The ethical and lawful implications of Data collection and its usage by companies in Europe

Academic Dissertation

Format: Social Issues.

By Robin TAILLEPIED

Under the supervision of Professor Rebecca DRY

ISCOM Paris

Master 1 International Communication

Academic year 2022-2023

EXECUTIVE SUMMARY

I chose this dissertation topic after a realisation: Many people around me were underestimating the stakes related to the processing and use of their personal data. It motivated me to treat this subject in detail, with the objective of enlightening the readers. This subject, often neglected in our daily life, but affects us all through our connected lives. Yet, it is often not challenged.

In this dissertation, I position myself as a user in the heart of a major social issue, so I have decided to focus on the most significant aspects of personal data processing. However, given the complexity and breadth of this topic, I had to make selective choices to maintain an appropriate dissertation size. My goal is to offer a useful reflection on the processing of personal data with a focus on Europe to provide information and analysis that will help readers better understand the issues and protections that are in place to protect their personal data. I have therefore chosen to focus on the aspects most relevant to regular users, to address a social issue, while providing in-depth information and analysis to help stimulate awareness. This exercise offered me the opportunity to dive deeper into a fascinating subject. It fed my natural curiosity through studies and research conducted by others, which can be found in the sources and appendices that have considerably enriched this thesis.

I would like to thank my thesis director and head of the Master 1 International Communication, Professor Rebecca DRY, who allowed me to write this thesis and pushed me to do meticulous research with her advice. I would also like to thank all the professors and contributors of this Master 1. Their encouragement and kindness allowed me to complete this thesis to the last detail. I would like to thank my friends for their unfailing support, which opened my mind to go further in my reflection. Finally, I thank my mother, my father, and my sister, whom I love and who always had the right words to support me, to motivate me, but also to remotivate me at certain points.
To all those who will read this dissertation, I thank you.

Table of contents

Introduction

"Privacy is not an option, and it shouldn't be the price we accept for just getting on the internet." - Gary Kovacs[1]

With these words in mind, this academic dissertation delves into the subject of the ethical and lawful implications of data collection and usage by companies in Europe. In an age where data has become an extremely valuable commodity, there is a growing concern about the ways in which personal information is collected, processed, and exploited by companies or organisations. This work examines the regulatory frameworks that have been put in place to protect individuals' rights in Europe, with a particular focus on the General Data Protection Regulation (GDPR). The dissertation also explores the challenges faced by companies in complying with these regulations, as well as the potential harms that can arise from unethical data collection and usage practices in the future. Through a combination of analysis and empirical research, this dissertation aims to shed light on the complex issues surrounding data privacy and to contribute to the ongoing discussion about how best to balance the need for innovation and growth, while keeping our fundamental right to privacy.

Firstly, we will become aware of the economic and social stakes behind the collection, analysis, and exploitation of personal data from the point of view of companies or organisations, and then from that of consumers, who are at the centre of this revolution. Subsequently, through the conducted quantitative study, we aim to comprehend the growth of individuals' concerns regarding the industry they contribute to through their daily connected interactions.

Secondly, we will look at the regulatory measures implemented in Europe, to understand the protection we benefit from. First, the question of consent will be treated, then we will focus on the General Data Protection Regulation (GDPR), addressing its successes and failures, allowing us to see in detail the grey areas that

[1] Gary Kovacs, former CEO of Mozilla Corporation, during his keynote speech at the TED Global 2012 conference in Edinburgh, Scotland.

complicate the understanding of a legislation, despite being intended to serve the consumer… creating loopholes that are exploited by some groups.

Finally, the future: How will data shape the world of tomorrow? We will go into detail about the future in terms of technological developments, but also the potential risks to which we are exposed, and finally, an opening axis on the role that humans must play in this revolution, which is only beginning.

Part 1: The Ethical Issues

Ethics is a moral code and like any moral code, it evolves over time. The code becomes blurred when it comes to data, and this leads to a certain complexity. Proper comprehension of the subject of data and how it works is key to understanding the ethical challenges that we are dealing with in today's world.

It is only recently that users became aware of the collection and use of their data purposes through the media and their personal curiosity. It would be inaccurate to say that users are fully aware of what happens to their data.

In fact, in the quantitative study that was conducted, out of 161 adults, 66,4% of answerers said they were "Not very aware of the data collected but do think it is a major concern." or "Not aware at all". 27,3% were "Somewhat aware, but do not take active measures", while only 6,2% were very aware of the data collected and took steps to manage their privacy settings.

Could data be used to infringe on privacy violations? Discriminate against groups or individuals? Create security breaches? Many people are curious about the real usage of data, some people care, some people don't, some people are even scared… while some people take advantage of it.

Section 1: The collection of data

"Data is the new oil. It's valuable, but if unrefined it cannot really be used. It has to be changed to create a valuable entity that drives profitable activity; so must data be broken down, analysed for it to have value." - Clive Humby, UK mathematician and data science pioneer.[2]

We all use our smartphones to handle texts, calls, photos, searches. Nowadays, almost everyone in the western world has connected home devices, TV, wearable devices… but how does that translate, in terms of data collection? Multiplying that by 5 billion users results in way more than our brains can handle.

1: The question of big data

Every minute, 347.000 tweets are posted on Twitter, over 231 million emails sent, and over 5.9 million user queries on Google...[3] We can say that Data has become the black gold of the 21st century.

The COVID-19 pandemic has further accelerated the adoption of Big Data technologies as organisations seek to leverage data insights for decision-making. It refers to the massive amounts of information that traditional computing systems cannot process. So large that, according to Statista[4], the amount of digital data created or replicated globally has increased more than 30-fold in the past decade, from 2 zettabytes in 2010 to 64 zettabytes in 2020. To put this into perspective, one zettabyte is equivalent to one trillion gigabytes, and the amount of data generated globally in 2020 (64.2 ZB) is equivalent to about 6.42 x 10^19 bytes. The amount of data generated in 2023 (463 exabytes per day) is equivalent to about 4.63 x 10^20 bytes per day.

The volume of data generated worldwide is projected to exceed 180 zettabytes by 2025 and reach the astronomical threshold of 2,000 zettabytes by 2035. Fortunately, with recently developed hardware, tools, and algorithms, all this data can be transformed into useful data. The insights derived from this information can be used

[2] The Guardian, "Data is the new oil", March 6, 2006.

[3] Domo "Data Never Sleeps 10.0", 2022, https://www.domo.com/data-never-sleeps

[4] Statista, "Data Growth Statistics - Size of the digital universe worldwide from 2010 to 2025", 2021.

to develop and improve decision making, efficiency, reduce costs, and increase profits. Data is at the heart of companies' strategies to optimise its use to effectively target their market, improve their products and services, and enhance their customer relationships.

To put these astronomical numbers in perspective, according to Forbes, 90 percent of all the data in the world has been generated over the last two years.[5]

Now, the collection of data has become increasingly important in recent years and the proliferation of digital technologies. But what are we exactly talking about?

A widely cited definition of big data was widely attributed to Doug Laney[6], a former analyst at Gartner, a global research and advisory firm., which defines big data as *"high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making."*

- Volume: it describes our increasingly nomadic situation and our exponential growth of data we generate, driven by factors such as the increasing use of smartphones and social networks, as well as the proliferation of Internet of Things (IoT) devices, which encourages us to leave behind more digital data, both professional and personal. More and more data are generated in the world.

- Speed or Velocity: This data generated is evolving at a considerable speed. They circulate more and more rapidly, coming from independent sources but connected to each other, in networks that are less and less acting in silos. They are evolving so fast that they need to be processed in real time to be able to exploit the information and make decisions accordingly.

- Variety: the sources of this data are diverse. It comes not only from internal sources but also from the environment around it, and from the growing number of connected objects: from smartphones to tablets, from watches to connected cars, Smart TVs, or speakers. These objects can track their users and send them precise and personalised information, even when they are not in use. This data is sent by as

[5] Forbes, "Big Data Statistics - How Much Data is Out There?", February 26, 2021.
[6] Gartner, "Big Data: A Three-Dimensional View.", 2012

many different languages, codes, and formats: there are structured data (csv files, geolocation, HTML5...), semi-structured (EDI documents, RSS feeds, XML files...) and unstructured (dynamic content, emails, photos, SMS, social media, blogs, voice recognition...). The principle of multiple sources is a key principle in the "Big Data" process. It is unthinkable to build a business intelligence project based on a single source of data and claim that it holds the truth. We must match several data from different sources to be able to deduce anything.

The interest of many Big Data projects lies in the ability to decipher and interpret a large amount of information. Data has become more accessible and can be used to help improve marketing strategies and many more. Of course, not all data is created equal, nor needs to be analysed. But data that is well sorted, sectioned and exploited can be a reliable source of information. Behind this marketing term lies a simple concept and many opportunities. Analysing this data allows us to identify the leads to exploit and provide growth to businesses.

2: But… how is data collected?

It is essential to understand how data is collected before going any further. There are nearly countless methods of collection, so we will review the most effective ones, which also generate the most common discussions in social ethics. At the same time, we will not yet address the potential risks and problems, as we will only provide factual information about data collection. Also keep in mind that this academic dissertation covers the "unknowing" aspect of data collection, which means that surveys, customer forms, loyalty programs, or any other form of fully consenting data collection method will not be discussed.

First, let’s talk about the Internet of Things (IoT), coined in 1999 by British technology entrepreneur Kevin Ashton[7], who envisioned a world where physical objects could be tracked and monitored in real-time using RFID (radio-frequency identification). A technology which refers to a network of physical objects that are connected to the internet in real time, allowing them to communicate and exchange data with other devices and systems. These objects are everyday items that we all have such as TV,

[7] RFID Journal, "That 'Internet of Things' Thing" by Kevin Ashton, June 22, 2009.

vehicles, smart speakers, but also more complex systems such as industrial machinery and smart city infrastructure. The concept of IoT has been around for several decades, it has gained significant attention and momentum in recent years with the proliferation of connected devices and the increasing availability of low-cost sensors and wireless networks. By 2025, 75.44 billion IoT devices will be installed worldwide.[8] IoT devices relying on mobile apps and cloud services to function are the most common. They collect personal data such as device usage and overall user preferences, sensors, and tracking devices (where user's physical activity, location, and behaviour are collected), cameras and microphones (where audio and video data are collected, including recordings of conversations and video footage of people and their activities.) When we asked in the quantitative survey "How aware are you of the data collected by your IoT connected devices, such as your smartwatch, smart speaker or cloud services...?" 59% felt "Slightly concerned, with some doubts about the security and privacy, but no major concerns."

The other common method of large data collection relies on two components, starting with marketing school's favourites: social media. Platforms such as Facebook, Twitter, and Instagram generate gigantic amounts of data every day. This includes user interactions such as posts, comments, likes, and shares, as well as user profile data and other metadata. This data can be used to analyse user behaviour, sentiment, preferences, demographics, and many others. An easy way to truly understand the depths of it (as a marketing student) is to see how precise Facebook pixel gets. According to a study by Kenshoo, Facebook's targeting is 86% accurate in predicting user behaviour, compared to Google AdWords' targeting accuracy of 83%.[9]

The other component is user input. This is when users provide information, filling out a form or sign up for a service, checking the "I read" box when asked about their rights. This is key, because if the user agrees, it leaves an open door to whatever he agreed about. When we asked 161 adults how often they read the terms of services of websites, 72,9% answered that they never do, while 24,5% occasionally do it,

[8] TechJury, "How Many IoT Devices Are There in 2023?", April 19, 2023 https://techjury.net/blog/how-many-iot-devices-are-there/

[9] Socialbakers, "Q1 2021 Social Media Trends", 2021

when they are concerned about their privacy. Only 2,6% answered that they always do. User input brings the question of consent, which is a deeper, more lawful subject that will be covered in the second section of this work. In the same category arises cookies and tracking pixels, small pieces of code that websites use to track user behaviour. Cookies are stored on the user's computer when they visit a website, it contains information about the user's activity on the website, such as pages visited, items purchased, and preferences selected, meaning this information is stored on the user's device and can be accessed by the website during future visits, this has become the norm, as study by Ghostery, a privacy-focused browser extension, found that the average website uses 12 third-party cookies to track user behaviour.[10] Tracking pixels on the other hand, are placed on a website or an email, and can be used to track user behaviour across multiple websites, containing information about the user's activity, such as their IP address and the type of device they are using. We found out in our quantitative study that half of the answerers (50,3%) didn't have any significant concerns about the collection of their personal data through cookies or content dialogues, while the other half had concerns related to the potential risks to their privacy (21,7%) and how their data is used for targeted advertising (54%)

It's worth noting that data collection and usage practices can vary widely depending on the company and the industry. Some companies are more transparent about their data collection practices than others, and some industries such as healthcare are subject to stricter regulations regarding data privacy. For some, the exploitation of data is perceived as an invasion of privacy, for others, it offers great opportunities for the future. In the age of cookies, connected devices or of behavioural targeting, consumers leave more and more traces behind, usually without realising it, the data. With the ever more frequent use of social networks and IoT, we expose more information about our lives, often publicly. A question arises: This data is now in an extremely large quantity on top of being very easily accessible... but can it have a monetary value?

[10] Ghostery, "Tracking the Trackers 2020: Web Tracking's Opaque Business Model of Selling Users", 2020,

3: Data brokers & data mining

According to a report by the World Economic Forum in 2020 estimated that the global data broker industry was worth $330 billion in 2020. This is a multi-billion-dollar industry is earned through the sale of our personal data to other businesses.
Acxiom, a US-based company that operates in Europe and provides a range of data services, including data collection, analysis, and marketing. It is one of the biggest actors in the Data Brokers world. They collect and maintain data on more than 2.5 billion people worldwide[11], including information about their demographics, interests, behaviours, and purchases. Then, processes and analyses this data using advanced algorithms and machine learning techniques to identify patterns, trends, and insights into consumer behaviour. These pieces of data are then used to create targeted advertising and marketing campaigns for their clients, helping them to reach their desired audience more effectively.

These data brokers make money by collecting, aggregating, analysing, and selling personal information about individuals and households to third-party clients, such as advertisers, marketers, and other businesses. What those companies do is gather all our personal data without us knowing: Do we travel often? Have a young child? Have allergies? Two bathrooms? Where do we shop? They are negotiating that data with other companies for enormous profits. So, how do they do it?
Let's talk about data mining, which is the process of analysing large amounts of data to extract useful insights, patterns, and relationships. It involves the use of statistical and machine learning techniques to identify patterns and relationships in the data that may not be immediately obvious.[12] It can be used in a variety of fields including business, healthcare, and science. In business, data mining can be used to analyse customer behaviour and preferences, identify trends, and improve decision-making. In healthcare, data mining can be used to identify risk factors for diseases, predict patient outcomes, and improve clinical decision-making. In science, data mining can be used to identify patterns in complex data sets, such as those generated by

[11] Acxiom, "About Us", accessed March 8, 2023, https://www.acxiom.com/about-us/

[12] Han, J., Kamber, M., & Pei, J. *Data Mining: Concepts and Techniques (3rd ed.).* Amsterdam, 2011, Elsevier. 744 p. p. 19.

genomic sequencing or particle accelerators. It is one of the key technologies that underpins the digital transformation of European industry.
Is it even legal? Yes, it is these people have agreed to handle their information. They're the ones who clicked “accepted and agreed” on the website popups, which requested the use of their personal data. They're the ones who left their social media pages public. They're the ones who like those Facebook pages or posted about their thoughts and interests.

We are in a new era, where individuals may not fully understand what they are agreeing to, yet their digital signatures remain in someone else’s possession. Their focus is to obtain as much data as possible, as each data point is valuable.

Section 2: It’s not our data anymore.

1: Becoming the product

Nowadays, people give out their information willingly, we are often required to provide personal information to create an account, the things we buy, like, our hobbies, relationships…. As of January 2023, there were 4.76 billion active social media users worldwide13, which is 59.4% of the entire population, according to data from Meltwater and We Are Social. This suggests that a large portion of people are willingly providing personal information, often without realising that this data can be extracted and sold. 35,4% of our study answerers said that they were not aware of third-party companies’ activity regarding this practice and 52,2% were indeed aware but didn’t take any measures to protect their privacy. Only 12,4% were indeed aware and took measures to protect it.

A former Google employee explained in an article for QUARTZ[14] that Google and Facebook create for each user a digital double which is fed by all our searches, our purchases, our likes and by dint of this, it becomes like us: the person who

[13] We Are Social and Hootsuite, "Digital 2023: Global Overview Report," 26, January 2023, https://datareportal.com/reports/digital-2021-global-overview-report
[14] Garcia, A., Google and Facebook are creating a dystopian nightmare, 2017, Retrieved from https://qz.com/1034972/google-and-facebook-are-creating-a-dystopian-nightmare/

resembles us most in the world is therefore stored in a computer somewhere in the world. It is for example possible to determine if a woman got pregnant: from the purchases she makes, if she's looking for more natural products, increased her calcium intake... these are all signals for the algorithm.

One of the first data lessons I learned at ISCOM Paris was from an article of Forbes[15], in which a dad complained that his daughter, still in high school, was getting emails from Target stores offering pregnancy products. The father had made a scandal, because it is disgusting to encourage a girl so young to have a baby... Except it was true. The store knew about it before the father even knew about it. With all this information, it is possible to anticipate the date of birth and follow the mother's needs at all stages of her life, which can give tragic situations.

Profiling is permanent, as everything is analysed: Google search, GPS location, likes, posts, even private conversations on Messenger. Until 2018, Google was even reading emails.[16] The lesson to keep is that Data at its core is based on human behaviour.

All this data, once sorted, can be looked at and identified as patterns. From there, anyone can be categorised depending on the activity that they have online. Someone's political leanings can be inferred from the content you save and like on social media, such as Instagram and Facebook. A person's purchase history can tell us a lot about them, including their age, their interest in dieting, and whether they have children. What we buy on a regular basis will reveal our age, our interest in fitness, and the kinds of products we like. An individual's credit score may be used to infer their health status, religious leanings, immigration status, and even their intention to leave the country. By utilising predictive algorithms through datasets, profiles are generated for everyone, transforming these individuals into the product. Such profiles can provide comprehensive information about one's personality, thought process, and propensity to act a certain way. The potential of this approach is boundless if one is ready to invest in acquiring such information. The product

[15] Forbes, "How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did", 2012

[16] Clubic "Google n'espionnera plus Gmail pour cibler sa pub", 27 June, 2017, https://www.clubic.com/messagerie-email/gmail/actualite-832480-google-espionnera-gmail-cibler-pub.html

comprises a neatly packaged bundle containing an individual's behavioural patterns. After creating these profiles and audience lists, the information can then be sold.

2: Selling our identity

Who does our data get sold to? This is the primary question… The primary customers for Data brokers are marketing companies, as it makes targeting anyone easier. Suppose we possess a record of individuals who prioritise their health, which is something discernible from their purchase behaviour we previously mentioned. In that case, some companies such as fitness centres, nutritionists, meal delivery services that prioritise healthy eating, virtual fitness programs, and so on, all of which may express interest in obtaining such data.

The more targeted an advertising campaign is, the more likely it is to be directed at specific individuals or groups, thus increasing the chances that users will react to it in the desired way.

But it goes deeper than the typical marketing company who wants to retarget a campaign, in fact, many other factors can influence how much our identity is worth online when being sold.

“Data brokers sell lists of people suffering from mental health diseases, cancer, and hundreds of other illnesses," (...) "lists of people who live in or near trailer parks so that these undesirable consumers can be targeted for suppression… often to those who make predatory offers to those in financial trouble. They sell lists of people who are impulse buyers or 'eager senior buyers.' All in all, there are millions of lists."[17]

Pam Dixon, founder, and executive director of the World Privacy Forum made this statement during her testimony almost ten years ago. It highlighted the ways in which data brokers can use personal information in harmful ways, such as targeting vulnerable individuals with predatory offers or discrimination based on health status or socioeconomic status. It brought light and calls for greater transparency and regulation of the data broker industry to protect consumers' privacy and prevent abuses of personal information.

[17] Pam Dixon Executive Director, World Privacy Forum, Before the Senate Committee on Commerce, Science, and Transportation, 18, December, 2013

By having access to an individual's personal information, it is possible to infer their lifestyle. This information is of great interest to organisations such as insurance companies that are keen to evaluate the level of risk involved. For example, banks and financial institutions collect data on customers' credit card purchases or buying behaviour to determine interest rates on loans. People who search websites that charge fees wanting to access personal information like names, addresses, and family relationships. Or even better, governments having a strong incentive to track down individuals, especially those who might be of interest to them…
It's the same thing with insurance companies that can scan a car's GPS location to see if the person is driving too fast or if they are parking in areas that are not so safe. As a result, one's price of insurance could be increased. These are all upsetting examples. However, we have 'knowingly' agreed to their service terms.

The price tag of our data gets more expensive, the more detailed and specific it is. An interesting case is one of computer consultants, Frederico Zannier, who decided that he would sell his data himself. For a month, he recorded his screen, his webcam, his mouse activity, his GPS location... then he put it all up for sale on KickStarter for the price of two euros a day. In only one month, he collected 2400€![18]

We can only imagine how much value, information like a person's political affiliations, their buying decisions or their health conditions are worth... It's hard to put a price on it, as this industry is built in the shadows.

[18] Frederico Zannier, "A bit(e) of me", KickStarter, October, 2013, https://www.kickstarter.com/projects/1461902402/a-bit-e-of-me/faqs?lang=en

3: Should we even care?

Anyone who is born in the world of surveillance technology like us Gen Z, hearing about how we became the product, and how our data is sold, isn't bothering as much as we may want it to, as long as we receive a quality service… Results from our quantitative study shows that most are either neutral (36%) or somewhat agree (26,1%) to this 'exchange of data for a service that benefits' state.

These results might be understandable if we consider the very common sentiment of: "I have nothing to hide, therefore, nothing to fear."
Or even "If you aren't doing anything wrong, you have nothing to worry about."
While we enjoy free access to various websites, apps, or services, is it ethically acceptable for our information to be sold? Data can, after all, be used to find criminals, pandemics, manage illnesses, and many others. But what if the bad guy obtains that data?

A concrete example to illustrate this point, which is not European but has led to action in Europe, is the Cambridge Analytica scandal, where the political consulting firm had collected the personal data of millions of Facebook users without their consent and used this data to influence the opinion of voters in the 2016 U.S. presidential election.[19] They had obtained the data through a third-party application called "This is your digital life", in exchange for a $3 pledge, which was launched after the user logged in via Facebook. This application collected information about the user's Facebook data... and friends. Only 270,000 people took the quiz, but with an average of 330 Facebook friends in the U.S. That's 87,000,000 Americans who had their data sold to Cambridge Analytica[20], which used it to create individual voter profiles and target them with personalised political ads.

What if a person or organisation bought our information without knowing their true purposes? What if a powerful government used that information to target people who were openly opposed to them? Then, would we care, or would it already be too late?

[19] The Guardian, "The Cambridge Analytica Files", 2018, https://www.theguardian.com/news/series/cambridge-analytica-files.
[20] The Guardian, "Cambridge Analytica scandal: what you need to know", 21 March 2018,

Is it already too late because our data isn't ours anymore? In the end, many of the free services, applications, and websites we use rely on the collection or sale of our data to make profit. Although it may sound dreadful, this industry does present a challenge: is it really a problem if many of the free websites and apps that we use depend on our data to remain profitable? Does it really matter if it's just being used for marketing purposes? We asked this question to 161 adults in our quantitative study : 68,3% of them answered that they would not be willing to pay a subscription fee and would prefer to continue using these services for free. 19,9% weren’t so sure while 11,8% would agree.

What regulations exist that can help us understand, or better yet, face those problems?

`

Part 2: Regulatory Frameworks

The modern concept of privacy law originated in the 1960s, from an academic study. Dr. Alan Weston, who was at Columbia University, wrote his doctoral dissertation for which he later got funding to turn into a book called "*Privacy and Freedom*."[21]
He established a universally accepted definition of privacy in that book. "*The claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others."*[22]
By the 1990s, every country had followed suit. During those years, the emergence of the internet and the widespread use of personal computers led to a significant increase in the collection and processing of personal data, which in turn sparked the first concerns about consent regarding privacy and data protection. In response, the European Union (EU) had adopted the Data Protection Directive in 1995, which established a comprehensive framework for the protection of personal data in the EU. In 2012, the EU began the process of updating its data protection framework to account for advances in technology and the increasing importance of data protection in the digital age. The General Data Protection Regulation (GDPR) was adopted in 2016 and came into effect in May 2018, replacing the Data Protection Directive. Although personal data laws have evolved over the years, some concerns and questions remain unanswered by the law. One of these concerns is the issue of consent. While laws require companies to obtain consent from individuals before collecting and processing their personal data, there is still some ambiguity around what constitutes valid consent.

Section 1: The issue of consent

"Accepting legitimate needs of law enforcement and public interest, control of information must rest with the person himself." Said William Sapphire, in the New York Times during the 1990s.[23]

[21] Weston, Alan F., *Privacy and Freedom*, 1967, Atheneum, 322 p.
[22] Weston, Alan F., *Privacy and Freedom*, 1967, Atheneum, 322 p., p. 7.
[23] Safire, William. "Nosy Parker Lives," New York Times, Sept. 23, 1999, at A29.

Keep in mind that we are providing much of the information we are talking about voluntarily. Millions upon millions of texts are being exchanged every single second, and countless photos and videos are being uploaded every single minute.
Challenging consent seems totally counterintuitive in the world of privacy because privacy is closely linked to us and our autonomy. For three reasons we are going to outline, we are going to understand if it is, or if it isn't, both practical or desirable that we focus on consent, and how it could explain why privacy laws are in the dreadful state they are today.

1: The complexity of privacy notices

Privacy notices, also known as privacy policies, are documents that explain how an organisation collects, uses, processes, and protects personal information. They're an important tool for ensuring transparency and accountability in the handling of personal data. Reality is users see them all the time and ignore them. In fact, of the 2,105 people surveyed by Addictivetips.com, a massive 87% said that they don't read through privacy policies before agreeing to hand over their data.[24] And that's okay, almost nobody does, unless it's a lawyer who gets paid to read them. When someone visits the doctor, they get a privacy notice. Logging onto a website? We'll get a cookie privacy notice that's required by the GDPR European law. This is why we get it.
For example, PayPal privacy notice in France is approximately 9,300 words long, Facebook privacy policy comes to around 10,000.
One well-known 2008 study conducted by the Carnegie Mellon University's CyLab calculated that to read the privacy policies of the 40 most popular websites in the world would take an individual 30 full working days a year[25]. While the study is now over a decade old and the number of websites and the length of their privacy policies have likely changed, the study's findings have remained a popular reference point in

[24] Addictivetips.com, "Our Attitudes Towards Privacy Policies", 2018, https://irishtechnews.ie/our-attitudes-towards-privacy-policies/
[25] Carnegie Mellon University CyLab, "Usable Privacy and Security: A Guide to Relevance and Research," 2008, p. 5.

discussions about the complexity and length of privacy policies. These notices are complex because the overall thing's complex, they are difficult to understand, and we often just pass them.

Secondly, they are often inaccessible. For example, imagine a user is presented with a privacy notice from a social media platform that allows the company to collect and share their personal information with third-party advertisers. The user wants to opt-out of this data sharing, but the option to do so is buried deep within the privacy notice and requires them to navigate through multiple pages of settings. The opt-out checkbox is already pre-checked, making it appear as though the user has already agreed to the data sharing. The user may not notice this or may assume that they cannot opt-out without agreeing to the company's terms. Companies also use design tactics that make it difficult for the user to make an informed choice about their data. This can explain why many people argue that current consent mechanisms do not provide users with meaningful control over their personal information, so how do we manage consent in a world in which data's being inferred about us or collected as part of a group?

Third point is that consent regarding data has been proven incredibly ineffective, mainly because people just ignore it. A quote from The Federal Trade Commission Chairman, (F.T.C), which is the largest privacy regulator in the United States) said back in 2009: *"We all agree that consumers don't read privacy policy."*[26], coming from the person who has done more than anyone else on earth to make us have those privacy policies, it is quite concerning.

2: Is the consent we give an illusion?

In most cases, we have no choice. When we try to update our phone software for example, it comes out every couple of weeks. If we don't update it quickly, we're prompted to update it. Then it starts forcing it, saying, it is going to update it for us automatically. Then, the first thing it does when we update it is showing their privacy

[26] Federal Trade Commission Chairman (F.T.C) John Liebowitz, 2009.

policy. We can download it, we can email it, we can agree to it, but we cannot not agree to it. The alternative being, ‘would you like us to turn your very expensive phone into a brick?’ Which is the alternative if we say no.

There's a huge burden on individuals of all these consent opportunities. While consent is often talked about as a right, sometimes even a human right, it's realistically much more of a duty. It's much more of a burden on individuals. Remember, when we make that choice, it has the legal effect of shifting the liability from the data processor to us. It's just like if someone drives in a garage and takes a ticket that says “We have no liability for anything. We can crush and melt your car. We can throw it out the edge of the garage. We're not liable for anything…” Meaning that, while we got the right to consent to that by driving in the garage, it was the imposition of a burden. That burden is quite significant at times because of the legal significance that can attach to those.
Another problematic that rises from consent, is the one regarding that choices often serve as an enormous disservice both to individuals and society. Think about press coverage. Do we really want the president to have a right to privacy right now that only his consent will allow coverage of what he's been up to? Fraud prevention, crime detection… Do we want to wait for the criminal’s consent so that we can use their data? They are unlikely to give their consent for the use of their data, so relying on consent in these situations would effectively prevent law enforcement from using data that could be crucial for preventing or solving crimes. What we do in these cases is we override the consent. Research often depends on being able to use past data, often in an anonymized fashion without going back and getting individual consent. The strict application of consent can be an obstacle to progress.

The real challenge may be that consent can lead to lousy privacy protection: it's not the same thing clicking ‘I agree to’, as ‘I get privacy now’. We agree to terms that often eliminate our privacy. We often agree to broad terms that appear to have no limit and are asked to comply with things that we could never be asked to do in other consumer protection settings.

3: The stewardship of Data.

If our data is used and something goes wrong that causes harm, there must be responsible management and protection of data, throughout its lifecycle, from collection to disposal. Stewardship involves ensuring that data is collected and used in a way that is ethical, legal, and respectful of individuals' privacy rights. Just as a lawyer act in the best interests of their client, those who collect and manage our data should act in our best interests by treating our data with care and ensuring that it is used in a responsible and ethical manner. Similarly, a banker and a doctor have a duty of care to their clients, they are expected to always act responsibly and professionally.
Why wouldn't we say, “If you're using my most personal information, you should be held to the same requirements?” This includes implementing appropriate security measures to protect data from unauthorised access or disclosure, as well as ensuring that data is accurate, up-to-date, and only used for legitimate purposes. The concept of data stewardship is increasingly important in today's data-driven world, where data breaches and misuse are becoming more common.

When sensitive information is accessed or disclosed without authorization, it is known as a data breach. Data breaches can take many different forms, including physical theft, malware attacks, hacking, and human error. When a data breach occurs, sensitive information such as confidential, personal, financial, or business information can be compromised, resulting in potential harm to individuals or organisations. It can have serious repercussions, such as identity theft, financial loss, reputational damage, and legal repercussions. According to a report by RiskBased Security, there were over 36 billion records exposed in data breaches in 2021 alone.[27] This highlights the need for effective stewardship of data to protect sensitive information.

Now, what are the things we agree that can be done with data in normal circumstances and what shouldn't be done with it? Starting by taking stalking or fraud out of the equation, but some things should be in activities such as bill collection,

[27] RiskBased Security. “2021 Mid Year Report Data Breach QuickView Report,” 2022, https://pages.riskbasedsecurity.com/download-the-2021-mid-year-data-breach-quickview-report-today

fraud detection, research are some things we might be able to slide over into the ‘generally permitted’ category if we use good security and not have to burden people by telling them the bleeding obvious.

We might think more about redress because no matter how much care is taken, something will go wrong at some point. It is therefore crucial to have a system in place that provides remedies when things go wrong. Individuals are often left in the cold or uninformed when their data privacy is breached. They may not even be aware of the breach until it is reported in the media. A study by the French data protection authority (CNIL) in 2018 found that only 34% of individuals who had experienced a data breach were informed by the company responsible, with the remaining 66% only becoming aware through media coverage or other means.[28]

There should be a more transparent and accountable system in place that informs individuals about the breach and offers remedies for the harm caused. Using consent in all these other settings has the unintended effect of making us tend to ignore it. When we could make meaningful, effective choices. That would protect our privacy. The overuse of the concept of "consent" can have unintended consequences, which may lead to people ignoring the importance of their privacy. When the idea of "consent" is used too frequently or inappropriately, it can desensitise people to the significance of privacy-related choices. People should be encouraged to make meaningful and effective choices regarding their privacy, rather than being bombarded with consent forms or agreements that they may not fully understand. By making more informed decisions about how their data is used and who has access to it, individuals can take a more active role in protecting their privacy. Finally, effective privacy protection does not solely depend on consent. It is important to implement measures and regulations that safeguard personal information and limit access to it, regardless of whether consent has been obtained or not.

[28] CNIL (Commission nationale de l'informatique et des libertés), "Baromètre annuel de la protection des données personnelles", 2018,

Section 2: Focus on GDPR

GDPR'S original goal was to improve the privacy of EU citizens in three main ways: First, before collecting any personally identifiable data, companies had to clearly explain what they would collect, how they would use that data, and who they would share it with. And they could then typically only go on with the collection if they had received explicit and informed consent for it. Second, once the company collected the data, the user could then ask to see, export, and delete said data and if it was compromised through a Data breach, the company now had to let the relevant authorities know within 72 hours[29]. Third, GDPR mandated that the companies and individual EU member states had put the processes and resources in place to be able to enforce these new rules, including member states establishing dedicated data protection authorities and companies naming dedicated data privacy officers if they were large enough, taking responsibility for the practices of their contractors. Now, most would only have to comply with one set of privacy rules for the entire continent going forward, not 31 individual ones.

In other words, while GDPR did not outright ban the tracking of users by itself, it did give people real information, real decision-making powers, and a legal system to turn to if needed, making it both surprisingly ambitious and well-intentioned. But intentions are one thing, and reality is another thing entirely.

1: The successes of GDPR

Five years have passed since the GDPR came into power, back in May 2018. We can now take a look at what happened, the successes, and where it sometimes went wrong.

- Perhaps the most obvious success of the GDPR is how it impacted data breach warnings. Last year alone, around 109,000 data breaches were reported across the EU according to law firm DLA Piper[30], while before the introduction of the GDPR, those numbers were around 18000 to 20,000 reports a year, meaning that firms are

[29] EU GDPR, Article 33, "Notification of a personal data breach to the supervisory authority", https://www.privacy-regulation.eu/en/33.htm

[30] DLA Piper, “GDPR fines and data breach survey,” January 2023, https://www.dlapiper.com/en-ae/insights/publications/2023/01/dla-piper-gdpr-fines-and-data-breach-survey-january-2023/

at least five times as likely to report breaches across the E.U. now that there are strict rules for it.
GDPR's requirement that organisations must report a data breach to the relevant supervisory authority within 72 hours of becoming aware of it also means that there should be no undue delay for reports in the, unlike how Equifax in the US for example, took around two full months to notify the authorities of their massive data breach concerning over a hundred million people[31]

There are now also clear fines being imposed when breaches were the result of inefficient security measures, with at least 122 fines being issued so far, including to Amazon (winner of the biggest fine for breaking European data laws, for $886.6m), Instagram (who were fined €405m over children's data privacy)[32], British Airways, hospitals, and even public authorities like the Bulgarian National Revenue Agency, for example.[33]

Maximum amounts allowed under GDPR, are up to €20 million or 4% of a company's global annual revenue, whichever is higher, for the most serious violations of the regulation, but Data breach warnings are overall undoubtedly a huge success of GDPR.

Another area of success is the ability for users to view, download, and delete the data a company has on them: "right to be forgotten"[34]. For example, places like Facebook, Google, and Amazon have automated portals that let users download or request a deletion for all of user's data, and while many have rolled this out to non-EU citizens as well, we almost certainly have GDPR to thank for that too.

[31] TechCrunch, "Equifax breach was 'entirely preventable' had it used basic security measures, says House report", 10, December, 2018, https://techcrunch.com/2018/12/10/equifax-breach-preventable-house-oversight-report/?guccounter=1

[32] BBC News. "Instagram owner Meta fined €405m over children's data privacy". September 15, 2022. https://www.bbc.com/news/technology-62800884

[33] https://www.enforcementtracker.com/

[34] European Union, "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)", 2016, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679.

Before GDPR, many services simply refused to delete our data. Personal experience of this, back in the day when my dad spent over three months and multiple hours of support calls trying to get an Apple ID deleted without success, or the countless stories of online newspaper subscribers who wanted to delete their email address from the newspaper's system as they were concerned about their privacy, wanting to remove their personal information from the database, but later find themselves never having an answer, without having the possibility to take concrete legal action. These cases set a precedent for people to hold tech companies accountable for how they handle personal data, and it showed how GDPR protects EU citizens' privacy rights.

- A third success is bringing more information and transparency, as the GDPR requires companies to show individuals every single company they could share their data with as well as how it will be used.[35] While this can be a lengthy list and may not be practical for individuals to go through for every website they visit, it does provide a tool for researchers, journalists, and privacy organisations to identify companies that do not prioritise privacy and to highlight the reasons why. This increased transparency can help to expose the darker side of data collection and hold companies accountable for their data practices.

2: The failures of GDPR

Some websites or applications claim to give users a choice about their data privacy, but they do not provide a fair and realistic option. The cookie banners and content dialogues that these websites display can even be of help, for us to understand the weaknesses of the GDPR regulation.

The cookie banner is a pop-up or banner that appears on a website or application that informs users about the use of cookies or other tracking technologies. An excellent implementation of the content banner can unsurprisingly be found on the official websites of the European Union.[36]

[35] Europa, Your Europe, "Providing transparent information", 2022, https://europa.eu/youreurope/business/dealing-with-customers/data-protection/data-protection-gdpr/index_en.htm#shortcut-11

[36] European Commission, "Privacy policy for websites managed by the European Commission", https://commission.europa.eu/privacy-policy-websites-managed-european-commission_fr

The banner is small and can be ignored without affecting the site's functionality, rejecting it is just as easy as accepting it, the only cookie that gets dropped is a technical one that stores information about the user's browser settings. And all other trackers the EU has are clearly explained in their policy and seem reasonable too. This example is very GDPR compliant.

One step below that is GitHub[37], a well-known website for hosting and sharing code repositories, doesn't have a cookie banner like the EU's official website does. Instead, GitHub places cookies on the browser that they consider to be "strictly necessary" for storing users' preferences and login information. Additionally, they only use first-party analytics, meaning that they are only collecting data on their own website and not sharing it with any third-party companies. They can confidently claim that whatever data they have on a user is either "necessary for the performance of a contract to which you are a party" or in their "legitimate interests", both of which allow them to collect the data without additional consent. GitHub's approach to cookies and data collection is clean and nice because they are collecting data within the bounds of what is considered necessary for their business operations and are not sharing it with others without consent.

Let's dive deeper into even shadier policies. Instagram[38] has no clear option to reject cookies on its platform. It only mentions that the user can opt-out after reading the cookie policy, but the cookie policy does not provide a real solution. Clicking on "I want to learn more" does not provide any useful information either. Moreover, the user even must accept all tracking first, to be able to read the cookie policy and understand what information Instagram will track. This lack of transparency and choice goes against the spirit of GDPR, which aims to give individuals control over their personal data. Instagram is owned by Meta and can access data from other Facebook-owned apps and websites, as well as data from third-party partners. This is alarming because Instagram users may not realise just how much of their data is being shared with other companies and entities.

[37] Github, https://github.com/

[38] Instagram, https://www.instagram.com/

If we click on the details[39], we get a page which shows that by default, users have given access to all their data to Meta and Google, as well as any member of the IAB (Interactive Advertising Bureau). The IAB is an industry association that represents many major advertisers and ad networks online, so allowing them access to personal data can potentially result in targeted advertising and tracking across a wide range of websites and apps.

Under GDPR, the people collecting the data are also the people who are designing the consent forms for that collection. It's kind of like telling criminals to design their own prisons: even if we instruct them not to leave the doors, they probably will just leave all the doors open! Contrast that with how Apple designed an iOS system prompt third party app makers must use if they want to track users across other apps, we can immediately see how their design is way less ambiguous than the one we've seen with the previous examples. Here, the answer is just a simple standard YES / NO question.

Users can easily understand and recognize the big tech companies such as Google, Facebook, Apple, and Amazon, and have a certain level of familiarity with their data practices. However, when it comes to publishing giants and programmatic ad networks, which operate behind the scenes, there is no way the average user is going to understand their data practices and how their data is being used by these faceless publishing giants and programmatic ad networks, this is just outside of the normal person's world completely.

3: Loopholes & conflicts of interest

Advertisers have found the perfect loophole out of GDPR, completely in the form of 'legitimate interest'. Under GDPR, there are six legal grounds for processing personal data, and one of them is "legitimate interests pursued by the controller or by a third party." This means that companies can process personal data without the user's

[39] Instagram, "Privacy Policy, What is the Privacy Policy and what does it cover?" https://privacycenter.instagram.com/policy/?annotations[0]=1.ex.37-OurPartnersBusinessesAnd

explicit consent if they have a legitimate interest in doing so. What exactly is a legitimate interest?
This is the million-dollar question all the companies are testing the limits of. The British Information Commissioner's Office (ICO) says “Companies can use legitimate interest whenever they process people's data ‘in ways that they would reasonably expect’ or where there is a “compelling justification of the processing.”

It says that commercial interests can count as legitimate too, which is vague. And we might also wonder what ‘commercial interests’ count as ‘compelling justification for data processing’ without even asking for consent under the law.

Different companies are using the concept of "legitimate interest" to justify tracking users' data. Some are arguing that most of their ad network companies do not have a legitimate interest in tracking users, but there are a few exceptions that they are unable to opt out from, in the list.
On the other hand, some argue that all their partners have a legitimate interest in tracking users, but they are not enforcing this in practice. This creates confusion because, according to GDPR regulations, users must give consent for their data to be processed based on legitimate interest, but they are not given the option to do so. In other words, companies are using legitimate interest as a loophole to track users' data without obtaining their consent, which goes against the spirit of GDPR.

Facebook in their policy[40] goes furthest and claims that their legitimate interests include, “providing an innovative, personalised, safe, and profitable service” They claim providing a ‘profitable service’ is a ‘legitimate interest’ under the GDPR. As a reminder, legitimate interest allows for the processing of user data without even having any user consent. So, if Facebook is right, and operating a profitable service count as legitimate interest, then basically any for-profit ad company would have legitimate interest to process basically any user data, without any user consent: which would be the mother of all loopholes! Keep in mind we are discussing a hypothetical scenario based on Facebook's interpretation of legitimate interest, and

[40] https://www.facebook.com/about/privacy/previous

the actual interpretation and application of GDPR is subject to ongoing debate and legal challenges.

There is still a lot of debate around legitimate interest, and while the European courts have made a few judgments around it, the exact premises aren't clear yet. However, combined with the other two weaknesses, such as companies designing their own consent and the lack of clarity around publishers and ad networks, it is clear to say that the GDPR has not been a complete success. There is still a lot left to be done, it is a significant first step into the right direction, but clearly there are still many loose ends that need to be tied up, and most of the EU would probably agree with this statement. Moving forward, the future of data privacy and protection remains uncertain, but it's important for individuals to stay informed and aware of their rights and options when it comes to their personal data.

Part 3: The future of Data

Section 1. Practical benefit and security

One thing is for sure: digital technologies are changing the lives of citizens. Notably, the EU Digital Agenda aims to ensure that this transformation benefits citizens and businesses, while helping the EU reach its goal of climate neutrality by 2050.[41] The Commission is determined to make the coming decade Europe's "*digital decade*". Europe wants to strengthen its digital sovereignty and set standards instead of following those of others, with a clear focus on data, technology, and infrastructure.

1: Data centralisation

The idea of centralisation of data refers to the practice of storing data in a single, centralised location, such as a server or database. In this model, all data is collected and managed in one place, and users access it remotely through the internet or a network. Centralization has shown to provide several advantages, including ease of access, security, and management of data.

With the amount of data generated by individuals and organisations growing at an unprecedented rate, the rise of cloud computing, companies demanding more advanced data analytics and insights, coupled with the growing concern about data privacy… centralisation of data promises to answer to all these goals.

A concrete example of this upcoming Digital Decade? In 2024, all European countries will have to provide their citizens with a Digital ID Wallet.[42] Thanks to this electronic wallet, they will be able to store and manage their digital identity and share their verified personal attributes from their electronic terminals. Usable throughout Europe, it will put citizens in control of their personal data and promises to guarantee a very high level of security. Renting an apartment, buying a car, opening a bank account, taking out a bank loan, giving consent for a medical procedure, enrolling in

[41] European Commission, "A Europe fit for the digital age, Empowering people with a new generation of technologies", 2020, https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age_en

[42] European Commission, "Digital Identity for all Europeans, a personal digital wallet for EU citizens and residents", June, 3, 2021, https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-digital-identity_en

a university... Today, these are all dematerialized procedures that require filling in dozens of forms, proving one's identity, authenticating oneself on a multitude of digital services (e-commerce site, public services...) and sometimes transmitting captures of paper documents by email, without any particular security, during transmission and then storage, and often containing much more information than is necessary... According to a survey by Onfido, a digital identity verification provider, 53% of consumers abandon online applications if they take too long, and 41% of respondents said they found digital identity verification frustrating.[43] So many complex and time-consuming operations for all users, especially those for whom the protection of personal rights and data and the mastery of digital tools and security solutions are not obvious, centralization seems to be the ideal compromise. The centralization of these data is explained by the offer that the EU wishes to allow all its citizens in a literal way, by allowing him to contain all his life on the same wallet: the verified elements, of the biometric national identity card (fingerprints, photographs, textual information), authenticated and certified personal attributes (driver's licence, birth certificate, bank card, proof of address, tax form, diplomas, pay slips, health documents...). By asking member states to offer their residents such a digital wallet, the EU wants to guarantee that each citizen can regain control and use of their identity. Thus, thanks to wallet technologies, they will be able to limit the sharing of their data to that which is necessary for the delivery of the service. Retrieve data created and verified by third parties in the form of Qualified Electronic Attribute Attestations (QEAA), store them in the wallet, and communicate them when necessary.

Currently, all personal data of each citizen is stored in private and public databases. It will soon be decentralised and managed individually by each European citizen. By becoming sovereign of his data, he will be able to consciously choose to share it without fear. The EU is betting the future of data on the fluidity and simplicity of its implementation, to convert all Europeans. This entire system seems flawless... but is it really?

[43] Onfido, "Customer attitudes to digital identity", November, 2020 https://onfido.com/report/customer-attitudes/

2: The race for personal data

In terms of AI and Big Data, the European Union at the moment lags far behind China and the United States. In 2019, the United States had invested €29 billion in AI, while the EU invested only €3.2 billion. In terms of patents filed in the field of AI, the EU accounted for only 10% of the global total in 2018, while China and the United States accounted for 48% and 31%, respectively. China is home to 7 of the world's top 10 companies in terms of the number of AI patent applications filed, while the EU has none.[44]

The two superpowers benefit from the delay because they have unrestricted access to data on European citizens, but GAFAM's rapid expansion is made possible by this data. These data-driven businesses are creating products like the current generation of robust social media algorithms and cutting-edge AI/Machine Learning (ML) systems (such as ChatGPT). In response, European Commission President Ursula von der Leyen announced a plan in September 2020 during a press conference to help the old continent reclaim its technological lead.[45]

The European Commission wants "*To fuel our economy and find European solutions for the digital age through digital transformation*". First, by the year 2030, the European Commission hopes to have established a "*single European data market*" for Big Data. Data must have the same freedom to move around the EU as does the movement of people and goods. As a result, European businesses will have access to all that data for the first time in history. To better compete with the likes of the American GAFAMs and the Chinese BATX, more EU businesses will be able to adopt a data-driven approach. To facilitate this change, the EU plans to establish a governance framework for data access and reuse, with the goal of promoting data sharing while upholding *"European values and rights"* like data protection and fair competition. Open Data is another EU-wide initiative with significant significance. A cloud infrastructure will be built to support massive data reuse, and high-value datasets will be made freely available for reuse.

[44] European Commission, "Digital Economy and Society Index (DESI) 2020", June, 2020, https://eufordigital.eu/library/digital-economy-and-society-index-desi-2020/

[45] European Parliament Plenary, Ursula von der Leyen, "State of the Union Address", https://state-of-the-union.ec.europa.eu/state-union-2020_fr

The European Union plans to invest more than 20 billion euros annually over the next few years to *"be a world leader by 2025"*[46] in artificial intelligence and to spur the development of this emerging technology.
Nonetheless, ethical behaviour should always be prioritised. Europe promises that it wants artificial intelligence (AI) systems to be open, trackable, and overseen by humans so that potential abuses can be avoided. They will also "need to be trained on unbiased data to eliminate the possibility of bias."[47] Facial recognition and other forms of biometric authentication are already highly debated topics... However, the National Assembly in France has given the green light to the application of intelligent video surveillance, an automated monitoring system, to secure the Paris Olympics in 2024.

A question arises: A more secure, simpler, and overall, more convenient future regarding our data, but under the condition that we have to let more and more of our personal data be in Europe's hands... Does this pose a risk for our future, especially for our freedoms?

Section 2. The risks

1: Security vs. individual freedom

Benjamin Franklin: "They that give up essential liberty to obtain a little temporary safety deserve neither liberty nor safety."

The debate between security and individual liberty is a highly controversial topic in Europe and around the world. On the one hand, proponents of security argue that governments must take measures to protect their citizens from threats, even if this means restrictions on individual freedoms. On the other hand, proponents of individual freedom argue that governments should not sacrifice individual rights to ensure security, and that this may lead to a violation of human rights. The debate has intensified in Europe since the 2015 Paris terrorist attacks, which led to increased security and surveillance measures.

[46] European Commission "Digital Europe Program", Press Release, June 6, 2018.
[47] European Commission's High-Level Expert Group on Artificial Intelligence (AI HLEG) "Ethics Guidelines for Trustworthy AI", April, 2019

In this part of the dissertation, we can only speculate, as the future is uncertain, despite many signs pointing in the same direction. To have a critical mind on the future of data for everyone, it is essential to refer to philosophical, technological and behavioural analyses. Here is a selection of authors' conferences accompanied by a synthesis of the reflection provided, which can enrich the reflection of our security against our freedom.

- George Orwell novel '1984' includes the criticism of surveillance and manipulation of the individual by the state, as well as the defence of freedom of thought and expression. It describes a totalitarian society in which the government uses surveillance and propaganda to maintain control over the population. This book has inspired many debates about security versus personal freedom, as well as how technology can be used to monitor and control people. The main character, Winston Smith, works for the government and gradually becomes aware of how he is being manipulated and monitored. "Freedom is the freedom to say that two plus two make four. If that is granted, all else follows."[48] The issues surrounding data surveillance are not without meaning. The great technological advances and the incomprehension of most people faced with the increasing technicality of information management systems lead to a mistrust of the practices of large companies. The unknown is frightening, and this is legitimate in a climate where scandals such as the Cambridge Analytica affair are coming to light, denouncing surveillance programs through the collection of information on the Internet. These revelations create a movement of distrust towards political practices. This book introduced the concept of 'Big Brother' describing a totalitarian system governed by the 'thought police', using data from its 'telecoms' (televisions recording every move of people) to predict crimes against the ruling party. Although published in 1949, this novel of anticipation resonates with even more force in the eyes of all and has greatly influenced the link that has been made between Big Data and Big Brother. Today, there are organisations who are trying to counterbalance a system where regulations are still too unresponsive to the rapid evolution of digital technologies. Their means is to denounce the dubious practices of certain organisations. For example, they organise the annual "Big

[48] Orwell, George. 1984. Signet Classic, 1950, p. 84.

Brother Awards", listing organisations that violate users' freedoms by processing their data.

- French philosopher Michel Foucault has examined how power operates in society, particularly through the mechanisms of surveillance and control. In his work, he explored how institutions use surveillance to control the behaviours and attitudes of individuals. “The 'enlightenment', which discovered the liberties also invented the disciplines that put them in parentheses.”[49]

In the case of personal data collection, this can translate into a sense of constant surveillance of our online activities, in our freedom of choice and privacy. Ultimately, Foucault's work reminds us of the importance of protecting our privacy and our right to self-determination. It is a matter of balancing our individual freedoms with our rights as citizens when it comes to the collection of personal data. It is more important than ever to be aware of how our data is used, to protect our privacy and our fundamental freedoms.

- Shoshana Zuboff, in her book The Age of Surveillance Capitalism, has written about the impacts of technology on privacy and individual rights. The American author examines how companies use personal data to create patterns of behaviour and influence consumer choices, implications for their privacy and individual rights. "Surveillance capitalism is not simply a lucrative market. It is a mutation of capitalism, a new form of wealth accumulation in the hands of a few that is taking over society and everyday life."[50]

The collection of massive data can allow companies to make decisions that affect our lives without our consent or knowledge. A company could refuse to offer someone a job based on data collected about their online behaviour, without that person being informed or having the opportunity to challenge the decision. In the future, such analyses could be the source of much discrimination and intrusion into the personal and professional lives of individuals. Zuboff's work underscores the importance of strong regulation of data collection and privacy for a desirable future.

[49] Foucault, Michel. Discipline and Punish: The Birth of the Prison. New York, Vintage Books, 1995, p. 228.

[50] Zuboff, Shoshana, "The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power", New York, NY, 2019, PublicAffairs, 704, p.73.

2: The threat landscape: danger, security

As we understood previously, centralising data implies an immense wealth of information in a single point, combined with the amount of data breaches, cyber-attacks, and insider threats increasing every year, the risk factor is now at an all-time high. We can therefore expect these risks are expected to be more prevalent in 2033 than they are in 2023, which is both undesirable and alarming. Attackers have more information available to them when the Internet of Things is factored in. We're gathering a lot of data from previously not yet digitised devices. For artificial intelligence (AI) start-ups, which are blooming as this dissertation is written. Data is required to train machine learning algorithms and is often the key differentiator between competitors. One of the risks associated with the use of personal data in machine learning algorithms is the possibility of re-identification attacks, which refers to a situation where anonymized data can be re-identified and linked back to an individual, revealing their personal information. As machine learning algorithms become more advanced, the risk of re-identification attacks also increases.

The unknown is frightening, and this is legitimate in a climate where scandals such as the Cambridge Analytica affair are coming to light, denouncing the misuse of personal data for political purposes. These revelations create a movement of mistrust towards companies.

What kinds of threats should we be concerned about? A rise in supply chain attacks targeting the weaker links such as third-party vendors, to gain access to a more secure network is a good start, but more importantly, state-sponsored attacks (by a government or a state agency for political or military purposes.). However, as more systems become accessible and brought online, the same types of attacks will develop because of the same errors made both in the present and in the past. This means that even as attack types evolve, the fundamental mistakes that make these attacks possible will remain the same. This highlights the importance of improving the security of data systems to prevent both present and future attacks.

When I questioned (CONFIDENTIAL) about the kinds of challenges that need to be addressed to ensure that centralization of personal data is done in a responsible and ethical way: he predicts that individual agreement with transparency, worldwide harmonised regulations with proper controls and limitation of external threats are the answer.

What do we think people's reaction will be to this amount of data, to this new reality? Will they close themselves? We can even go behind cognitive analysis and we can envision a reality where everything is recorded and with the interaction we may have, monitoring us with cameras installed in the streets, they’ll build patterns and know exactly how we feel, how to motivate us and how to make us do what they want to decide… the possibilities are endless.

Conclusion:

The lessons learned shed new light on the technologies we use every day, which have demonstrated a real change in the way digital data is apprehended, particularly from an ethical point of view. However, the control of data by companies is not without risk for the general interest. The digital traces left by users are richer than ever in information and the deduction capacities of our most recent algorithms leave little room for mystery. This ambivalence leads to the growth of jobs that combine marketing thinking and technical knowledge in computer language. As we have seen, in the data era, Data Scientists are kings: when companies manage to integrate these changes, they gain an undeniable competitive advantage.

We have also observed that the European Union has adopted a proactive approach to protect the fundamental rights of individuals. Most companies have complied with strict data security and privacy standards. However, despite these advances, challenges remain to ensure full protection of personal data. New data collection technologies raise new ethical and legal issues. It is therefore essential that regulators continue to monitor and update data protection laws and standards to respond to changing technologies and business practices.

The goal of this dissertation is for it to still be relevant 15-20 years from now, and to provide perspective on the times we live in, as opposed to those to come. As individuals, we have the power to shape the future by how we use and process data. Being well-prepared to make good decisions with data, rather than using it in harmful or dangerous ways, should be key in the years to come. Despite technological advances and the growing importance of data in various fields, humans will always remain in control of how data is used and should take responsibility for ensuring its ethical and responsible use through education and ongoing monitoring of information.

Bibliography and webography

1. [1] Gary Kovacs, former CEO of Mozilla Corporation, during his keynote speech at the TED Global 2012 conference in Edinburgh, Scotland.
2. [2] The Guardian, "Data is the new oil", March 6, 2006.
3. [3] Domo “Data Never Sleeps 10.0”, 2022, https://www.domo.com/data-never-sleeps
4. [4] Statista, “Data Growth Statistics - Size of the digital universe worldwide from 2010 to 2025", 2021.
5. [5] Forbes, "Big Data Statistics - How Much Data is Out There?", February 26, 2021.
6. [6] Gartner, "Big Data: A Three-Dimensional View.", 2012
7. [7] RFID Journal, "That 'Internet of Things' Thing" by Kevin Ashton, June 22, 2009.
8. [8] TechJury, “How Many IoT Devices Are There in 2023?”, April 19, 2023 https://techjury.net/blog/how-many-iot-devices-are-there/
9. [9] Socialbakers, "Q1 2021 Social Media Trends", 2021
10. [10] Ghostery, "Tracking the Trackers 2020: Web Tracking's Opaque Business Model of Selling Users", 2020,
11. [11] Acxiom, "About Us", accessed March 8, 2023, https://www.acxiom.com/about-us/
12. [12] Han, J., Kamber, M., & Pei, J. *Data Mining: Concepts and Techniques (3rd ed.).* Amsterdam, 2011, Elsevier. 744 p. p. 19.
13. [13] We Are Social and Hootsuite, "Digital 2023: Global Overview Report," 26, January 2023, https://datareportal.com/reports/digital-2021-global-overview-report
14. [14] Garcia, A., Google and Facebook are creating a dystopian nightmare, 2017, Retrieved from https://qz.com/1034972/google-and-facebook-are-creating-a-dystopian-nightmare/
15. [15] Forbes, "How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did", 2012
16. [16] Clubic “Google n’espionnera plus Gmail pour cibler sa pub”, 27 June, 2017, https://www.clubic.com/messagerie-email/gmail/actualite-832480-google-espionnera-gmail-cibler-pub.html
17. [17] Pam Dixon Executive Director, World Privacy Forum, Before the Senate Committee on Commerce, Science, and Transportation, 18, December, 2013
18. [18] Frederico Zannier, "A bit(e) of me", KickStarter, October, 2013, https://www.kickstarter.com/projects/1461902402/a-bit-e-of-me/faqs?lang=en
19. [19] The Guardian, "The Cambridge Analytica Files", 2018, https://www.theguardian.com/news/series/cambridge-analytica-files.
20. [20] The Guardian, "Cambridge Analytica scandal: what you need to know", 21 March 2018,

21. [21] Weston, Alan F., *Privacy and Freedom*, 1967, Atheneum, 322 p.
22. [22] Weston, Alan F., *Privacy and Freedom*, 1967, Atheneum, 322 p., p. 7.
23. [23] Safire, William. “Nosy Parker Lives,” New York Times, Sept. 23, 1999, at A29.
24. [24] Addictivetips.com, “Our Attitudes Towards Privacy Policies”, 2018, https://irishtechnews.ie/our-attitudes-towards-privacy-policies/
25. [25] Carnegie Mellon University CyLab, “Usable Privacy and Security: A Guide to Relevance and Research,” 2008, p. 5.
26. [26] Federal Trade Commission Chairman (F.T.C) John Liebowitz, 2009.
27. [27] RiskBased Security. “2021 Mid Year Report Data Breach QuickView Report,” 2022, https://pages.riskbasedsecurity.com/download-the-2021-mid-year-data-breach-quickview-report-today
28. [28] CNIL (Commission nationale de l'informatique et des libertés), "Baromètre annuel de la protection des données personnelles", 2018,
29. [29] EU GDPR, Article 33, "Notification of a personal data breach to the supervisory authority", https://www.privacy-regulation.eu/en/33.htm
30. [30] DLA Piper, “GDPR fines and data breach survey,” January 2023, https://www.dlapiper.com/en-ae/insights/publications/2023/01/dla-piper-gdpr-fines-and-data-breach-survey-january-2023/
31. [31] TechCrunch, “Equifax breach was ‘entirely preventable’ had it used basic security measures, says House report", 10, December, 2018, https://techcrunch.com/2018/12/10/equifax-breach-preventable-house-oversight-report/?guccounter=1
32. [32] BBC News. "Instagram owner Meta fined €405m over children's data privacy". September 15, 2022. https://www.bbc.com/news/technology-62800884
33. [33] https://www.enforcementtracker.com/
34. [34] European Union, "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)", 2016, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679.
35. [35] Europa, Your Europe, “Providing transparent information”, 2022, https://europa.eu/youreurope/business/dealing-with-customers/data-protection/data-protection-gdpr/index_en.htm#shortcut-11
36. [36] European Commission, “Privacy policy for websites managed by the European Commission”, https://commission.europa.eu/privacy-policy-websites-managed-european-commission_fr
37. [37]Github, https://github.com/

38. [38]Instagram, https://www.instagram.com/
39. [39]Instagram, “Privacy Policy, What is the Privacy Policy and what does it cover?” https://privacycenter.instagram.com/policy/?annotations[0]=1.ex.37-OurPartnersBusinessesAnd
40. [40] https://www.facebook.com/about/privacy/previous
41. [41] European Commission, “A Europe fit for the digital age, Empowering people with a new generation of technologies”, 2020, https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age_en
42. [42] European Commission, “Digital Identity for all Europeans, a personal digital wallet for EU citizens and residents”, June, 3, 2021, https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-digital-identity_en
43. [43] Onfido, “Customer attitudes to digital identity”, November, 2020 https://onfido.com/report/customer-attitudes/
44. [44] European Commission, "Digital Economy and Society Index (DESI) 2020", June, 2020, https://eufordigital.eu/library/digital-economy-and-society-index-desi-2020/
45. [45] European Parliament Plenary, Ursula von der Leyen, "State of the Union Address”, https://state-of-the-union.ec.europa.eu/state-union-2020_fr
46. [46] European Commission “Digital Europe Program”, Press Release, June 6, 2018.
47. [47] European Commission's High-Level Expert Group on Artificial Intelligence (AI HLEG) “Ethics Guidelines for Trustworthy AI”, April, 2019
48. [48] Orwell, George. 1984. Signet Classic, 1950, p. 84.
49. [49] Foucault, Michel. Discipline and Punish: The Birth of the Prison. New York, Vintage Books, 1995, p. 228.
50. [50] Zuboff, Shoshana, "The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power", New York, NY, 2019, PublicAffairs, 704, p.73.