy protection tools which can retain more value from collected data (Layton, 2019).

Third, priority could be given to the development of better and faster tools for verification of GDPR compliance. And fourth, certification of data management and vetting processes could go a long way toward reducing costs. The government can either certify such tools or help facilitate such certification. In addition, the use of certified tools should be taken into account when assessing liability, and presumptions based on the use of reasonable tools should be created. This, in turn, could significantly reduce the risks involved in data sharing.

A final suggestion relates to the structuring of mandatory data-sharing obligations under other laws in a way which is sensitive to the fact that it has become more difficult for small or new firms to grow and enjoy significant data synergies by obtaining data from external sources. To illustrate, under the Public Sector Information Directive (Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013), some types of governmental data must be shared. The regulation does not differentiate between sharing with firms that already possess much data and with those that do no. It must thus be worth considering the option of asymmetric sharing of data, so that in certain circumstances the obligation to share data will relate mainly to sharing it with small or new entities. In line with this suggestion, it is worth exploring whether more flexible mechanisms for obtaining user content, such as opt-out rather than opt-in, should be applied with regard to certain types of data, the benefits of which are undeniable (Goldfarb and Tucker, 2015).

Thus, privacy policy is interlinked with competition and the resultant data-based innovation in more ways that have yet been recognized. In particular, the GDPR raises the transaction costs of sharing data between different data controllers. Recognizing such effects should probably lead to a re-evaluation of the balance reached, and to the adoption of tools to ensure that there is an overall increase in welfare.

Chapter 3

GDPR and the importance of Data to AI Startups

3.1 Big Data and AI Startups

As described in the AI Index 2018 Annual Report (Shomhan et al., 2019), AI has advanced rapidly over the past decade. Many scholars believe that AI has the potential to boost human productivity and economic growth (Furman and Seamans, 2019). Scholars also worry that these gains may come at a cost, potentially including labor displacement, income inequality and loss of privacy. AI algorithms rely on lots of data, often including data on individuals. In an effort to protect consumers' privacy, a number of regulators have passed or considered laws restricting use and sharing of data, including the European GDPR and the California Consumer Privacy Act (CCPA). Even though this increased regulation is intended to protect consumers' privacy, the legislation may negatively impact firms that need data to develop AI products. These burdens could be particularly costly for startups in Europe (Jia et al., 2019).

AI relies on large quantities of data. These data are used to train and tune algorithms. Certain types of algorithms, such as neutral network and ensemble learning algorithms, support more complex tasks and require more training data. Also, certain technologies are more difficult to develop. For example, a startup that wants to train the underlying natural language comprehension capabilities of a chatbot would benefit from using neural networks and relatively large amounts of training data, as compared with other less sophisticated technologies or algorithms. It is apparent through numerous contests, including those leading up to the prestigious Loebner Prize for most "human-like" AI chatbots, that data is a key ingredient for success. The need for data does not diminish with firm size; larger and smaller firms targeting the creation of similar AI products require similar data resources. However, large firms may be able to access data more easily from supplier and customer relationships as they benefit from a breadth of supplier relationships and a more developed customer ecosystem. Additionally, larger firms could benefit from complementary business models which provide data as an externality of normal business operations. So, firms that have user-based platforms as part of another business line may be able to reuse that customer chat data to develop their chatbot.

In addition to greater access to data through relationships, larger firms also have access to additional capital to hire computer scientists and engineers (Athey and Luca, 2019). It is even difficult for high-growth potential startups to raise capital (Nanda, 2016). Though startups benefit from cloud computing and other variable cost IT resources (Jin et al., 2018), larger firms could have an over-abundance of these IT resources. Slack resources, such as excess cloud computing capabilities, could be used to run valuable experiments (Thomke, 2003; Varian, 2014, 27-31) or to develop infrastructures that capture and store large amounts of customer data. More so, high technology firms may be able to avoid inertial tenancies to use outdated or aging IT resources.

There is an apparent tradeoff between access to training data and consumer privacy. GDPR and other types of data regulation make it harder for firms to collect, store and analyze certain types of data, especially personally identifiable or employment data. Also, these regulations may impact the willingness of other firms to enter into data sharing collaboration.

Further to a first-round survey of AI startups in how the AI product – of these startups – impacts labor (Bessen et al., 20018), Bessen et al. (2020) made a second-round survey of AI startups, which includes an additional 7 questions on data importance and the impact of the GDPR.

The survey was designed to address two questions. First, the impact of the GDPR and data regulation on AI startups. Second, the importance of data to AI product development. Bessen et al. (2020) found that training data is important for AI startups that rely on natural nets and ensemble learning algorithms. They also found that firms with customers in Europe are significantly more likely to create a new position to handle GDPR-related issues or to reallocate firm resources due to GDPR. This implies that the GDPR imposes costs, perhaps substantial costs, on startup AI firms.

Data privacy and protection continue to be a point of intense debate in research, policy and mass media. Several high visibility data breaches, from Equifax and Facebook/Cambridge Analytica in 2017 to the alleged hacking of Jeff Bezos's mobile phone more recently, have received significant attention in the news. Consumers continue to pressure regulators and legislations to more effectively safeguard their data and privacy. However, this increased regulation creates tradeoffs between safeguarding personally identifiable information and data access for entrepreneurial activities.

The European Union passed the GDPR in 2016, but the United States and other similarly advanced countries have yet to pass substantial regulatory policies. Given the interconnectedness of world economies, many firms are compliant with GDPR regardless of their headquarters location because they have customer operations in the EU. GDPR is being used as a rubric to frame similar legislation in other countries. More recently, CCPA focuses on the right to access your personal data from technology providers, to know what personal data is being collected and stored by employers, to delete one's data, and to prevent the sale of one's data.

Even though other states lack similarly exhaustive policies, firms that conduct interstate business or sell their products online often adhere to the most stringent state guidelines.

Prior research argues that GDPR increases the costs of collecting and using customer data. Revenues from online sales for EU firms, impacted by GDPR enforcement in 2018, dropped by 10% (Goldberg et al., 2019). Additionally, GDPR has asymmetrically impacted smaller firms in some industries. Enforcement of GDPR led to a reduction in the number of smaller web technology vendors used, leading to an increased concentration of more established, larger firms in the web technology industry (Johnson and Shriver, 2020). Rates of venture capital funding of startups in the EU also declined during this time period in comparison with the United States (Jin et al., 2019). After GDPR was legislated, many websites even outside of the EU were less likely to share personal data with web technology providers. At the same time, Google increased market concentration while smaller firms lost significant share, raising concerns over possible negative externalities to competition (Batikas et al., 2020).

The use of Big Data raises numerous critical questions for regulation and policies focused on safeguarding personal data (Boyd and Crawford, 2012). Until GDPR and CCPA, there was little guidance for firms on managing data privacy. The topic of data privacy and protection is intertwined with that of AI due to the necessity of data in product development. Often, personally identifiable information is intermingled with other firm data prompting many legal scholars to discuss data ownership and exclusion rights. Privacy concerns arise when firms may analyze data that include information about specific individuals. Also, this more personal data could be used in a way that leads to biased decision-making (Cowgill and Tucker, 2019). Computer scientists and programmers may not have the training method needed to create models that are less biased. Data that is less suited to the task could further exacerbate issues of bias.

There is some concern that GDPR and other data regulations, specifically in Europe, adversely affect entrepreneurial ventures (Jia et al, 2018). Even though some smaller firms with less than $1M in revenue are exempt from GDPR, this regulation has become – de facto – standard.

Ultimately, these startups are targeting swift revenue growth and investors want to know how they can quickly be compliant to regulatory policies. Also, systems need to be initially designed correctly to enable adherence. For example, the ability to delete customer data requires the ability to search all the data sources by a single customer’s name or identifier.

3.2 Survey on AI Startups

Artificial Intelligence is likely to drastically change the economy in a decade or so. That is, if 47% of jobs will soon be at risk from AI, then surely jobs will be at risk already at those firms using AI applications. A look at the cutting-edge applications of the technology provides a window into likely outcomes over the next decade or so. To gain a peek through that window, a global survey of startup firms developing and selling commercial applications based on AI has been carried out. The survey questionnaire covers topics about the startup firms and their markets and customers, about their technologies, their use of data, jobs, and other required skills.

The survey was administered online using Qualtrics from May to September 2018. The 22 questions (excluding name and email questions) were a preview sampling on the academics of roughly a half dozen firms with interviews. Potential respondents were contacted via email.

In addition to basic facts about the business of AI, the survey provides evidence that bears on several major questions that have been raised in the literature, including the impact of AI on jobs and occupations, and the role of data as a critical resource for AI-enabled startups.

The first question concerns the impact of AI on jobs. Frey and Osborne (2017, 254-280) base their estimate on a technical evaluation of "automatability", the technical feasibility of automation. The literature highlights several economic and technical reasons this might not be a sufficient metric for understating the impact of AI on jobs. One reason is that automating some or even most tasks an occupation performs does not necessarily eliminate the occupation. Historically, most automation has been partial. Bessen (2016) finds that of 270 detailed occupations listed in the 1950 Census, only one can be described as having been eliminated due to automation, namely the job of elevator operator. Furthermore, partially automating a job can increase employment in that occupation or industry as well as decrease employment (Acemoglu and Restrepo, 2018). This is because automation tends to decrease prices, driving greater demand. When demand is elastic enough, greater demand will offset the labor-saving effect of automation. For example, during the 19th century, the textile industry was heavily automated, yet employment rose (Bessen, 2017). More recently, the automatic teller machine (ATM) automated some of the work of bank tellers, yet their employment grew as well (Bessen, 2016). Also, the effect of AI is not just to automate tasks which replace humans, but also to enhance human capabilities at both performing new tasks and old tasks more effectively.

A second question that the survey addresses is which occupation will be affected.

Frey and Osborne (2017, 254-280) argue that lower wage occupations will experience greater job losses than higher wage occupations. The McKinsey Global Institute projects, somewhat differently, that high wage occupations will grow while mid-wage occupations shrink. Kaplan (2015) posits that "automation is blind to the color of your collar" and many professions will be devastated. Susskind and Susskind (2015) argue that new technology will lead to the decline of the professions. The survey provides some evidence about which occupations are growing and which are losing jobs in response to AI. Other recent papers that take the task-or-ability-based approach include Brynjolfsson, Mitchell and Rock (2018) and Felten, Raj and Seamans (2018, 54-57).

A third question the survey addresses is the extent of entry barriers into AI markets.

It appears that investment in AI is currently dominated by large firms, especially a few large tech firms (Bughin et al., 2017). In particular, some observers, such as Strucke and Grunes (2016), argue that the combination of data and network effects creates substantial entry barriers in online markets. For example, because Google has an enormous amount of search data, it might be hard for a new competitor to compete in the search engine market. Others contend that data, by itself, is not likely to pose an entry barrier (Lambrecht and Tucker, 2015; Sokol and Comerford, 2016). It may be that Google's advantage comes not from the amount of data, but from the results of the product-focused experiments. Varian (2010, 1-10) reports that Google conducted 6,000 experiments on its search engine in 2008. Moreover, while large amounts of data are typically needed for machine learning, there may be diminishing returns to the amount of data beyond a certain point. For instance, Bajari et al. (2018) find that increasing the number of online products that Amazon tracks does not significantly improve machine learning prediction accuracy after a certain point, implying that data quantity provides only a low barrier to entry. Amazon and other large technology companies benefit from the breadth of the data captured. For example, startups are more likely to sell a subset of the broader product offering of a larger company and to engage with small to medium size customers. In any case, the survey provides basic information about startups' access to data and circumstantial evidence on relative entry barriers in different industries.

A fourth question the survey addresses is the use of data protections by AI startups. Data protection and data privacy has become a flashpoint in the media, due in part to high profile data breaches such as at Equifax in 2017 and in part to high profile exposure of Facebook user's personal data to Cambridge Analytica in 2016 and 2017. Partially in response to these events, government regulators have instituted tighter rules on data protection. This has most notably manifested in Europe in the form of General Data Protection Regime (GDPR). There has been some concern that the increased data protection required under GDPR may constitute a barrier to entry for startups, and early research suggests that GDPR may in part be contributing to a slowdown in VC investment in European based startups (Jia, Jin and Wagman, 2018).

The surveyed startups sell their products in all major regions. The distribution of customers is generally broader, with over 50% of startup firms selling in the United States, in Asia and in Europe. Almost half of the startups (46%) sell to firms in the middle category of 51-1000 employees, but these firms make up only 26% of the market as measured by employment size. The survey firms sell to both smaller and larger customers, but proportionately less than would be expected given the distribution of firms in the US, as indicated by the distribution of firms in the LBD.

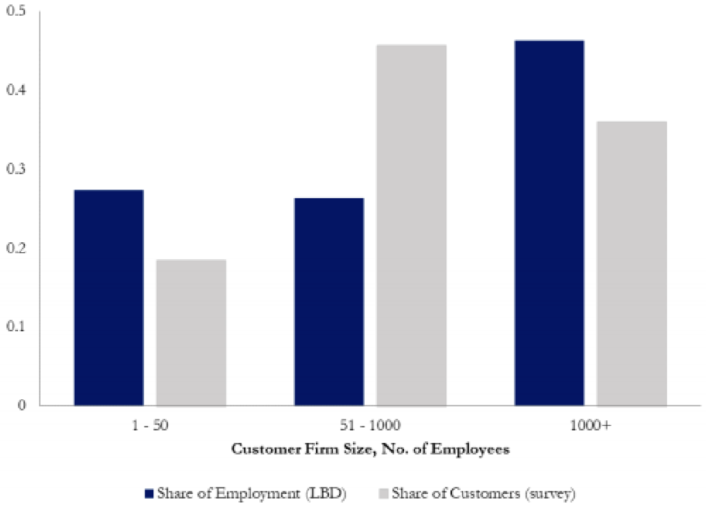


Figure 5: Startups Market Disproportionate to Mid-Sized Firms (Bessen et al., 2018)

The survey respondents sell to customers in all major industry sectors as well as to individual customers, as seen in Figure 5. In this sense, machine learning is a general-purpose technology that can be used in a variety of applications across a variety of industries (Cockburn et al., 2018, 115-146).

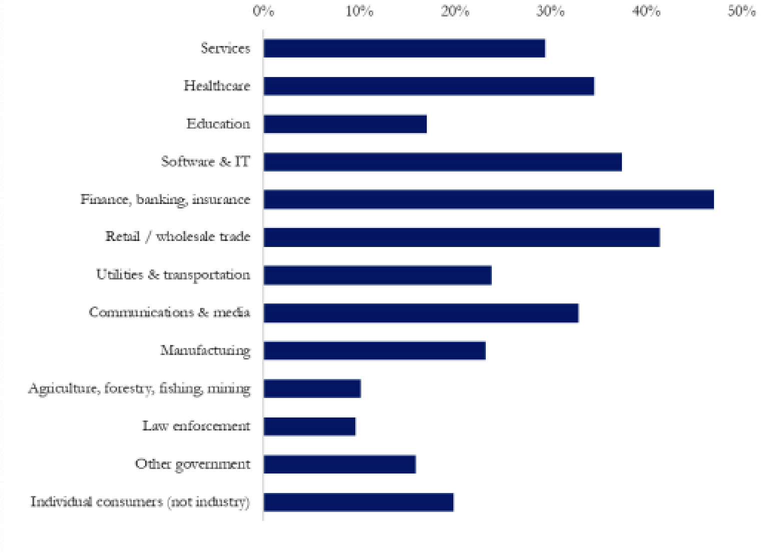


Figure 6: Share of Firms Selling to Different Industries (Bessen et al., 2018)

This broad distribution across industries suggests that the startup environment is healthy and many opportunities for entry exists. However, some industries may have relatively higher entry barriers than others. If there are no entry barriers, then both large and small firms will invest more in those industries that have greater technological opportunity. This means that without entry barriers, the distribution of AI development spending should look the same for small firms as it does for large ones. If, on the other hand, startups face significant entry barriers in an industry, then large firms would spend proportionately more in that industry. Comparing startup funding in different industries relative to total investment in AI in those industries helps identify those industries with possible entry barriers.

For what concerns technology, almost all the startups report that their products are cloud-based (97%) although 33% of the firms additionally provide software on premises. Only 3% of the firms provide software on premises only. Most of the firms (68%) provide a commercial application using AI. Some use the AI in their own products and services (43%) and 12% provide developer tools for AI applications.

In building their products, the surveyed firms use the technology to perform a variety of functions. Some of these are developed internally and some are purchased from outside vendors. The most commonly used technology is natural language understating and text analysis (62% of firms), followed by natural language classification and decision management (both at 56%). These are followed by visual recognition, including image, face and video (45% of firms) and sentiment/emotion analysis (43% of firms). Other technologies are used by a smaller percentage of firms.

Overall, many more firms develop their own software for the most commonly used technologies rather than purchase them from an external vendor. Only in two areas do firms rely more on outside vendors: speech recognition with 19% using external products while 13% develop their own, and natural language translation, with 17% using external software and 13% develop their own.

Regarding the type of algorithm used, 76% of reporting firms use neural networks including recurrent, convolutional, and generative adversarial neural networks. The next most common methods are clustering algorithms (59% of firms) and Bayesian or other methods of probabilistic inference (58% of firms). Other methods are used less frequently.

Thus, while startups use a wide variety of technologies to perform a variety of functions, a typical firm uses neural networks to do natural language text analysis for a cloud-based product.

Startups also use a wide variety of data. 57% of the firms use unstructured text, 44% use transaction data, 38% use image, 37% use administrative data or other structured records. A smaller share of firms use audio, video, or other types of data. These data are mainly used to train algorithms. Consequently, the algorithms are re-trained as more data accumulates. Roughly a quarter of firms report refreshing their models daily, weekly, or monthly. 13% of firms report having models that are not refreshed with new data.

Startup firms generally use other people's data. The most common source of data is from customers. Indeed, 80% of the startup firms report using customer data, including data about their customers and users as well as other data. 63% use data from third parties, including government data, data scraped from the Internet, and public benchmarking data. 51% of firms report using their own proprietary data. Most of this use is with data from other sources, only 6% of firms rely only on their own data.

To protect their access to data, startup firms who use customer data retain secondary reuse rights 52% of the time. To control the use of proprietary data between the firms and its customers, 83% of the firms use legal contracts that specify data uses. Additionally, firms use a variety of technical means to protect and control data access, including de-identification encryption, passwords, access logs, and application program interfaces.

Only 22% of firms report that the GDPR has impacted sales and marketing to non-EU countries. That figure is 27% for firms headquartered in Europe, excluding the United Kingdom. Given that the GDPR went into effect during the survey period, these figures might change as firms have more experience with the regulation. The survey respondents were asked to select all types of data protection used. As reported in Figure 7, across all types of data protection, startup firms with customers in the EU report using data protections more intensively than startup firms without customers in the EU. This could reflect the impact of the GDPR and also different customer sensitivities.

On the other hand, there appears to be limited differences in data protection according to firm size. Startup firms with ten or fewer employees represent close to 40% of the total survey respondents. The small startup firms appear slightly less likely to use legal contracts, de-identification, data encryption and password protected access. However, small startup firms are equally likely to use logged access and application program interface as large startups.

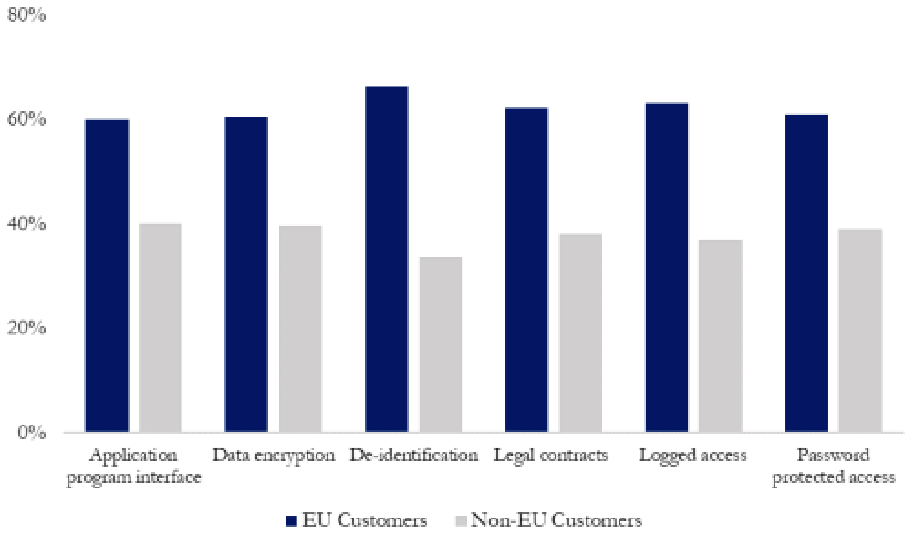


Figure 7: Data Protection by Customer Location (Bessen et al., 2018)

The analysis above suggests that most firms are oriented to enhancing customer capabilities rather than reducing customer labor costs, especially in the broad service sector. This suggests that AI might have some job-creating potential, especially for the professional, managerial, sales and marketing occupations that tend to be the users of these products. On the other hand, AI also reduces labor costs for some customers. Moreover, these effects might differ across occupations because the use of AI differs dramatically across occupations.

Clearly, AI is not all about destroying jobs. In particular, those occupations that are most likely to use AI are also most likely to see jobs created. At the same time, many jobs will be eliminated, especially in the three occupational groups that use AI relatively less. In many cases, jobs will be created in some occupations and jobs will be destroyed in other occupations at the same affected firms. Of the firms that responded to this question, almost half (46%) reported that their products both create and destroy jobs at customer firms. 26% report only creating jobs and 28% report only eliminating jobs. It is possible, of course, that survey respondents, perhaps sensitive to publicity about job losses, shaded their answer to reflect better job outcomes than is actually the case. Nevertheless, Figure 8 reveals dramatic relative difference between occupational groups. Moreover, the results here are roughly consistent with the evidence above that only about half of the firms report labor cost reductions as a substantial customer benefit.

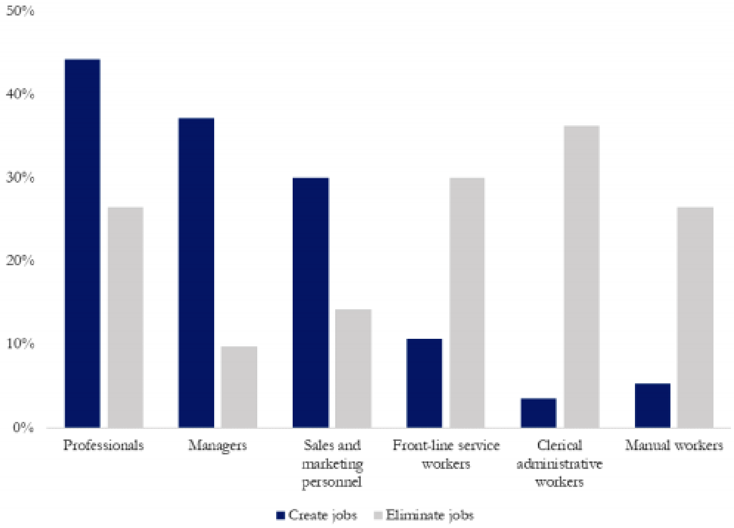


Figure 8: Job Creation and Destruction by Occupational Group (Bessen et al., 2018)

3.3 Constitutional democracy and technology in the age of AI

The thread of human rights, democracy and the rule of law are the core elements of western and liberal constitutions. These principles are the supreme law of the land – all actions of government, legislators and indeed social reality are measured against them (Kumm, 2013). Given the foreseeable pervasiveness of Artificial Intelligence in modern societies, it is legitimate and necessary to ask the question of how this new technology must be shaped to support the maintenance and strengthening of constitutional "Trinitarian formula" rather than weakening it.

The principle of rule of law, democracy and human rights by design in AI is necessary because on the one hand the capabilities of AI, based on big data and combined with the pervasiveness of devices and sensors of the IoT, will eventually govern core functions of society, reaching from education via health, science and business right to the sphere of law, security and defense, political discourse and democratic decision making. On the other hand, it is also high time to bind new technologies to the basic constitutional principles, as the absence of such framing for the Internet economy has already led to a widespread culture of disregard of the law and put democracy in danger, the Facebook Cambridge Analytics scandal being only the latest wake-up call in that respect.

In the same way the “Greening of GE” and generally of industry came about after environmental protection legislation incentivized and forced innovation in the direction of environmental sustainability, so now will the GDPR of the EU drive innovation for a way of collecting and processing personal data with respect individual rights and the importance of privacy in democracy (Nemitz, 2018).

It is clear that GDPR will always apply to AI when it processes personal data. GDPR contains important rights for users relating to any processing of their personal data as well as obligations of processors which will shape the way AI will be developed and applied. The principles of privacy and data protection by design and default set out in the GDPR are certainly becoming very important for AI as the limitations to automated processing and the related rights to meaningful information on the logic involved, the significance and the envisaged consequences of processing personal data with AI for those concerned. No new law is necessary in this respect.

On the other hand, in democratic discourse, it is important to know whether one’s counterpart in discussion is a human or a machine. If machines could participate in the political discourse without being identified as such or even impersonating humans without sanction, this would amount to an important distortion of discourse, untenable in democracy. No law secures that people are made aware if machines enter into dialogue with them in the political context. As transparent political discourse among humans is the key to democracy, the principle of essentiality prescribes that transparency must be created by law as to whether a machine or a human is speaking. Nontransparent machine speech and a fortiori impersonation should be sanctioned, and those who maintain major infrastructures of political discourse should be held responsible to ensure that there is full transparency regarding machine speech on their infrastructures. This will require new law.

3.4 A focus on AI Startups: Elaisian case study

Elaisian is an innovative Italian startup founded in Rome in 2016 by Damiano Angelici and Giovanni Di Mambro. In 2017, Elaisian was selected as one out of ten best projects by StartupBootCamp FoodTech for bringing real time IOT monitoring to olive trees in order to increase yields and reduce maintenance costs. Simultaneously, the startup started the first test with several manufacturers including Monini and partnered with Flos Olei.

During 2017, several tests were made to highlight the benefits of the product offered. Then, in 2018 Elaisian entered the market and launched a crowdfunding campaign that in 24 hours led the startup to raise the € 80,000 target budget. In 2019, the startup began to expand internationally, with a focus on the Spanish market. Indeed, they opened the first headquarters in Cordoba and then in Madrid.

Over the past 12 months, Elaisian's growth has been significant enough to double the team, without which they would not have achieved important results. Indeed, despite the Covid-19 pandemic crisis, they were able to adapt and react immediately in order to not be impacted by the situation, managing to maintain an annual growth of 250%.

Today the startup has 20 employees and its technology is used by over 1,000 agricultural holdings in 10 countries: Italy, Spain, Greece, Portugal, Cyprus, Morocco, Algeria, Chile, Uruguay and the United States.

The startup adopts two different business models. The first one, SaaS (Software as a Service), consists in renting the hardware and the user pays the platform and the software, while the second one consists in the sale of the hardware with the software bundled for 3 years, so that the user buys the station and then he has the service available for 3 years.

Three are three types of services offered: (i) without the installation of hardware, the user receives information from nearby weather stations, proprietary or not; (ii) station installed in an olive grove that collects data on temperature and humidity, at an affordable price; (iii) station in vineyards or olive groves, that collects data on temperature, humidity, pressure and rain. The station can be sold to the user or given on loan. All data is then processed by the software and users have access to the application and can see the data in real time and receive notifications and alerts on which interventions are required in the fields.

The difference in these three models lies in the accuracy of the data, which depends on the positioning of the station: the closer the station to the field, the more accurate the data.

A legal problem with which the company has to deal with concerns the relationship with some foreign investors, who wanted to invest in the startup but found Italian bureaucracy rather slow and cumbersome.

With respect to the ongoing legislation process for the GDPR, it places responsibilities on data controllers, including an obligation to implement appropriate technical and organizational measures to ensure and to be able to demonstrate that processing is performed in accordance with this Regulation (GDPR, Art. 24(1)).

Elaisian, having an international perspective, has been immediately compliant with the GDPR. However, the startup has had to support greater costs. The average value of the investment is in the amount of approximately € 300,000, represented by both HR costs arising from the role of the DPO, consultants' costs and IT investments resulting from the need to be compliant with the legislation.

Regarding the Spanish market, December 6, 2018, the Official Gazette of Spain published the Organic Law 3/2018, of December 5, on the Protection of Personal Data and the Guarantee of Digital Rights.

According to Article 1, this law has a double object. First, it adapts the Spanish legal system to the GDPR and further provides specifications or restrictions of its rules as explained in the GDPR. In this sense, the law states that the fundamental right to data protection of a natural person, under Article 18.4 of the Spanish Constitution, shall be exercised under the GDPR and this law. Second, the law guarantees the digital rights of citizens and employees, beyond the GDPR. For example, it includes the right to internet access, the right to digital education, the right to correction on the Internet and the right to digital disconnection in the workplace (Recio, 2019).

The GDPR has had a number of unintended negative consequences for the EU's competitiveness in AI. Indeed, it has become clear that because the GDPR was initially drafted in 2014, before awareness of machine learning was widespread, policymakers did not properly consider its impact on AI. In many ways, it would have been better to have delayed the GDPR process by a year or two, as that would have given drafters more insight into the algorithmic economy. Nevertheless, this oversight has made the GDPR unfit for the emerging algorithmic economy. In particular, the GDPR has created artificial scarcity of data by making it more difficult for organizations to collect and share data. In addition, it has made it more difficult for companies and startups that use AI applications that automate decision-making regarding individuals using personal information. As a result, the GDPR has put the EU at a competitive disadvantage in the development and use of AI (Chivot and Castro, 2019).

3.5 Fostering a European approach to Artificial Intelligence

AI will have an enormous impact on the way people live and work in the coming decades. This reasoning is the basis of the European strategy on AI, which was launched in April 2018 and has been confirmed since. The potential benefits of AI for our societies are manifold, from less pollution to fewer traffic deaths, from improved medical care and enhanced opportunities for people with disabilities and older people to better education and more ways to engage citizens in democratic processes, from swifter adjudication to a more effective fight against terrorism and crime, online and offline, as well as enhancing cybersecurity.

Faced with the rapid technological development of AI and a global policy context where more and more countries are investing heavily in AI, the EU must act as one to harness the many opportunities and address challenges of AI in a future-proof manner. Starting with the launch of the European AI strategy in April 2018, the Commission's two-pronged policy has been to make the EU a world-class hub for AI, while ensuring that AI is human-centric and trustworthy. Published in February 2020, the Commission's White Paper on AI set out a clear vision for AI in Europe: an ecosystem of excellence and an ecosystem of trust for AI.

Today's AI package represents a key milestone in both policy dimensions. To promote the development of AI and address the potential high risks it poses to safety and fundamental rights equally, the Commission is presenting both a proposal for a regulatory framework on AI and a revised coordinated plan on AI.

Given AI's potential, the European Union is promoting its development and deployment. Through the Digital Europe and Horizon Europe programs, the Commission plans to invest 1 billion euros per year in AI and mobilize additional investments from the private sector and the Member States to reach 20 billion euros per year over the course of this decade.

Strengthening Europe's AI capabilities is a key element of the wider strategy of making Europe fit for the digital age and turning the next 10 year into the Digital Decade, as set out in the Digital Compass (2030 Digital Compass: the European way for the Digital Decade, COM (2021) 118). In particular, the promotion of AI-driven innovation is closely linked to the implementation of the European Data Strategy, including the recent proposal for the Data Governance Act, since AI can only thrive when there is smooth access to data. Especially small and medium-sized enterprises will need fair access to data to make a broad uptake of AI in the EU economy possible. The proposed regulatory framework on AI will also work in tandem with applicable product safety legislation and in particular the revision of the Machinery Directive (COM (2021) 202), which addresses – among others – the safety risks of new technologies, including the risks emerging from human-robot collaboration, cyber risks with safety implications, and autonomous machines. Equally, the framework is an addition to the EU Security Union strategy, the new cybersecurity strategy, the Digital Education Action Plan 2021-2027 and the recently proposed Digital Services Act and Digital Markets Act as well as the European Democracy Action Plan.

Finally, the proposed framework will be complemented by legislation to adapt the EU liability framework, such as revising the Product Liability Directive, in order to address liability issues related to new technologies, including AI, and by a revision of the General Product Safety Directive.

The newly adopted Recovery and Resilience Facility (RRF) will enable Europe to raise its ambition and become a first mover in adopting AI. The RRF, which will be the centerpiece of the EU recovery plan, will make an unprecedent 627.5 billion euros in loans and grants available to support reforms and investments by Member States for the crucial first year of the recovery. At least 20% of the available funding will be allocated to measures fostering the digital transition, amounting to up 134 billion euros in the life cycle of the RRF.

The RRF can be used to expected to boost Member States' investments in AI and support leading research, innovation and testing capacities, so that the accelerated development and use of AI can contribute to economic and social recovery and improve competitiveness in the longer term. The opportunity is all the greater, since the RRF funding comes on top of the Digital Europe and Horizon Europe programs, as well as substantial innovation funding under the Cohesion Policy programs.

To harness the benefits of AI, Europe can build upon its existing strengths. Europe has a world-leading position in robotics and competitive industrial ecosystems. These assets, together with an increasingly performant computing infrastructure (e.g. high-performance computers) and large volumes of public and industrial data, put Europe in a position of being able to create world-leading AI capabilities on the back of its excellent research centers and in increasing number of innovative startups. To leverage these strengths with available funding, EU Member States and the Commission will pool expertise, coordinate actions and jointly mobilize additional resources. For this purpose, building on the cooperation it has developed with the Member States since 2018, the Commission is today presenting a revised coordinated plan on AI.

At the same time, the use of AI also creates risks that need to be addressed. Certain characteristics of AI, such as the opacity of many algorithms that makes investigating causal relationships difficult, pose specific and potentially high risks to the safety and fundamental rights that existing legislation is unable to address or in view of which it is challenging to enforce existing legislation. For example, it is often not possible to determine why an AI system has arrived at a specific result. As a consequence, it may become difficult to assess and prove whether someone has been unfairly disadvantages by the use of AI systems, for example in a recruitment or promotion decision or an application for a public benefit scheme. The use of AI systems may leave affected people with significant difficulties to correct erroneous decisions. Facial recognition in public spaces can have a very intrusive effect on privacy unless properly regulated. In addition, poor training and design of AI systems can result in significant errors that may undermine privacy and non-discrimination. AI-enabled robots and intelligent systems must be engineered and designed to meet the same high standards of safety and protection of fundamental rights for traditional technologies provided for by European law.

Reacting to these challenges in AI, the European Parliament and the European Council have repeatedly called for legislative action to ensure a well-functioning internal market for AI systems, where both the benefits and the risks of AI are appropriately addressed in a manner that will stand the test of time. The Commission's proposal for a regulatory framework on AI represents a key juncture in the journey towards protecting safety and fundamental rights and hence ensuing trust in the development and uptake of AI.

The coordinated plan and the proposal for a regulatory framework are part of the European Union's efforts to be an active player in international and multilateral fora in the field of digital technologies and a global leader in the promotion of trustworthy AI, and to ensure consistency between the EU's external actions and its internal policies. On the global stage, AI has become an area of strategic importance at the crossroads of geopolitics, commercial stakes and security concerns. Countries around the world are choosing to use AI as a means of technical advancement due to its utility and potential. AI regulation is in its infancy and the stakes are high for the EU to spearhead the development of new ambitious global norms, AI-related international standardization initiatives and cooperation frameworks, in line with the rules-based multilateral system and the values it upholds. In line with the Joint Communication on strengthening the EU's contribution to rules-based multilateralism, the EU intends to deepen partnership, coalitions, and alliances with third countries – notably like-minded partners – as well as multilateral and regional organizations. It also intends to engage in issue-based cooperation with other countries, and to push back where those values are threatened.

3.6 The proposal for a regulatory framework of AI

As set out in the White Paper on AI, and largely confirmed by the public consultation that followed, the use of AI creates a number of specific high risks for which existing legislation is insufficient. While there is already a solid framework of legislation in place at the EU and national levels to protect fundamental rights and ensure safety and consumer rights, particularly including the General Data Protection Regulation and the Law Enforcement Directive, certain specific features of AI technologies (e.g. opacity) can make the application and enforcement of such legislation more challenging and generate high risks for which a tailored regulatory response is needed. Therefore, the proposal introduces a set of harmonized rules applicable to the design, development and use of certain high-risk AI systems, as well as restrictions on certain uses of remote biometric identification systems.

By earning people's trust, the envisaged risk-based legislation should foster the uptake of AI across Europe and boost Europe's competitiveness. The Commission's proposal therefore pursues the twin objectives of addressing the risks associated with specific AI applications in a proportionate manner and of prompting the uptake of AI. The proposed legal framework is designed to intervene only where this is strictly needed and in a way that minimizes the burden of economic operators, with a light governance structure.

The proposed AI regulation puts forward rules to enhance transparency and minimize risks to safety and fundamental rights before AI systems can be used in the European Union. Its architecture is based on a number of core components, which, as a whole, build a proportionate and risk-based European regulatory approach. Firstly, it provides for a technology-neutral definition of AI systems that is future-proof, to the extent that it can cover techniques and approaches which are not yet known or developed.

Secondly, to avoid regulatory overreach, the proposal focuses on so-called "high risk" AI use cases, i.e. where the risks that AI systems pose are particularly high. Whether an AI system is classified as high-risk depends on the intended purpose of the system and on the severity of the possible harm and the probability of its occurrence. To ensure that the rules are future-proof and can be adjusted to emerging uses and applications of high-risk AI systems, the possibility exists to classify new AI systems as high-risk with certain predefined areas of use.

Thirdly, the proposal provides that high-risk AI systems need to respect a set of specifically designed requirements, which include the use of high-quality datasets, the establishment of appropriate documentation to enhance traceability, the sharing of adequate information with the user, the design and implementation of appropriate human oversight measures, and the achievement of the highest standards in terms of robustness, safety, cybersecurity and accuracy. High-risk AI systems must be assessed for conformity with these requirements before being placed on the market or put into service. For the smooth integration with existing legal frameworks, the proposal takes into account – where relevant – of the sectorial rules for safety, ensuring coherence between the legal acts and simplification for economic operators.

The proposed draft regulation lays down a ban on a limited set of uses of AI that contravene European Union values or violate fundamental rights. The prohibition covers AI systems that distort a person's behavior through subliminal techniques or by exploiting specific vulnerabilities in ways that cause or are likely to cause physical or psychological harm. It also covers general purpose social scoring of AI systems by public authorities.

Under the proposed regulation, other uses of AI systems are only subject to minimal transparency requirements, for example in the case of chatbots, emotion recognition systems, or deep fakes. This will allow people to make informed choice or withdraw from a given situation. Finally, the proposed regulation will encourage the use of regulatory sandboxes establishing a controlled environment to test innovative technologies for a limited time, access to Digital Innovation Hubs and access to testing experimentation facilities, which will help innovative companies, SMEs and startups to continue innovating in compliance with the new draft regulation.

Thus, the proposed regulation on AI combines greater safety and fundamental rights protection while supporting innovation, enabling trust without preventing innovation.

3.6.1 Reasons for and objectives of the proposal

Artificial Intelligence is a fast-evolving family of technologies that can bring a wide array of economic and societal benefits across the entire spectrum of industries and social activities. By improving prediction, optimizing operations and resource allocation, and personalizing service delivery, the use of AI can support socially and environmentally beneficial outcomes and provide key competitive advantages to companies and the European economy. Such action is especially needed in high-impact sectors, including climate change, environment and health, the public sector, finance, mobility, home affairs and agriculture. However, the same elements and techniques that power the socio-economic benefits of AI can also bring about new risks or negative consequences for individuals or the society. In light of the speed of technological change and possible challenges, the EU is committed to strive for a balanced approach. It is in the Union interest to preserve the EU's technological leadership and to ensure that Europeans can benefit from new technologies developed and functioning according to Union values, fundamental rights and principles.

This proposal delivers on the political commitment by President von der Leyen, who announced in her political guidelines for the 2019-2024 Commission "A Union that strives more", that the Commission would put forward legislation for a coordinated European approach on the human and ethical impressions of AI. Following on that announcement, on 19 February 2020 the Commission published the White Paper on AI – A European approach to excellence and trust (2020). The White Paper sets out policy options on how to achieve the twin objective of promoting the uptake of AI and of addressing the risk associated with certain uses of such technology. This proposal aims to implement the second objective for the development of an ecosystem of trust by proposing a legal framework for trustworthy AI. The proposal is based on EU values and fundamental rights and aims to give people and other users the confidence to embrace AI-based solutions, while encouraging businesses to develop them. AI should be a tool for people and be a force for good in society with the ultimate aim of increasing human well-being. Rules for AI available in the Union market or otherwise affecting people in the Union should therefore be human centered, so that people can trust that the technology is used in a way that is safe and compliant with the law, including the respect of fundamental rights. Following the publication of the White Paper, the Commission launched a broad stakeholder consultation, which was met with a great interest by a large number of stakeholders who were largely supportive of regulatory intervention to address the challenges and concerns raised by the increasing use of AI.

The proposal also responds to explicit requests from the European Parliament and the European Council, which have repeatedly expressed calls for legislative actions to ensure a well-functioning internal market for AI systems where both benefits and risks of AI are adequately addressed at the Union level. It supports the objective of the Union being a global leader in the development of secure, trustworthy and ethical artificial intelligence as stated by the European Council (European Council, 2020) and ensures the protection of ethical principles as specifically requested by the European Parliament.

In 2017, the European Council called for a "sense of urgency to address emerging trends" including "issues such as artificial intelligence…, while at the same time ensuring a high level of data protection, digital rights and ethical standards" (European Council, 2017). In its 2019 Conclusions on the Coordinated Plan on the development and use of artificial intelligence Made in Europe, the Council further highlighted the importance of ensuring that the rights of European citizens are fully respected and called for a review of the existing relevant legislation to make it fit for the purpose of new opportunities and challenges raised by AI (European Council, 2019). The European Council has also called for a clear determination of the AI applications that should be considered high-risk.

The most recent Conclusions from 21 October 2020 further called for addressing the opacity, complexity, bias, a certain degree of unpredictability and partially autonomous behavior of certain AI systems, to ensure their compatibility with fundamental rights and to facilitate the enforcement of legal rules.

The European Parliament has also undertaken a considerable amount of work in the area of AI. In October 2020, it adopted a number of resolutions related to AI, including ethics, liability and copyrights. In 2021, those were followed by resolutions on AI in criminal matters and in education, culture and the audio-visual sector. The EP Resolution on a Framework of Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies specifically recommends to the Commission to propose legislative actions to harness the opportunities and benefits of AI, but also to ensure protection of ethical principles for the development, deployment and use of AI, robotics and related technologies. In accordance with the political commitment made by President von der Leyen in her Political Guidelines as regards resolutions adopted by the European Parliament under Article 225 TFEU, this proposal takes into account the aforementioned resolution of the European Parliament in full respect of proportionally, subsidiary and better law-making principles.

Against this political context, the Commission puts forward the proposed regulatory framework on Artificial Intelligence with the following specific objectives: (i) ensure that AI systems placed on the Union market and used are safe and respect existing laws on fundamental rights and Union values; (ii) ensure legal certainty to facilitate investment and innovation in AI; (iii) facilitate the development of a single market for lawful, safe and trustworthy AI applications, and prevent market fragmentation.

To achieve these objectives, this proposal presents a balanced and proportionate horizontal regulatory approach to AI that is limited to the minimum necessary requirements to address the risks and problems linked to AI, without unduly constraining or hindering technological development or otherwise disproportionately increasing the cost of placing AI solutions on the market. The proposal sets a robust and flexible legal framework. On one hand, it is comprehensive and future-proof in its fundamental regulatory choices, including the principle-based requirements that AI systems should comply with. On the other hand, it puts in place a proportionate regulatory system centered on a well-defined risk-based regulatory approach that does not create unnecessary restrictions to trade, whereby legal intervention is tailored to those concrete situations where there is a justified cause for concern or where such concern can reasonably be anticipated in the near future. At the same time, the legal framework includes flexible mechanisms that enable it to be dynamically adapted as the technology evolves and new concerning situations emerge.

The proposal sets harmonized rules for the development, placement on the market and use of AI systems in the Union following a proportionate risk-based approach. It proposes a single future-proof definition of AI. Certain particularly harmful AI practices are prohibited as contravening Union values, while specific restrictions and safeguards are proposed in relation to certain uses of remote biometric identification systems for the purpose of law enforcement. The proposal lays down a solid risk methodology to define “high-risk” AI systems that pose significant risks to the health and safety of the fundamental rights of people. Those AI systems will have to comply with a set of horizontal mandatory requirements for trustworthy AI and follow conformity assessment procedures before those systems can be placed on the Union market. Predictable, proportionate, and clear obligations are also placed on providers and users of those systems to ensure the safety and respect of existing legislation protecting fundamental rights throughout the whole AI systems lifecycle. For some specific AI systems, only minimum transparency obligations are proposed, in particular when chatbots or “deep fakes” are used.

The proposed rules will be enforced through a governance system at the Member State level, building on already existing structures, and a cooperation mechanism at the Union level with the establishment of a European Artificial Intelligence Board. Additional measures are also proposed to support innovation, in particular through AI regulatory sandboxes and other measures to reduce the regulatory burden and to support Small and Medium-Sized Enterprises (SMEs) and startups.

3.6.2 Consistency with existing policy provisions in the policy area

The horizontal nature of the proposal requires full consistency with existing Union legislation applicable to sectors where high-risk AI systems are already used or likely to be used in the near future.

Consistency is also ensured with the EU Charter of Fundamental Rights and the existing secondary Union legislation on data protection, consumer protection, non-discrimination and gender equality. The proposal is without prejudice and complements the GDPR and the Law Enforcement Directive with a set of harmonized rules applicable to the design, development and use of certain high-risk AI systems and restrictions on certain uses of remote biometric identification systems. Furthermore, the proposal complements existing Union law on non-discrimination with specific requirements that aim to minimize the risk of algorithmic discrimination, in particular in relation to the design and the quality of data sets used for the development of AI systems complemented with obligations for testing, risk management, documentation and human oversight throughout the AI systems lifecycle. The proposal is without prejudice to the application of Union competition law.

As regards high-risk AI systems which are safety components of products, this proposal will be integrated into the existing sectoral safety legislation to ensure consistency, avoid duplications and minimize additional burdens. In particular, as regards high-risk AI systems related to products covered by the New Legislative Framework (NLF) (e.g. machinery, medical devices, toys), the requirements for AI systems set out in this proposal will be checked as part of the existing conformity assessment procedures under the relevant NLF legislation. With regard to the interplay of requirements, while the safety risks specific to AI systems are meant to be covered by the requirements of this proposal, NLF legislation aims at ensuring the overall safety of the final product and therefore may contain specific requirements regarding the safe integration of an AI system into the final product. The proposal for a specific type of Machinery, which was adopted on the same day that this proposal was made, fully reflects this approach. As regards high-risk AI systems related to products covered by relevant Old Approach legislation (e.g. aviation, cars), this proposal would not directly apply. However, the ex-ante essential requirements for high-risk AI systems set out in this proposal will have to be taken into account when adopting relevant implementing or delegated legislation under those acts.

As regards AI systems provided or used by regulated credit institutions, the authorities responsible for the supervision of the Union's financial services legislation should be designated as competent authorities for supervising the requirements in this proposal to ensure a coherent enforcement of the obligations under this proposal and the Union's financial services legislation where AI systems are to some extent implicitly regulated in relation to the internal governance system of credit institutions. To further enhance consistency, the conformity assessment procedure and some of the providers' procedural obligations under this proposal are integrated into the procedures under Directive 2013/36/EU on access to the activity of credit institutions and the prudential supervision.

This proposal is also consistent with the applicable Union legislation on services, including intermediary services regulated by the e-Commerce Directive 2000/31/EC and the Commission's recent proposal for the Digital Service Act (DSA).

In relation to AI systems that are components of large-scale IT systems in the Area of Freedom, Security and Justice managed by the European Union Agency for the Operational Management of Large-Scale IT Systems (eu-LISA), the proposal will not apply to those AI systems that have been placed on the market or put into service before one year has elapsed from the date of application of this Regulation, unless the replacement or amendment of those legal acts leads to a significant change in the design or intended purpose of the AI system or AI systems concerned.

For what concerns consistency with other Union policies, the proposal is part of a wider comprehensive package of measures that address problems posed by the development and use of AI, as examined in the White Paper on AI. Consistency and complementarity are therefore ensured with other ongoing or planned initiatives of the Commission that also aim to address those problems, including the revision of sectoral product legislation (e.g. the Machinery Directive, the General Product Safety Directive) and initiatives that address liability issues related to new technologies, including AI systems. Those initiatives will build on and complement this proposal in order to bring legal clarity and foster the development of an ecosystem of trust in AI in Europe.

The proposal is also coherent with the Commission's overall digital strategy in its contribution to prompting technology that works for people, one of the three main pillars of the policy orientation and objectives announced in the Communication "Shaping Europe's digital future" (COM/2020/767). It lays down a coherent, effective and proportionate framework to ensure AI is developed in ways that respect people's rights and earn their trust, making Europe fit for the digital age and turning the next ten years into the Digital Decade.

Furthermore, the promotion of AI-driven innovation is closely linked to the Data Governance Act, the Open Data Directive and other initiatives under the EU strategy for data, which will establish trusted mechanisms and services for the re-use, sharing and pooling of data that are essential for the development of data-driven AI models of high quality.

The proposal also strengthens significantly the Union's role to help shape global norms and standards and promote trustworthy AI that is consistent with Union values and interests. It provides the Union with a powerful basis to engage further with its external partners, including third countries, and at international fora on issues relating to AI.

Conclusions

The advent of Big Data and Artificial Intelligence has brought with them numerous benefits as productivity, safety, more useful data, information, and knowledge for people and organizations. However, there are some drawbacks that have to be solved as personal privacy, over-hyped expectations, and increasing technological complexity. The issue of consumer data protection has been taken into account by the European Parliament. The most feasible solution seemed to be the GDPR, which aims to provide a set of standardized data protection laws across all the 28 EU member states. The three fundamental principles on which the GDPR is based are transparency, compliance, and punishment. From these, three key pillars that the GDPR is built on originate: (i) a new transparency framework, (ii) a new compliance journey, and (iii) a new punishment regime. Since it came into force, companies based in the EU are automatically bound to the regulation of the GDPR. Furthermore, even non-EU-based companies are subject to this regulation, if they sell to people in the European Union.

Data protection and consumer rights have a cost that companies must bear to be GDPR compliant. In particular, the cost of privacy compliance is more than financial, because it is operational, and it is an ongoing cost. These higher costs are principally due to consulting services and technological solutions. There are costs of having the best employees and top decision-makers engaged in GDPR compliance rather than actually running the business. And yet, despite the effort, only 51% of organizations self-reported themselves as GDPR-compliant by the May 25, 2018 deadline.

The analysis conducted in the course of the work highlights that the gap in compliance between large business groups and SMEs is significant. The first ones cover 75% of total security and privacy expenditure where the main item is dedicated to plans to adapt to the GDPR. On the other hand, only 18% of SMEs are at a good level in this adjustment process. This is because SMEs and startups are structured from an organizational point of view to be cost-oriented, so the hypothesis of investing and allocating budgets for privacy is experienced as a burden difficult to overcome, especially in this context of Covid-19 pandemic crisis. Even if the number of employees does not exceed the GDPR threshold, if processing data is a core activity, the startup must appoint a DPO as long as there is no conflict of interest with his/her current role. Moreover, once a DPO is in place, organizations are required to carry out Data Protection Impact Assessments (DPIAs) if their proposed activities are likely to result in a high risk for the rights and freedoms of individuals, especially through the use of new technologies. Or, even if a startup already complies with the DPO, it must review its current policies to verify whether it complies with the new provisions, especially in terms of privacy notices, consent and accountability.

In addition to the costs, the problems encountered by companies, of any size, in adapting to the GDPR are considerable, such as lack of awareness of the employees on data protection, adoption of ineffective technical and organizational solutions, shortage of experienced professionals in the sector, and collection and mapping of data in a regulatory manner. Moreover, the GDPR does not provide a list of requirements to be met that guarantees the compliance, but gives generic indications leaving complete freedom to the Data Controller, that has only to prove that is has taken all the appropriate measures in accordance with the accountability principle of the GDPR. It is therefore necessary to identify systems for assessing compliance with the model and to have competent organizations, structured processes and professional skills suitable for the task.

What is more, for organizations and startups which are naturally prone to innovation, concepts such as privacy by design, profiling and data portability provide the opportunity not only to innovate, but also to build consumers' trust and confidence. Also, ensuring compliance within the company has become a synonym for the quality of the entire organization. Indeed, with particular reference to SMEs and startups, ensuring accountability is an added value that allows them to stand out from competition and gain an advantage in their sector, especially in the case where the company core business is based on the collection and processing of data provided by customers.

Therefore, although the GDPR is a concern for startup firms, it is definitely something that is more relevant than ever. Thus, despite the initial effort, the cost to bear is worth the final result.

Executive summary

This work seeks to address the impact of being GDPR compliant for startups, particularly those using Artificial Intelligence, through the analysis of the Elaisian case study.

As far as back as 2001, **the new Big Data phenomenon** was beginning to emerge. Big Data is a term applied to data sets whose size is beyond the ability of commonly used software tools to capture, manage and process within a tolerable elapsed time. At the same time, **Artificial Intelligence** was advancing rapidly, both in terms of the amount of resources devoted to it and also in terms of output. AI is used to define a technical discipline that researches and develops theories, methods, technologies and application systems for simulating the extension and expansion of human intelligence. Its main goal is for machines to perform some complex tasks that require intelligent humans to complete.

Subsequently, these two elements merged into **a synergic relationship**, where AI is useless without data and data is unsurmountable without AI. The latter depends heavily on the former for success, while also helping organizations unlock the potential in their data stores in ways that were previously cumbersome or impossible. There is a virtuous cycle in evidence: the more data are put through the machine learning models, the better they become. There are three key ways that AI can deliver **better insights** with Big Data. Firstly, AI creates new and enhanced methods for analyzing data. Secondly, data analytics is becoming less labor-intensive. So, the value of Big Data sets is related to data quality and the ones that are of inferior quality are of little or no worth for the organizational decision-making process. Finally, analytics become more predictive and prescriptive than in the past.

Moreover, in the age of Big Data, companies have more consumer data, and more kinds of data, available to them than ever before. As companies fall further and further behind, they miss opportunities to learn from data and apply what they learn to how they connect. AI closes the gap by moving far past human limitations to consume and analyze data on a scale that previously was only imagined. In fact, the "intelligence" in Artificial Intelligence is exactly what it sounds like: the ability to think independently, to become more knowledgeable from being exposed to more information and to adapt and adjust when things change.

These technological developments allow firms to transform themselves into **data-driven organizations**. The need for data does not diminish with firm size; larger and smaller firms targeting the creation of similar AI products require similar data resources. However, large firms may be able to access data more easily from supplier and customer relationships as they benefit from a breadth of supplier relationships and a more developed customer ecosystem. Additionally, larger firms could benefit from complementary business models which provide data as an externality of normal business operations.

Data are expected to become **the fuel of the digital economy** as they can be used to reduce information asymmetries, improve resource management, and identify casual relationships using AI and statistical analyses. The lower cost of data collection and the adoption of digital communication networks have dramatically increased volumes of data collected. So, there is certainly not a lack of data available. However, the quality of that data still leaves much to be decided.

As the organizations and institutions in data markets are rapidly evolving, new regulations as **the European General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA)** have been implemented.

The GDPR replaces the Data Protection Initiative of 95/46/EC. Agreed upon by the European Parliament and Council in April 2016, it came into force on May 25, 2018 with the aims of **strengthening the data protection framework** within the EU, providing **a uniform regulatory data protection environment** and ensuring **the free movement of data**.

The global nature of Internet means that this regulation applies to **both EU and non-EU companies** that collect, process, or store information on EU citizens as well as on non-citizens while they reside in the EU. Indeed, one survey suggest that 52 percent of U.S. companies possess data on EU citizens, which make them liable for implementing the required privacy practices.

Article 5 summarizes the **inspiring criteria** of the entire personal data protection principle, listing all the general principles. These are as follows: lawfulness, fairness and transparency; purpose limitation; data minimization; accuracy; storage limitation; integrity and confidentiality; accountability.

Any data collection and processing must comply with the above principles.

One of the most remarkable novelties within the GDPR is the right to data portability. **Data Portability** is the ability and capacity to export data collected or stored digitally concerning a data subject and the ability to receive data concerning the data subject and to allow another controller to receive portable data. It is regulated by the article 20 of the GDPR. Firstly, it states that data portability is a right of the data subject to receive a subset of the personal data processed by a data controller concerning him or her, and to store those data for further personal use. Such storage can be on a private device or on a private cloud, without necessarily transmitting the data to another data controller. In this regard, data portability complements the right of access. One specificity of data portability lies in the fact that it offers an easy way for data subjects to manage and reuse personal data themselves. The data should be received "in a structured, commonly used and machine-readable format".

Secondly, Article 20(1) provides data subject with the right to transmit personal data from one data controller to another data controller **"without hindrance"**. In accordance with Article 20(2), data can also be transmitted directly from one data controller to another on request of the data subject and where it is technically feasible. In this respect, recital 68 encourages data controllers to develop interoperable formats that enable data portability but without creating an obligation for controllers to adopt or maintain processing systems which are technically compatible. The GDPR does, however, prohibit controllers from establishing barriers to the transmission.

The impact of the right to data portability is relevant both for businesses, in particular for e-business involved in the digital market, and for individual users (data subjects). From the business perspective, this impact is tangible in several fields, it is both a challenge to the traditional system of competition law and a problematic opportunity in terms of interoperability of systems. From the user perspective, this impact is evident both in terms of control of personal data and in terms of a more user-centric interrelation among services.

The importance of the GDPR cannot be overstated. It seeks to protect consumers and users from harm resulting from unauthorized and excessive use of their personal data, in ways that might negatively affect human dignity and well-being, including but not limited to price discrimination, other forms of discrimination, black-mail, intangible nuisance, identity theft and harm to autonomy. The GDPR also seeks to change the balance of power between data subjects and data controllers, potentially enabling data subjects to enjoy a larger portion of the fruits from sharing their data. Additionally, it seeks to ensure the free flow of data between EU member states - inter alia - by eliminating differences among such states regarding data processing. Furthermore, it seeks to strengthen the trust users have that their personal data will not be used in ways that do not conform to their reasonable expectations, which is necessary for the efficient working on the market and for society to realize the value of technology.

However, **compliance with the GDPR is a particularly onerous task**, especially for small- and medium-sized companies, as it places a particularly heavy burden on their resources. Indeed, they must put in place consent gathering mechanisms, provide detailed information regarding their data processing activities, implement technical and organizational measures to ensure compliance with the GDPR, monitor and document GDPR compliance, carry out Data

Protection Impact Assessments and have a designed DPO. Thus, companies with data presence in the EU have been required to spend millions to comply with the GDPR.

The human resources and capital costs involved in ensuring compliance with the GDPR disproportionately burden small- and medium-sized companies, which are limited on both financial resources and personnel. Indeed, while a big company has dozens if not hundreds of experts working on GDPR compliance, most ad tech companies do not have the lawyers, data experts and programmers necessary to comply with the GDPR in a smooth and effective process. Additionally, compliance with the burdensome requirements of the GDPR, such as adopting technical and organizational measures and monitoring and documenting data flows, exhibits economies of scale and scope, which tend to create a competitive advantage for large organizations.

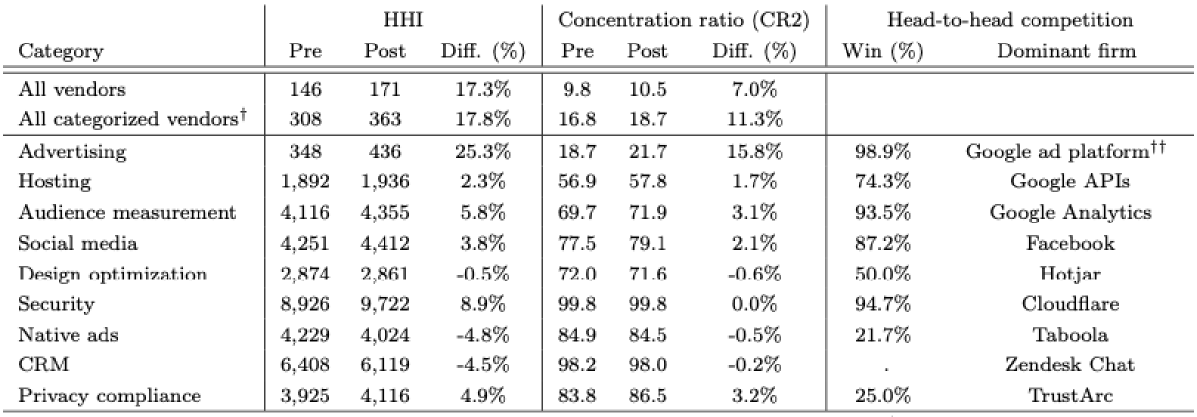
Moreover, in the longer term, the GDPR may affect firms' performance by changing the efficiency of resource allocation among firms and sectors. The capability of the firm to cost-effectively meet the requirements of the GDPR affects its financial performance and, in the end, whether or not the firm can operate in the given market or industry. **The GDPR may thus affect market structure and competition.**

When constructing the legal regime, many effects were disregarded. The magnitude and breadth of such effects may well constitute and unintended and unheeded welfare-reducing consequence. Indeed, **the GDPR limits competition and increases concentration** in data and data-related markets, and potentially strengthens large data controllers. It also further reinforces the already existing barriers to data sharing in the EU, thereby potentially reducing data synergies that might result from combining different datasets controlled by separate entities.

Seven main parallel and cumulative market dynamics that may limit competition and increase market concentration have been identified so far.

First, the costs of organizing a dataset in a way which complies with the GDPR may be high and are characterized by economies of scale. Accordingly, some small entrants might find it unprofitable to collect data. Second, the GDPR prohibits or makes it more difficult to engage in some methods of data collection, creating comparative advantages to some data controllers. Third, the GDPR reduces the economic incentives of firms to share any data collected. This is because those sharing data are still liable for monitoring its use by anyone with whom the data is shared. This, in turn, further reduces the number of data suppliers. Fourth, even where data is shared, the GDPR may limit its use. Fifth, the costs of non-compliance are high. The less severe infringements could result in a fine up to € 10 million, or 2% of the firm's worldwide annual revenue from the preceding financial year, whichever amount is higher; while the most serious types of infringements could result in a fine up to € 20 million, or 4% of the firm's worldwide annual revenue from the preceding financial year, whichever amount is higher. Sixth, the GDPR creates uncertainty, which may impose higher costs on smaller players, and might also enable large firms to use such uncertainty strategically, limiting the sharing of their data based on broad interpretations of the GDPR. Finally, the GDPR, and especially the discussion surrounding it, could have an indirect effect on data subjects, who might be more willing to provide their data to larger, more reputable firms, or to firms with which they must interact, at least until the trust of data subjects in the actual enforcement of data protection obligations is increased. Thus, the cumulative effect of such dynamics causes a decline in competition in data and in data-based markets.

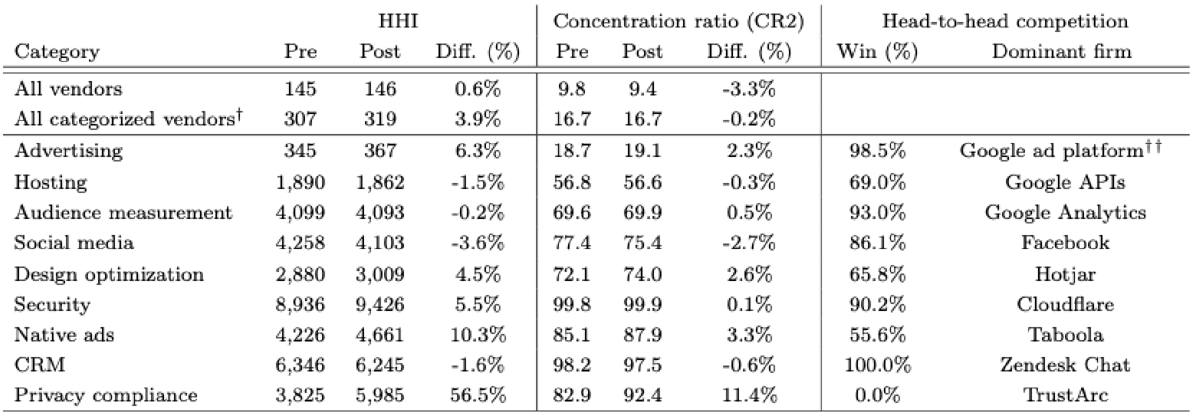
From **the GDPR short-run concentration impact analysis**, it can be noted that sites prefer to keep the dominant firm over alternatives.



	HHI			Concentration ratio (CR2)			Head-to-head competition	
Category	Pre	Post	Diff. (%)	Pre	Post	Diff. (%)	Win (%)	Dominant firm
All vendors	146	171	17.3%	9.8	10.5	7.0%		
All categorized vendors[†]	308	363	17.8%	16.8	18.7	11.3%		
Advertising	348	436	25.3%	18.7	21.7	15.8%	98.9%	Google ad platform[††]
Hosting	1,892	1,936	2.3%	56.9	57.8	1.7%	74.3%	Google APIs
Audience measurement	4,116	4,355	5.8%	69.7	71.9	3.1%	93.5%	Google Analytics
Social media	4,251	4,412	3.8%	77.5	79.1	2.1%	87.2%	Facebook
Design optimization	2,874	2,861	-0.5%	72.0	71.6	-0.6%	50.0%	Hotjar
Security	8,926	9,722	8.9%	99.8	99.8	0.0%	94.7%	Cloudflare
Native ads	4,229	4,024	-4.8%	84.9	84.5	-0.5%	21.7%	Taboola
CRM	6,408	6,119	-4.5%	98.2	98.0	-0.2%	.	Zendesk Chat
Privacy compliance	3,925	4,116	4.9%	83.8	86.5	3.2%	25.0%	TrustArc

Short-term GDPR impact on concentration (1 week) (Garrett et al., 2021)

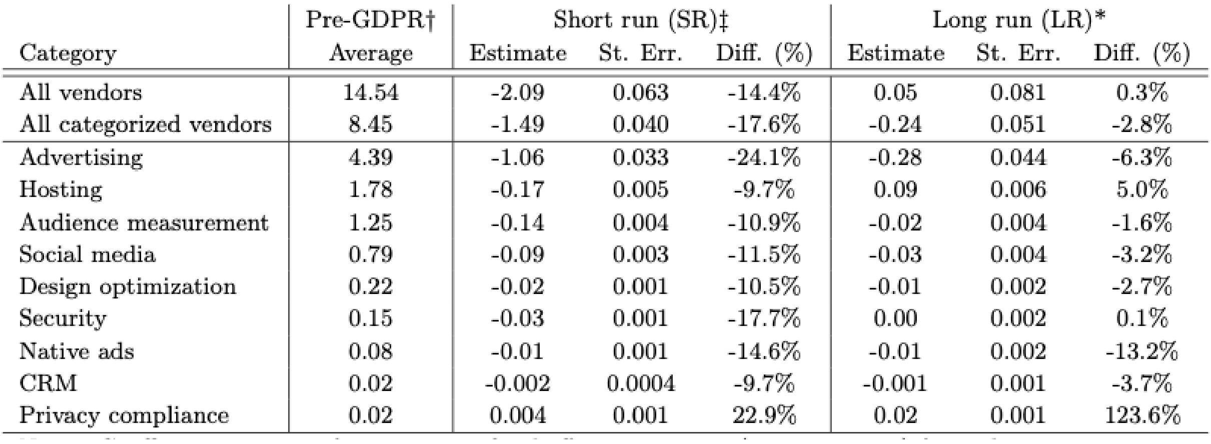
While, concentration effects appear to dissipate over **the long run**, though increased concentration in the advertising technology market persists to some extent. So, this model again suggests that this dissipation is consistent with publishers' declining beliefs about regulatory enforcement in the absence of enforcement actions.



Category	HHI			Concentration ratio (CR2)			Head-to-head competition	
	Pre	Post	Diff. (%)	Pre	Post	Diff. (%)	Win (%)	Dominant firm
All vendors	145	146	0.6%	9.8	9.4	-3.3%		
All categorized vendors†	307	319	3.9%	16.7	16.7	-0.2%		
Advertising	345	367	6.3%	18.7	19.1	2.3%	98.5%	Google ad platform††
Hosting	1,890	1,862	-1.5%	56.8	56.6	-0.3%	69.0%	Google APIs
Audience measurement	4,099	4,093	-0.2%	69.6	69.9	0.5%	93.0%	Google Analytics
Social media	4,258	4,103	-3.6%	77.4	75.4	-2.7%	86.1%	Facebook
Design optimization	2,880	3,009	4.5%	72.1	74.0	2.6%	65.8%	Hotjar
Security	8,936	9,426	5.5%	99.8	99.9	0.1%	90.2%	Cloudflare
Native ads	4,226	4,661	10.3%	85.1	87.9	3.3%	55.6%	Taboola
CRM	6,346	6,245	-1.6%	98.2	97.5	-0.6%	100.0%	Zendesk Chat
Privacy compliance	3,825	5,985	56.5%	82.9	92.4	11.4%	0.0%	TrustArc

Long-run GDPR impact on concentration (27 weeks post) (Garrett et al., 2021)

For what concerns **the GDPR impact on vendor usage**, this falls in all categories but privacy compliance, where it increases.



Category	Pre-GDPR†	Short run (SR)‡			Long run (LR)*		
	Average	Estimate	St. Err.	Diff. (%)	Estimate	St. Err.	Diff. (%)
All vendors	14.54	-2.09	0.063	-14.4%	0.05	0.081	0.3%
All categorized vendors	8.45	-1.49	0.040	-17.6%	-0.24	0.051	-2.8%
Advertising	4.39	-1.06	0.033	-24.1%	-0.28	0.044	-6.3%
Hosting	1.78	-0.17	0.005	-9.7%	0.09	0.006	5.0%
Audience measurement	1.25	-0.14	0.004	-10.9%	-0.02	0.004	-1.6%
Social media	0.79	-0.09	0.003	-11.5%	-0.03	0.004	-3.2%
Design optimization	0.22	-0.02	0.001	-10.5%	-0.01	0.002	-2.7%
Security	0.15	-0.03	0.001	-17.7%	0.00	0.002	0.1%
Native ads	0.08	-0.01	0.001	-14.6%	-0.01	0.002	-13.2%
CRM	0.02	-0.002	0.0004	-9.7%	-0.001	0.001	-3.7%
Privacy compliance	0.02	0.004	0.001	22.9%	0.02	0.001	123.6%

GDPR impact on average vendor use by category (Garrett et al., 2021)

There is an **apparent tradeoff** between access to training data and consumer privacy. GDPR makes it harder for firms to collect, store, and analyze certain types of data, especially personally identifiable or employment data. In order to address this issue, a survey on AI startups has been carried out. The survey was designed to address two questions. First, the impact of the GDPR ad data regulation on AI startups. Second, the importance of data to AI product development. The survey results show that training data is important for AI startups that rely on natural nets and ensemble learning algorithms. In addition to that, firms with customers in Europe are significantly more likely to create a new position to handle GDPR-related issues or to reallocate firm resources due to GDPR. This implies that the GDPR imposes costs, perhaps substantial costs, on startup AI firms.

Evidence of this is **the Elaisian case study**. Elaisian is **an innovative Italian startup** founded in Rome in 2016 by Damiano Angelici and Giovanni di Mambro. In 2017, it was selected as one out of the ten best projects by StartupBootCamp FoodTech for bringing real time IOT monitoring to olive trees in order to increase yields and reduce maintenance costs. During 2017, several tests were made to highlight the benefits of the product offered. Then, in 2018 Elaisian entered the market and in 2019 it started to expand internationally, with a focus on the Spanish market. Despite the Covid-19 pandemic crisis, the startup was able to adapt and react immediately in order to avoid being impacted by the situation, managing to maintain **an annual growth of 250%**. Today Elaisian has 20 employees and its technology is used in over 1,000 agricultural holdings in 10 countries such as The United States, Portugal and Morocco.

The startup adopts **two different business models**. The first one, Saas (Software as a Service), consists in renting the hardware and the user pays the platform and the software; while the second one consists in the sale of the hardware with the software bundled for 3 years, so that the user buys the station and then has the service available for 3 years.

It offers three different services, differing in the accuracy of the data, which depends on the position of the station: the closer the station to the field, the more accurate the data.

Elaisian had to face **a legal problem** concerning the relationships with some foreign investors, who wanted to invest in the startup but found Italian bureaucracy rather slow and cumbersome. With respect to the ongoing legislation process for the GDPR, Elaisian was **immediately compliant with the GDPR**. However, the startup had to support **greater costs**. The average value of the investment is in the amount of approximately **€ 300,000**, represented both HR costs arising from the role of the DPO, consultants' costs and IT investments resulting from the need to be compliant with the legislation.

On the contrary, the entry into force of the GDPR **did not have much impact** on dominant firms such as **Google and Facebook**. Google's market dominance allows it to easily obtain user consent to the collection of personal data. And while some users are unhappy about the way consent is sought, the perceived essential nature of the services provided by Google, combined with the lack of credible alternatives, outweigh their concerns about their data being collected. Indeed, with the introduction of GDPR, Google increases its market share whereas all other firms that supply web technology either do not see a change in market share or suffer losses.

Thus, the GDPR has had a number of **unintended negative consequences** for the EU's competitiveness in AI. Indeed, it has become clear that because the GDPR was initially drafted in 2014, before awareness of machine learning was widespread, policymakers did not properly consider its impact on AI. In many ways, it would have been better to have delayed the GDPR process by a year or two, as that would have given drafters more insight into the algorithmic economy. Nevertheless, this oversight has made the GDPR unfit for the emerging algorithmic economy. In particular, the GDPR has created artificial scarcity of data by making it more difficult for organizations to collect and share data. In addition, it has made it more difficult for companies and startups that use AI applications that automate decision-making regarding individuals using personal information. As a result, the GDPR has put the EU at a competitive disadvantage in the development and use of AI.

Several suggestions could be made for creating a more welfare-enhancing equilibrium.

Most importantly, competition law should give more weight to factors which might balance the negative effects of the GDPR on competition and innovation. This implies, for example, a more lenient policy towards mergers or joint ventures between small or medium-sized data controllers, which would enable them to reach economies of scale and scope in data analysis and compete more effectively with those who already enjoy such economies. It also implies that when at least one data controller in a proposed merger or joint venture already possesses strong comparative advantages in data analysis, a careful balance is required between the benefits of increased data synergies and the need to ensure the ability of other firms to effectively complete, in light of the increased hurdles to data collection and processing resulting from the GDPR.

Secondly, assessments of market power and potential competition should take into account the actual competitive effects of the GDPR. No longer can it be assumed that new players seeking to accumulate large volumes of data face only low barriers, as was the case in several earlier Commission decisions, especially where separate entities collect different parts of the dataset.

In addition, other suggestions that might go some way toward making the GDPR more welfare-enhancing could be made.

Firstly, where uncertainty regarding how to meet GDPR legal obligations contributes to concentration, it may be useful to consider ways of limiting such uncertainty.

Secondly, governments might support investment in the development of better privacy protection tools which can retain more value from collected data.

Third, priority could be given to the development of better and faster tools for verification of GDPR compliance. And fourth, certification of data management and vetting processes could go a long way toward reducing costs. The government can either certify such tools or help facilitate such certification. In addition, the use of certified tools should be taken into account when assessing liability, and presumptions based on the use of reasonable tools should be created. This, in turn, could significantly reduce the risks involved in data sharing.

Thus, privacy policy is interlinked with competition and the resultant data-based innovation in more ways that have yet been recognized. In particular, the GDPR raises the transaction costs of sharing data between different data controllers. Recognizing such effects should probably lead to a re-evaluation of the balance reached, and to the adoption of tools to ensure that there is an overall increase in welfare.

Given the potential of AI, the European Union is prompting its development and deployment. **The White Paper on AI**, published in February 2020 by the Commission, sets out a clear vision for AI in Europe: an ecosystem of excellence and an ecosystem of trust for AI.

Through the Digital Europe and Horizon Europe programs, the Commission plans to invest one billion euros per year in AI and mobilize additional investments from the private sector and the member states to reach 20 billion euros per year over the course of this decade.

In particular, the promotion of AI-driven innovation is closely linked to the implementation of **the European Data Strategy**, including the recent proposal for the Data Governance Act, since AI can only thrive when there is smooth access to data.

The proposed regulatory framework on AI will also work in tandem with applicable product safety legislation and in particular the revision of the Machinery Directive, which addresses the safety risks of new technologies, including the risks emerging from human-robot collaboration, cyber with safety implications, and automaton machines. Equally, the framework is an addition to the EU Security Union strategy, the new cybersecurity strategy, the Digital Education Action Plan 2021-2027 and the recently proposed Digital Services Act and Digital Markets Act as well as the European Democracy Action Plan.

Finally, the proposed framework will be complemented by legislation to adapt the EU liability framework, such as revising the Product Liability Directive, in order to address liability issues related to new technologies, including AI, and by a revision of the General Product Safety Directive.

The Commission’s White Paper pursues the twin objectives of addressing the risks associated with specific AI applications in a proportionate manner and of prompting the uptake of AI.

Thus, the proposed AI regulation puts forward rules to **enhance transparency and minimize risks** to safety and fundamental rights before AI systems can be used in the European Union. Its architecture is based on a number of core components, which, as a whole, build a proportionate and risk-based European regulatory approach.

Firstly, it provides for a technology-neutral definition of AI systems that is future-proof, to the extent that it can cover techniques and approaches which are not yet known or developed.

Secondly, to avoid regulatory overreach, the proposal focuses on so-called “high risk” AI use cases. Whether an AI system is classified as high-risk depends on the intended purpose of the system and on the severity of the possible harm and the probability of its occurrence. To ensure that the rules are future-proof and can be adjusted to emerging uses and applications of high-risk AI systems, the possibility exists to classify new AI systems as high-risk with certain predefined areas of use.

Thirdly, the proposal provides that high-risk AI systems need to respect a set of specifically designed requirements, which include the use of high-quality datasets, the establishment of appropriate documentation to enhance traceability, the sharing of adequate information with the user, the design and implementation of appropriate human oversight measures, and the achievement of the highest standards in terms of robustness, safety, cybersecurity and accuracy. High-risk AI systems must be assessed for conformity with these requirements before being placed on the market or put into service. For the smooth integration with existing legal frameworks, the proposal takes into account – where relevant – of the sectorial rules for safety, ensuring coherence between the legal acts and simplification for economic operators.

The Commission puts forward the proposed regulatory framework on Artificial Intelligence with the following **specific objectives**: (i) ensure that AI systems placed on the Union market and used are safe and respect existing laws on fundamental rights and Union values; (ii) ensure legal certainty to facilitate investment and innovation in AI; (iii) facilitate the development of a single market for lawful, safe and trustworthy AI applications, and prevent market fragmentation.

To achieve these objectives, this proposal presents **a balanced and proportional horizontal regulatory approach to AI** that is limited to the minimum necessary requirements to address the risks and problems linked to AI, without unduly constraining or hindering technological development or otherwise disproportionately increasing the cost of placing AI solutions on the market. The proposal sets a robust and flexible legal framework. On one hand, it is comprehensive and future-proof in its fundamental regulatory choices, including the principle-based requirements that AI systems should comply with. On the other hand, it puts in place a proportionate regulatory system centered on a well-defined risk-based regulatory approach that does not create unnecessary restrictions to trade, whereby legal intervention is tailored to those concrete situations where there is a justified cause for concern or where such concern can reasonably be anticipated in the near future. At the same time, the legal framework includes flexible mechanisms that enable it to be dynamically adapted as the technology evolves and new concerning situations emerge.

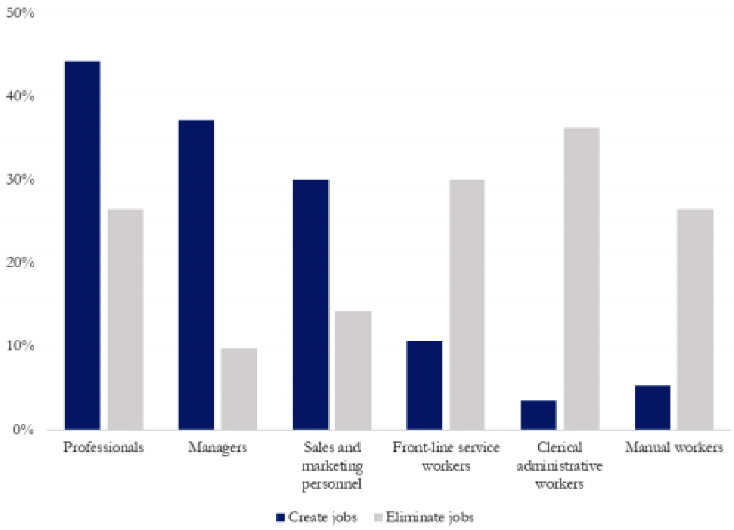
This horizontal nature requires **full consistency** with existing Union legislation applicable to sectors where high-risk AI systems are already used or likely to be used in the near future.

The proposal is without prejudice and complements the GDPR and the Law Enforcement Directive with a set of harmonized rules applicable to the design, development and use of certain high-risk AI systems and restrictions on certain uses of remote biometric identification systems.

For what concerns **consistency with other Union policies**, the proposal is part of a wider comprehensive package of measures that address problems posed by the development and use of AI, as examined in the White Paper on AI. Consistency and complementarity are therefore ensured with other ongoing or planned initiatives of the Commission that also aim to address those same problems, including the revision of sectoral product legislation (e.g. the Machinery Directive, the General Product Safety Directive) and initiatives that address liability issues related to new technologies, including AI systems.

However, Artificial Intelligence is likely to drastically change the economy in a decade or so. That is, **if 47% of jobs will soon be at risk due to AI**, then surely jobs will be already at risk at those firms using AI applications. To gain a peek through that window, a global survey of startup firms developing and selling commercial applications based on AI was carried out from May to September 2018. This suggests that AI might have some job-creating potential, especially for the professional, managerial, sales and marketing occupations that tend to be the users of these products. On the other hand, AI also reduces labor costs for some customers. Moreover, these effects might differ across occupations because the use of AI differs dramatically from one occupation to another.

Clearly, AI is not all about destroying jobs. In particular, those occupations that are most likely to use AI are also most likely to see jobs created. At the same time, many jobs will be eliminated, especially in the three occupational groups that use AI relatively less.



Job Creation and Destruction by Occupational Group (Bessen et al., 2018)

As data constitutes the building block of information and markets are becoming more concentrated and dominated by data-savvy firms, a regulation that governs consumers' personal data is thus necessary. That is the GDPR, under which firms have to manage their data in a transparent, responsible, and accountable way.

Currently, it is natural to wonder if the GDPR is worth the cost. Regulations always have costs, which are meant to be recovered with the expected benefits. Indeed, **data privacy comes at a price**. But, despite the fact that the GDPR is a concern for firms, especially for SMEs and startups, the costs of protecting the rights of data subjects **are worth it**.

Bibliography

- Acemoglu, D. and Restrepo, P. (2018), "Artificial Intelligence, Automation and Work", NBER Working Paper
- Acquisti, A., and Grossklags, J. (2005), "Privacy and Rationality in Individual Decision Making", IEEE Security & Privacy, 3, 26–33.
- Acquisti, A., Taylor, C., Wagman, L. (2016), "The economics of privacy", (54/2) Journal of Economic Literature 442
- Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., Tao, Q., Sun, Z. and Xia, L. (2020), "Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases", Radiology 2020
- Akerlof, G. A. (1970), "The market for "lemons": quality uncertainty and the market mechanism", Quarterly Journal of Economics, 84(3), 488–500.
- Albrecht, J. P. (2013), "#EUdataP State of the Union", Speech at the Chaos Communication Congress
- Allen, G., And Taniel Chan, T. (2017), "Artificial Intelligence and National Security", Belfer Ctr., Harvard Kennedy School
- Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P.N., Inkpen, K. and Teevan, J., (2019), "Guidelines for human-AI interaction", In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 3
- Arnold, M., Bellamy, R.K., Hind, M., Houde, S., Mehta, S., Mojsilović, A., Nair, R., Ramamurthy, K.N., Olteanu, A., Piorkowski, D. and Reimer, D., (2019), "Increasing trust in AI services through supplier's declarations of conformity", IBM Journal of Research and Development, 63(4/5)
- Arrow, K. J. (1962), "Economic welfare and the allocation of resources for invention", in Universities-National Bureau Committee for Economic Research and Committee on Economic Growth of the Social Science Research Council (eds), The Rate and Direction of Inventive Activity: Economic and Social Factors. Princeton University Press: Princeton, NJ, pp. 609–626
- Article 29 Data Protection Working Party (2017), "Guidelines for identifying a controller or processor's lead supervisory authority", 4
- Athey, S., and Luca, M. (2019), "Economists (and Economics) in Tech Companies", Journal of Economic Perspectives, 33(1), 209-230
- Autoriteit Persoonsgegevens (2019), "Ap: veel websites vragen op onjuiste wijze toestemming voor plaatsen tracking cookies"

- Bajari, P., et al. (2018), “The Impact of Big Data on Firm Performance: An Empirical Investigation”, NBER Working Paper No. 24334
- Bakos, J. T. (1991), “A strategic analysis of electronic marketplaces”, MIS Quarterly, 15(3), 295–310.
- Bamberger, K., and Mulligan, D. K. (2015), “Privacy on the Ground”, MIT Press
- Barth, S., and de Jong, M. D. (2017), “The Privacy Paradox: Investigating Discrepancies between Expressed Privacy Concerns and Actual Online Behavior—A Systematic Literature Review”, Telem- atics and Informatics, 34, 1038–1058.
- Batikas, M., Bechtold, S., Kretschmer, T. and Peukert, C. (2020), “European Privacy Law and global markets for Data”, Centre for Economic Policy Research, London
- Batikas, M., Claussen, J., and Peukert, C. (2019), “Follow the Money: Online Piracy and Self-regulation in the Advertising Industry”, International Journal of Industrial Organization, 65, 121– 151
- Bennett, C.J. (1992),” Regulating Privacy: Data Protection and Public Policy in Europe and the United States”, Cornell University Press
- Beslay, L., and Sanchez, I. (2018), “The right to data portability in the GDPR: Towards user-centric interoperability of digital services”, Computer Law & Security Review, 193-203
- Bessen, J. (2016), "How Computer Automation Affects Occupations: Technology, Jobs, and Skills", Boston University School of Law Working Paper No. 15-49
- Bessen, J. (2017), "Automation and Jobs: When Technology Boosts Employment", Boston University School of Law, Law & Economics No. 17-09
- Bessen, J. E., Impink, S.M., Reichensperger, L. and Seamans, R. (2018), “The Business of AI Startups”, Boston University School of Law, Law and Economics Research Paper, No. 18-28
- Borgman, C. L. (2012), “The conundrum of sharing research data”, Journal of the American Society for Information Science and Technology, 63(6), 1059–1078.
- Boyd, D. and Crawford, K. (2012), “Critical questions for Big Data: Provocations for a cultural, technological and scholarly phenomenon”, Information, Communication & Society, 15(5), 662-679
- Bradford, A. (2012), “The Brussels Effect”, Northwestern Unviersity Law Review, 107, 1–68
- Bragazzi, N.L. , Dai, H., Damiani, G., Behzadifar, M., Martini, M. and Wu, J. (2020), “How Big Data and Artificial Intelligence Can Help Better Manage the COVID-19 Pandemic”, International Journal of Environmental Research and Public Health

- Bresnahan, T. F., and Trajtenberg, M. (1995), “General purpose technologies “Engines of growth?”, Journal of Econometrics, 65(1), 83–108
- Brill, J. (2018), “Microsoft’s commitment to GDPR, privacy and putting customers in control of their own data”, Microsoft Blog
- Brynjolfsson, E., Mitchell, T., and Rock, D. (2018), “What can Machines Learn and What does it mean for Occupations and Industries”, AEA Papers and Proceedings, 108
- Bughin, J., Hazan, E., Ramaswamy, S., Chui, M., Allas, T., Dahlström, P., Henke, N. and Trench, M. (2017), “Artificial intelligence–the next digital frontier”, McKinsey Global
- Cadwalladr, C. (2020), “Fresh Cambridge Analytica leak ‘shows global manipulation is out of control”, The Guardian
- Campbell, J., Goldfarb, A., and Tucker, C. (2015), “Privacy Regulation and Market Structure”, Journal of Economics & Management Strategy, 24, 47–73
- Carnelley, P., Schwenk, H., Cattaneo, G., Micheletti, G., and Osimo, D. (2016), “Europe’s data marketplaces - current status and future perspectives”, European Data Market SMART 2013/0063 D.39. IDC
- Cassey, K. (2019), “How Big Data and AI work together”, The Enterprises project, available at https://enterprisersproject.com/article/2019/10/how-big-data-and-ai-work-together.
- Chan, J. (2020), “How Big Data and AI will work together”, ISM Guide
- Chattu, V.K., Nanda, A., Chattu, S.K., Kadri, S.M., Knight, A.W. (2019), “The Emerging Role of Blockchain Technology Applications in Routine Disease Surveillance Systems to Strengthen Global Health Security”, Big Data Cogn. Comput.
- Cheatham, B. Javanmardian, K. Samandari, H. (2019), “Confronting the risks of artificial intelligence,” McKinsey Global
- Chebli, O., Goodridge, P., and Haskel, J. (2015), “Measuring activity in big data: new estimates of big data employment in the UK market sector”, Imperial College Business School Discussion Paper, Imperial College Business School, London, 28
- Chivot, E., and Castro, D. (2019), “The EU needs to reform the GDPR to remain competitive in the algorithmic economy”, Center for Data Innovation
- Cockburn, I.M., Henderson, R. and Stern, S. (2018), “The Impact of Artifical Intelligence on Innovation: An Explanatory Analysis”, The Economics of Artifical Intelligence: An Agenda, 115-146, National Bureau of Economic Research
- Commission Nationale de l’Informatique et des Libertés (2020), “Cookies et autres traceurs: la cnil publie des lignes directrices modificatives et sa recommandation”

- Cowgill, B. and Tucker, C.E. (2019), "Economics, fairness and algorithmic bias", Journal of Economic Perspectives
- Crockett, K., Stoklas, J., O'Shea, J., Krügel, T., and Khan, W. (2020), "Reconciling Adapted Psychological Profiling with the New European Data Protection Legislation," Computational Intelligence, Eds: Sabourin, C., Mereio, J., Barranco, N., Madani, K., Warwick, K. Springer
- Custers, B., and Ursic, H. (2016), "Big Data and Data Reuse: a taxonomy of data reuse for balancing big data benefits and personal data protection", International Data Privacy Law, 9
- Data Protection Commission (2020), "Report by the data protection commission on the use of cookies and other tracking technologies", Technical report, Data Protection Commission
- Davenport, T.H., and Ronanki, R. (2018), "Artificial Intelligence for the Real World. Don't start with moon shots.", Harvard Business Review
- Davies, J. (2018), "The Google Data Protection Regulation': GDPR is strafing ad sellers", Digiday, available at https://digiday.com/media/google-data-protection-regulation-gdpr-strafing-ad-sellers/.
- De Streel and Tombal T. (2019), "The Fifty Shades of Data Sharing and the Law"
- Doshi-Velez, F., Kortz, M., Budish, R., Bavitz, C., Gershman, S., O'Brien, D., Schieber, S., Waldo, J., Weinberger, D. and Wood, A. (2017), "Accountability of AI under the law: The role of explanation"
- Economides, N., and Lianos, I. (2019). "Restrictions on Privacy and Exploitation in the Digital Economy: A Competition Law Perspective.", Working Paper, SSRN–ID 3474099
- Eisenmann, T. R., Parker, G., and Van Alstyne, M.W. (2006), "Strategies for two-sided markets", Harvard Business Review, 84(10), 92–101
- Elkin-Koren, N., and Gal, M. S. (2019), "The chilling effects of governance by Data", 86 Chi.L.Rev.
- European Commission (2008), "Flash Eurobarometer 226: Data protection in the European Union: Data controllers' perceptions", available at https://data.europa.eu
- European Commission (2012), "Impact Assessment Accompanying the document Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Process- ing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation) and Directive of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data by Competent Authorities for the Purposes of Preven- tion, Investigation, Detection or Prosecution of

Criminal Offences or the Execution of Criminal Penalties, and the Free Movement of Such Data.” SEC(2012) 72 final

- European Commission (2017), “Enter the Data Economy”, 21 EPSC Strategic Notes
- European Commission (2019), “Data protection rules ad a trust-enabler in the EU and beyond – tacking stock”, Communication from the Commission to the European Parliament and the Council
- European Commission, “The GDPR: new opportunities, new obligations – What every business needs to know about the EU’s General Data Protection Regulation”, available at https://ec.europa.eu/info/sites/info/files/data-protection-factsheet-sme-obligations_en.pdf, 2
- European Council (2017), “European Council meeting (19 October 2017) – Conclusions”, EUCO 14/17, 8
- European Council (2019), “Artificial intelligence b) Conclusions on the coordinated plan on artificial intelligence – Adoption”, 6117/19
- European Council (2020), “Special meeting of the European Council (1 and 2 October 2020) – Conclusions”, EUCO 13/20, 6
- European Data Protection Board (2019), “Guidelines 3/2018 on the Territorial Scope of the GDPR (Article 3)”
- European Data Protection Board (2020), “Contribution to the EDPB to the evaluation of the GDPR under Article 97”, Technical Report
- European Data Protection Board (2020), “Guidelines 05/2020 on consent under Regulation 2016/679. Version 1.0”, 4 May 2020, paragraph 86
- European Union (1995), “Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data”, Official Journal, L 281, 31–50
- European Union (2002), “Direcitve 2002/58/EC of the European Parliament and of the Council of 12 July 2002 Concerning the Processing of Personal Data and the Protection of Privacy in the Electronic Communications Sector (Directive on Privacy and Electronic Communications, lasted amended by Directive 2009/136/EC”, Official Journal, L 201, 37–47
- European Union (2016), “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)”, Official Journal, L 119, 1–88

- Farboodi, M., et.al. (2019), “Big Data and Firm Dynamics,” The National Bureau of Economic Research Working Paper No. 25515
- Felten, E. W., Raj, M., and Seamans, R. (2018), “A Method to Link Advances in Artificial Intelligence to Occupational Abilities”, AEA Papers and Proceedings, 108: 54-57
- Frey, C. B. and Osborne, M.A. (2017), "The future of employment: How susceptible are jobs to computerisation?", Technological Forecasting and Social Change, 254-280
- Furman, J., and Seamans, R. (2019), “AI and the Economy”, Innovation Policy and the Economy, 19(1), 161-191
- Fuster, G. (2014), “The Emergence of Personal Data Protection as a Fundamental Right”, Springer, 164–166
- Gal, M. S., and Aviv, O. (2020), “The competitive effects of the GDPR,” Journal of Competition Law and Economics
- Gal, M. S., and Rubinfeld, D. (2019), “Data Standardization”, 94 NYU Law Rev. 101
- Gans, J. S., and Stern, S. (2010), “Is there a market for ideas?”, Industrial & Corporate Change,
- 805–837
- Garrett A. Johnson, Scott K. Shriver & Samuel G. Goldberg (2021), “Privacy & market concentration: Intended & unintended consequences of the GDPR”
- Gellman, R. (2017), “Fair Information Practices: A Basic History”, Cornell University Press
- Geradin, D., T. Karanikioti, and D. Katsifis (2020), “Gdpr myopia: how a well-intended regulation ended up favouring large online platforms - the case of ad tech”, European Competition Journal, 1–46
- Gill, P., Erramili, V., Chaintreau, A., Krishnamurthy, B., and Papagiannaki, D. (2013), “Follow the Money ? Understanding Economics of Online Aggregation and Advertising”, Proceedings of the 2013 ACM Internet Measurement Conference
- Goddard, M. (2017), “The EU General Data Protection Regulation (GDPR): European regulation that has a global impact”, International Journal of Market Research Vol. 59 Issue 6
- Goldberg, S., Johnson, G. and Shriver, S. (2019), “Regulating privacy online: the early impact of the GDPR on European Web Traffic and E-commerce outcomes”, Boston University Working Paper
- Gonzalez, O. (2018), “Here’s how much information Facebook and Google have on you”, Inverse
- Graef, I. et al., (2018), “Feedback to the Commission’s proposal on a Framework for the Free Flow of Non-Personal Data”

- Graef, I., Verschakelen, J., And Valcke, P. (2013), "Putting the Right to Data Portability into a Competition Law Perspective", The Journal of the Higher School of Economics, Annual Review, 53–63
- Hilbert, M., and Lopez, P. (2011), "The world's technological capacityto store, communicate, and compute information", Science, 332(6025), 60–65
- Holt, T. J. and E. Lampke (2010), "Exploring stolen data markets online: products and market forces", Criminal Justice Studies, 33–50
- Hoofnagle, C.J. (2014), "The Origin of Fair Information Practices: Archive of the Meetings of the Secretary's Advisory Committee on Automated Personal Data Systems (SACAPDS)", Cornell University Press
- Information Commissioner's Office (2019), "Update report into adtech and real time bidding", Technical report
- Irwin, L. (2021), "How much does GDPR compliance cost in 2021?", IT Governance European Blog
- Jia, J., Jin, G.Z., Wagman, L. (2018), "The short-run effects of GDPR on technology venture investment", National Bureau of Economic Research, No. w25248
- Jin, W., and McElheran, K. (2019). "Economies Before Scale: Survival and Performance of Young Plants in the Age of Cloud Computing", Rotman School of Management Working Paper No. 3112901
- Johnson, G. and Shiver, S. (2019), "Privacy & market concentration: intended and unintended consequences of the GDPR", Boston University Working Paper
- Kaminski, M. E. (2019), "Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability", Southern California L.Rev. 1529
- Kaplan, J. (2015), "The Age of the Robot Worker Will Be Worse for Men", The Atlantic
- Kenton, W. (2020), "Concentration Ratio", Investopedia
- Koch, R. (2019), "GDPR fines after one year: key takeaways for businesses", available at https://gdpr.eu/gdpr-fines-so-far/
- Koops, B.J., Newell, B., Timan, T., Škorvánek, I., Chokrevski, T., Galič, M. (2017), "A Typology of Privacy", University of Pennsylvania Journal of International Law 483
- Koski, H. and Valmari, N. (2020), "Short-term Impacts of the GDPR on Firm Performance", ETLA Working Paper No. 77

- Kostov, N., and Schechner, S. (2019), “GDPR Has Been a Boon for Google and Facebook”, The Wall Street Journal, 17 , available at https://www.wsj.com/articles/gdpr-has-been-a-boon-for-google-and- facebook-11560789219
- Kottasova, I. (2018), “These companies are getting killed by GDPR”, CNN Business
- Koutroumpis, P., A. Leiponen and L. D. W. Thomas (2019), “The nature of data”, Innovation and Entrepreneurship Working Papers, Imperial College Business School, London
- Kuchler, H. (2018), “US small businesses drop EU customers over new data rule”, Financial Times
- Kumm, M. (2013), “The cosmopolitan turn in constitutionalism: an integrated conception of public law”, 20 Indiana J Global Legal Studies 605
- Laney, D. (2001), “3-D data management: Controlling data volume, velocity and variety”, Application Delivery Strategies by META Group Inc.
- Layton, R. (2019), “10 problems with the GDPR”, Written Testimony, in Senate Hearings
- Lee, I. (2017), “Big data: Dimensions, evolution, impacts, and challenges”, School of Computer Sciences, Western Illinois University, 1 University Circle, Macomb, IL, U.S.A.
- Leemans, V. and Molony, D. (2018), “Are GDPR fines insurable?”, Risk Management, Vol. 65 Iss. 9, New York
- Lerner, A., Simpson, A. K., Kohno, T., and Roesner, F. (2016), “Internet jones and the raiders of the lost trackers: an archeological study of web tracking from 1966 to 2016”, Security Symposium 16
- Lewandoski, M. (2015), “Designing the Business Models for Circular Economy- Towards the Conceptual Framework”, Adam Jabłonski Institute of Public Affairs, Faculty of Management and Social Communication, Jagiellonian University, Lojasiewicza 4, Krakow
- Libert, T. (2015), “Exposing the Hidden Web: An Analysis of Third-Party HTTP Requests on One Million Websites”, International Journal of Communication, 9, 3544–3561
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Byers, A.H. (2011), “Big data: the next frontier for innovation, competition, and productivity”, in M. G. Institute (ed.), McKinsey Global Institute Report, 1–156
- Markl, V. (2014), “Project Final Report: Data Supply Chains for Pools, Services and Analytics in Economics and Finance”, TU Berlin, Berlin, Germany
- Markman, J. (2018), “GDPR Is Great News For Google And Facebook, Really”, Forbes, available at https://www.forbes.com/sites/jonmarkman/2018/05/22/gdpr-is-great-news-for-google-and-facebook- really/#fac153448f63

- Mashamba-Thompson, T.P. and Crayton, E.D. (2020), “Blockchain and Artificial Intelligence Technology for Novel Coronavirus Disease-19 Self-Testing”, Diagnostics 2020, 198
- Mayer-Schonberger, V. and Cukier, K. (2013), “Big Data: A Revolution That Will Transform How We Live, Work and Think,” John Murray Publishers, London, UK
- Mayer-Schonberger, V., And Ramge, T. (2018), “Reinventing capitalism in the age of Big Data”, Basic Books
- McCall, B. (2020), “COVID-19 and artificial intelligence: Protecting health-care workers and curbing the spread”, Lancet Digit. Health, 166–167
- Merrill Corporation (2018), “GDPR burdens hinder M&A transactions in the EMEA Region, according to Merrill Corporation survey”
- Murgia, M. (2017), “Adtech funding drops in face of Facebook-Google duopoly”, Financial Times
- Nanda, R. (2016), “Financing high-potential entrepreneurship”, IZA Word of Labor, 05530
- Neil, C. (2019), “China Knows How to Take Away Your Health Insurance”
- Nemitz, P. (2018), “Constitutional democracy and technology in the age of Artificial Intelligence”, Phil. Trans. R. Soc. A 376: 20180089
- Niedermann, A. (2019), “Freely-Given and Informed Consent? The User’s Perspective”, Presentation of the results of the Allensbach survey, DLD Europe
- OECD (2015), “Data-driven innovation: big data for growth and well-being”
- Oostveen, M. (2016), “Identifiability and the applicability of data protection to big data”, International Data Privacy Law, Vol. 6, No. 4
- Parker, G., and Van Alstyne, M.W. (2005), “Two-sided network effects: a theory of information product design,” Management Science, 51(10), 1494–1504
- Parmar, R., Mackenzie, I., Cohn, D. and Gann, D. M. (2014), “The new patterns of innovation”, Harvard Business Review, 92(1/2), 86–95
- Perry, B. and Uuk, R. (2019), “AI Governance and the Policymaking Process: Key Considerations for Reducing AI Risk”, Big Data and Cognitive Computing, Vol3(2):26
- Peukert, C., Bechtold, S., Barikas, M., and Kretschmer, T., (2020), “European Privacy Law and global markets for Data”, CEPR, DP 14475
- Price, W. L. et al. (2019), “Shadow Health Records Meet New Data Privacy Laws”, 363(6426) Science 448
- Recio, M. (2019), “Spain’s new data protection law: more than just GDPR implementation”, IAPP (International Association of Privacy Professionals)

- Reinsel, D., Gantz, J., Rydning, J. (2017), "Data Age 2025: The Evolution of Data to Life-Critical", IDC: Framingham, MA., 25.
- Romer, P. M. (1990), "Endogenous technological change", Journal of Political Economy, 98(5, Part 2), S71–102
- Roosendaal, A. (2012), "We Are All Connected to Facebook... by Facebook!", In Gutwirth, S.,
- Leenes, R., DeHert, P., and Poullet, Y. (Eds.), European Data Protection: In Good Health?, 3–19, Springer, New York
- Schechner, S., and Kostov, N. (2018), "Google and Facebook likely to benefit from Europe's privacy crackdown", WSJ
- Schomm, F., Stahl, F., and Vossen, G. (2013), "Marketplaces for data: an initial survey", SIGMOD Record, 42(1), 15–26
- Schwartz, P. M., and Peifer, K.N. (2017), "Transatlantic Data Privacy Law", Georgetown Law Journal, 106, 115-179
- Schwartz, P. M., and Solve, D.J. (2019), "Information privacy concerns the collection, use and disclosure of personal information", Aspen
- Scott, M., Cerulus, L., and Kayaly, L. (2018), "Six month in, Europe's privacy revolution favors Google, Facebook", Politico
- Shields, R. (2018), "Investment in Ad Tech grows increasingly scarce, with Forrester predicting a 75% drop in venture capital", Adweek
- Shiller, B., Waldfogel, J., and Ryan, J. (2018), "The effect of ad blocking on website traffic and quality", The RAND Journal of Economics 49(1)
- Shoham, Y., Perrault, R., Brynjolfsson, E., Clark, J., Manyika, J., Niebles, J. C., ... and Bauer, Z. (2018), "The AI Index 2018 annual report", AI Index Steering Committee, Human- Centered AI Initiative, Stanford University, Stanford, CA
- Shulman, A. (2010), "The underground credentials market", Computer Fraud & Security, 5–8
- Soh, C., Markus, M. L., and Goh, K. H. (2006), "Electronic marketplaces and price transparency: strategy, information technology, and success," MIS Quarterly, 30(3), 705–723
- Strandburg, K. J. (2014), "Monitoring, Datafication, and Consent: Legal Approaches to Privacy in the Big Data Context," in Julia Lane et al., eds, "Privacy, big data, and the public good: frameworks for engagement," 10–12, Cambridge
- Stucke, M. E., and Grunes, A. P. (2016), "Big Data and Competition Policy", Oxford University Press
- Susskind, R. and Susskind, D. (2015), "The Future of the Professions", Oxford University Press

- Teece, D. J. (1986), “Profiting from technological innovation: implications for integration, collaboration, licensing”, Research Policy, 15(6), 285–305
- Thomas, L. D. W. and Leiponen, A. (2016), “Big data commercialization”, IEEE Engineering Management Review, 44(2), 74–90
- Thomke, S. H. (2003), “Experimentation matters: unlocking the potential of new technologies for innovation”, Harvard Business Press
- Tobin, O. (2019), “GDPR: 3 Areas of Ambiguity”, Privacy & Data Protection, 20(2), 15–16
- Utz, C., Degeling, M., Fahl, S., Schaub, F., and Holz T., (2019), “Studying GDPR consent notices in the field”, arXiv preprint arXiv:1909.02638
- Varian, H. R. (2010), “Computer Mediated Transactions”, American Economic Review, 1-10
- Varian, H.R. (2014), “Beyond big data”, Business Economics, 49(1), 27-31
- Vinocur, N. (2019), ‘We have a huge problem’: European tech regulator despairs over lack of enforcement”, Politico
- Weissman, C. G. (2019), “One year in, GDPR seems to have helped Google and Facebook”, FastCompany
- Wong, Z.S.Y., Zhou, J., and Zhang, Q. (2019), “Artificial Intelligence for infectious disease Big Data Analytics.”, Infect. Dis. Health, 24, 44–48
- Yan, B. and Li, J. (2019), “The policy effect of the General Data Protectoin Regulation (GDPR) on the digital public health sector in the European Union: An empirical investigation”, International Journal of Environmental and Public Health 16, 1-15
- Yeung, J. (2020), “What is Big Data and what Artificial Intelligence can do?”, Towards Data Science
- Younas, M. (2019), “Research challenges of big data”, SOCA 13, 105-10
- Zhuo, R., Huffaker, B., Claffy, K., and Greenstein, S. (2019), “The impact of the General Data Protection Regulation on internet interconnection”, Working paper
- Zingales, N. (2018), “Data protection consideration in EU Competition Law: funnel or straitjacket for innovation”, in The Role of Innovation in Competition Analysis (Nihoul, P., and Van Cleyenbreugel, P., eds., 2018)