Big Data and Privacy

Alexandra Yakimova

Bachelor's thesis
2023

Tomas Bata University in Zlín
Faculty of Applied Informatics

Univerzita Tomáše Bati ve Zlíně
Fakulta aplikované informatiky
Ústav počítačových a komunikačních systémů

Akademický rok: 2022/2023

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení:	**Alexandra Yakimova**
Osobní číslo:	**A20358**
Studijní program:	**B0688A140008 Informační technologie v administrativě**
Forma studia:	**Prezenční**
Téma práce:	**Big data a soukromí**
Téma práce anglicky:	**Big Data and Privacy**

Zásady pro vypracování

1. Proveďte literární rešerši na téma Big data.
2. Popište výhody a nevýhody využití Big data na Internetu.
3. Zaměřte se na legislativu spojenou s uživatelským soukromím na Internetu.
4. Rozeberte jakým způsobem s uživatelskými daty pracují vybrané internetové společnosti.
5. Zpracujte uživatelská doporučení k ochraně osobních dat.

Forma zpracování bakalářské práce: **tištěná/elektronická**

Seznam doporučené literatury:

1. TORRA, Vincenc. Data Privacy: Foundations, New Developments and the Big Data Challenge. 00028. 2017. ISBN 9783319573564. Dostupné také z: https://search.ebscohost.com/login.aspx?direct=true&db=edsebk&an=1520240&scope=site
2. GUERRIER, Claudine. Security and privacy in the digital era. Volume 1. London: ISTE, 2016, 1 online zdroj (xxvii, 252 stran). Innovation and Technology set. ISBN 9781119347750. Dostupné také z: https://proxy.k.utb.cz/login?url=https://onlinelibrary.wiley.com/doi/book/10.1002/9781119347750
3. MATEJKA, Ján. Internet jako objekt práva: hledání rovnováhy autonomie a soukromí. Praha: CZ.NIC, z.s.p.o., 2013, 1 online zdroj (262 stran). CZ.NIC. ISBN 978-80-904248-7-6. Dostupné také z: https://knihy.nic.cz/files/edice/internet_jako_objekt_prava.pdf
4. YARALI, Abdulrahman, Randall JOYCE a Brandon DIXON. Ethics of Big Data: Privacy, Security and Trust. In: 2020 Wireless Telecommunications Symposium (WTS) [online]. IEEE, 2020, 2020, s. 1-7 [cit. 2022-11-30]. ISBN 978-1-7281--4695-9. Dostupné z: doi:10.1109/WTS48268.2020.9198734
5. GOEL, Parth, Radhika PATEL, Dweepna GARG a Amit GANATRA. A Review on Big Data: Privacy and Security Challenges. In: 2021 3rd International Conference on Signal Processing and Communication (ICPSC) [online]. IEEE, 2021, 2021-5--13, s. 705-709 [cit. 2022-11-30]. ISBN 978-1-6654-2864-4. Dostupné z: doi:10.1109/ICSPC51351.2021.9451749.

Vedoucí bakalářské práce: **doc. Ing. Jiří Vojtěšek, Ph.D.**
Ústav řízení procesů

Datum zadání bakalářské práce: **2. prosince 2022**
Termín odevzdání bakalářské práce: **24. května 2023**

UNIVERZITA TOMÁŠE BATI VE ZLÍNĚ
Fakulta aplikované informatiky
L.S.

doc. Ing. Jiří Vojtěšek, Ph.D. v.r.
děkan

doc. Ing. Petr Šilhavý, Ph.D. v.r.
garant oboru

Ve Zlíně dne 8. prosince 2022

I hereby declare that:

- I understand that by submitting my Bachelor´s Thesis, I agree to the publication of my work according to Law No. 111/1998, Coll., On Universities and on changes and amendments to other acts (e.g. the Universities Act), as amended by subsequent legislation, without regard to the results of the defence of the thesis.
- I understand that my Bachelor´s Thesis will be stored electronically in the university information system and be made available for on-site inspection, and that a copy of the Bachelor´s Thesis will be stored in the Reference Library of the Faculty of Applied Informatics, Tomas Bata University in Zlín, and that a copy shall be deposited with my Supervisor.
- I am aware of the fact that my Bachelor´s Thesis is fully covered by Act No. 121/2000 Coll. On Copyright, and Rights Related to Copyright, as amended by some other laws (e.g. the Copyright Act), as amended by subsequent legislation; and especially, by §35, Para. 3.
- I understand that, according to §60, Para. 1 of the Copyright Act, TBU in Zlín has the right to conclude licensing agreements relating to the use of scholastic work within the full extent of §12, Para. 4, of the Copyright Act.
- I understand that, according to §60, Para. 2, and Para. 3, of the Copyright Act, I may use my work - Bachelor´s Thesis, or grant a license for its use, only if permitted by the licensing agreement concluded between myself and Tomas Bata University in Zlín with a view to the fact that Tomas Bata University in Zlín must be compensated for any reasonable contribution to covering such expenses/costs as invested by them in the creation of the thesis (up until the full actual amount) shall also be a subject of this licensing agreement.
- I understand that, should the elaboration of the Bachelor´s Thesis include the use of software provided by Tomas Bata University in Zlín or other such entities strictly for study and research purposes (i.e. only for non-commercial use), the results of my Bachelor´s Thesis cannot be used for commercial purposes.
- I understand that, if the output of my Bachelor´s Thesis is any software product(s), this/these shall equally be considered as part of the thesis, as well as any source codes, or files from which the project is composed. Not submitting any part of this/these component(s) may be a reason for the non-defence of my thesis.

I herewith declare that:

- I have worked on my thesis alone and duly cited any literature I have used. In the case of the publication of the results of my thesis, I shall be listed as co-author.
- That the submitted version of the thesis and its electronic version uploaded to IS/STAG are both identical.

In Zlín; dated: 23.05.2023

Alexandra Yakimova
Student´s Signature

ABSTRAKT

Tato práce zkoumá vztah mezi Big data a obavami o soukromí, analyzuje jejich vnímaná rizika a výhody a také současný stav legislativy na ochranu soukromí. S tím, jak se svět stále více spoléhá na technologie, objem denně produkovaných dat rychle roste. Zvýšené shromažďování, ukládání a analýza osobních dat však vede k rostoucím obavám o soukromí. S využitím dostupných statistických analýz tato rešerše zkoumá úrovně důvěry zákazníků ve schopnost korporací chránit svá data a zdůrazňuje, jak důležité je tyto obavy řešit.

Praktický aspekt tohoto výzkumu si klade za cíl zpracovat uživatelská doporučení k ochraně soukromí uživatelů internetu a porovnat přístupy vybraných technologických společností k práci s daty.

Klíčová slova: Big data, soukromí, osobní údaje, sběr dat, ochrana spotřebitelských dat.

ABSTRACT

This paper explores the relationship between big data and privacy concerns, analyzing its perceived risks and advantages, as well as the present state of privacy legislation. With the world becoming more reliant on technology, the volume of data produced daily is rapidly increasing. However, increased collection, storage, and analysis of personal data have led to growing concerns about privacy. Utilizing available statistical analysis examines the trust levels of customers in the ability of corporate entities to protect their data, highlighting how important it is to address these concerns.

The practical aspect of this research aims to compile measures that protect user privacy online and compare the data handling practices of selected tech companies.

Keywords: Big data, privacy, data privacy laws, data collection, data breaches.

ACKNOWLEDGEMENTS

Acknowledgements, motto and a declaration of honour saying that the print version of the Bachelor's thesis and the electronic version of the thesis deposited in the IS/STAG system are identical, worded as follows:

I hereby declare that the print version of my Bachelor's thesis and the electronic version of my thesis deposited in the IS/STAG system are identical.

CONTENTS

INTRODUCTION 7

I THEORY 8

1 WHAT IS BIG DATA? 9

1.1 CHARACTERISTICS OF BIG DATA 9

1.2 HISTORY OF BIG DATA 10

1.3 BIG DATA CLASSIFICATION 11

1.4 BIG DATA USE CASES 12

2 WHAT ARE THE BENEFITS OF BIG DATA? 15

3 WHAT ARE THE RISKS OF BIG DATA? 18

3.1 SECURITY ISSUES 18

3.2 ETHICAL ISSUES 19

3.3 MISUSE 20

3.4 BIGGEST DATA BREACHES 20

4 THE EVOLUTION OF ONLINE CONSUMER PRIVACY 23

5 DATA PRIVACY LAWS 24

5.1 U.S. DATA PRIVACY LAWS 24

5.2 EUROPE PRIVACY LAWS 26

5.3 EU VS. US PRIVACY LAWS COMPARISON 26

5.4 PRIVACY LAWS ON AN INTERNATIONAL SCOPE 29

6 PUBLIC AWARENESS 32

7 PROACTIVE STEPS FOR COMPANIES 36

II RESEARCH 37

8 TECH GIANTS AND DATA HANDLING 38

8.1 INFORMATION THE COMPANIES COLLECT 38

8.2 DATA PROCESSING AND LEGAL CASES 42

9 DIGITAL PRIVACY PROTECTION MEASURES 45

9.1 VIRTUAL PRIVATE NETWORKS 46

9.2 SECURE EMAIL PROVIDERS 49

9.3 PRIVACY-ORIENTED BROWSERS 50

9.4 BROWSER PRIVACY SETTINGS 54

9.4.1 Google Chrome 54

9.4.2 Safari 57

9.4.3 Firefox 59

9.5 RECOMMENDATIONS 61

CONCLUSION 63

BIBLIOGRAPHY 64

LIST OF FIGURES 72

LIST OF TABLES 73

INTRODUCTION

In recent times, the world has experienced substantial changes, particularly within the domain of Information Technology. This has enabled the widespread availability of extensive amounts of information to the general populace while simultaneously permitting corporations to collect substantial quantities of data on individuals. The notion of Big Data has emerged as a subject of discussion, as it offers benefits while also eliciting concerns relating to personal privacy.

Every day, we produce a considerable amount of personal data regarding our activities, social connections, preferences, and transactions. This data possesses substantial worth for commercial entities as it allows them insight into our distinctive requirements and desires to adapt their advertising and promotional efforts accordingly. As a result, this helps businesses to increase their sales.

When amassing a copious amount of data, there is a significant probability that it may encompass confidential and personal details. This situation can occur even without the involvement of cyber criminals or thieves, as anyone with malicious intentions could use this sensitive information. This includes various harmful entities, such as unethical corporations and institutions.

The purpose of this paper is to explore the advantages and disadvantages of Big Data while also delving into the historical context of online consumer privacy and data privacy legislation implemented globally. By understanding these aspects, businesses can find a balance between protecting customer privacy and achieving financial success.

In addition to that, I will also conduct research on how big companies compare at handling user data utilizing publicly accessible resources and provide recommendations for individuals to protect their privacy in the digital realm.

I. THEORY

1 WHAT IS BIG DATA?

Big data is often defined as any data that is large in volume, has a wide variety, and arrives with a fast rate velocity. These attributes are commonly referred to as the three V's. [1]

Big data is a term used to describe large, complex data sets that businesses commonly deal with. These data sets are often collected from new data sources and are so large that it can be difficult to manage and analyze them using traditional business intelligence tools. As a result, big data often requires special data processing software to be effectively managed. [1]

1.1 Characteristics of Big Data

As previously mentioned, the three main dimensions of big data are volume, variety, and velocity. Here are some more specific descriptions of those characteristics.

Volume. As big data becomes increasingly prevalent, there is a growing need for efficient processing of large volumes of unstructured data. This data can come from a variety of sources, ranging from telemetry data to business customer data collection. It is likely that even more data collections will be generated in the future. [1]

The current issue with data, however, is not storage but rather identifying relevant data and using it effectively. [2]

Velocity. The velocity of a data stream is the measure of how fast data is coming in. High-velocity data streams directly into memory rather than being written to a disk. [1]

An increasing number of devices are able to transmit data at high speeds, creating a large and fast-moving flow of information. This flow is known as the velocity vector. One of the challenges in data analytics is finding efficient ways to collect, process, and utilize large data sets as they are generated. [1], [2]

Variety. Variety refers to the vast number of existing data types. Big data can vary significantly depending on the source, and a lot of it is unstructured or semistructured data, which can be more difficult to process and extract meaning from than traditional, structured data. [1]

Although those three points are the widely accepted core of the definition of big data, in recent years, another two characterization items became important enough to be noted: veracity, value, and volatility. [3], [4]

Veracity. The quality of a data set is determined by how accurate it is in reflecting reality. Data that is more precise and accurate is more reliable and, therefore, more valuable. When determining the accuracy of data, it is often helpful to examine the source. However, when combining sources of data with the purpose of increasing variety, it may be difficult to track the interaction across different sets of data. [3]

The veracity of big data is judged on a scale from high to low. Data on the high end of the scale is more suitable for further processing, while data on the low end of the scale is less reliable and contains a higher percentage of inaccurate information. [5]

If the data used for analysis is inaccurate, the resulting conclusions will be invalid. In a world where automation is increasingly relied upon to make decisions, it is essential that the data used to inform those decisions is accurate. [2]

It is important to keep in mind that veracity and interpretability are not the same thing. Just because data is accurate does not mean it cannot be misinterpreted and lead to false conclusions. The interpretability of data depends on the analytic methods used, not on the data set itself. [3]

Value. Data value is typically defined by the potential economic and social value it might create. However, this is only a general concept, as high-value data can be left unused and not generate any value if it is not processed or applied. [3]

Volatility. Data volatility is a common issue in the field of big data. Changes in data can lead to inaccurate results, especially in businesses with rapidly changing data, such as the stock markets. Organizations need to be able to retrieve information rapidly to avoid issues of data invalidity and low veracity. [4]

1.2 History of Big Data

The concept of big data has existed since the 1970s, albeit the terminology used to describe it is comparatively new. The evolution of big data can be classified into three key phases when examining its development over time [6]:

Phase 1 (1970-2000). The origins of data analytics and Big Data can be traced back to the development of database management. The relational database, data warehousing, data mining, and statistical analysis are the foundations of modern data analysis.

Phase 2 (2000-2010). As the internet and web become increasingly rich with data, the need for improved data analysis techniques has become more pressing. Companies who wish to understand customer behavior need to be able to store and process large volumes of data, which is where Hadoop and Spark frameworks come in. These frameworks have made it possible to deal with big data more effectively, contributing to its growth.

Phase 3 (2010 – present day). The potential for big data analytics is growing as organizations strive to better understand their customers and optimize their products.

The generation of mobile data, location-based analytics, and person-centric analysis are all made possible by sensor-based devices. Mobile and sensor-based internet-enabled devices are playing an increasingly important role in this process by collecting data that can be used to track user behavior and make predictions about future trends.

1.3 Big Data Classification

There are various ways of describing big data, such as its source, content, and how it is stored and organized. The classification of big data can be illustrated in the hierarchical structure as shown in Figure 1. [4]

There are several ways to classify Big Data, including by its source. Data can come from a variety of sources, such as web and social media, machine-generated data, or human-generated data. Knowing the range of data sources is important for assessing the usefulness of the data from a business perspective. [7]

By format, data can be classified as either structured or unstructured. Structured data is usually human-readable and can be indexed, while unstructured data is more chaotic and difficult to process. An example of structured data is a database, while examples of unstructured data include source code, videos, images, and text documents. [7]

It is usually best to store different types of data in different databases rather than relying on a single data store. Although a single database system can often support multiple storage models, storage models can be broadly categorized into key/value storage, document databases, graph databases, and relational database management systems (row/column-oriented storage). [8]

Depending on which stage of processing the data is, it can be cleaned, transformed, and normalized. Data cleaning is a process of identifying and removing errors, inaccuracies, and inconsistencies from data. Data transformation is a process of converting data from one format to another, typically in order to make it more readable or accessible. Different data attributes often have different scales, so data normalization is needed in order to map the data onto a common scale. [9]

Big Data can be classified according to the type of analysis that is performed on it, either in real-time or as a batch process. In some cases, a combination of both types of analysis may be necessary. [7]

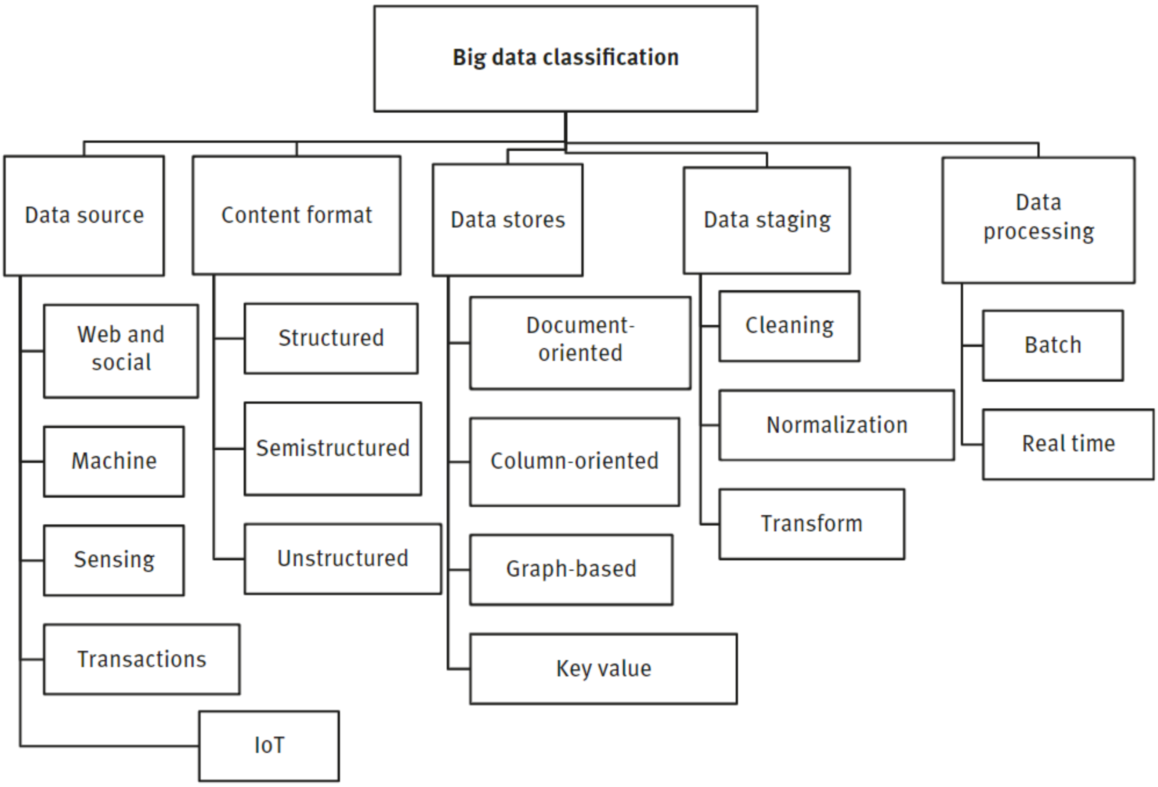


Figure 1 Big data classification [4]

1.4 Big data use cases

The use of big data is becoming increasingly widespread across a variety of industries and sectors. Its use cases are diverse, ranging from improving customer experience to enhancing product development, optimizing supply chain management, and detecting fraud.

The development of new products at large companies relies heavily on data analytics in order to anticipate customer demand. By utilizing predictive models, they are able to get a good idea of how successful a new product might be, based on customer reaction to similar products. [1]

Organizations can save money on maintenance and maximize equipment uptime by analyzing both structured and unstructured data. This data analysis can help predict mechanical failures and take preventive measures. [1]

The use of big data allows companies to track all interactions with customers, from social media posts to website visits to phone calls. This data can help them improve their service delivery and keep customers satisfied. [1]

The challenge of security requires experts to work together to stay ahead of the constantly changing landscape. Big data can help identify patterns of fraud and make it easier to comply with regulatory requirements. [1]

As we move towards more highly concentrated forms of urban living, we face significant challenges in how we manage resources, dispose of waste, and address inequality. However, arguably the most pressing issue is transportation management. Big data could be extremely helpful in improving transportation planning, although it is often not used effectively. The problem with extracting and processing transportation big data is that it is complex and privacy-sensitive. Most cities lack the infrastructure to track traffic flows effectively and capturing and mining video data requires a robust platform that can handle different formats. Privacy concerns are a major obstacle to the wider adoption of big data in transportation, but both issues are addressable, and the potential positive changes that the wider adoption of big data in transportation could create are significant. [10]

Machine learning is highly sought-after at present due in large part to the abundance of big data. With so much data available, we can train machines instead of programming them, making machine learning more versatile and efficient. [10]

Operational efficiency is often overlooked, but it is an increasingly important area for big data. By analyzing production, customer feedback, and other factors, big data can help reduce outages and anticipate future demands. This information can also be used to improve business decision-making in line with current market conditions. [10]

Building automation systems can generate a lot of data that can be used to improve energy efficiency. Analyzing past and present energy usage data, in combination with data on different building components, can help improve building design for better energy efficiency. [11]

The clinical research industry is finding that big data is helpful in making better decisions. Big data allows researchers to see patterns and trends that could potentially lead to new ways of treating illnesses. Additionally, big data is assisting researchers in developing new drugs and therapies. By looking at data on patient characteristics and genetic information, for example, researchers can identify potential targets for new drugs. Big data can also help researchers verify that their findings are correct and reliable. [12]

Behavioral Data Science is a field of study that uses a combination of big data and behavioral insights to improve decision-making and solve problems in various areas such as healthcare, finance, and public policy. The application of behavioral economic principles to big data can provide a deeper understanding of consumer behavior and decision-making. [13]

Big data can be used to improve decision-making by studying the relationships between humans, institutions, entities, and processes. By understanding these relationships, businesses can gain insights into what customers want and how to price products and services to stay competitive in the market.

Overall, big data is playing an increasingly critical role in today's digital economy, transforming the way businesses operate and deliver value to customers.

2 WHAT ARE THE BENEFITS OF BIG DATA?

There are many benefits to big data, which can impact businesses, industries, and society in a positive way. In this chapter, we will explore some of the key advantages that big data provides and how it can be used to drive innovation and growth.

Data regarding customers can originate from numerous different sources. Some of these sources are extensive data sets that can unveil observations about customers on an individual or group level. Other sources of customer data comprise purchase and support records, financial transactions and credit reports, social media activity, and internal and external surveys. Clickstream analysis of e-commerce activity can also be beneficial in comprehending how customers navigate a company's website to locate products and services. [14]

The analysis of online customer behavior can provide valuable insights for businesses, helping them to better tailor their products and services. Similarly, analyzing video footage of customers in physical stores can also give businesses useful information about customer preferences and behavior. [14]

The application of big data within businesses can help to improve understanding of patterns of customer behavior, as well as market dynamics. Social media provides a rich source of market intelligence, which can be used to inform decisions about marketing campaigns. Big data can help businesses to learn about customer preferences and experiences with products, giving them a competitive edge. [14]

In today's interconnected world, disruptions to supply chains can have widespread consequences. By integrating data from different sources, big data systems can help predict disruptions and choose appropriate action. [14]

Customer intelligence and real-time pricing can benefit not only large businesses but also small e-commerce businesses. By providing insights into stock levels, risk reduction, and staffing needs, these tools can help optimize business decisions and improve overall efficiency. [14]

With big data, we have come to rely on recommendation engines more and more to help us make choices. These engines used to rely on simple associations between products and customers to provide suggestions. They are still present on some e-commerce websites, suggesting customers who have already bought certain items might also be interested in buying related items. [14]

The newer generation of recommendation systems is more intelligent and takes into account a variety of factors, such as customer demographics and behavior. These systems are not just limited to e-commerce but can also be seen in other industries, such as hospitality. For example, a waiter's recommendations may be based on data from a point-of-sale system that takes into account popular menu items, combos, and trends. When customers share pictures of their meals on social media, this provides even more data for these systems to analyze. [14]

Streaming content providers use more sophisticated techniques to keep viewers engaged. They may use a combination of customer preferences and big data analysis to determine what to play next, even before the current selection is finished. This helps to keep viewers binge-watching by always having something new and interesting offered to them. [14]

Innovation is not only driven by original ideas but also by hard work to identify areas where new efforts and experiments can be successful. Data and technology can be used to improve research and development. In some cases, data – once it has been cleaned, prepared, and governed for sharing – can become a product. The London Stock Exchange, for example, now generates more revenue from selling data and analysis than from securities trading. [14]

Data is not enough to produce new insights by itself, but when data is combined with human imagination, it can lead to discoveries. Having data stored in one place (like a Hadoop cluster or cloud data lake) can help people see trends that would be difficult to spot otherwise. This can be beneficial for data scientists and BI analysts. [14]

In the era of big data, it is possible to store all of the raw data in a data lake and only apply data models when they are needed for particular analytics applications. This enables a high degree of flexibility in the number and types of applications that can be run against the same data set. [14]

The potential benefits of big data analytics for businesses are numerous and significant, ranging from improved process efficiency to reduced costs and increased productivity. By tracking and analyzing data on employee performance, shipments, and machine performance, businesses can optimize their delivery routes, better understand their staff, improve the effectiveness of hiring and employee management, and more effectively manage risks. [15]

Big data analytics systems can also play a role in fraud prevention by quickly identifying anomalous patterns that may indicate fraudulent activity. [14]

The application of big data analytics to physical operations can result in considerable improvements. For example, big data can be used to create predictive maintenance schedules that keep equipment running smoothly and avoid costly downtime. Big data can help to integrate all relevant factors, such as age, condition, location, and warranty details, and optimize equipment maintenance. [14]

According to a survey by NewVantage Partners, the use of tools like Hadoop and Spark has led to an increase in productivity for 59.9% of businesses. These tools are able to accurately and quickly analyze data, which in turn leads to increased productivity at both the individual and organizational levels. [16]

The use of big data by various government departments, such as the police and fire department, has allowed for the development and implementation of new policies and procedures for public safety. By tracking incidents in real-time, law enforcement is sometimes able to prevent crime by reacting to incidents before they occur. This proactive approach can even be used to counter terrorist threats and improve the personal safety of public figures. [16]

3 WHAT ARE THE RISKS OF BIG DATA?

Data accumulation can improve customer care by providing more detailed information about customer behavior. However, if data security is not handled properly, the enormous amount of data can pose privacy hazards. Organizations must take active measures to ensure the safety of their data against potential threats from outside sources. Implementing more efficient data management techniques can save space and reduce costs associated with data storage.

The increased availability of cloud-based storage has made it easier for businesses to collect and access data, but it also poses some risks to privacy and security. This is because the security measures that are in place might not be able to effectively handle dynamic data, and routine safety checks might not be able to pinpoint all potential concerns. To help reduce these risks, it is important to take measures to protect privacy and security full-time when streaming data in cases where sensitive information might be transmitted.

When transmitting sensitive data over a network, it is important to monitor transaction logs closely to ensure that the information is being transmitted securely. If transaction levels vary, however, it may be difficult to track how safely the transaction is being conducted. The handling of all storage devices may be made automated for convenience as the quantity of information grows, but this also makes the data more vulnerable to threats since it is harder to keep track of where it is stored.

Organizations working with large databases should prioritize security precautions when collecting and storing information and implement robust security measures to protect that information from threats and vulnerabilities. Threats and weaknesses can develop over time, and past security measures may not be enough to protect against them. Therefore, organizations must regularly evaluate and update their security measures to stay ahead of potential threats.

3.1 Security issues

A third-party data breach occurs when an unauthorized party accesses sensitive information or systems by compromising a vendor or other organization the victim conducts business with. Third-party data breaches are becoming more common as businesses become interconnected and supply chains grow more complex. Organizations may not be aware of where their data goes and how it is being used, making it vulnerable to being shared with unauthorized parties. The large volume, variety, and variability of Big Data can increase the risk of security and privacy issues. According to the Risk-Based Security Mid-Year Data Breach report [17], 4.1 billion records were exposed through data breaches in the first half of 2019. Data security is, therefore, a critical concern for organizations, but one that poses significant challenges in terms of cost and practicality. [18], [19]

As businesses increasingly adopt cloud data storage to speed up their operations, they face increased security risks. To minimize these risks, tech companies have begun implementing both offline and cloud data storage strategies. Less sensitive data is kept in the cloud for easy access, while more sensitive data is stored on local devices. [19]

Access control is an essential part of protecting data integrity and privacy, but it can be challenging to manage, especially in organizations with a large number of employees. [19]

There is an ongoing debate about whether organizations should be storing vast amounts of data, particularly sensitive information such as personal bank details. Organizations such as governments, social media giants, insurance companies, and healthcare providers have access to large amounts of data. Although these organizations are subject to data protection laws, the increasing number of data breaches in recent years suggests that more needs to be done to protect data. [19]

3.2 Ethical issues

The use of Big Data technologies can have a significant impact on privacy, including but not limited to profiling, data discrimination, and automated decision-making. [20]

It is understood that businesses need to be profitable, and this is typically achieved through interactions with the public. Organizations typically expect transparency from their clients and prospects, but it would also be beneficial if organizations using Big Data were transparent about their procedures and made sure that these procedures are easily accessible and well-known to the public. Taking an ethical perspective would drive innovation and boundary setting, considering the individual’s need for privacy and dignity. [20]

There is some ambiguity about how data can be used by companies who have obtained it legally, despite the existence of data protection laws. The concept of the Right to Be Forgotten has emerged in response to the ethical concerns surrounding data consent and privacy. Some jurisdictions, including Argentina [21], the European Union [22], and the Philippines

[23], have enacted new laws to protect people's privacy by establishing the right to have their personal data removed from public records. [19]

3.3 Misuse

The risks associated with data theft are a major concern that has not been adequately addressed. The misuse of sensitive information by third parties is a possible danger of big data. Hackers could use this data to exploit it in ways that could lead to serious criminal activity, such as phishing, bank fraud, or insurance scams. [19]

The spread of misinformation and fake news has also been linked to big data. Organizations can use data to target ads or fake news that aim to influence people's beliefs and voting behavior. The success of fake news is often due to its well-targeted and customized content, which can be gleaned from data. [19]

There is also still a lack of clarity about how certain problems will develop over time in data science, so it is important to explore potential issues now before the technology becomes more widespread. [19]

Machine learning can assist in making sense of large data sets, which is vital for mitigating potential problems. However, it introduces risks that can occur inadvertently. Machine learning algorithms must be created by humans in the first place, which can introduce human bias and lead to inaccurate insights. This can have negative consequences if those insights are used to make critical decisions. [19]

3.4 Biggest data breaches

As digital transformation progresses, businesses are becoming increasingly vulnerable to large-scale cyberattacks. These attacks can have a devastating impact on millions or even billions of people, depending on their size. It is still unclear whether future attacks will be even larger in scale, but businesses need to be prepared for the possibility. [24]

The largest data breaches have typically resulted from hacking attacks, although there have been notable exceptions. For example, one of the earliest reported data breaches, which impacted AOL and compromised 92 million records, was reported as an inside job. An Apple breach in 2011 resulted from the accidental publishing of sensitive data. A data breach affecting the UK Revenue and Customs in 2006 is also noteworthy, as 25 million records were compromised as a result of lost or stolen media. [25]

Many companies choose to invest in data breach insurance as it provides protection against unauthorized access or exposure of confidential data.

This graph from Statista (Figure 2) demonstrates the number of data records exposed worldwide, providing a better understanding of the scale of the problem. [26]

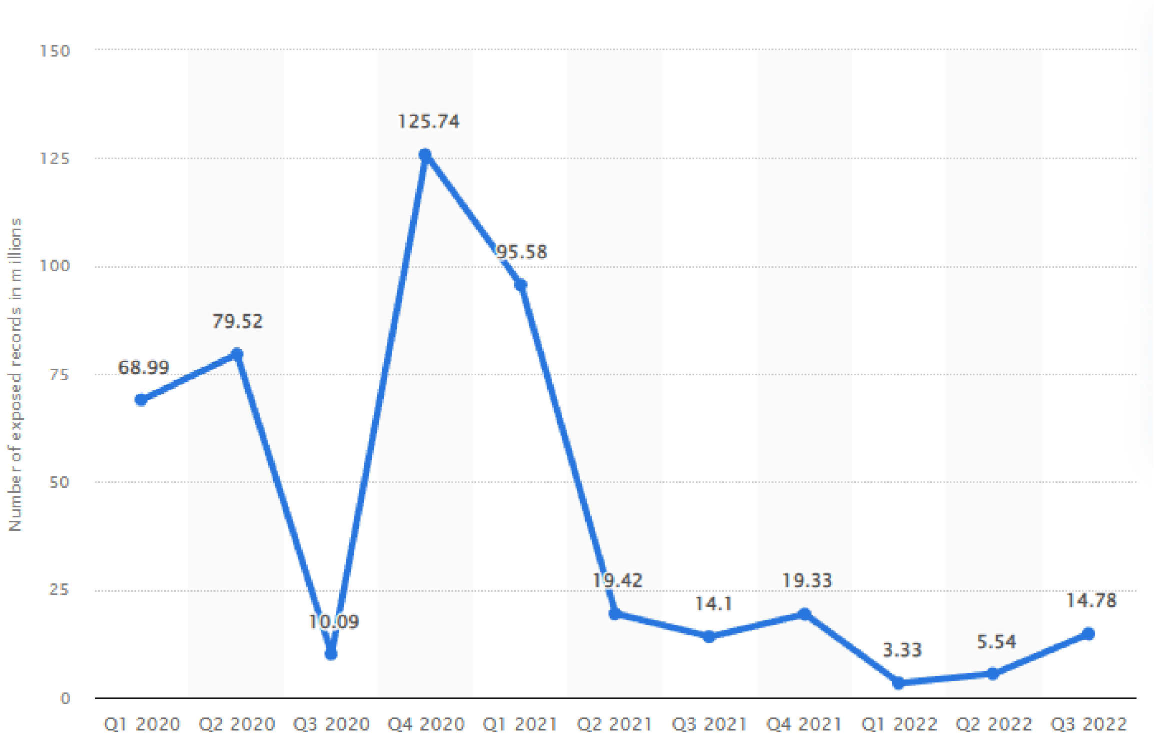


Figure 2 Global number of breached data sets 2020-2022 [26]

It is clear that it would be impossible to cover all data breaches in the presented time period. However, there is a large number of different lists consisting of the most impactful ones that can be found.

The following five examples are from a list of the largest data breaches in recent memory, compiled by Michael Hill and Dan Swinhoe for CSO.

Yahoo experienced a data breach in August 2013, which affected 3 billion accounts. In December 2016, the company announced that a hacking group had accessed the account information of more than a billion of its customers in 2013. This revelation led to a lower purchase price for Verizon, which acquired Yahoo in 2017. The investigation found that the hackers did not get access to passwords or payment information. [24]

In January 2018, a large database of Indian citizens' personal information was breached, exposing the data of 1.1 billion people. This included their names, addresses, photos, and phone numbers. Additionally, since the database also contained information about people's bank accounts, it also became a major credit breach. The data leak occurred on a system run by a state-owned utility company, and authorities only became aware of it in March 2018, after the data had been sold for $7. [24]

In November 2019, the Chinese shopping website Alibaba experienced a data leak that affected 1.1 billion users. The leak occurred when a developer working for an affiliate marketer scraped customer data, including usernames and mobile numbers, from the website using crawler software. [24]

In June 2021, a hacker posted data associated with 700 million LinkedIn users. This data included email addresses, phone numbers, geolocation records, genders, and other social media details. While LinkedIn argued that no sensitive, private personal data was exposed, the UK's National Cyber Security Centre warned that the leak could lead to social engineering attacks. [24]

The exposure of two Facebook datasets in April 2019 put the personal data of over 530 million users at risk. This included their phone numbers, account names, and Facebook IDs. Two years later, this data was posted online for anyone to access, which may have been done with criminal intent. [24]

4 THE EVOLUTION OF ONLINE CONSUMER PRIVACY

The way we think about online privacy has changed significantly since the early days of the internet. In the past, the internet was seen as a vast area of potential, but we have since realized that not everyone online can be trusted. Today, the issue of privacy is still evolving as people become more informed and technology becomes ubiquitous. [27]

The prevalence of online privacy concerns has increased in recent years. Large-scale attacks on companies such as Experian, Adobe, and eBay have made non-expert internet users aware of the risks, especially since breaches can go undetected for months or even years. [27]

As digital technology continues to evolve, data privacy laws are changing in order to protect consumers. The General Data Protection Regulation (GDPR), the California Consumer Privacy Act (CCPA), and other similar regulations are putting pressure on companies to reevaluate their methods for handling, storing, and protecting customer information. This new regulatory environment is forcing companies to be more transparent and upfront about data breaches and other security issues. With technology becoming more integrated into our lives, it is essential to have these laws in place to prevent our information from being accessed and used without our permission. [27]

The large dataset collected from smartphone usage has allowed marketers to develop profiles of our behavior with high accuracy. For instance, many smartphones have GPS tracking, which allows companies to monitor our location and the stores we visit. This information, in addition to data from social media activity and the kinds of content we consume, provides marketers with a comprehensive understanding of our preferences and interests. [27]

Although it is common knowledge that social media platforms collect data on individuals, many people are unaware that this data is then used to influence their behavior. According to research conducted by Privacy International, even if you do not have a Facebook account, there are apps that share your data with Facebook. This means that opting out of social media does not necessarily protect your privacy. [27]

The privacy of consumers is declining as we adopt more smart home and mobile technologies. This gives companies more opportunities to collect our personal data, even though we may benefit from increased convenience. However, we are actually sacrificing much more than we gain in return, as the companies we are sharing our information with have a proven track record of using profiling technologies. [27]

5 DATA PRIVACY LAWS

As internet access has been recognized as a basic human right in some countries, it is only logical that we should also have a say in how that access is used and controlled. [28]

Data privacy laws are constantly changing and evolving, which can make it difficult for websites to keep up with compliance. However, it is essential to protect user privacy, and website professionals must stay informed of new developments in order to successfully help their organizations avoid legal action. [28]

Data privacy laws are designed to protect people's personal information from being mishandled, disclosed, or misused. Some data privacy laws have stricter standards than others, and in recent years, many countries have implemented new data privacy laws to better protect people's information. [28]

Until recently, companies have been able to collect our data online without our consent or even our knowledge, but that is changing. People are becoming more aware of how they are tracked online, by whom, and for what purposes. This is a positive development, as it gives individuals more control over their personal data. [28]

5.1 U.S. data privacy laws

The data privacy of individuals in the United States is currently protected by a patchwork of state laws, as the federal government has yet to pass a comprehensive law on the matter. However, many states are considering data privacy laws in the wake of California, Utah, Virginia, and Colorado enacting legislation with national impact. [29]

The Federal Trade Commission is a government agency that protects consumers from unfair and deceptive business practices. When the agency finds violations, it can issue warnings or take legal action against the company, such as seeking financial penalties. The agency also has the authority to investigate suspected illegal activities and refer cases to the Department of Justice for prosecution. In addition to enforcement, the FTC also provides educational resources for consumers and operates a complaint process that allows consumers to report companies they believe are engaging in illegal activities. [30]

Health Insurance Portability and Accountability Act (HIPAA) governs how healthcare providers, insurers, and other entities use and disclose patient medical information. [29]

Children's Online Privacy Protection Act (COPPA) requires websites and online services to obtain verifiable parental consent before collecting personal information from children under 13. [29]

Gramm-Leach-Bliley Act (GLBA) requires financial institutions to disclose to customers how they use and protect their personal information and to give customers the opportunity to opt out of certain information-sharing practices. [29]

The passage of the California Privacy Rights Act (CPRA) in 2020 is a major milestone in the US data privacy landscape. This law applies to businesses that operate in California and meet certain criteria. It gives California residents the right to know what personal information is being collected about them, the right to request deletion of that data, and requires websites to get explicit permission from consumers before selling or sharing data obtained from cookies. The CCPA also gives consumers the right to opt out from non-essential cookies. [29]

The Virginia Consumer Data Protection Act 2021 is designed to protect consumers by requiring businesses to get customers' consent before using their personal data and to put in place reasonable security measures to protect it. The act also gives consumers the right to access their data and request corrections or deletions of any inaccuracies. Finally, it allows consumers to take legal action against businesses that violate the act's provisions. [29]

Both the California Privacy Rights Act (CPRA) and the Virginia Consumer Data Protection Act (VCDPA) went into effect on January 1, 2023. [29]

The Colorado Privacy Act (CPA), which was signed into law in July 2021, provides certain rights to Colorado residents with regard to their personal data and requires businesses to disclose any data security breaches that may occur. This law is expected to have a wide-ranging impact on the way businesses collect and use consumer data in Colorado. [29]

The UCPA has similarities to the statutes in Colorado and Virginia with respect to exemptions for certain types of personal data; however, the UCPA is broader in scope in terms of both the entities covered and the types of data covered. [29]

The Data Privacy Law in Connecticut is the first to exempt payment transaction data specifically. Consumers have the option to opt out of data processing for targeted advertisements, sales to third parties, and profiling. [29]

In July 2019, New York enacted the Stop Hacks and Improve Electronic Data Security (SHIELD) Act. The SHIELD Act is a data security and breach notification law that applies to any person or business that owns, licenses or maintains computerized data that includes private information of New York residents. The law expands the definition of private information to include biometric data, such as fingerprints and facial recognition data and requires businesses to implement reasonable data security measures to protect such information. [29]

5.2 Europe privacy laws

The General Data Protection Regulation (GDPR) is a set of regulations designed to give individuals in the European Union (EU) greater control over their personal data. GDPR replaces the outdated Data Protection Directive and applies to any organization that handles the data of EU residents, regardless of where the organization is located. Organizations that are found to be in violation of GDPR can be subject to heavy fines. [31]

The General Data Protection Regulation provides individuals with a number of rights with respect to their personal data, including the right to access it, the right to request that it be erased, and the right to object to its processing. This means that organizations must obtain explicit consent from individuals before using or storing their data. The goal of the GDPR is to safeguard the privacy and personal information of EU citizens while creating a more transparent and accountable data protection framework. [31]

The Privacy and Electronic Communications Directive or ePrivacy Directive is a legal instrument that regulates privacy in the digital environment. It requires websites to obtain consumer consent before using cookies for marketing purposes. The EU is currently considering reforming the ePrivacy Directive, but even before that happens, the directive, along with GDPR, creates one of the strictest privacy protections in the world. [32]

GDPR remains the main privacy law in the EU, but a few new data privacy laws have also been passed recently. Most notably, the Digital Services Act and the Digital Markets Act were both passed in early 2018. There are also several proposals that could become law in 2023, so it is important to stay up to date. [32]

5.3 EU vs. US privacy laws comparison

In the US, most privacy laws exempt de-identified data from their definition of personal data, although some special requirements may still apply to it. In the EU's GDPR, the term pseudonymized data is used instead of de-identified data, and it is stated that any pseudonymized data that could be attributed to an individual with the use of additional information should be considered personal data. [33]

Most laws require people to give their permission for their personal data to be processed by opting in or opting out. This consent is typically obtained through a strictly opt-out system, a strictly opt-in system, or a hybrid system that includes some aspects of both. [34]

The most common data collection system in the U.S. is opt-out, meaning that organizations are automatically allowed to collect certain information unless consumers take action to opt out of having their data sold. [33]

The other option is a strictly opt-in approach. This is the approach used in GDPR, but it is not common in the US. An opt-in regime requires obtaining active consent from a consumer before collecting or processing data. [33]

When it comes to the discussed problem, the opt-in and opt-out regimens affect the approach taken when obtaining consumer agreement with the terms of use and privacy policy, as well as cookie placement.

It would be helpful to explain what cookies are and how they are related to data privacy, as they have been mentioned in the discussion of both US and EU laws. An article by Richie Koch for GDPR.eu (a repository for organizations and individuals researching the GDPR, not an official EU resource) is a relatively short but informative read that explains cookies and their categorization.

Cookies are small files that are sent to a user's web browser from the websites they visit. These files are stored on the user's device in order to collect personal data. ***Essential cookies*** are necessary for a website to function properly and do not require user consent, while ***non-essential cookies*** are not typically necessary and, because of that, usually require user consent in some way. [35]

There are multiple approaches to classifying cookies, although some may fall into more than one category or may not fit well into any one of them. [35]

Cookies can be categorized into two types based on their **origin and source** [35]:

- ***First-party cookies*** are cookies that are placed directly on a device by the website being visited.

- ***Third-party cookies*** are installed on your device by a company other than the website being visited. Websites can incorporate elements from different domains, like images, scripts, or analytics systems. Whenever your browser seeks these resources, the associated domains' servers can also transmit cookies that will be connected to their domain.

The categorization of cookies according to **duration** encompasses two types, namely session cookies and persistent cookies [35]:

- ***Session cookies*** are temporary and expire when the browser is closed, or the session ends.

- ***Persistent cookies***, on the other hand, are kept in the computer's memory up until deletion or expiration. Despite the ePrivacy Directive dictating that persistent cookies should expire within 12 months, they may remain on the user's device beyond this timeframe unless steps to ensure otherwise are taken.

The categorization of cookies based on their intended **purpose** is as follows [35]:

- ***Strictly necessary cookies*** are crucial for users to access and use the features of the website. Typically, these cookies are first-party session cookies, and while obtaining consent for these cookies is not mandatory, their function should be explained to the user.

- ***Marketing cookies*** are designed to keep track of what users do online to provide advertisers with pertinent data for serving customized advertisements to the user or regulating the frequency of a specific advertisement presented to them. These are persistent cookies and typically come from third-party sources.

- ***Statistics cookies***, also referred to as "performance cookies," are designed to gather data on website usage, such as page visits and link clicks, with no ability to personally identify the user. Their only objective is to improve the functionality of the website. They may originate from third-party analytics services, provided that they are exclusively utilized by the proprietor of the visited website.

- ***Functionality cookies***, also called preferences cookies, are utilized by websites to retain specific choices that users have previously made. They can store information such as your region, preferred language, or login credentials.

5.4 Privacy Laws on an international scope

The present study also aims to conduct an analysis of the privacy laws implemented in the other most prominent nations globally, including Australia, Brazil, Canada, China, and Russia. Furthermore, the privacy laws of Belarus are addressed towards the conclusion, as this is the author's country of origin.

For countries like China, Russia, and Belarus, which are known for their level of government surveillance, the concerns for governmental access to private data of the citizens in relation to their laws are discussed as well.

The Australian Privacy Act of 1988 is a law regulating the collection, use, storage, and sharing of personal information by organizations in Australia. The law applies to most private sector organizations with an annual income of over $3 million, as well as most Australian Government agencies.

The Lei Geral de Proteção de Dados, or LGPD, is a set of laws in Brazil that protect people's data. This law applies to anyone in Brazil, regardless of where the company collecting the data is based. The law establishes ten legal bases for how data can be collected and used, as well as accountability requirements and mandatory breach notifications. Anyone violating this law may be subject to significant penalties. [36]

The Personal Information Protection and Electronic Documents Act (PIPEDA) is a federal law in Canada that establishes rules for how private sector organizations can collect, use, and disclose personal information. The Act applies to businesses that are federally regulated and offer commercial services. [36]

The following are the three primary laws in China that regulate data privacy. [36]

The Personal Information Protection Law is the primary legislation in China concerning data protection, with the aim of ensuring the proper usage of personal information. The law mandates opt-in consent for the handling of sensitive personal information. The Data Security Law was established to establish uniform data processing standards, whereas China's Cybersecurity Law offers directives on cybersecurity prerequisites for upholding the protection of Chinese cyberspace. [36]

The aforementioned regulations establish criteria for processing and accessing data by both individuals and organizations while also allowing the government to obtain personal information without any restrictions. Despite the inclusion of the phrase "in accordance with law" in legislation such as cybersecurity law, there are no limitations on government data access explicitly stated within the law itself. [37]

Russian Federal law "On Personal Data" N 152-ФЗ covers processing, storage, and access to personal data and requires operators to delete personal data upon user request. Additionally, consent is necessary for personal data processing. The Federal Law "On Information, Information Technologies and Information Protection" N 149-ФЗ establishes a register of violators of personal data subjects' rights. The Code on Administrative Offenses and the Criminal Code of the Russian Federation also include penalties for any violations related to the management and handling of personal data. In contrast to the GDPR, the Russian legal framework does not mandate the requirement of disclosing data breaches. [37], [38]

The primary legislation in Belarus that regulates the handling of personal information is the Law on Personal Data Protection N. 99-Z. This law is applicable to all forms of data processing, regardless of whether it is done manually or automatically. The law includes a definition of personal data and covers the transfer of data across borders. Additionally, it outlines the steps for forming an authorized organization responsible for ensuring the protection of individuals' rights relating to their personal data. [39]

The newly enacted legislation entails an extensive list of circumstances under which the government is authorized to procure personal data without obtaining consent from the individual. Among these scenarios are legal proceedings, national security concerns, and electoral procedures. [39]

This is not a new concept for Belarus, as from 2010 onwards, the government has been utilizing a system that enables them to automatically acquire communication data from various sources, including landline phones, mobile networks, and internet service providers. With the increasing usage of Telegram in 2020, security agencies heightened their monitoring of messenger chats and closed groups. Later, Google removed an application utilized by authorities to observe protestors in Belarus. [40]

As a conclusion for this chapter, it is worth noting that 120 countries across the world have taken measures to protect the privacy and security of their citizens' data through the implementation of legislation. The following map provided by Securiti experts (Figure 3) demonstrates that the majority of regions have enforced data privacy laws, although there are some outliers: [29]

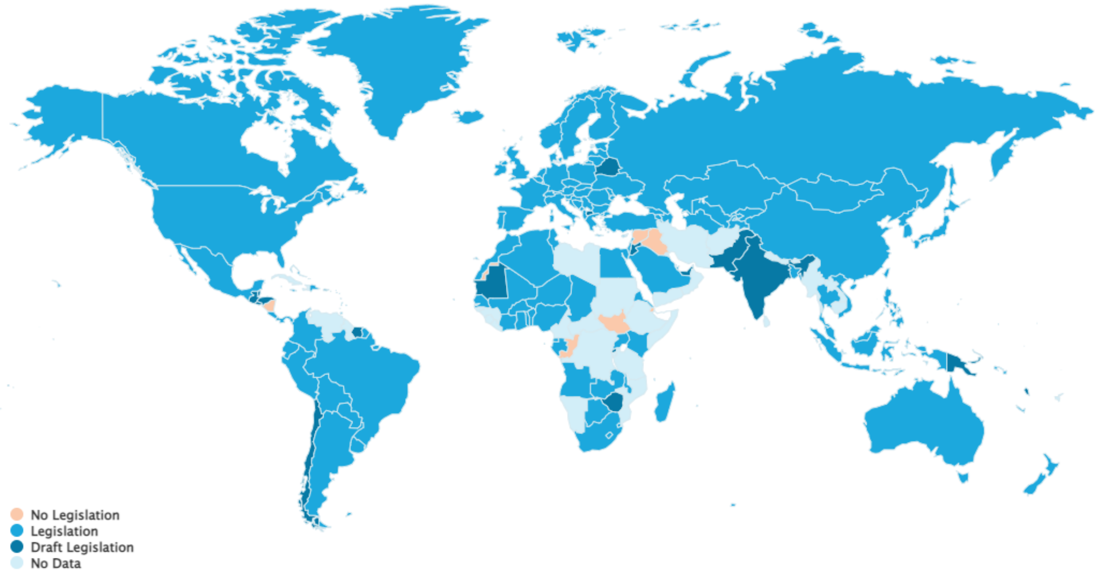


Figure 3 Privacy Laws Around the World [29]

6 PUBLIC AWARENESS

The increasing use of digital technology by consumers creates a lot of data, which can provide businesses with insights into how to improve their engagement with consumers, create more personalized marketing campaigns, and optimize their products and services. However, enterprises must take care to protect this data and safeguard consumers' privacy, as implementing data protection measures is essential for gaining trust and customer loyalty.

As companies become more aware of consumers' consideration for their data privacy, they are beginning to realize that investing in efficient means of data protection can be beneficial for their business. This is clearly reflected in the CrunchBase statistic that overviews the amount of money invested into security and privacy companies all over the world (Figure 4). The investment volume increased almost six times from 2010 to 2019. [41]

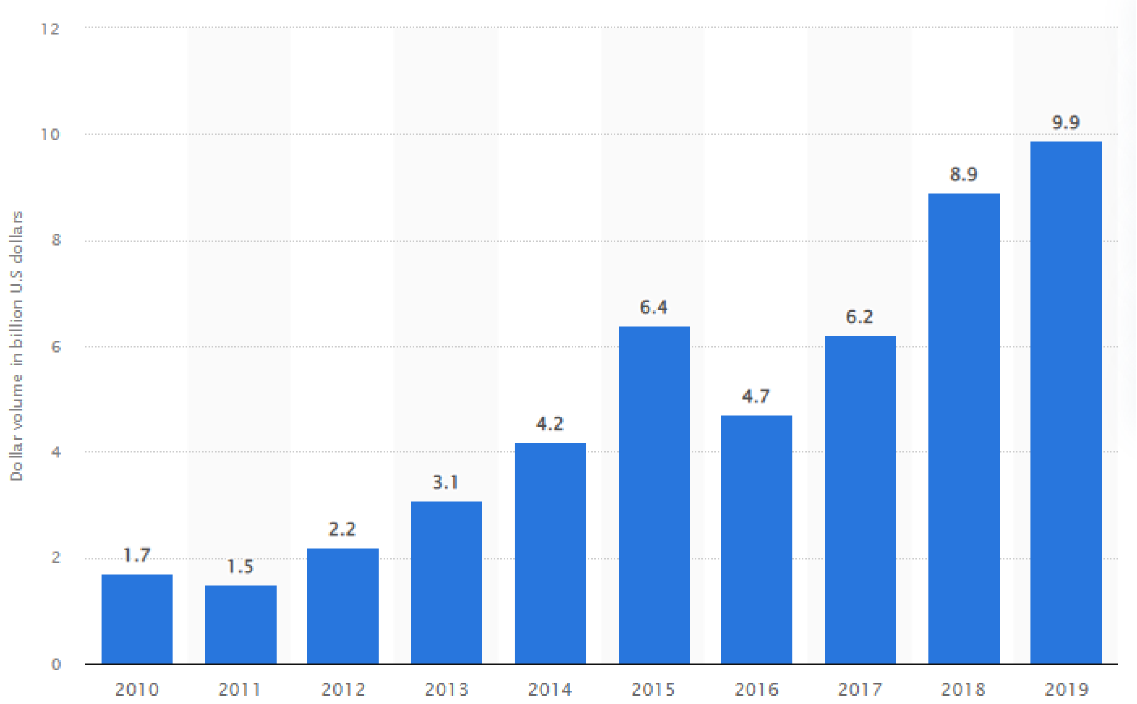


Figure 4 Dollar volume of investments in privacy and security companies worldwide from 2010 to 2019 (in billion U.S dollars) [41]

People are becoming more cautious about data collection requests and are more likely to only share personal information when it is necessary for their interactions with businesses. According to The Privacy and E-commerce Report 2022 from DataGrail, worldwide consumers' awareness of companies selling data to third parties increased from 62% in 2020 to 75% in 2022. [42]

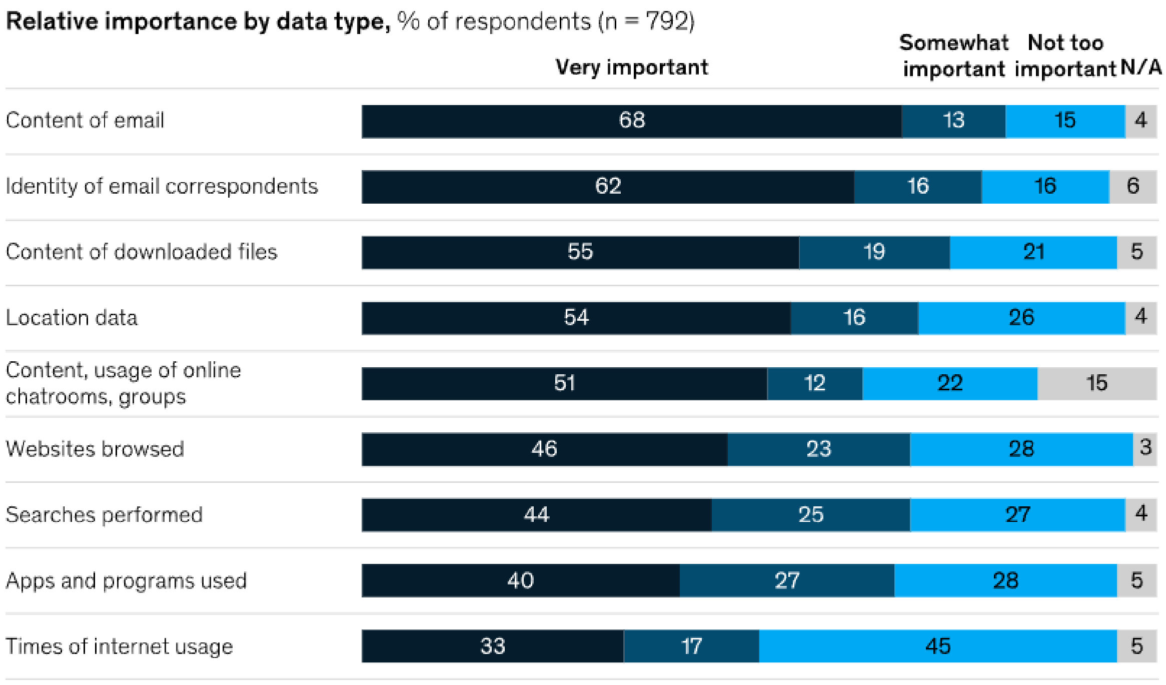
The number of cases of data violation due to cyber-attacks has increased in the past years, as illustrated by the Identity Theft Resource Center 2022 Annual Data Breach Report (Table 1). [43] The recent increase in data breaches has made many people lose trust in the companies that manage their data. Even those who have not been personally touched by the breaches have seen how companies have responded and are concerned about the potential consequences.

Characteristic	2020	2021	2022
Healthcare	306	330	344
Financial services	138	279	268
Manufacturing and utilities	70	222	249
Professional services	144	184	224
Education	42	125	100
Technology	67	79	86
Government	47	66	74
Non-profit/NGO	31	86	71
Retail	53	102	65
Transportation	21	44	36
Hospitality	17	33	34
Unknown	-	4	-
Other	172	308	251

Table 1 Number of cases of data violation due to cyber attacks in the US from 2020 to 2022 by industry [43]

Almost three-quarters of 2020 McKinsey survey respondents indicated that they would not do business with a company if they had concerns about its security system. Additionally, over 70% of respondents said they would not do business with a company that gave away personal data without prior agreement. [44]

Different types of digital data can have different implications for consumer privacy and protection, and McKinsey has found that concerns vary depending on the type of data involved (Figure 5) [44]:



Relative importance by data type, % of respondents (n = 792)

Source: Internet & American Life Project, Pew Research Center

Figure 5 Privacy concern by data type [44]

Approximately 50% of respondents said they would trust a company more if it was more responsible with data management - for example, only seeking information concerning its products or only asking for a limited amount of personal details. [44]

Companies that react quickly to data breaches and proactively disclose such incidents to the public are more likely to earn the trust of consumers. Promoting privacy for a company's products, having trustworthy management, and the presence of industry regulations can also affect consumers' confidence in an organization. [44]

2023 Norton Cyber Safety Insights Report conducted by Harris Poll (consulting and market research firm) between November and December 2022 found that 68% of consumers felt more vulnerable to identity theft at the time of the survey than they did a few years ago (Figure 6). [45]

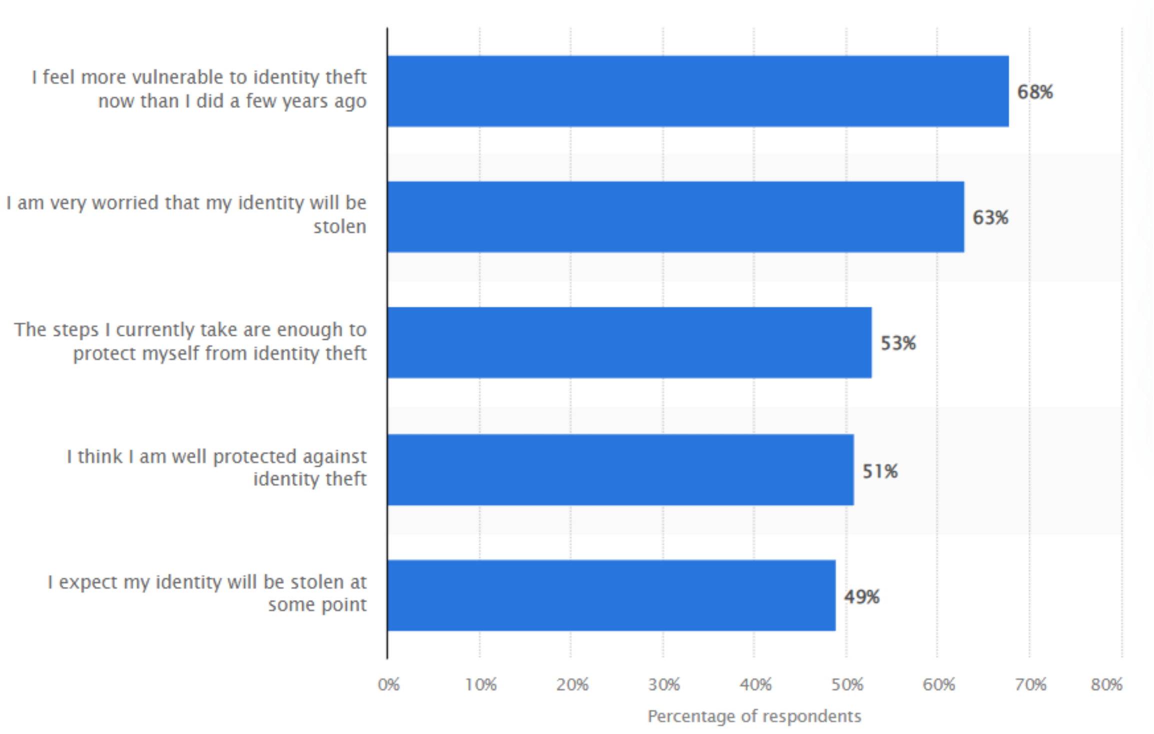


Figure 6 Worldwide consumer confidence about personal online data security, December 2022 [45]

From the above statistics, we can conclude that as the digital environment changes and becomes more profitable, companies are quick to implement all sorts of Big Data solutions to increase marketability. However, as the amount of data harvested grows, so does the public's concern. More people are starting to pay attention to their privacy on the internet, and the laws worldwide are changing and evolving to accommodate that.

It is slowly becoming evident that companies that handle consumer data and privacy responsibly can set themselves apart from the competition and even gain an advantage in the market, as consumers are more likely to do business with companies that are transparent about their data handling practices, have strong security measures in place, and respect their privacy rights.

7 PROACTIVE STEPS FOR COMPANIES

There is a number of cataloging and mapping data tools available that can help categorize large amounts of customer data effectively. This data then needs to be analyzed in order to make use of it. However, companies should focus specifically on supervised-learning algorithms to minimize safety risks.

Organizations should be aware of the potential damage that could be caused by insiders and take precautions to prevent it. This includes putting controls in place to ensure that people can only see the data they need and that no one has access to all data. Even with good identity and access management practices, some breaches can still happen, so it is essential to also have activity monitoring in place. If a breach does occur, it is necessary to have a plan in place for emergency response.

In order to comply with regulations and quickly remove or transfer data at the consumer's request, companies need to have transparent and consistent procedures in place. These procedures should be designed to quickly and easily locate data stored within the company and across its affiliated third parties and streamline or automate the process as much as possible. When working with third parties, it is vital to understand how and where their data is stored in order to avoid any potential problems in the future.

Companies typically look for ways to set up their infrastructure so that it can handle large data sets. Having fewer systems in which data is stored reduces the likelihood of data breaches.

Organizations need to be transparent with their customers about why they are collecting data and what it will be used for. Many companies are making consumer privacy protections a part of their standard business practices and ensuring that security and privacy are the default settings for customers.

II. RESEARCH

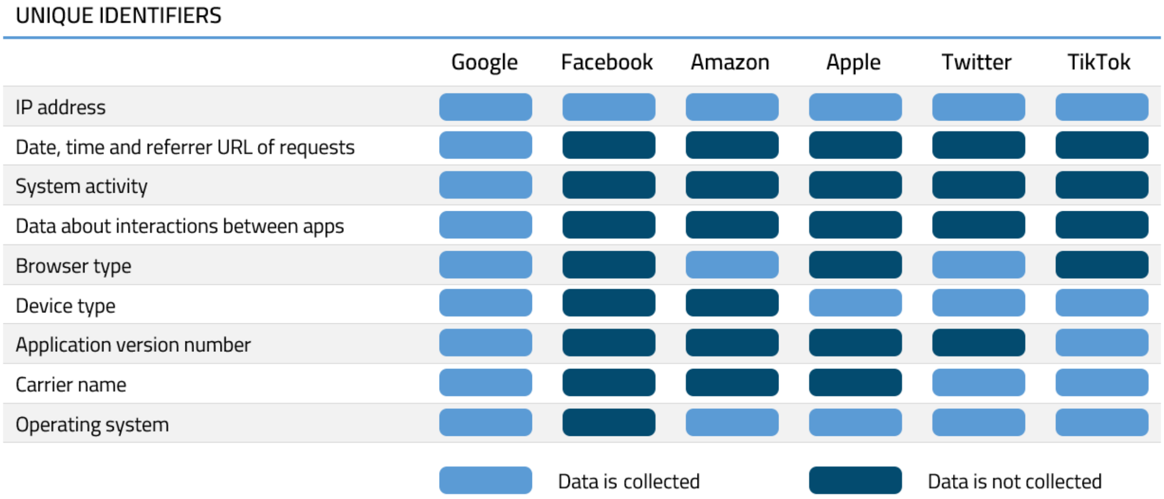
8 TECH GIANTS AND DATA HANDLING

This chapter consists of two parts. The first part compares the data collection and sharing practices of six well-known tech companies: Google, Facebook, Amazon, Apple, Twitter, and TikTok. The second part focuses primarily on third-party user data sharing and any issues that the companies face with data privacy laws.

8.1 Information the companies collect

To begin, let's consider the type of information that companies collect about their users.

When it comes to collecting data on unique user identifiers, Google logs almost everything, with the exception of crash reports. Facebook only collects data on IP addresses, while Amazon makes use of information on the browser and operating system in addition to IP addresses. Apple logs data on users' IP addresses, the type of device, and the operating system, while Twitter collects data on IP addresses, browsers, operating systems, device type, and carrier name. TikTok logs the user's IP and information about the device, its operating system, and mobile carrier. The comparison is depicted in Figure 7. [46], [47]



UNIQUE IDENTIFIERS

	Google	Facebook	Amazon	Apple	Twitter	TikTok
IP address	🟦	🟦	🟦	🟦	🟦	🟦
Date, time and referrer URL of requests	🟦	⬛	⬛	⬛	⬛	⬛
System activity	🟦	⬛	⬛	⬛	⬛	⬛
Data about interactions between apps	🟦	⬛	⬛	⬛	⬛	⬛
Browser type	🟦	⬛	🟦	⬛	🟦	⬛
Device type	🟦	⬛	⬛	🟦	🟦	🟦
Application version number	🟦	⬛	⬛	⬛	⬛	🟦
Carrier name	🟦	⬛	⬛	⬛	🟦	🟦
Operating system	🟦	⬛	🟦	🟦	🟦	🟦

🟦 Data is collected ⬛ Data is not collected

Figure 7 Data collection by company, unique identifiers [46]

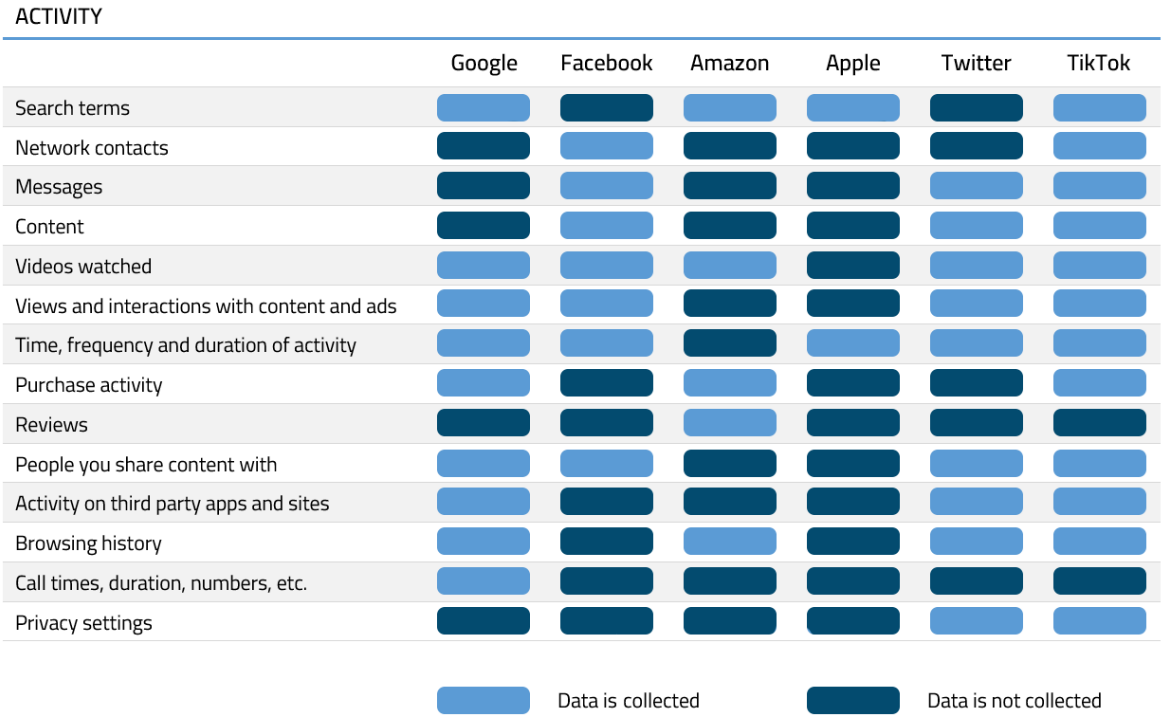
Google tracks users' search terms, video views, and interactions with content and ads. They also keep track of consumers' purchase activity and track their activity on third-party sites that use Google services. Google also tracks users' browsing history if they use Chrome. If a consumer uses Google to make calls or text, the company will collect the details of those communications, including the numbers, call duration, times and dates of calls, and SMS. [46]

Facebook gathers data on users' interactions and activity on the site, such as the content they share and how they interact with it, as well as with different advertisements. Amazon collects data on what consumers search for, what videos they watch on Prime, and their purchase activity. However, Amazon is less reliant on advertising revenue than other websites, so it does not need to collect as much activity data. Apple stores less user activity data than other companies do. They only keep track of search terms and the time, frequency, and duration of user activity. [46]

Twitter tracks a wide range of user activity, including tweets, messages, interactions with content and ads, and video and audio information. They monitor the time, frequency, and duration of activity on Twitter as well as the people users communicate and share content with. Twitter tracks user activity on third-party sites and apps along with their browsing history, although the latter is never linked to any personally identifiable information. [46]

TikTok collects information about user activities on other websites and apps or stores obtained from its partners. This information is matched to the TikTok account by mobile identifier, email addresses or phone numbers, and cookie identifiers. TikTok automatically collects browsing and search history, information about watched videos, messages content, and phone and social network contacts. [47]

The information regarding user activity collection presented above is illustrated in Figure 8:



ACTIVITY

	Google	Facebook	Amazon	Apple	Twitter	TikTok
Search terms	🟦	⬛	🟦	🟦	⬛	🟦
Network contacts	⬛	🟦	⬛	⬛	⬛	🟦
Messages	⬛	🟦	⬛	⬛	🟦	🟦
Content	⬛	🟦	⬛	⬛	🟦	🟦
Videos watched	🟦	🟦	🟦	⬛	🟦	🟦
Views and interactions with content and ads	🟦	🟦	⬛	⬛	🟦	🟦
Time, frequency and duration of activity	🟦	🟦	⬛	🟦	🟦	🟦
Purchase activity	🟦	⬛	🟦	⬛	⬛	🟦
Reviews	⬛	⬛	🟦	⬛	⬛	⬛
People you share content with	🟦	🟦	⬛	⬛	🟦	🟦
Activity on third party apps and sites	🟦	⬛	⬛	⬛	🟦	🟦
Browsing history	🟦	⬛	🟦	⬛	🟦	🟦
Call times, duration, numbers, etc.	🟦	⬛	⬛	⬛	⬛	⬛
Privacy settings	⬛	⬛	⬛	⬛	🟦	🟦

🟦 Data is collected

⬛ Data is not collected

Figure 8 Data collection by company, user activity [46], [47]

Although it is not surprising given their business model, Google is known for collecting and storing large amounts of data on users. This includes information like location, browsing history, and activity on third-party sites. [46]

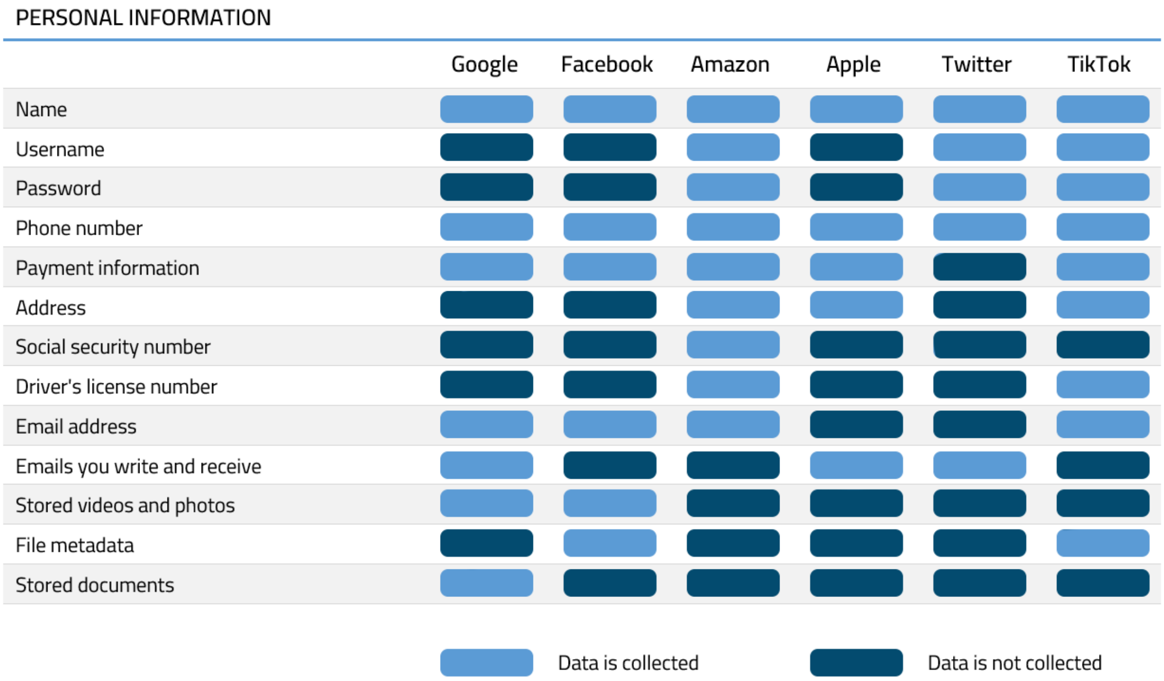
Given that Facebook is a social networking application, it has access to a considerable amount of personal information about its users. This includes data about the individuals with whom consumers interact, the groups to which they belong, and their private messages. [46]

Amazon collects a variety of personal information from its users, including names, phone numbers, payment information, shipping addresses, and email addresses. In addition, the company also collects social security and driver's license numbers from US citizens. [46]

Apple does not retain much personal information aside from names, phone numbers, payment information, shipping address, and email, all necessary to maintain user accounts. [46]

Twitter is not very strict when it comes to personal information and only requires users to provide their names, phone numbers, and email addresses. [46]

TikTok collects a range of data from users, including contact lists, calendar information, hard drive contents, and the contents of the device’s clipboard, as well as purchase information and shipping address. [47]



PERSONAL INFORMATION

	Google	Facebook	Amazon	Apple	Twitter	TikTok
Name	🟦	🟦	🟦	🟦	🟦	🟦
Username	⬛	⬛	🟦	⬛	🟦	🟦
Password	⬛	⬛	🟦	⬛	🟦	🟦
Phone number	🟦	🟦	🟦	🟦	🟦	🟦
Payment information	🟦	🟦	🟦	🟦	⬛	🟦
Address	⬛	⬛	🟦	🟦	⬛	🟦
Social security number	⬛	⬛	🟦	⬛	⬛	⬛
Driver's license number	⬛	⬛	🟦	⬛	⬛	🟦
Email address	🟦	🟦	🟦	⬛	⬛	🟦
Emails you write and receive	🟦	⬛	⬛	🟦	🟦	⬛
Stored videos and photos	🟦	🟦	⬛	⬛	⬛	⬛
File metadata	⬛	🟦	⬛	⬛	⬛	🟦
Stored documents	🟦	⬛	⬛	⬛	⬛	⬛

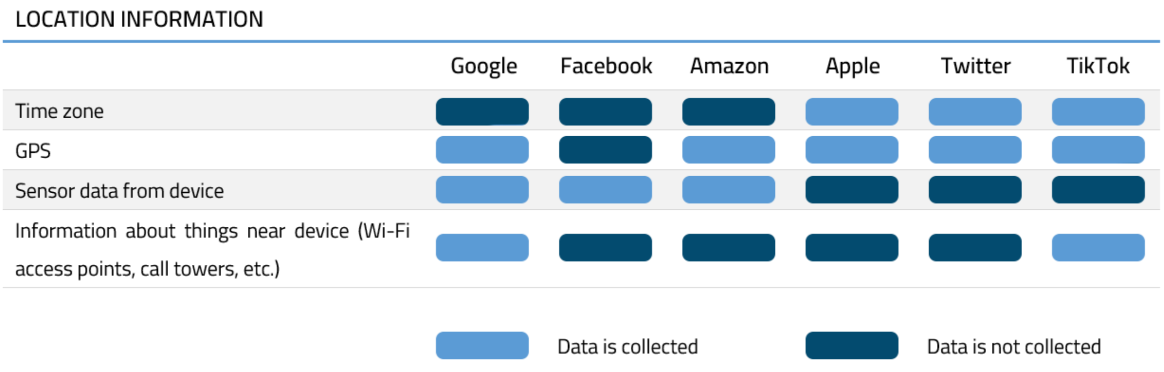
🟦 Data is collected ⬛ Data is not collected

Figure 9 Data collection by company, personal information [46], [47]

The graphical representation depicted in Figure 9 presents a comparison of the logging of personal information of customers.

Google tracks its users' locations by GPS, sensor data, and information about nearby Wi-Fi access points, cell towers, or Bluetooth-enabled devices. Facebook and Amazon use sensor data from user devices to log their location. Apple and Twitter track the time zone and GPS information from user devices. TikTok collects location information based on the user's SIM card and IP address. [46], [47]

Figure 10 illustrates the comparison of location data collection.



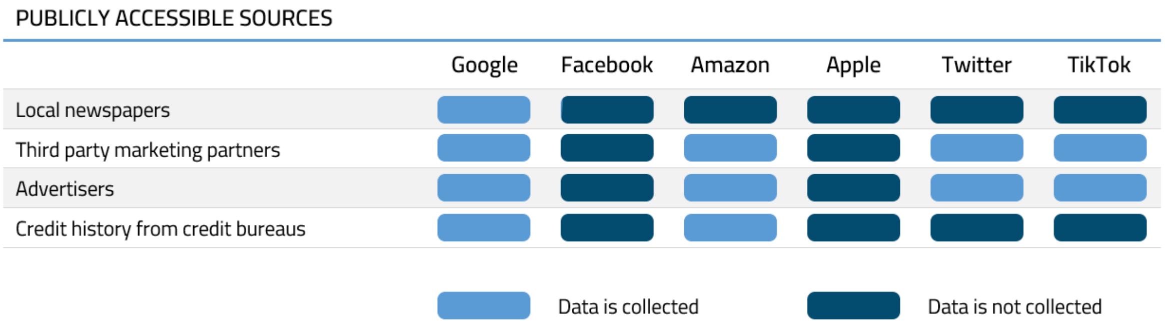
LOCATION INFORMATION

	Google	Facebook	Amazon	Apple	Twitter	TikTok
Time zone	■	■	■	□	□	□
GPS	□	■	□	□	□	□
Sensor data from device	□	□	□	■	■	■
Information about things near device (Wi-Fi access points, call towers, etc.)	□	■	■	■	■	□

□ Data is collected ■ Data is not collected

Figure 10 Data collection by company, location information [46], [47]

Google may collect additional user data from newspapers, marketing partners, or advertisers. The data Facebook relies on for its user base is not derived from public sources but rather from the users themselves. Amazon, on the contrary, gets its data from third-party marketing partners, advertisers, and credit bureaus. Apple does not obtain any information about users from publicly accessible sources. Twitter records user data from third-party marketing partners and advertisers. TikTok obtains information about its users from its partners and affiliated entities within its corporate group. Its privacy policy also states that it may collect information from other publicly available sources. [46], [47]



PUBLICLY ACCESSIBLE SOURCES

	Google	Facebook	Amazon	Apple	Twitter	TikTok
Local newspapers	□	■	■	■	■	■
Third party marketing partners	□	■	□	■	□	□
Advertisers	□	■	□	■	□	□
Credit history from credit bureaus	□	■	□	■	■	■

□ Data is collected ■ Data is not collected

Figure 11 Data collection by company, publicly accessible sources [46], [47]

8.2 Data Processing and legal cases

Google's CEO Sundar Pichai has stated that the company will never sell personal user data to third parties. Data collection is used solely for the purpose of improving user experience, and only a small amount of data is used for customized advertising. [48]

Although it is not stated explicitly, it appears that Google's real-time bidding system exploits a loophole in the definition of the word "sell." Real-time bidding (RTB) is the process of buying and selling digital advertising inventory, which happens in less than a second. [48]

Although the company does not sell personal data to third parties, it does share this data with advertisers. Advertisers then use this data to target ads to a specific audience. By doing so, Google discloses sensitive information such as geolocation, device IDs, identifying cookies, and browsing history to numerous third parties. [48]

Some users were concerned about their privacy after Google was criticized in May 2021 for selling personal information through real-time bidding. They filed a lawsuit against the company, claiming that it had violated its policy. [48]

Aside from that, Google has been fined on multiple occasions for not adhering to GDPR standards. In 2019, the French data protection regulator, Commission Nationale de l'Informatique et des Libertés (CNIL), issued a €50 million fine to Google for failing to properly disclose to users how data is collected for targeted advertising purposes. Then, in 2020, the CNIL issued a €60 million fine to Google LLC and a €40 million fine to Google Ireland for non-compliance with the French cookie consent law. [49]

Facebook uses the data it collects on its users to improve its own products and to serve targeted ads to those users. Facebook makes most of its money from advertising, so it provides third parties with user personal details by default. Facebook also provides data to researchers and academics, as well as law enforcement agencies, if requested. [49]

For example, as some states implement abortion bans, law enforcement is increasingly turning to social media to identify women who seek abortions or abortion-inducing medication. Google and Facebook are providing data from online pharmacies and social media posts and communications that can be used to prosecute women for seeking abortion. [50]

The Irish Data Protection Commission found in January 2023 that Meta, the parent company of Facebook and Instagram, was violating European Union privacy laws by requiring users to consent to its data processing practices. This means that users were effectively forced to agree to have their data used for behavioral and personalized advertising. [51]

The DPC said that Meta's personalized services, which are customized for users by processing their data, are not necessary to provide the service. The DPC found that Meta Ireland did not provide clear information about what would happen to users' data when they consented to its use. The DPC also found that Meta relied on the assumption that users were entering into a contract by using its services, which allowed Meta to process data without consent. [51]

Since Amazon works with millions of Marketplace sellers, user information is shared with third-party service providers, co-branding partners, and other businesses. While Amazon might send its customers promotional offers on behalf of other businesses, they would not share names and addresses without user consent. Users can also choose to opt out of those practices. Amazon may release user data when required to do so by law. [46]

In 2020, a privacy rights group filed a complaint against Amazon in Germany, claiming that the company's internal email security did not encrypt emails sent between its third-party partners and their customers. In the same year, CNIL fined Amazon €35 million for using cookies on users’ devices without prior user consent. [49]

Apple states that consumer information is not shared with third parties for marketing purposes. Although it may be shared with customers or delivery services, as well as anyone else who needs it for legal or public purposes. [46]

Despite its public image as a privacy-conscious company, Apple has been the subject of a number of investigations for possible privacy violations since 2018. France Digitale, a French lobby group, has filed a complaint with CNIL over Apple's use of personalized ads, alleging that Apple collects user data for marketing services and shares it with collaborating companies without explicit permission from users. [49]

In 2018, the Irish DPC began an investigation into how Apple processes users' personal data for targeted advertising and whether the company's privacy policy adequately informs users about this. From 2019 Apple has been investigated for possible GDPR violations stemming from a customer's access request. [49]

Twitter shares its users' data with advertisers, service providers, law enforcement, or the government as necessary, but it does not share users' personal information, such as names, phone numbers, usernames, or email addresses, with advertisers. [46]

Twitter has not been subject to much regulatory scrutiny under GDPR, but in 2020 Spanish Agencia Española de Protección de Datos (AEPD) fined Twitter $30,000 for violating cookie regulations. In the same year, Ireland's DPC fined Twitter €450,000 for not taking quick action to report and document a data breach. [49]

A study conducted by mobile marketing company URL Genius published in January 2022 found that YouTube and TikTok collect more user data than any other social media apps. TikTok had 14 network contacts, 13 of which were from third parties. [52]

Even when researchers opted out of data tracking in the application settings, third-party tracking still occurred. The report's authors also note that users are unable to find out what data is being shared or how it will be used by third-party entities. [52]

TikTok's privacy policy, however, does state that the service shares users' personal data with third parties that are not involved in its operation. [52]

9 DIGITAL PRIVACY PROTECTION MEASURES

The measures for protecting online privacy outlined in this chapter have been formulated with consideration for the abilities and constraints of an average user. Fundamental privacy tools, such as a virtual private network and anti-tracking features on browsers, as well as the use of secure browsers and email, can be employed to reduce the amount of personal information gathered by companies.

A comprehensive explanation of the aforementioned tools will be provided in subsequent sections of this chapter. However, there are also less complex steps that warrant equal attention and should not be underestimated.

- It is advisable to periodically delete cookies to hinder ad trackers from continuously monitoring your online activities across multiple devices.
- While third-party cookies are not the exclusive method of monitoring online activity, their deactivation in a browser's configuration renders advertisers incapable of retaining data on browsing behavior by utilizing browser files. Generally, the relevant setting is available through browser preferences in a privacy section.
  It is important to note that certain websites necessitate the use of third-party cookies for optimal functionality. To accommodate such instances, the website can be added to exceptions in the same settings.
- To minimize interest-based advertising, it is advised to opt out whenever possible. Many technology companies, including Google, Twitter, and Facebook, offer instructions on how to opt out of interest-driven ads. This measure is not capable of eliminating the advertisements displayed on online platforms; instead, it decreases their relevance to the individual user.
- Another uncomplicated measure for protecting privacy entails the implementation of an ad-blocking extension for the desktop browser. The author's recommendation would be **uBlock Origin**, free and open-source software that, along with blocking advertisements, offers effective anti-tracking capabilities. Users can choose which filter lists they wish to implement in the extension settings, as well as block specific elements on the page using the element picker. The uBlock Origin extension also demonstrates relatively low memory usage and CPU consumption. [53]
- There is a number of privacy-focused search engines that allow users to avoid data collection while maintaining their anonymity, such as **DuckDuckGo**, **Startpage**,

SearX, **MetaGer**, **Qwant**, etc. Typically, these search engines generate revenue by means of affiliations with partners, contextual advertising (sponsored links appearing adjacent to search results), and donations.

There are two distinct categories of private search engines. The first category is the search engines that utilize their own web crawlers. The second category, commonly referred to as meta-search or proxy search engines, is the engines that function as an intermediary between users and search engines such as Google. [54]

9.1 Virtual Private Networks

A Virtual Private Network (VPN) enables users to access the public internet through a secure and private network connection. VPN services provide enhanced privacy by concealing the IP address from third parties and establishing an encrypted connection. [55]

VPN serves as the primary safeguard against intrusive surveillance and monitoring; however, it is very important to choose the provider carefully.

Two crucial things that need to be considered before settling on a provider are the No-Logs Policy and VPN's Headquarters location.

VPNs can be subject to the laws of different governments, which is why it is best to choose providers that have implemented strict **no-logs policies**. These policies ensure that the VPN provider will not gather or distribute any of your information while using their servers. By following a no-logs policy, a VPN company is making a commitment not to monitor your online activities, including the websites you browse, the duration of your visits, your downloads, and your search history.

The significance of the **headquarters location** lies in the potential for your VPN provider to be obligated to disclose user data to government authorities if it is based in a country associated with international intelligence alliances.

The Five Eyes (FVEY), Nine Eyes, and Fourteen Eyes are intelligence alliances between states formed with the intention of preserving national security by monitoring and exchanging information about private citizens. Initially, only five countries collaborated on intelligence, but over time, the Five Eyes Alliance grew to encompass more nations. At present, there are three principal alliances, each with differing levels of cooperation. While all countries work together, those within the Five Eyes have a stronger partnership than those in the Nine Eyes or Fourteen Eyes groups.[56], [57] Figure 12 contains the list of nations that are partaking in the Eyes alliances.

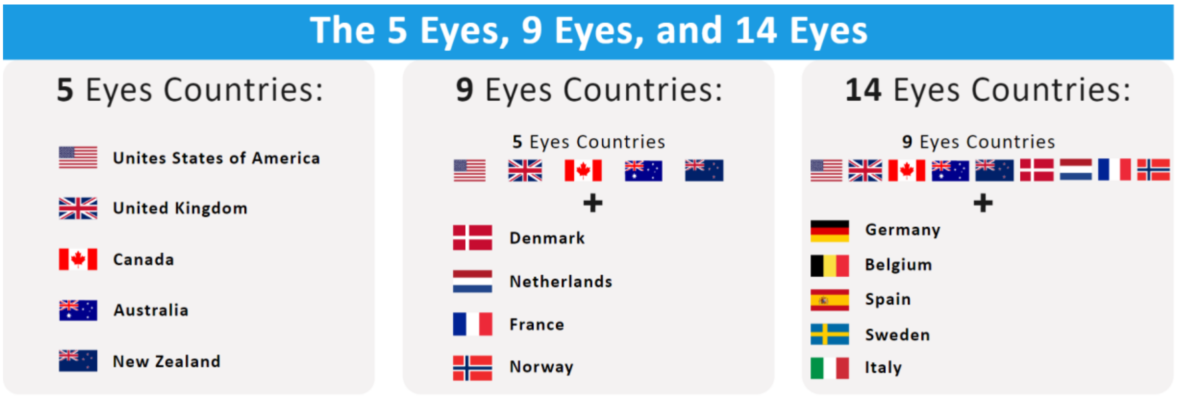


Figure 12 The Eyes Programs member countries [56]

Many VPN providers offer a kill switch function developed to immediately cut your internet connection in the event of a VPN disconnection.

A VPN employs encryption and obfuscation techniques to conceal the user's internet activities from ISP, websites, and government agencies. VPNs also can serve as a means to circumvent internet resource restrictions enforced by governments. Nonetheless, certain countries are implementing measures to prevent the use of non-governmental VPNs. Presently, only VPNs that possess strong obfuscation technology are capable of circumventing the VPN blocks imposed by China and Russia. Despite making attempts to block some proxy services in the past, Belarus currently permits VPNs to function.

The following three reputable providers were chosen for comparison in this paper: Express VPN, Proton VPN, and Nord VPN (Table 2). All three providers are situated outside of the Eyes Alliances that have been referenced earlier.

	EXPRESS VPN	**PROTON VPN**	**NORD VPN**
Based in	British Virgin Islands	Switzerland	Panama
No-Logs Policy	Yes, confirmed by an external audit by KPMG in 2022 [58]	Yes, confirmed by an external audit by Securitum in 2022	Yes, confirmed by an external audit by Deloitte in 2022
Support	24/7 live chat	Email	24/7 live chat
Worldwide servers	3000+ servers in 94 countries	1800+ servers in 66 countries	5500+ servers in 60 countries
Kill switch	Yes	Yes	Yes
Specialty servers	Obfuscated, Split-tunneling	Secure Core, P2P, Streaming, Tor over	Obfuscated, Double VPN, Onion Over VPN, Dedicated IPs

	EXPRESS VPN	PROTON VPN	NORD VPN
		VPN, Split-tunneling	
Price	From $6.67/month	From $4.99/month	From $3.29/month
Tunneling protocols	OpenVPN, IKEv2/IPsec, Lightway	OpenVPN, IKEv2/IPsec, WireGuard, Stealth	OpenVPN, IKEv2/IPsec, WireGuard (NordLynx)
Avg. download speed (April 10 to April 21, Cybernews speed test)	239.60 Mbit/s	600.61 Mbit/s	768.26 Mbit/s

Table 2 VPN comparison

The following section presents a comparative analysis of VPN download speeds. The data was gathered by Cybernews through the utilization of the Cybernews VPN speed test tool, which procures live data from global servers through standard VPN accounts. A single provider is tested every five minutes in each location to obtain an accurate average result that closely resembles real-world performance. Australia, Canada, the Netherlands, the United Kingdom, Germany, the United States, Singapore, Sweden, France, and Japan are among the ten countries where data for the Cybernews VPN speed test tool is acquired. [61]

Despite the higher price, Express VPN seems to have relatively low download speeds on this speed test. Both Nord and Proton VPN provide a high connection speed on average.

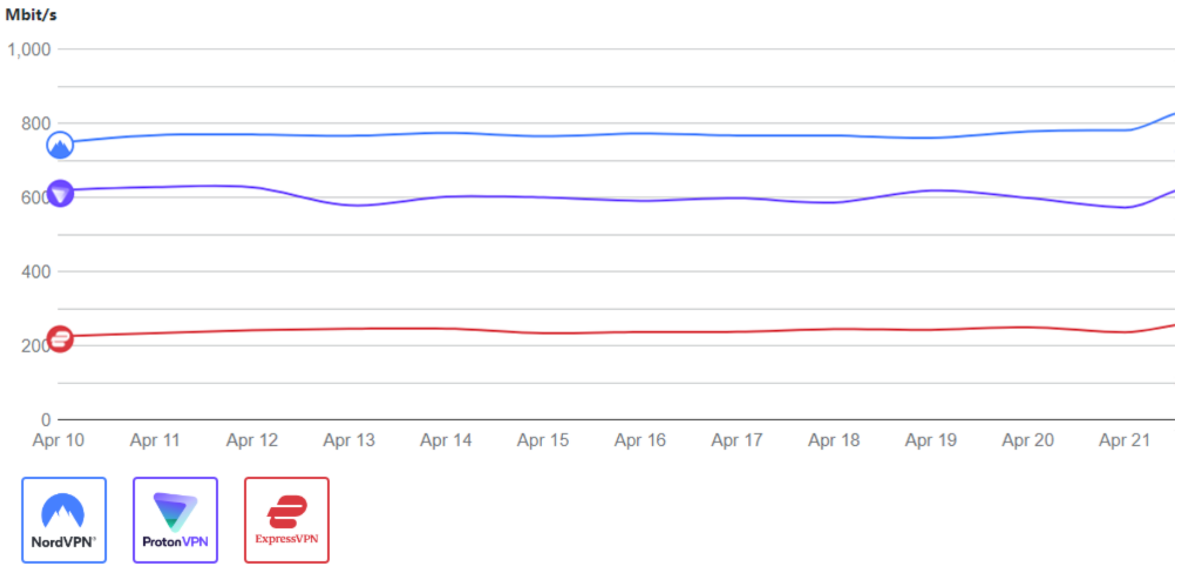


Figure 13 VPN speed test comparison [61]

Each of the VPNs that were listed provides both free and paid options; however, the speeds that are provided by the free alternatives are insufficient for convenient internet usage. This holds true for all other free VPN service providers as well. Furthermore, the free alternatives frequently have inadequate privacy policies.

9.2 Secure Email Providers

The transmission of personal information through email is deemed unsafe due to the lack of end-to-end encryption, which renders the data susceptible to interception and collection. Furthermore, as previously demonstrated in this study, Gmail permits third-party entities to examine private emails for the purpose of displaying personalized advertisements. Consequently, certain email service providers have implemented encryption features to enhance communication security.

Currently, there are a multitude of secure email providers that are available, which can make it difficult for individuals to determine the most appropriate choice. Certain users may prefer a service that provides the highest level of security or strong encryption, while others may prioritize simplicity and convenience.

There are several factors that should be taken into account when transitioning to a secure email provider:

- **Features**. When assessing the platform, users may wish to take into account certain features such as contact management, calendars, file storage, inbox search options, and collaborative tools.
- **Privacy and Security.** The security standards and policies of the provider with regard to user data carry significant importance. This encompasses their logging of user data, the reasons behind it, the methods employed, and the duration of its storage.
- **Pretty Good Privacy (PGP) encryption program support**. Certain email service providers offer support for PGP encryption, whereas others refrain from using it due to its vulnerabilities.
- **Contacts Import feature** availability.
- **Encryption.** Prior to transitioning to a service provider, it is imperative to verify that electronic communications, along with any attached documents, are encrypted through the entirety of their transmission.

- **Jurisdiction**. Users should take note of the service’s location and the storage location of their data, as these aspects are directly correlated with the protection of user confidentiality. The selection of the headquarters location should adhere to the same principles as those applied to VPNs.

The following is a compilation of selected secure email providers, accompanied by relevant information about each:

- **ProtonMail**, a secure email provider based in Switzerland, operates on both iOS and Android platforms. It is an open-source service with end-to-end encryption and a principle of zero knowledge and zero access. This means that staff and email servers cannot access or read your emails. However, the search function is limited to subject lines only, and the free version supports 500 MB of email storage and 150 messages a day. ProtonMail does not encrypt email subject lines. [62], [63]
- **Tutanota**, an email service based in Germany, employs AES and RSA encryption standards to guarantee secure communication between senders and recipients. The encryption key remains confidential and inaccessible to anyone else. However, it should be noted that there is no search function available for previous emails, and PGP and IMAP are not supported. To compensate for the lack of IMAP support, Tutanota has introduced desktop clients for Windows, Linux, and macOS. [62], [63]
- **Mailfence** is a reliable email service provider based in Belgium and equipped with a variety of features such as a calendar, file storage, and PGP encryption support. The free version of Mailfence provides users with 500 MB of file storage. However, it is important to note that Mailfence's software is not open source, hence it cannot be examined. Additionally, Mailfence stores the private encryption keys of its users on its servers but guarantees that these keys are encrypted using the users' passphrases, thereby making them inaccessible to unauthorized parties. [62], [63]

Each of the aforementioned providers offers paid versions of their service; however, free versions suffice for an individual of average usage. [62], [63]

9.3 Privacy-oriented browsers

In comparison to the widely known Incognito or Private Browsing feature in standard web browsers, private browsing options provide numerous benefits, given that the former only removes browsing history from a particular session. After a user ends a private session, browsing history, cookies, and passwords are erased; however, this mode does not provide protection against website tracking.

Google Chrome has dominated the browser market for the past few years and today accounts for more than 70% of the browser market [64]. However, despite its popularity, it is not necessarily the most advanced browser in terms of functionality and other parameters. Other browsers, such as Firefox, Opera, and Tor, offer certain unique features that are not available in Google Chrome.

By utilizing web browsers that are specifically designed to prioritize the protection of personal data, individuals can ensure the preservation of their security, privacy, and online identity. Given that the majority of internet activity is conducted through a browser, it is crucial that they are equipped to handle private information securely. An optimal browser should balance privacy and security while retaining ease of use and functionality. Ultimately, the choice of the most suitable browser will depend on the specific needs and preferences of the user.

The browser overviews provided here will reference such things as browser fingerprinting, HTTPS protocol and tracking protection.

The core idea behind browser fingerprinting is gathering information particular to a device, referred to as the device's "fingerprint," for the purposes of identification or increased security. Such data may include the operating system, hardware, and browser configuration. Nonetheless, the precise extent of this concept is subject to constant change due to the limitations imposed by current web browser technology. Browser fingerprinting is categorized as entirely stateless, unlike other means of identification, such as cookies, which rely on a distinct identifier that is stored within the browser itself. Since no data is stored, it does not have any discernible trace. [65]

The core protocol used for communication between webpages and web browsers is called the Hypertext Transfer Protocol (HTTP). However, data shared through that protocol is unsecured and transmitted in plain text, rendering it susceptible to being viewed or manipulated by malicious entities. This vulnerability is addressed by utilizing HTTPS, which employs the Transport Layer Security (TLS) protocol to create a secure and encrypted connection between the browser and the website. [66]

Various browser modes, such as HTTPS-Only or HTTPS Everywhere, may have differing labels based on the browser, but they share the same fundamental principle. They prioritize secure connection (HTTPS) while establishing connection with websites and only switch to unsecured connection (HTTP) when secure connection is unavailable. Moreover, these modes ensure the security of all elements of a website, including images and scripts, by upgrading them to HTTPS if possible. [66]

If a website employs HTTPS but contains some unencrypted segments, it may encounter functionality issues in HTTPS-Only Mode and not be displayed correctly. In such instances, the user can temporarily disable HTTPS-Only Mode by clicking the lock icon next to the website address in the URL bar. [66]

Private web browsers may include an integrated tracking protection feature, which prevents content from domains that monitor user activity from being loaded. Such domains typically include third-party advertising and analytics sites. However, other website components relying on these trackers will also be impacted when tracking protection is active. In such cases, users can disable tracking protection in their browser settings.

The following three web browsers are among the potential alternatives a user might want to consider.

- **Brave** is based on Chromium and was made with a heavy emphasis on privacy. It comes equipped with features like built-in HTTPS Everywhere, tracking protection, ad and script blocking, and anti-fingerprinting. Recent studies indicate that Brave is the most secure browser when it comes to telemetry data collection and transmission to its developers. Moreover, Brave disables third-party cookies by default and offers the option to turn off cookie consent messages. [67], [68]

  As Brave is built upon Chromium, its users are able to use standard Chrome browser extensions. [67], [68]

  Additionally, Brave offers some features that may be considered controversial. For instance, Brave Rewards provides users with the opportunity to earn Brave's proprietary cryptocurrency by choosing to view advertisements from commercial partners. [67], [68]
- **Firefox** is an open-source browser that has been security audited. It prioritizes the privacy of its users by incorporating features that enable better tracking avoidance and resistance to fingerprinting. However, in order to utilize the fingerprinting resistance option, the user must enable it in their browser configuration. Additionally,

Firefox's flexibility is enhanced by its compatibility with a diverse array of third-party add-ons, some of which provide notable privacy benefits. [69], [70], [71]

There is, however, evidence that Firefox tags telemetry data with IDs unique to each browser instance; however, it is possible to deactivate telemetry functionality in Firefox.

Mozilla is continuously working towards improving Firefox. They have recently implemented Total Cookie Protection as a default feature and provided the option to remove tracking parameters from URLs. Total Cookie Protection effectively limits the scope of cookies to their originating website, preventing tracking entities from using cookies to monitor user activity on different websites. Firefox is currently the only significant free and open-source competitor to Chrome, and the decline in its market share over the years is alarming. [69], [70]

- **Chromium**, the foundation upon which Google Chrome was established, is an open-source framework that, despite being associated with Google, remains independent from the company's data gathering strategies. Concerns have been raised about the possibility of hidden Google code in the extensive Chromium code base. Nonetheless, one key benefit of Chromium is its ability to utilize Chrome extensions. **Error! Reference source not found.**

  Chromium undergoes frequent updates, with a new release being made available on a daily basis. Although this is advantageous in promptly addressing any potential vulnerabilities, it does require end-users to manually install them. **Error! Reference s ource not found.**

Presented below is a table with the results of benchmarking tests of those three browsers without any additional extensions installed (Table 3).

BENCHMARK	**BRAVE**	**MOZILLA FIREFOX**	**CHROMIUM**
JetStream 2.1	164	102	154
MotionMark 1.2	1249	792	1007
Speedometer	186	139	132

Table 3 Browser benchmark comparison (higher score is better)

JetStream 2.1 is a web-based tool that combines a variety of JavaScript and Web Assembly benchmarks to evaluate a browser's capacity to carry out complex tasks. These benchmarks gauge separate workloads and generate individual scores for each. JetStream then aggregates these scores to determine a comprehensive score. [73]

MotionMark is a web-based performance assessment tool that emphasizes graphics processing capabilities over JavaScript. It employs various rendering elements that are representative of commonly utilized web techniques; the variations among these elements prevent browser optimizations utilizing cache. To provide a singular score, MotionMark calculates the geometric mean of the results of each module. [74]

Speedometer is a performance metric used to assess the level of responsiveness of web browsers through the measurement of simulated user interactions across a range of workloads, with the aim of replicating the real-world web experience as closely as possible. [75]

Superior performance is indicated by higher scores across all three benchmarks, with Brave browser demonstrating the greatest speed when compared to the other two browsers.

9.4 Browser privacy settings

Per the data provided by W3Counter, the three top positions in the Web Browser Market Share as of April 2023 are occupied by Google Chrome, Safari, and Firefox. Given that these browsers are widely utilized by users, this section is dedicated to suggested strategies for enhancing their privacy.

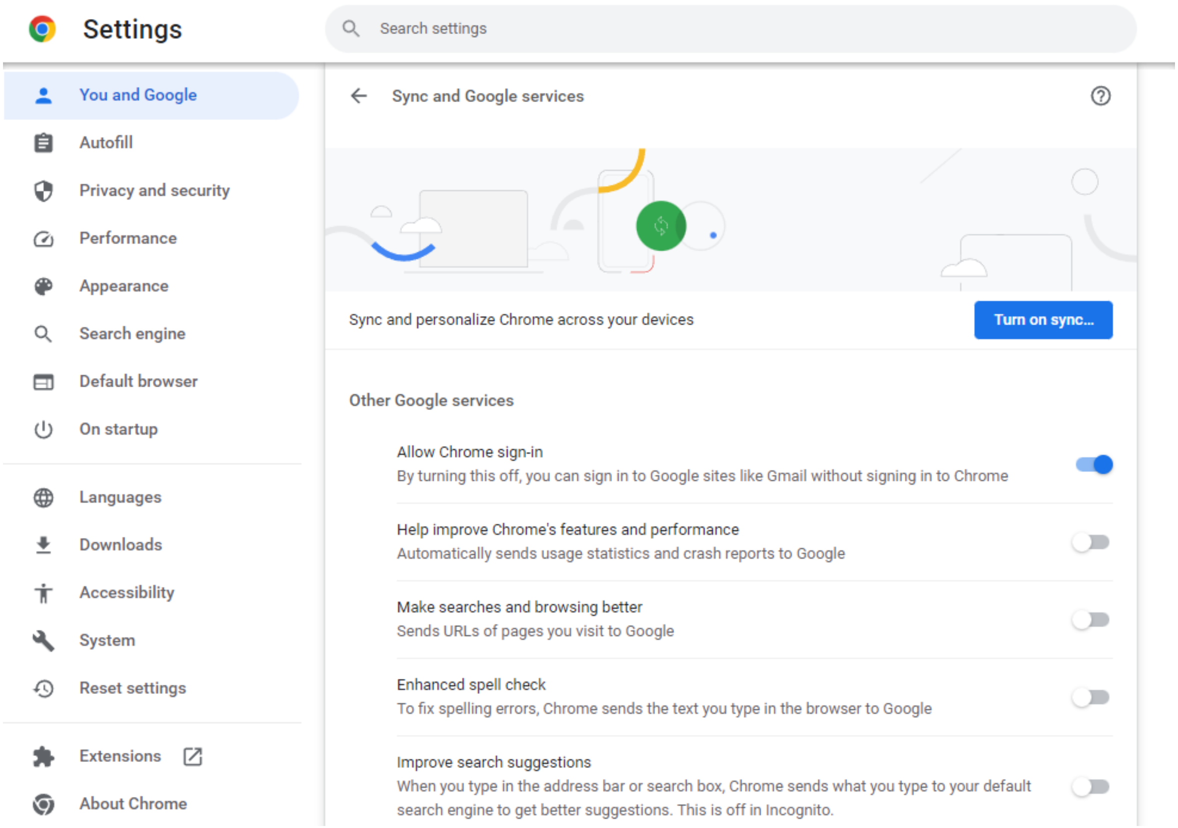
9.4.1 Google Chrome

Chrome provides various functionalities that entail transmitting information to Google's servers, and although it has implemented certain privacy configurations, they are not activated by default. Users have the option to adjust these settings according to their preferences.

Following is the recommended approach:

- Choose *Settings > Privacy and Security > Cookies and other site data* and turn on *Send a "Do Not Track" request with your browsing traffic.*
- In *Settings > Privacy and Security > Site settings,* users can view websites that requested permission to use location, microphone, camera, motion sensors, etc., and modify those permissions if needed.
- Choose *Settings > Search engine.* Here you should either choose DuckDuckGo from the available options or click *Manage search engines and site search* and set whatever other safe search engine you prefer.

- Choose *Settings > Privacy and security > Security* and turn on *Always use secure connections.*
- Choose *Settings > You and Google > Sync and Google services,* then click *Control how your browsing history is used to personalize Search and more.* This will open a new window and redirect you to *Google Account Activity Controls.* Here you can turn off *Web Activity*.
- Choose *Settings > You and Google* (Figure 14), and turn off *the following options:*
    - Help improve Chrome's features and performance
    - Make searches and browsing better
    - Enhanced spell check
    - Improve search suggestions



Sync and Google services

Sync and personalize Chrome across your devices

Other Google services

Allow Chrome sign-in ●
By turning this off, you can sign in to Google sites like Gmail without signing in to Chrome

Help improve Chrome's features and performance ○
Automatically sends usage statistics and crash reports to Google

Make searches and browsing better ○
Sends URLs of pages you visit to Google

Enhanced spell check ○
To fix spelling errors, Chrome sends the text you type in the browser to Google

Improve search suggestions ○
When you type in the address bar or search box, Chrome sends what you type to your default search engine to get better suggestions. This is off in Incognito.

Figure 14 Google Chrome Settings

- While in the *Google Account Activity Controls,* you can also choose to turn off *Location History.*
- Choose *Settings > Extensions.* This page allows you to see extensions that are already installed. From here click on *Main Menu* in the top left corner and then open *Chrome Web Store.* Here you can choose to install privacy-focused extensions (see recommendations below).

- To clear browsing data go to *Settings > Privacy and Security > Clear browsing data,* check the options you want and click *Clear data.*

Extensions recommended for Google Chrome (based on functionality and ratings):

- The Cleaner - delete Cookies and Cache. Available from: *https://chrome.google.com/webstore/detail/the-cleaner-delete-cookie/ogfjgagnmkiigilnoiabkbbajinanlbn*
  Offers quick access to cleaning cookies, cache, downloads and history. Has an option of automatic cleaning at set times.
- Privacy Badger. Available from: *https://chrome.google.com/webstore/detail/privacy-badger/pkehgijcmpdhfbdbbnkijodmdjhbjlgp*
  An open-source browser extension that automatically discovers trackers based on their behavior and blocks them.
- DuckDuckGo Privacy Essentials. Available from: *https://chrome.google.com/webstore/detail/duckduckgo-privacy-essent/bkdgflcldnnnapblkhphbgpggdiikppg*
  Automatically stops trackers from loading, enforces the use of HTTPS connection, blocks most email trackers, removes tracking elements from URLs, helps prevent browser fingerprinting, etc.
- Disconnect. Available from: *https://chrome.google.com/webstore/detail/disconnect/jeoacafpbcihiomhlakheieifhpjdfeo*
  An open-source extension focused on blocking trackers.
- uBlock Origin. Available from: *https://chrome.google.com/webstore/detail/ublock-origin/cjpalhdlnbpafiamejdnhcphjbkeiagm*
  A content blocker, the user can set which blocking lists to activate. Blocks trackers, advertisements, etc. A user can additionally set custom rules.
- AdGuard AdBlocker. Available from: *https://chrome.google.com/webstore/detail/adguard-adblocker/bgnkhhnnamicmpeenaelnjfhikgbkllg*

A content blocker for common third-party tracking systems, spyware and adware. Sets protections against malware and phishing.

- ClearURLs. Available from: *https://chrome.google.com/webstore/detail/clearurls/lckanjgmijmafbedllaakclka-icjfmnk*
  Automatically removes tracking elements from URLs.

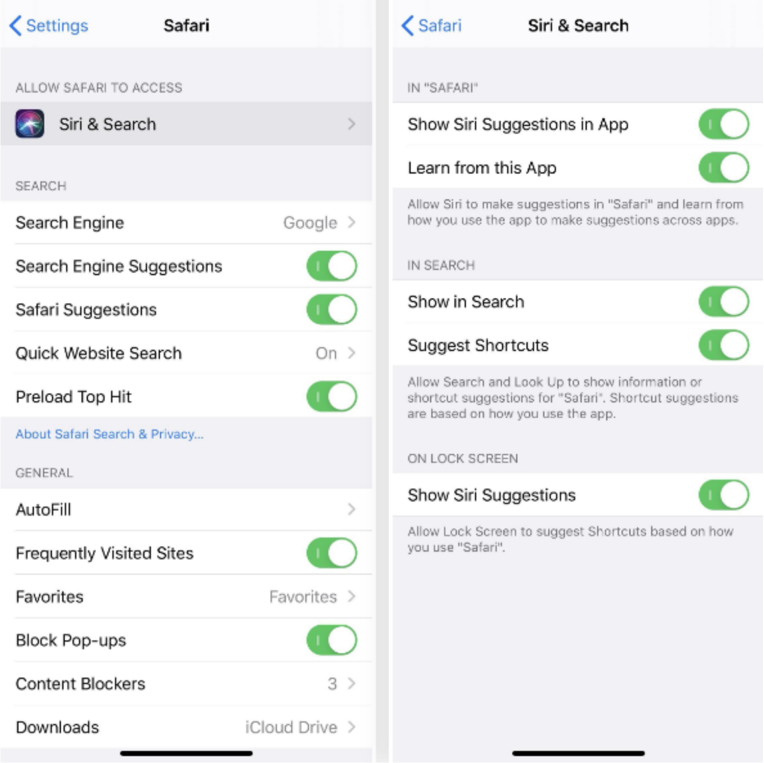
It should be noted that the browser's speed tends to decrease as more extensions are added. Therefore, it is advisable to install only those extensions that serve the user's specific needs and to limit the number of extensions that perform the same function to a single one. For example, there is no need to install multiple content blockers. Additionally, the use of extensions may, in some instances, cause web pages not to be displayed properly or function incorrectly.

9.4.2 Safari

Apple emphasizes privacy as a fundamental part of its marketing strategy in its advertising and product launches; however, in order to achieve optimal privacy on Safari, users must modify certain browser settings. The recommended approach is as follows:

- Choose *Safari* > *Settings* > *Search*, then change the default search engine choice from Google to a more confidential alternative such as DuckDuckGo.
- Choose *Safari* > *Settings* > *Privacy*, and turn on *Prevent Cross-Site Tracking.* The browser uses machine learning algorithms to identify where websites and companies get user data, subsequently preventing data from being shared on other platforms. Furthermore, it restricts other companies from accessing any user data that is gathered by Apple.
- Choose *Safari* > *Settings* > *Privacy*, and turn off *Privacy Preserving Ad Measurement.* This will deactivate the tool for advertisers.
- Choose *Safari* > *Settings* > *Privacy* and turn on Hide IP address (this option is only available to iCloud+ subscribers).
- Choose *Settings* > *Safari* > *Privacy* and set *Camera*, *Microphone*, and *Location* to *Ask*. This allows users to choose whether to grant individual websites access or not.
- Choose *Settings* > *Safari* > *General*, then click on *Extensions.* This is where you can install extensions for Safari.

- To access a record of data trackers that have been prevented by Safari and the websites from which they originated, choose *Safari > Privacy Report*.
- To clear history and website data go to *Settings > Safari* and click *Clear History and Website Data,* then choose *Clear History and Data* in the pop-up.
- Choose *Settings > Safari >Siri* (Figure 15), and turn off the following six options:
    - *Show Siri Suggestions in App;*
    - *Learn from this App;*
    - *Show in Search;*
    - *Show App;*
    - *Suggest Shortcuts;*
    - *Show Siri Suggestions.*



< Settings Safari

ALLOW SAFARI TO ACCESS

Siri & Search >

SEARCH

Search Engine Google >

Search Engine Suggestions ●

Safari Suggestions ●

Quick Website Search On >

Preload Top Hit ●

About Safari Search & Privacy...

GENERAL

AutoFill >

Frequently Visited Sites ●

Favorites Favorites >

Block Pop-ups ●

Content Blockers 3 >

Downloads iCloud Drive >

< Safari Siri & Search

IN "SAFARI"

Show Siri Suggestions in App ●

Learn from this App ●

Allow Siri to make suggestions in "Safari" and learn from how you use the app to make suggestions across apps.

IN SEARCH

Show in Search ●

Suggest Shortcuts ●

Allow Search and Look Up to show information or shortcut suggestions for "Safari". Shortcut suggestions are based on how you use the app.

ON LOCK SCREEN

Show Siri Suggestions ●

Allow Lock Screen to suggest Shortcuts based on how you use "Safari".

Figure 15 Siri Settings

Suggested Safari extensions (descriptions added only for the extensions not mentioned previously):

- Consent-O-Matic. Available from:
  https://apps.apple.com/nl/app/consent-o-matic/id1606897889
  An extension that answers cookie consent pop-ups automatically based on preferences set by the user.

- DuckDuckGo Privacy for Safari. Available from: *https://apps.apple.com/nl/app/duckduckgo-privacy-for-safari/id1482920575?mt=12*
- StopTheMadness. Available from: *https://apps.apple.com/nl/app/stopthemadness/id1376402589?mt=12*
  Removes trackers from URLs and blocks trackers.
- Norton AntiTrack. Available from: *https://apps.apple.com/nl/app/norton-antitrack/id1610227172*
  Prevents advanced browser fingerprinting and blocks trackers.
- Cookie DNT Privacy for Safari. Available from: *https://apps.apple.com/nl/app/cookie-dnt-privacy-for-safari/id1594049656*
  Automatically cleans cookies at a time interval configured by the user.
- PrivacyScan. Available from: *https://apps.apple.com/nl/app/privacyscan/id494950833?mt=12*
  Erases cache, browsing history, cookies, and temporary files.
- AdGuard for Safari. Available from: *https://apps.apple.com/nl/app/adguard-for-safari/id1440147259?mt=12*
- 1Blocker - Ad Blocker. Available from: *https://apps.apple.com/nl/app/1blocker-ad-blocker/id1365531024*
  A content blocker for trackers and advertisements.
- AdBlock Pro: Browser AdBlocker. Available from: *https://apps.apple.com/nl/app/adblock-pro-browser-adblocker/id1018301773*
  An extension that blocks advertisements, third-party trackers and bypasses anti-adblock detectors.

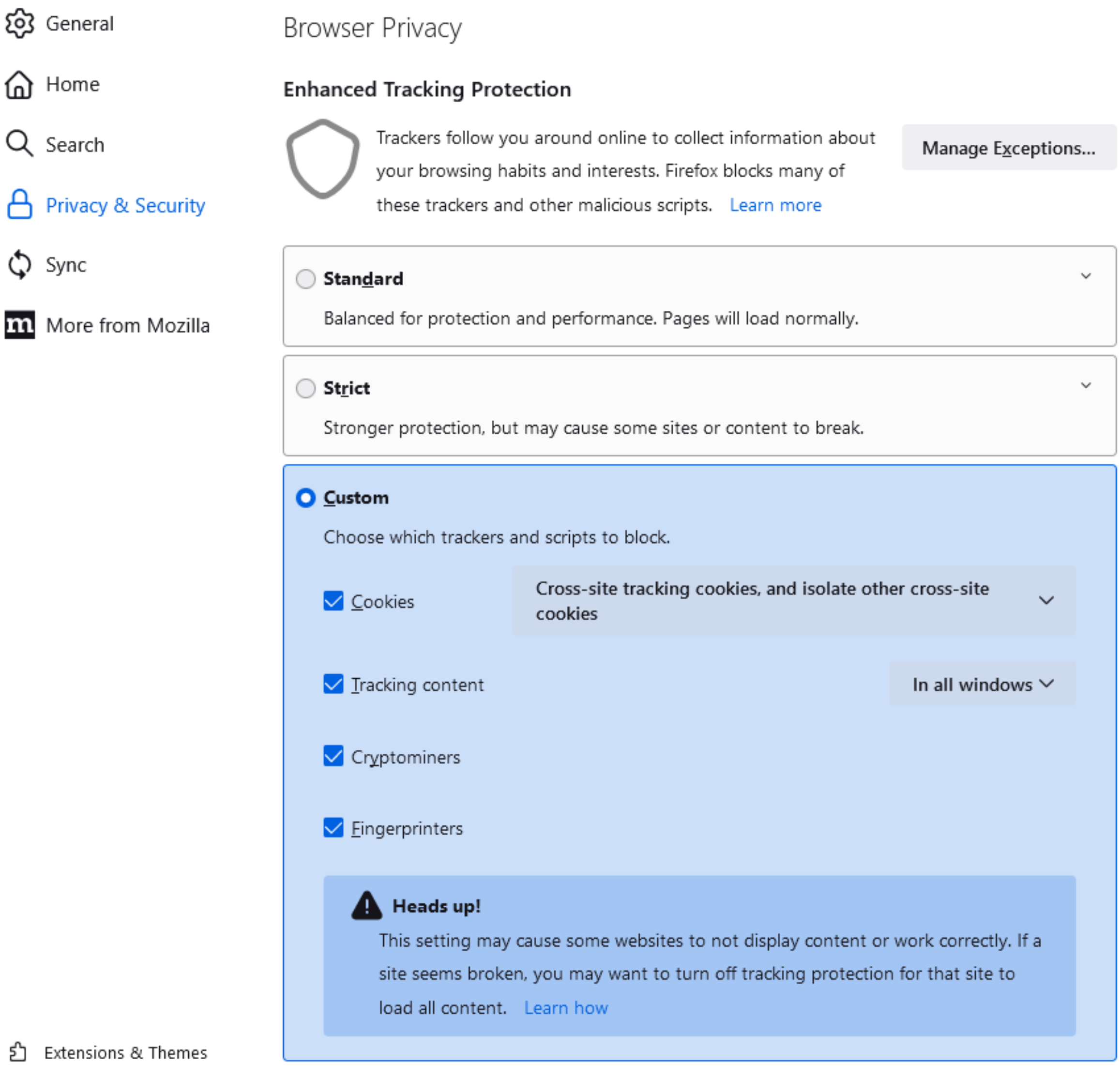
9.4.3 Firefox

The default privacy settings of Firefox are comparatively more robust than Chrome, and the browser provides a wider range of privacy options that can be modified.

Below are the proposed recommendations:

- Choose *Settings > Search*. In this tab, change *Default Search Engine* to DuckDuckGo or other private engine of your preference.
- Choose *Settings > Privacy & Security > HTTPS-Only Mode* then toggle *Enable HTTPS-Only Mode in all windows.*

- Choose *Settings > Privacy & Security > Browser Privacy.* In *Enhanced Tracking Protection,* choose *Custom* and check all of the available boxes. Then next to *Cookies,* make sure that *Cross-site tracking cookies and isolate other cross-site cookies* option is chosen, and next to *Tracking content > In all windows* (Figure 16).



Enhanced Tracking Protection

Trackers follow you around online to collect information about your browsing habits and interests. Firefox blocks many of these trackers and other malicious scripts. Learn more

○ Standard

Balanced for protection and performance. Pages will load normally.

○ Strict

Stronger protection, but may cause some sites or content to break.

● Custom

Choose which trackers and scripts to block.

☑ Cookies: Cross-site tracking cookies, and isolate other cross-site cookies

☑ Tracking content: In all windows

☑ Cryptominers

☑ Fingerprinters

Heads up!

This setting may cause some websites to not display content or work correctly. If a site seems broken, you may want to turn off tracking protection for that site to load all content. Learn how

Figure 16 Firefox Privacy & Security

- Same as with Chrome you can change permissions granted to websites in *Settings > Privacy & Security > Permissions.*
- Go to *about:config* through the address bar. If a warning appears choose *Accept the Risk and Continue.* Search for *privacy.resistFingerprinting* and set it to *True*.
- In the same *about:config* menu search for *privacy.query_stripping.enabled.pbmode* and set it to *True* to enable stripping URLs of tracking parameters.

- Choose *Settings* > *Extensions and Themes* > *Extensions.* Again, this is where you can find and install privacy-enhancing extensions.
- To clear site data go to *Settings* > *Privacy & Security* > *Cookies and Site Data* and click *Clear data,* check the necessary boxes and click *Clear*.

Recommended Firefox extensions:

- Cookie AutoDelete. Available from:
  https://addons.mozilla.org/en-US/firefox/addon/cookie-autodelete/
  Automatically deletes any cookies that aren't being used after a tab is closed unless otherwise configured by the user.
- DuckDuckGo Privacy Essentials. Available from:
  https://addons.mozilla.org/en-US/firefox/addon/duckduckgo-for-firefox/
- Privacy Badger. Available from:
  https://addons.mozilla.org/en-US/firefox/addon/privacy-badger17/
- ClearURLs. Available from:
  https://addons.mozilla.org/en-US/firefox/addon/clearurls/
- Facebook Container. Available from:
  https://addons.mozilla.org/en-US/firefox/addon/facebook-container/
  An extension that isolates Facebook pages into a separate container effectively preventing the tracking of user activity on external websites through third-party cookies.
- AdGuard AdBlocker. Available from:
  https://addons.mozilla.org/en-US/firefox/addon/adguard-adblocker/
- AdBlocker Ultimate. Available from:
  https://addons.mozilla.org/en-US/firefox/addon/adblocker-ultimate/
  An extension that blocks trackers, advertisements and malware.
- uBlock Origin. Available from:
  https://addons.mozilla.org/en-US/firefox/addon/ublock-origin/

9.5 Recommendations

To conclude this chapter, presented below is a compilation of recommendations derived from all the preceding advice:

- Use a secure browser when possible (sometimes individuals are required to use a specific browser for their work; in those cases, having two browsers installed is a suitable option);
- Clear website cookies periodically;
- When presented with a choice only accept essential cookies;
- Opt out of personalized advertisement;
- Install browser extensions meant for blocking trackers and advertisements:
    - for Google Chrome: DuckDuckGo Privacy Essentials and uBlock Origin;
    - for Safari: DuckDuckGo Privacy, Consent-O-Matic and 1Blocker;
    - for Firefox: Cookie AutoDelete, DuckDuckGo Privacy Essentials and uBlock Origin;
- Use a privacy-focused search engine;
- Enable privacy features in the browser settings;
- When using mobile applications, modify their privacy setting to suit you;
- For an additional layer of privacy, use a VPN.

CONCLUSION

The aim of the research was to examine the interconnection between the benefits provided by Big Data and the right of online consumers to maintain their privacy.

Big data has revolutionized the way we collect, analyze, and utilize information. Nevertheless, as the quantity of data being collected continues to increase, apprehensions about privacy have gained more prominence. The emergence of digital technologies has simplified the process of monitoring and tracing individuals' online actions, which could result in the improper use of sensitive information.

It appears that over time, this matter may increasingly gain attention and be more widespread. Therefore, it seems improbable that this subject will lose relevance in the near future.

The significance of digital privacy and its importance for governments, corporate entities, and users was emphasized throughout this work. By analyzing the privacy policies and data handling strategies of major tech companies, it was found that there is still room for improvement despite the recent public concern for online privacy. It is important to note that the practical approach of businesses may differ from their marketing campaigns, and legislation on privacy standards may vary depending on the region.

Several methods exist for users to uphold their privacy, which are outlined in the practical segment of this thesis through a set of recommendations.

BIBLIOGRAPHY

[1] *What is Big Data?* Online. Oracle. [n.d.]. Available from: https://www.oracle.com/big-data/what-is-big-data/. [viewed 2023-04-27].

[2] UW EXTENDED CAMPUS. *What is Big Data?* Online. University of Wisconsin. 2015-05-18. Available from: https://uwex.wisconsin.edu/stories-news/what-is-big-data/. [viewed 2023-04-27].

[3] CRISTOBAL, Samuel. *Two more V's in Big Data: Veracity and Value*. Online. Datascience.aero. 2020-06-17. Available from: https://datascience.aero/big-data-veracity-value/. [viewed 2023-04-27].

[4] GUPTA, Shibakali, Indradip BANERJEE, and Siddhartha BHATTACHARYYA (eds.). *Big Data Security*. Online. De Gruyter, 2019. ISBN 9783110606058. Available from: https://doi.org/10.1515/9783110606058. [viewed 2023-04-27].

[5] *Veracity in Big Data*. Online. BigDataScalability. 2023-04-21. Available from: https://bigdatascalability.com/veracity-in-big-data/. [viewed 2023-04-27].

[6] BANSAL, Sumeet. *Big Data Analytics: Key Aspects One Must Know*. Online. Analytixlabs. 2020-11-04. Available from: https://www.analytixlabs.co.in/blog/big-data-analytics/. [viewed 2023-04-27].

[7] MYSORE, Divakar, Shrikant KHUPAT, and Shweta JAIN. *Introduction to big data classification and architecture*. Online. IBM Developer. 2013-09-16. Available from: https://developer.ibm.com/articles/bd-archpatterns1/. [viewed 2023-04-27].

[8] *Understand data store models*. Online. Microsoft Learn. 2022-09-02. Available from: https://learn.microsoft.com/en-us/azure/architecture/guide/technology-choices/data-store-overview. [viewed 2023-04-27].

[9] *A Complete Guide for Processing of Data*. Online. JanbaskTraining. 2020-04-09. Available from: https://www.janbasktraining.com/blog/data-processing/. [viewed 2023-04-27].

[10] DUTKA, Vitalii. *Big Data and Transportation: Imperative Use Cases*. Online. Intellias. 2021-10-06. Available from: https://intellias.com/big-data-and-transportation-use-cases-urban-planning/. [viewed 2023-04-27].

[11] TAKER, Shayne. *Big Data for Building Energy Management*. Online. Buildings IOT. [n.d.]. Available from: https://www.buildingsiot.com/blog/what-big-data-can-do-for-building-energy-management-bd. [viewed 2023-04-27].

[12] *Relevance of Big data in Healthcare*. Online. James Lind Institute. 2023-01-11. Available from: https://jliedu.ch/blog/relevance-of-big-data-in-healthcare/. [viewed 2023-04-27].

[13] *Big Data and Implications of Behavioral Economics*. Online. James Lind Institute. 2023-02-01. Available from: https://jliedu.ch/blog/big-data-and-implications-of-behavioral-economics/. [viewed 2023-04-27].

[14] FARMER, Donald. *8 Benefits of Using Big Data for Businesses*. Online. TechTarget. 2022-02-23. Available from: https://www.techtarget.com/searchbusinessanalytics/feature/6-big-data-benefits-for-businesses. [viewed 2023-04-27].

[15] MARR, Bernard. *4 Ways Big Data Will Change Every Business*. Online. Forbes. 2015-09-08. Available from: https://www.forbes.com/sites/bernardmarr/2015/09/08/4-ways-big-data-will-change-every-business/. [viewed 2023-04-27].

[16] NARANG, Mounika. *Top 18 Advantages of Big Data*. Online. KnowledgeHut. 2023-01-23. Available from: https://www.knowledgehut.com/blog/big-data/advantages-of-big-data. [viewed 2023-05-08].

[17] *Data Breach QuickView Report*. Press release. Risk Based Security, 2019. Available from: https://f.hubspotusercontent-eu1.net/hubfs/24969859/Merlin%20Ventures/Merlin%20Ventures%20WebPages/Pdf/2019%20Q1%20Data%20Breach%20QuickView%20Report.pdf. [viewed 2023-04-27].

[18] PARR, Alastair. *Third-Party Data Breaches: What You Need to Know*. Online. Prevalent. 2022-10-11. Available from: https://www.prevalent.net/blog/third-party-data-breaches/. [viewed 2023-04-27].

[19] HILLER, Will. *Is Big Data Dangerous? The Risks Uncovered*. Online. CareerFoundry. 2022-12-15. Available from: https://careerfoundry.com/en/blog/data-analytics/is-big-data-dangerous/. [viewed 2023-04-27].

[20] IPHOFEN, Ron, and Dónal O'MATHÚNA (eds.). *Ethical Issues in Covert, Security and Surveillance Research*. Online. Emerald Publishing Limited, 2021. ISBN

9781802624144. Available from: https://doi.org/10.1108/s2398-6018202108. [viewed 2023-04-27].

[21] SREEHARSHA, Vinod. *Google and Yahoo Win Appeal in Argentine Case*. Online. The New York Times. 2010-08-20. Available from: https://www.nytimes.com/2010/08/20/technology/internet/20google.html. [viewed 2023-04-27].

[22] MANTELERO, Alessandro. *The EU Proposal for a General Data Protection Regulation and the roots of the 'right to be forgotten.'* Online. Computer Law & Security Review, vol. 29 (June 2013), no. 3, pp. 229–235. ISSN 0267-3649. Available from: https://doi.org/10.1016/j.clsr.2013.03.010. [viewed 2023-04-27].

[23] REPUBLIC OF THE PHILIPPINES. National Privacy Commission. *IMPLEMENTING RULES AND REGULATIONS OF REPUBLIC ACT NO. 10173*. Online. 2016-08-16. Available from: https://www.privacy.gov.ph/implementing-rules-regulations-data-privacy-act-2012/. [viewed 2023-04-27].

[24] HILL, Michael, and Dan SWINHOE. *The 15 biggest data breaches of the 21st century*. Online. CSO Online. 2022-11-08. Available from: https://www.csoonline.com/article/2130877/the-biggest-data-breaches-of-the-21st-century.html. [viewed 2023-04-27].

[25] DE GROOT, Juliana. *The History of Data Breaches*. Online. Digital Guardian. 2022-08-22. Available from: https://www.digitalguardian.com/blog/history-data-breaches. [viewed 2023-04-27].

[26] PETROSYAN, Ani. *Global number of breached data sets 2020-2022*. Image. 2022-11-29. Available from: https://www.statista.com/statistics/1307426/number-of-data-breaches-worldwide/. [viewed 2023-04-27].

[27] LIPMAN, Paul. *The Evolution Of Online Consumer Privacy*. Online. Forbes. 2019-03-29. Available from: https://www.forbes.com/sites/forbestechcouncil/2019/03/29/the-evolution-of-online-consumer-privacy/. [viewed 2023-05-07].

[28] KUSHMARO, Philip. *Why Data Privacy Is A Human Right (And What Businesses Should Do About It)*. Online. Forbes. 2021-06-07. Available from: https://www.forbes.com/sites/forbescommunicationscouncil/2021/06/07/why-data-privacy-is-a-human-right-and-what-businesses-should-do-about-it/. [viewed 2023-05-07].

[29] *Data privacy laws: What you need to know in 2023*. Online. Osano. 2022-12-14. Available from: https://www.osano.com/articles/data-privacy-laws. [viewed 2023-05-07].

[30] SAFANE, Jake. *The Federal Trade Commission Act sets the guidelines underpinning the FTC's consumer-protection enforcement*. Online. Business Insider. 2022-08-02. Available from: https://www.businessinsider.com/personal-finance/federal-trade-commission-act. [viewed 2023-05-07].

[31] MIDDLETON-LEAL, Matt. *What is GDPR: 10 Frequently Asked Questions*. Online. Netwrix Blog. 2023-03-17. Available from: https://blog.netwrix.com/2018/02/06/what-is-the-general-data-protection-regulation-gdpr-10-frequently-asked-questions/. [viewed 2023-05-07].

[32] BRACY, Jedidiah. *EU Council ambassadors agree to negotiating position on ePrivacy Regulation*. Online. International Association of Privacy Professionals. 2021-02-10. Available from: https://iapp.org/news/a/eu-council-ambassadors-agree-to-negotiating-position-on-eprivacy-regulation/. [viewed 2023-05-07].

[33] *Comparison Charts: U.S. State vs. EU Data Privacy Laws*. Online. Bloomberg Law. 2023-05-03. Available from: https://pro.bloomberglaw.com/brief/data-privacy-laws-in-the-u-s/. [viewed 2023-05-07].

[34] DEARIE, KJ. *Opt In vs. Opt Out*. Online. Termly. 2021-09-28. Available from: https://termly.io/resources/articles/opt-in-vs-opt-out/. [viewed 2023-05-07].

[35] KOCH, Richie. *Cookies, the GDPR, and the ePrivacy Directive*. Online. GDPR.eu. [n.d.]. Available from: https://gdpr.eu/cookies/. [viewed 2023-05-07].

[36] *Data Protection & Privacy Laws Around the World*. Online. Securiti. 2021-11-08. Available from: https://securiti.ai/data-privacy-laws/. [viewed 2023-05-07].

[37] CZARNOCKI, Jan, Flavia GIGLIO, Eyup KUN, Mykyta PETIK, and Sofie ROYER. *Government access to data in third countries*. Online. Brussels: Milieu Consulting SRL, 2021. Available from: https://edpb.europa.eu/system/files/2022-01/legalstudy_on_government_access_0.pdf. [viewed 2023-05-08].

[38] *Как защищены персональные данные россиян*. Online. Государственная Дума. 2021-01-14. Available from: http://duma.gov.ru/news/50497/. [viewed 2023-05-08].

[39] REPUBLIC OF BELARUS. The House of Representatives of the National Assembly. *Law on Personal Data Protection*. Online. Law No. 99-Z of 2021-05-07. Available from: https://etalonline.by/document/?regnum=h12100099&amp;q_id=6232166. [viewed 2023-05-08].

[40] *Belarus: Freedom on the Net 2021 Country Report*. Online. Freedom House. [n.d.]. Available from: https://freedomhouse.org/country/belarus/freedom-net/2021. [viewed 2023-05-08].

[41] STATISTA RESEARCH DEPARTMENT. *Investment in privacy and security companies worldwide 2019*. Image. 2023-03-31. Available from: https://www.statista.com/statistics/1123238/worldwide-privacy-and-security-companies-investment/. [viewed 2023-05-07].

[42] PETROSYAN, Ani. *Global consumer aware of companies selling personal data 2022*. Image. 2023-03-07. Available from: https://www.statista.com/statistics/1369055/consumer-awareness-global-private-data-companies-sell/. [viewed 2023-05-07].

[43] PETROSYAN, Ani. *U.S. data compromises by industry 2022*. Image. 2023-03-10. Available from: https://www.statista.com/statistics/1318379/us-number-of-private-data-compromises-by-industry/. [viewed 2023-05-07].

[44] ANANT, Venky, Lisa DONCHAK, James KAPLAN, and Henning SOLLER. *The consumer-data opportunity and the privacy imperative*. Online. McKinsey & Company. 2020-04-27. Available from: https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/the-consumer-data-opportunity-and-the-privacy-imperative. [viewed 2023-05-07].

[45] PETROSYAN, Ani. *Attitudes about identity theft worldwide 2022*. Image. 2023-03-10. Available from: https://www.statista.com/statistics/296700/personal-data-security-perception-online/. [viewed 2023-05-07].

[46] VIGDERMAN, Aliza, and Gabe TURNER. *The Data Big Tech Companies Have On You*. Online. Security.org. 2023-02-06. Available from: https://www.security.org/resources/data-tech-companies-have/. [viewed 2023-05-07].

[47] *TikTok Privacy Policy*. Online. TikTok. 2023-03-21. Available from: https://www.tiktok.com/legal/page/us/privacy-policy/en. [viewed 2023-05-07].

[48] *Google vs. Apple on User Privacy*. Online. Identity Review. 2022-08-03. Available from: https://identityreview.com/user-privacy/. [viewed 2023-05-07].

[49] KAVYA. *Big Tech vs GDPR*. Online. Cookie Law Info. 2023-03-28. Available from: https://www.cookielawinfo.com/big-tech-vs-gdpr/. [viewed 2023-05-07].

[50] TANGALAKIS-LIPPERT, Katherine. *Facebook, Google Give Police Data to Prosecute Abortion Seekers*. Online. Business Insider. 2023-03-05. Available from: https://www.businessinsider.com/police-getting-help-social-media-to-prosecute-people-seeking-abortions-2023-2. [viewed 2023-05-07].

[51] FISHER, Joe. *Irish data commission issues $400 million in fines to Meta*. Online. UPI. 2023-01-04. Available from: https://www.upi.com/Top_News/World-News/2023/01/04/meta-facebook-instagram-data-protection-commission-fine/8551672862793/. [viewed 2023-05-07].

[52] HUDDLESTON JR, Tom. *TikTok shares your data more than any other social media app: study*. Online. CNBC. 2022-02-08. Available from: https://www.cnbc.com/2022/02/08/tiktok-shares-your-data-more-than-any-other-social-media-app-study.html. [viewed 2023-05-07].

[53] @KRYSTIAN3W. *uBlock vs. ABP: efficiency compared*. GitHub. 2022-11-29. Available from: https://github.com/gorhill/uBlock/wiki/uBlock-vs.-ABP:-efficiency-compared. [viewed 2023-05-07].

[54] TAYLOR, Sven. *10 Best Private Search Engines in 2023*. Online. RestorePrivacy. 2023-03-02. Available from: https://restoreprivacy.com/private-search-engine/. [viewed 2023-05-08].

[55] SYMANOVICH, Steve. *What is a VPN?* Online. Norton. 2022-02-24. Available from: https://us.norton.com/blog/privacy/what-is-a-vpn. [viewed 2023-05-07].

[56] MARKS, Tove. *5 Eyes, 9 Eyes, 14 Eyes: Protect Yourself From Global Surveillance*. Online. VPNOverview.com. 2022-11-22. Available from: https://vpnoverview.com/privacy/anonymous-browsing/5-9-14-eyes/. [viewed 2023-05-07].

[57] *The 5 Eyes, 9 Eyes, and 14 Eyes Explained*. Image. [n.d.]. Available from: https://vpnstore.com/5-eyes-9-eyes-14-eyes-explained/. [viewed 2023-05-07].

[58] *KPMG and Cure53 Audit ExpressVPN Security*. Online. ExpressVPN Blog. 2022-10-26. Available from: https://www.expressvpn.com/blog/kpmg-privacy-policy-cure53-trustedserver-audit/. [viewed 2023-05-07].

[59] SZYMCZAK, Maciej, and Jakub DARECKI. *Security report: ProtonVPN's No-Logs policy*. Online. Proton VPN. 2022-03-24. Available from: https://protonvpn.com/blog/wp-content/uploads/2022/04/securitum-protonvpn-nologs-20220330.pdf. [viewed 2023-05-07].

[60] *Independent Reasonable Assurance Report*. Online. NordVPN. 2022-12-21. Available from: https://s1.nordcdn.com/nord/misc/0.58.0/vpn/brand/ISAE_3000-NordVPN_report_20dec2022.pdf. [viewed 2023-05-07].

[61] *VPN speed test comparison*. Online. Cybernews. [n.d.]. Available from: https://cybernews.com/vpn-speed-test/. [viewed 2023-05-07].

[62] QAMAR, Ali. The 20 Most Secure Email Providers to Use in May 2023. Online. PrivacySavvy. 2023-04-30. Available from: https://privacysavvy.com/email/best/secure-email-providers/. [viewed 2023-05-08].

[63] TAYLOR, Sven. *10 Best Private and Secure Email Services for 2023*. Online. RestorePrivacy. 2023-01-02. Available from: https://restoreprivacy.com/email/secure/. [viewed 2023-05-08].

[64] *W3Counter: Global Web Stats - April 2023*. Online. W3Counter. [n.d.]. Available from: http://www.w3counter.com/globalstats.php?year=2023&amp;month=4. [viewed 2023-05-10].

[65] LAPERDRIX, Pierre, Nataliia BIELOVA, Benoit BAUDRY, and Gildas AVOINE. *Browser Fingerprinting*. Online. ACM Transactions on the Web, vol. 14 (April 2020), no. 2, pp. 1–33. ISSN 1559-114X. Available from: https://doi.org/10.1145/3386040. [viewed 2023-05-10].

[66] KERSCHBAUMER, Christoph, Julian GAIBLER, Arthur EDELSTEIN, and Thyla van der MERWE. *HTTPS-Only: Upgrading all connections to https in Web Browsers.* Online. Workshop on Measurements, Attacks, and Defenses for the Web. MADWeb, 2021. ISBN 1-891562-67-3. Available from: https://dx.doi.org/10.14722/madweb.2021.23010. [viewed 2023-05-19].

[67] PITCHKITES, Max. *Brave Browser Review*. Online. Cloudwards. 2022-12-20. Available from: https://www.cloudwards.net/brave-review/. [viewed 2023-05-10].

[68] BIDWELL, Jonni, and Lindsay PIETROLUONGO. *Brave Browser Review*. Online. TechRadar. 2021-08-18. Available from: https://www.techradar.com/reviews/brave-web-browser. [viewed 2023-05-10].

[69] PIETROLUONGO, Lindsay. *Mozilla Firefox review*. Online. TechRadar. 2021-08-12. Available from: https://www.techradar.com/reviews/mozilla-firefox. [viewed 2023-05-10].

[70] PITCHKITES, Max. *Firefox Review*. Online. Cloudwards. 2022-12-20. Available from: https://www.cloudwards.net/firefox-review/. [viewed 2023-05-10].

[71] CLAUDIUS, Jonathan. *Mozilla VPN Security Audit*. Online. Mozilla Security Blog. 2021-08-31. Available from: https://blog.mozilla.org/security/2021/08/31/mozilla-vpn-security-audit/. [viewed 2023-05-10].

[72] PITCHKITES, Max. *Chromium Review 2023*. Online. Cloudwards. 2022-12-20. Available from: https://www.cloudwards.net/chromium-review/. [viewed 2023-05-16].

[73] *JetStream 2.1 In-Depth Analysis*. Online. Browser Benchmarks. [n.d.]. Available from: https://browserbench.org/JetStream/in-depth.html. [viewed 2023-05-20].

[74] *About MotionMark*. Online. Browser Benchmarks. [n.d.]. Available from: https://browserbench.org/MotionMark/about.html. [viewed 2023-05-20].

[75] *About Speedometer 2.0*. Online. Browser Benchmarks. [n.d.]. Available from: https://browserbench.org/Speedometer2.0/. [viewed 2023-05-20].

LIST OF FIGURES

Figure 1 Big data classification [4] 12
Figure 2 Global number of breached data sets 2020-2022 [26] 21
Figure 3 Privacy Laws Around the World [29] 31
Figure 4 Dollar volume of investments in privacy and security companies worldwide from 2010 to 2019 (in billion U.S dollars) [41] 32
Figure 5 Privacy concern by data type [44] 34
Figure 6 Worldwide consumer confidence about personal online data security, December 2022 [45] 35
Figure 7 Data collection by company, unique identifiers [46] 38
Figure 8 Data collection by company, user activity [46, 47] 39
Figure 9 Data collection by company, personal information [46, 47] 40
Figure 10 Data collection by company, location information [46, 47] 41
Figure 11 Data collection by company, publicly accessible sources [46, 47] 41
Figure 12 The Eyes Programs member countries [56] 47
Figure 13 VPN speed test comparison [61] 48
Figure 14 Google Chrome Settings 55
Figure 15 Siri Settings 58
Figure 16 Firefox Privacy & Security 60

LIST OF TABLES

Table 1 Number of cases of data violation due to cyber attacks in the US from 2020 to 2022 by industry [43] 33

Table 2 VPN comparison..48

Table 3 Browser benchmark comparison (higher score is better) 53