Vilnius University Faculty of Law

Department of Public Law

Anil Berk Gumus,

II study year, International and European Law Programme Student

Master's Thesis

Legal and Ethical Considerations in Personal Data Usage for Algorithmic Training

Teisiniai ir etiniai asmens duomenų naudojimo algoritminiam mokymui aspektai

Supervisor: Lekt. dr.Paulius Jurčys

Reviewer: Lekt. dr. Eglė Lauraitytė

Vilnius

2024

ABSTRACT AND KEY WORDS

This Master's thesis analyzes the intricate legal and ethical landscapes surrounding the use of personal data in the algorithmic training in a comparative analysis of two vital data protection laws, the GDPR of the European Union and the CCPA-CPRA of the United States. The study delves into how these regulations shape the collection, processing, and utilization of personal data for developing algorithms. The study approaches to ethical frameworks to acknowledge their implications for technology and personal data use. The analysis further demonstrates the application of theoretical concepts on the matter, evidenced by precedents from case law.

Key words: data privacy, personal data protection, algorithmic training, GDPR, CCPA,CPRA.

LIST OF ABBREVIATIONS

AI: Artificial Intelligence

CCPA: California Consumer Privacy Act

CPRA: California Privacy Rights Act

DPIA: Data Protection Impact Assessment

DPO: Data Protection Officer

ECJ: European Court of Justice

EU: European Union

GDPR: General Data Protection Regulation

GAN: Generative Adversarial Networks

IS: Information Systems

IT: Information Technology

IoT: Internet of Things

US: United States

PII: Personally Identifiable Information

SPI: Sensitive Personal Information

TABLE OF CONTENTS

INTRODUCTION

"The power of the Web is in its universality. Access by everyone regardless of disability is an essential aspect," said Tim Berners-Lee, the inventor of the World Wide Web. (Berners-Lee, T., 2013) These ideas are especially meaningful today, in a time when the internet has changed from being a new invention to a common part of life, impacting virtually every aspect of modern life. The development of particularly internet, has played a major role in transforming the world. It has redefined the communication, business, and entertainment by making the world a connected digital community. Yet, this technological advancement, highlights the critical issue of data. Data, in its different types, is vital for the digital world, to fuel innovations and build the future.

Personal data is the core of this data-driven world; as a subset of data that specific to an individual. Personal data encompasses an individual's digital footprint and can range from basic identity information to more sensitive data. Personal data is; in the location data that maps our daily commute, the shopping preferences tracked by online retailers, or the health indicators recorded by our fitness apps. It's the informations we share on social media, from our birthdays to our trip pictures, which slightly shapes the advertisements we see online. This data, while often shared unassumingly, creates a thorough digital profile that affects our digital interactions. From personalized movie recommendations on streaming platforms to customized news feeds on social platforms, these are all tailored experiences driven by our own data. As we transition to discussing its role in algorithmic training, it's essential to recognize this widespread influence of personal data in shaping our online experiences.

The use of personal data gains further complexity when applied to the realm of algorithmic training. Algorithms, the sophisticated sets of rules driving much of our digital world, learn and evolve based on the data they process. In essence, the more nuanced and comprehensive the data, the more accurate and effective the algorithms can become. This process, often unseen, subtly shapes the digital services and products we interact with daily basis. Understanding their function and the role of personal data in their development is a key aspect to explore ethical, legal and practical implications of their use.

Aim of the Research

This research aims to explore the ethical, legal considerations of using personal data in algorithmic training, with a focus on how this impacts individual privacy and data protection in the context of evolving digital technologies.

Objectives of the Research

1. Examine the impact of laws such as the General Data Protection Regulation (GDPR), California Consumer Privacy Act (CCPA) and the California Privacy Rights Act (CPRA) on the use of personal data in algorithmic training.

2. Examine ethical challenges associated with using personal data in algorithms, including issues of consent, privacy, and the balance between technological advancement and individual rights.

3. Analyze how personal data is used in algorithmic decision-making, particularly focusing on the transparency, fairness, and accountability of these systems.

4. Assess the importance and impact of personal data in the development and training of artificial intelligence systems, and how this affects privacy and data protection.

5. Evaluate current measures and practices implemented for protecting personal data in the context of AI and machine learning technologies.

Research Methods

A diverse methodological approach has been adopted. The thesis commences by analyzing and defining important terms and concepts, establishing a solid foundation for the study. This inital phase involved an in-depth exploration of the contemprary uses of relevant terminology., ensuring a robust and precise conceptual framework.

Following the establishment of this conceptual groundwork, the study progressed to the literature review. This phase involved a thorough examination of existing scholarly works, allowing for the identification of gaps in the current body of knowledge and the contextualization of the research within the wider academic field. The literature review served as a cornerstone for developing informed and relevant research questions and hypotheses.

As the thesis evolved, comparative method and the legislative analysis became a central methodology. This approach was instrumental in contrasting and comparing different legal frameworks, policies, and case studies. By examining these elements side by side, the research was able to draw nuances and understand the variances and similarities across different contexts, particularly in legal and regulatory environments.

Moreover, a case law analysis was conducted to bridge the gap between theoretical frameworks and their practical application. This method involved an examination of relevant legal cases and decisions, providing real-world examples of how laws and regulations are interpreted and enforced.

Originality

While several scholarly works have delved into the topic of personal data usage in algorithmic training, this thesis sets itself apart through its comprehensive approach, particularly in addressing both the legal dimensions. As well, it sets itself apart through its critical approach to ethical implications, assessing how these ethical considerations are manifested and navigated in various practical applications.

To elaborate further, thesis provides an in-depth analysis that goes beyond the surface-level exploration common in much of the existing literature. By precisely reviewing both the legal nuances, as seen in the evolving interpretations of laws such as the GDPR, CCPA and CPRA; and the ethical implications, which include concerns of transparency, fairness, and accountability in AI systems, the work offers a comprehensive perspective that is relatively unexplored in current academic discourse. Additionally, the research incorporates actual situations and examines how they were managed in accordance with these principles.

The Most Important Sources

The most important sources for this master's thesis regarding legal acts are GDPR, CCPA, and CPRA. The case laws from both US and EU zone. These documents form the backbone of the legal analysis in the thesis, providing insights into the complexities and evolution of data protection laws.

Particularly in regards to CCPA and CPRA, Jordan, S. (2022). "Strengths and Weaknesses of Notice and Consent Requirements under the GDPR, the CCPA/CPRA, and the FCC Broadband Privacy Order" constitues an important source as the work of scholar in the research.

CHAPTER I – NAVIGATION OF CONCEPTS

1.1. Personal Data

Throughout the history, "information" has been a key and transformative element. It holds significant relevance for both individuals and broader society, influencing sectors such as science, politics, education, and industry, and shaping their development. Highlighting its overarching significance, the term "information age" was first introduced by Richard Leghorn in the early 1960s to characterise our current era. (Pawlak, 2019 p. 1)

Leghorn's designation of the current period as the "information age" symbolises the evolution towards a society founded on information. While Richard Leghorn was the first to use this term, the inception of the "information age" is often attributed to Claude Shannon. Shannon's groundbreaking information theory, a cornerstone of this era, revolutionized our understanding and handling of communication and information processing. (Roberts, 2016)

The evolution of the 'data' concept is a crucial element in comprehending the information age. Unlike the more ancient concept of information, “data” as a significant concept has been used first time around the 1946. (Remenyi & Griffiths, 2022) During this era, the evolution of data and its processing has led to varied viewpoints on the creation and application of information. In this context, data is seen as raw information, whereas data that has been processed and tailored for specific uses is identified as information. (Rani, 2019)

Within the evolution of the concepts of information and data, the importance and impact of personal data has gradually increased, especially in recent years. Accordingly, personal data is defined in the GDPR as "personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person" (General Data Protection Regulation (GDPR) (Regulation 2016/679 Of the European Parliament and the European Council, on the protection of natural persons with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (General Data Protection Regulation), the CCPA-CPRA states that "The rights set forth in the CCPA apply to personal information, defined as information that identifies, relates to, describes, is capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household. The CCPA includes an illustrative list of items that fall within this definition, including names, physical addresses, email addresses, Internet Protocol addresses, geolocation data, Social Security numbers, telephone numbers, driver's licence numbers, account numbers, biometric identifiers, physical descriptions, medi_cal information, insurance information, financial information, employment information, purchase histories, and browser histories, as well as as inferences that can be drawn from the preceding items regarding consumer preferences, psychological trends, predispositions, behaviour, attitudes, intelligence, abilities and aptitudes." (Rothstein & Tovino, 2019) These differences in the definitions of "personal data" in these three legal frameworks reflect both jurisdictions' approaches to data protection and privacy. While the GDPR extends the individual's control over data, the CCPA-CPRA focuses on consumer rights.

1.2. Algorithm

The role and importance of algorithms in our age is noteworthy in terms of their relationship with data and, more specifically, with personal data. Although there is a "lack of satisfactory consensus" on the definition of algorithm even among computer scientists, it has a critical role in the information age, especially in data processing and analysis, which we will discuss in this paper. (Hill, 2016) Basically, algorithms are steps in a process that are designed to solve a specific problem or fulfill a task, and are activated when triggered. These steps are created by programs that automate the processes of processing data, analysing and drawing conclusions, thereby performing complex tasks quickly and efficiently. (Coleman, 2020)

When an algorithm is utilised partially or completely in the conclusion of an issue, it means that the algorithm makes a decision (Brkan & Bonnet, 2020). The feature to be noted here is whether there is human intervention in the decision-making process. Spam filters used in e-mail boxes, for example, are an example of "fully automatic" algorithmic decision making. When a bank employee uses the help of an algorithm in the process of evaluating a loan application, it is a "partially automated" decision-making process. In this case, the bank employee is in control, however the algorithm assists the employee. (Borgesius, 2020) (Aksoy, 2022)

The relationship between artificial intelligence and algorithms is critical to the understanding and evaluation of modern technology. Artificial intelligence, to start with the definition by purpose, is a technology that is used to imitate human intelligence and has a wide range of capabilities such as learning, problem solving, perception, comprehension and language processing. The process is based on the fact that artificial intelligence algorithms are trained on data sets to recognise certain patterns and adapt to new situations previously unencountered, without needing human intervention. An algorithm is essentially a series of instructions or rules; thus, AI comprises a collection of these algorithms which enable a computer to learn and make decisions based on its learning. Therefore, the algorithm plays an essential role in the execution of all these activities related to artificial intelligence. (Garza, 2023)

In fields such as artificial intelligence and its subset, machine learning, algorithms enable data-driven learning and decision-making processes. Machine learning is a type of AI that develops models based on data, using these models to make predictions or decisions. In AI applications, algorithms are trained on extensive data sets. This training allows them to recognize specific patterns, make decisions based on these patterns, and acquire the ability to deal with situations they have not encountered before. (Ledesma et al., 2018) This approach minimises human intervention in big data analyses and complex system investigations.

Delving deeper into the relationship between artificial intelligence and algorithms is essential. When viewed through a Venn diagram, it's evident that these two concepts are not mutually inclusive, nor is one broader than the other. Despite the close relationship between AI and algorithms, not every algorithm incorporates AI. Specifically, there exist simpler algorithms that lack AI elements and are designed for a specific function. These algorithms require more human interaction rather than AI intervention. This highlights that while algorithms are crucial for AI, not all algorithms are associated with AI.

On this broad scale, the interaction between AI and algorithms represents a complex issue that needs to be considered in depth from technological, ethical and legal perspectives. The growing impact of AI applications raises new and important questions about how to govern and regulate these technologies. These questions point to the significant impact of algorithms and AI on society, especially in the context of personal data use and protection.

Particularly in the realm of e-commerce, utilising personal data for tailored consumer recommendations presents privacy risks. The influence of algorithms on this data necessitates the protection of individuals' privacy rights and the security of their personal data. Additionally, inaccurate data analyses by algorithms could potentially infringe upon various rights and freedoms of individuals. Consequently, managing personal data used in the training of algorithms has emerged as a complex issue from both ethical and legal perspectives.

1.3. Personal Data Breach

It is necessary to evaluate the personal data breach within the context, especially the European Union GDPR and other relevant regulations endeavour to establish a legal framework for the protection of personal data and to standardise the issue. In this context, personal data breach is defined as "accidental or unlawful destruction, loss, alteration, unauthorised disclosure or access to personal data as a result of a breach of security." (Articles 5(1)(f) and 32 GDPR)

Based on the aforementioned definition, it will be seen that the definition includes not only the stored data but also the transferred or processed personal data. In addition, Articles 33 and 34 of the GDPR provide guidance on detecting data breaches and notifying the necessary authorities. Briefly mentioned, Articles 33 and 34 regulate the time period for applying to the competent authorities in case of breach. (Kiesow Cortez, 2020 p. 243-244)

A personal data breach can occur either intentionally or unintentionally. A basic example of an unintentional breach is sending a message with sensitive work-related information to the wrong recipient, leading to data leakage. In instances of intentional breaches, theft of a device containing data, resulting in cybercriminals acquiring the data, can be executed through obtaining access credentials using malware, which constitutes "unauthorised access." (Ong, 2023)

Finally, it must be noted that a personal data breach does not only affect individual data subjects, but it can also have consequences for companies, institutions, and governments, making it a significant issue in our times. Therefore, it is of great importance for both data subjects and entities processing, storing, etc., data to be aware of data protection and security, and to take necessary precautions. Legislation like the GDPR, CCPA, CPRA and other relevant data protection laws serve as a guide in our highly digitalised world, providing a framework for the collection, processing, use, and protection of personal data.

1.4. Connection Between Algorithm and Personal Data

From online shopping websites to the health sector, these two concepts play an active role in a wide range of sectors. In particular, algorithms function as decision mechanisms in these sectors, often relying on personal data in this process. A concrete example of this interaction is social media platforms.

The advertisements we are exposed to on social media in our daily lives clearly demonstrate how user data is used by algorithms for analysis and personalised content delivery. Algorithms use personal data, such as our location information, to serve us adverts tailored to our time, mood, needs and wants. In this process, personal data has a significant impact on the accuracy and efficiency of the algorithm, and the user's personalised user experience is made possible by this data.

In this context, it is possible to say that algorithms and personal data are essential for each other. However, legislation and legal frameworks to protect the rights of individuals and consumers regarding the protection and privacy of personal data may not always be at peace with algorithm-based decision-making systems. While algorithms need large data sets to make the most accurate decisions, the relevant legal frameworks limit this use and aim to keep the control in the individual.

It's crucial to present the interaction between algorithms and personal data creates a critical balance point in the modern world in terms of technological advances and individual privacy rights. This balance emphasises the importance of both technological innovation and the protection of individual rights.

1.5. Introduction to Legal Frameworks

As it has been mentioned more than once since the beginning of the article, it would be correct to address it superficially in this section, the GDPR is a comprehensive regulation enacted by the EU in 2018, aiming to protect the personal data of individuals. This regulation clarifies the rights of data subjects over their personal data and imposes certain responsibilities on organisations that collect and process data. The GDPR sets a global example in data protection and aims to standardise data security and privacy. (Van Ooijen & Vrabec, 2019)

In parallel with the GDPR owned by the European Union, the United States of America has also started to create a framework on the subject with the California Consumer Privacy Act (CCPA). Enacted in 2018, this law is based on the consumer, unlike the GDPR. In the CCPA, consumers have the right to refuse the collection of their data, to learn what their data is used for, and to request the deletion of their personal data. The CCPA, like the GDPR, strengthens the hand of the consumer by imposing various restrictions on data processing organisations. The obligations and rights introduced by the CCPA have had a significant impact, especially on large technology companies, and have played an important role in the global dialogue on data protection. (Moreira, 2023) Comparing these two regulations is important in terms of showing how data protection and privacy rights are handled in different geographies and the evolution of legal approaches on this issue.

The GDPR combines individual rights and regulatory oversight mechanisms to limit the impact of algorithmic decisions on individuals and ensure transparency in these processes. The GDPR imposes an obligation on data processing organisations to both protect the rights of individuals and provide comprehensive oversight over the algorithm and the people around it. Furthermore, the GDPR includes tools such as Data Protection Impact Assessment (DPIA) for algorithmic decision-making processes and obligations such as third-party audits and the appointment of Data Protection Officers (DPOs) (GDPR, Article 35-37). This approach of the GDPR represents a collaborative governance approach between the public and private sectors, both for the protection of individual rights and for algorithmic decision-making. (Blume, 2017 p. 1434-1435)

Unlike the GDPR, the CCPA does not explicitly define the concept of algorithms or the algorithmic decision-making process. CCPA focuses more on regulating the data and personal data that constitute the beginning to end of this process, and its consequences for consumers and stakeholders. (Byun, 2019)

The California Privacy Rights Act (CPRA) as well is a significant addition to the evolving landscape of data protection and privacy laws, complementing the GDPR and the CCPA. Enacted in November 2020 through Proposition 24, the CPRA came into effect on January 1, 2023. This law enhances and expands the CCPA, which was already a comprehensive consumer privacy legislation in California. (Wong, 2021 p. 311)

The GDPR and CPRA have taken important steps in the protection of personal data and expanded the rights of data subjects. As mentioned, these legal frameworks impose certain responsibilities on organisations that collect and process data, while at the same time protecting the rights of individuals over data. They require international companies to comply with different legal and ethical standards and set new standards for data protection and privacy.

1.6. Ethical Considerations

The use of algorithms and personal data in daily life also raises ethical and moral questions. These ethical and moral questions cover a wide range of areas, from how algorithms are programmed to what data they use. The most agenda example related to the process may be that the algorithm carries out its operations in a fair and non-discriminatory manner. (Giovanola & Tiribelli, 2023) At this point, an example of a constructive rather than destructive approach to the issue has been exhibited, and concepts such as creating ethical algorithms have been put forward with the motive of eliminating ethical risks. (Kearns & Roth, 2019) Discussions on this issue have also introduced the concept of "AI Ethical Alignment" or "Artificial Intelligence Ethical Alignment", which includes moral approaches involved in processes such as the design and use of artificial intelligence in our lives. (Ray, 2023)

The concept requires algorithms to be compatible not only with technical efficiency, but also with ethical standards and social values. In particular, it advocates the observance of ethical principles such as transparency, fairness, responsibility and confidentiality in artificial intelligence, algorithms, architectures and interfaces. Human rights and freedoms should be respected in the design and implementation of algorithms. (Zhou & Chen, 2022)

Another dimension of these ethical debates is embodied in Isaac Asimov's Laws of Robotics. Asimov addressed the ethical aspect of technological development through the rules related to robots. Although these rules are directly related to robots, they constitute a good example of the restrictive and regulatory ethical approaches of technology. Asimov's rules provide guidance that increases awareness of the relationship between technology and ethics and should be taken into account in the design of algorithms. (Asimov, 2004)

In conclusion, since the use of personal data and algorithms is an integral part of modern technology, this process should be ethically managed for the benefit of humanity as much as possible. These processes should provide experiences tailored to individuals' preferences and behaviours, while at the same time paying attention to principles of ethical compliance and fairness. This process requires ethical considerations and a balanced approach between data protection, privacy and individual freedoms.

CHAPTER II – LEGAL ASPECTS OF PERSONAL DATA USAGE

2.1. GDPR

The GDPR influences nations to regulate such modern concept “data” and represents an important turning point in data privacy and protection within Europe. This regulation introduces 'stricter data protection standards'. It is influential not only within the European Union but it has an influence in global aspect as well. It emphasizes the right of EU citizens to manage their own personal information, while also enforcing stringent guidelines and obligations on organizations regarding the collection, retention, and processing of such data. Additionally, the GDPR has a significant impact on market strategies and business operations, as companies are required to adapt their practices to comply with the new regulation. By doing so, the GDPR seeks a more 'harmonized legal framework within the EU', leading to greater legal clarity and consistency across member states. This unified approach to data protection is a step towards a more trustworthy and secure digital environment, setting a new global standard for data protection and privacy. (Albrecht, 2016 p. 288-289)

A fundamental rule of the GDPR is the need for clear consent. Organizations can't just assume they have permission or use unclear, broad agreements that ignore consent. According to articles 4(11) and 7 of the GDPR, consent has to be given freely, specific, informed, auditable, explicit, with unambigious wishes indicated and with a mechanism to withdraw; often needing a direct yes from the person whose data is being used. GDPR makes organizations really think about why they need the data they collect, making sure they only take what is needed for their purpose. (Breen, Ouazzane & Patel, 2020 p. 22)

It is also worth noting that the GDPR Article 32 mandates strict security measures for the protection of personal data. This includes a range of technical and organisational strategies, such as robust encryption and pseudonymisation techniques, which serve to hide or separate data from direct identifiers and thus reduce the risk of harm to individuals in the event of a data breach. In addition, organisations must ensure the ongoing confidentiality, integrity and resilience of their processing systems and services. (Minssen et al., 2020 p. 47)

Another critical component of GDPR is the obligation of organizations to report certain types of data breaches to the relevant supervisory authority within 72 hours of becoming aware of the breach, according to Article 33. This requirement seeks a culture of transparency and accountability, demanding that organizations not only take immediate action in the event of a data breach but also communicate these incidents effectively and timely. This swift response is important to protect data subjects from the potential impact. (Kahler, 2020 p. 9)

The principles of GDPR emphasise the importance of handling personal data with utmost care and responsibility. Firstly, acccording to Article 5, it is crucial that personal data is processed in a lawful, fair, and transparent way, ensuring the data subject is always considered. Secondly, the collection of personal data must be for clear, specific, and legitimate reasons, and its processing should align with these initial purposes. Thirdly, the data collected should be just enough, relevant, and not excessive for its intended use. Accuracy is another key principle, necessitating the data to be precise and current. Furthermore, the storage of personal data should be limited to the duration necessary for its intended purposes, ensuring the data subjects' identities are not kept longer than needed (GDPR, Article 5(1)(e)). Another essential aspect is the integrity and confidentiality of personal data, requiring it to be processed securely to prevent unauthorised access, damage, or loss. Lastly, accountability rests with the data controller, who must not only comply with GDPR but also be able to demonstrate this compliance. These principles collectively form the backbone of GDPR, guiding how personal data should be managed and protected. (Štarchoň & Pikulík, 2019 p. 305)

The operationalization of these GDPR principles presents substantial challenges, especially for smaller organizations with limited resources. Compliance demands not just a technical catchup but a fundamental change in organizational culture towards data protection. This encompasses regular data protection impact assessments, appointment of data protection officers in certain cases, and a continuous evaluation of data processing activities to align with GDPR standards. (Tomashchuk et al., 2020 p. 12-13)

GDPR's enforcement mechanism is characterized by penalties in Article 83. Non-compliance can result in substantial fines, calculated based on the severity of the breach and the organization's annual global turnover. This stringent penalty structure underlines the regulation's commitment to ensuring that data protection is not only a checkbox exercise but a fundamental business principle.

In summary, the GDPR's impact on data usage is not only compliance. It has made a fundamental shift in the relationship between individuals and their personal data, balancing the power dynamics and placing greater accountability on organizations that handle such data. The regulation not only protects individual rights but also encourages a safer, more transparent digital environment, fostering trust and confidence in the digital economy.

2.1.1. Compliance and Enforcement

The entry into force of GDPR was met with enthusiasm, especially from individuals and NGOs concerned about the increasing capacity of private companies and government bodies to collect and process private information. However, it has also presented challenges, especially for small and medium-sized businesses, which may have fewer resources for compliance. (Lindgren, 2016 p. 250)

The regulation requires companies to provide individuals' data in a structured, commonly used, and machine-readable format to facilitate data portability. (GDPR, Article 20) Despite the intended protections, some companies have found ways to circumvent the Regulation, such as by denying or restricting access to EU visitors or by using consent management pop-ups that complicate the refusal of consent. This approach, while complying with the letter of the law, can be seen as a way to avoid the spirit of GDPR, which aims to enhance data privacy and security. (Cara & Dumitrașciuc, 2021 p. 2)

GDPR presents several complexities for businesses as mentioned. It sets high-level principles open to interpretation, requiring companies to navigate the regulation's nuances. Another one for instance, the storage limitation principle dictates that data should not be kept longer than necessary for the purpose it was collected. (GDPR, Article 5 (1) (e)) This principle challenges traditional data management practices, where companies often store data indefinitely. Businesses must now actively engage in data deletion processes, balancing legal, tax, or compliance reasons with GDPR requirements. (Shah et al., 2019 p. 5-6)

Another critical aspect is the limitation around the purpose of data collection. Under GDPR, data must be collected for a specified, legitimate purpose, and once used for that purpose, it cannot be repurposed for other uses like marketing without additional consent. (GDPR, Article 6 (4)) This limitation has significant implications for how businesses disclose data collection purposes and use the data for various business activities. (Finck, 2021 p. 5)

GDPR's principles-based approach leaves much open to interpretation, which companies have to navigate carefully. Defining the scope of what constitutes personally identifiable information (PII) and understanding who is covered under GDPR (customers, employees, visitors, etc.) are challenges that businesses face. Ambiguities around what constitutes PII, especially in the case of IP addresses or clickstream information, add to the complexity.

Despite the challenges indicated, businesses should see following GDPR rules as a chance to get ahead in today's world where data is key, not just as a challenge. Tech companies aiming for global markets need to work harder to make sure their data, systems, products, and services meet GDPR standards. It's also a good idea for researchers and experts to look into GDPR-related issues and share what they learn. Information Systems (IS) and Information Technology (IT) are really important in many areas. For example, IS experts can create ways, methods, and designs that follow GDPR rules for taking back consent and getting rid of personal data that's spread far and wide. (Politou, Alepis & Patsakis, 2018 p. 4) They can also figure out how much it costs to follow GDPR, identify what affects GDPR compliance, explore how culture and the country's situation impact GDPR, and look into how GDPR affects how businesses operate and their money matters. (Li, Yu & He, 2019 p. 5)

2.1.2. Rights of Individuals

The GDPR strengthens individual rights to data protection but introduces several obligations for businesses that collect and process personal data. It has a significant effect on competition, innovation, marketing activities, and cross-border data flows. These rights and obligations create a complex landscape for businesses, balancing the protection of individual data with the operational and innovative capabilities of businesses in the digital economy.

2.1.2.1. Control Over Personal Data and Right to Data Protection

The idea of informational self-determination is a key part of GDPR. It's about people having control over their own personal data. This idea was first brought up by the German Federal Constitutional Court. (BVerfGE - 1 BvR 16/13, 87.) It says that people should be the ones to choose how information about them is shared and used. GDPR puts this into action by making sure that personal data is handled in a clear way, with the person's permission or for other valid reasons, and not just randomly by private companies. GDPR gives people a lot of power in this area, but in real life, it seems that most of the time, private companies process data based on their own interests, not because the person said it was okay. This means that in practice, people don't get to use their right to control their information as much as they should. (Thouvenin, 2021)

The reason why data protection is seen as different from other basic rights is because of its unique place in the EU Charter. Having control over one's personal data is a way to deal with problems like unequal power, the weakness of those whose data it is, and the chance of being discriminated against or manipulated through the use of their data.

2.1.2.2. EU Policy on Data Subject Control

EU policy papers indicate that controlling your own data is a modern way of looking at personal data protection. In 2011, Vivian Reding focused on giving people more power over their data. She talked about important parts of this control, like the right to be forgotten, being clear about data use, having privacy as the default setting, and protecting data no matter where it is. (Reding, 2011) The GDPR maintains control as an important underlying idea, with an enhanced focus on data subject control rights and updated provisions on consent. (Vrabec, 2021 p. 57)

2.1.2.3. Mechanisms Reflecting Control in the GDPR

The GDPR is the EU's main law for handling personal data. It shows the importance of controlling your own data through things like giving consent and having rights as the data subject. The goal of this regulation is to make the power between those who control the data and the individuals more balanced, and to strengthen the control that individuals have, which comes from their own independence and values. Data protection law is binary, addressing control and protection of personal data. The GDPR has strengthened provisions on data subject control and consent in response to the evolving data economy. (Tikkinen-Piri, Rohunen & Markkula, 2018 p. 135)

2.1.2.4. Catalogue of Control Rights under the GDPR

The GDPR lists eight key entitlements: the right to information (GDPR, Article 13-14), access (GDPR, Article 15), rectification (GDPR, Article 16), erasure (to be forgotten) (GDPR, Article 17), restriction of processing (GDPR, Article 17), data portability (GDPR Article 20), to object (GDPR, Article 21), and not to be subject to decisions based solely on automated processing, including profiling. (GDPR, Article 22)

Especially the right of access, allowing individuals to request and obtain their personal data from organizations, and the right to erasure (also known as the right to be forgotten), enabling individuals to request the deletion of their personal data should be pointed out. (Tamò & George, 2014) As well as the right to data portability allows individuals to transfer their data from one service provider to another; these rights aim to empower individuals by giving them greater control over their personal information. (De Hert et al., 2018 p. 194)

2.1.2.5. Situations of Data Collection and Information Provision

When data is collected directly from the data subject, such as signing up for a social media service, the data controller must provide information as listed in Article 13. This often takes the form of a privacy policy or terms and conditions. (Gil González & De Hert, 2019 p. 612)

In cases where data is obtained from third parties, like a hiring manager using social media for screening, data subjects must be informed about data processing. The obligation to inform may fall on both the hiring manager and the social media company. Additionally, when data is not directly collected from a data subject, there is a need to describe categories of data and sources if the data is from publicly accessible sources. (Tombal & Graef, 2023 p. 4)

Overall, GDPR has catalyzed a global shift in data protection norms, influencing not only European businesses or individuals but also prompting other countries and regions, like the U.S., to adopt similar data protection legislations. As businesses adapt to these changes, they must balance the rights of individuals with the operational and strategic implications of the GDPR.

2.2. US Data Protection Laws

In the United States, CCPA and the CPRA represent the milestones of legislative efforts to regulate the use of personal data. The CCPA, a pioneering privacy law, and its subsequent enhancement, the CPRA, have set significant precedents in data protection. These laws mandate comprehensive measures for data processing and handling, requiring businesses to maintain clear data processing agreements. These agreements are critical as they clearly set out the terms of processing, ensuring that both parties are aware of their rights and obligations. (Gamwell, 2022 p. 3)

2.2.1. Comparison Between CCPA and CPRA

2.2.1.1. Scope and Applicability

The CCPA applies to businesses with annual gross revenues exceeding $25 million, those that manage personal information of 50,000 or more consumers, or those earning more than half of their annual revenue from selling consumers' personal information.

The CPRA modifies these thresholds, increasing the criteria from 50,000 to 100,000 consumers or households and includes businesses that derive 50% or more of their annual revenue from selling or sharing personal data. (Mayfield, 2023 p. 9-10)

2.2.1.2. Consumer Rights

Under the CCPA, the primary focus was on consumer data privacy. The CCPA provided California residents with certain rights regarding their personal information, including the right to know what personal information is collected, the right to delete collected data, and the right to opt-out of the sale of their data. Businesses covered by the CCPA were required to inform consumers about these rights and their own privacy practices.

However, the CPRA, sometimes referred as "CCPA 2.0", expanded the scope of data privacy rights to include employees, marking a significant shift in the handling of personal data. ((Chang, Pei-Chuan. "Legislation trends: Latest data regulation observations in the US." (2023). P.4) Under the CPRA, from 2023, employees in California gained new rights similar to consumers. These are; right to delete personal information, right to opt-out of sale or sharing, right to correct inaccurate personal information, right to know what personal information is collected, right to limit use and disclosure of sensitive personal information, and non-discriminatory treatment for exercising Rights. The CPRA's extension of rights to employees necessitates businesses to adjust their data handling and privacy practices not only for consumers but also for their employees. (Blanke, 2022 p. 18-19)

2.2.1.3. Sensitive Personal Information

The CPRA introduces the concept of "Sensitive Personal Information" (SPI), a category of data that requires greater protection due to its sensitive nature. This category is an expansion over the personal data types identified in the CCPA.

Sensitive Personal Information under the CPRA includes a range of data types that are more likely to impact an individual's privacy if mishandled. This includes government identifiers like Social Security and driver's license numbers, and account log-in details, especially for financial accounts, which are coupled with necessary security codes or passwords. Precise geolocation information, which can pinpoint an individual's location within a small geographic area, is also classified as SPI. (Singh, Amritha & Sethumadhavan, 2022 p. 161)

Additionally, the CPRA covers personal characteristics such as racial or ethnic origin, religious or philosophical beliefs, and union membership. Communications content like emails, text messages, and postal mail are protected unless the business is their intended recipient. Other categories include genetic data and biometric information, which are particularly sensitive as they can uniquely identify an individual or reveal information about a person's health, sex life, or sexual orientation. (Walker, 2020 p. 40)

The CPRA also imposes additional limitations on the use and disclosure of SPI, extending beyond the limitations applicable to all personal information. Businesses are required to update their websites with specific links that allow consumers to limit the use of their SPI and to opt-out of the sale or sharing of their SPI. (Jordan, 2022 p. 164-165)

2.2.1.4. Penalties and Enforcement

The CPRA significantly strengthens the penalties and enforcement mechanisms compared to its predecessor, CCPA. One of the key areas where the CPRA enhances penalties is in regard to minors' privacy rights.

Under the CPRA, penalties for violations involving minors are substantially increased. Businesses can be fined up to $7,500 for each violation that involves children. This is a notable increase from the CCPA, which had a general maximum penalty of $2,500 for unintentional violations and $7,500 for intentional ones. The enhanced penalties under the CPRA demonstrate a heightened focus on protecting the privacy rights of minors. (Perumal, 2022 p. 114)

Additionally, the CPRA removes the 30-day cure period that was present under the CCPA. This cure period previously allowed businesses a timeframe of 30 days to address and rectify a violation after being formally notified, potentially avoiding penalties. With the CPRA, this grace period is eliminated, indicating a stricter approach to compliance and enforcement. Businesses no longer have this window to rectify their violations to avoid penalties. (Goldman, 2021 p. 3-7)

2.2.1.5. Data Minimization and Purpose Limitation

The CPRA introduces significant changes to the data privacy landscape, particularly regarding data minimization and purpose limitation, addressing some areas where the CCPA was not sufficient.

Under the CPRA, there's a clear emphasis on the principles of data minimization and purpose limitation, aligning more closely with the GDPR principles. Data minimization under the CPRA requires businesses to collect only personal information that is reasonably necessary for the purposes for which it is collected. (Anciaux et al., 2024 p. 4) This marks a departure from the CCPA, which did not explicitly include these principles.

Purpose limitation, another key aspect introduced by the CPRA, stipulates that businesses can only collect consumer's personal information for specific, explicit, and legitimate disclosed purposes. They are not allowed to further collect, use, or disclose consumers' personal information for reasons incompatible with those initially disclosed purposes. This change aims to restrict businesses from using personal data beyond the scope of the original intent of collection, something the CCPA did not explicitly regulate. (Blanke, 2022 p. 19)

The introduction of these principles under the CPRA means that businesses now need to be more intentional and transparent about their data collection and usage practices. They must ensure that personal data is only used for the purposes stated at the time of collection and that the amount of data collected is limited to what is necessary to fulfill those purposes.

These adjustments in the CPRA represent a significant shift towards more stringent data protection and privacy standards in California, reflecting global trends in data regulation. For businesses, this means adapting their data processing and privacy practices to meet these new requirements and ensure compliance.

2.2.1.6. New Administrative Agency

The establishment of the California Privacy Protection Agency (CPPA) under the CPRA represents a significant evolution from the CCPA. Unlike the CCPA, which relied on the California Attorney General for enforcement, the CPRA created the CPPA as a dedicated agency for enforcing data privacy laws. This change indicates a more focused approach to data privacy in California. The CPPA is empowered not only to enforce the CCPA as amended by the CPRA but also to engage in rulemaking, updating existing regulations and adopting new ones to broaden data protection. Furthermore, the CPPA has the authority to conduct administrative hearings, impose fines, and take civil actions for violations, enhancing the potential for stricter enforcement. (Jordan, 2022 p. 165)

2.2.2. Business Perspective

From the business perspective, the transition from the CCPA to the CPRA represents a significant shift in the data privacy landscape. The CPRA, building upon the foundations of the CCPA, introduces requirements and broadens the scope of CCPA, regarding obligations for businesses, particularly in the areas of data protection, contractual agreements, and enforcement mechanisms.

The CPRA mandates businesses to adopt a more proactive approach towards data protection. Unlike the CCPA, which does not explicitly require data protection by design, the CPRA requires businesses to limit the collection of personal and sensitive information to what is necessary for the disclosed purpose. This means businesses must now reassess their data collection and processing activities to ensure they do not collect additional data categories that are incompatible with the intended use. (Jordan, 2022 p. 134)

Additionally, the CPRA expands the contractual requirements for businesses. It requires not only service provider agreements but also contracts with contractors and third parties that use or process the personal information collected by the business. These contracts must stipulate that the personal information is used only for specified purposes and that these entities comply with CPRA's obligations regarding the protection of personal information and consumer rights. (Jordan, 2022 p. 133)

A major change introduced by the CPRA is the establishment of the CPPA, which is tasked with the exclusive enforcement of the CPRA. This represents a departure from the CCPA's enforcement mechanism, which relied on the California Attorney General's office. The creation of the CPPA signifies a more focused and specialized approach to data privacy enforcement, potentially leading to stricter and more consistent application of the law. (Jordan, 2022 p. 165)

Moreover, the CPRA underscores the importance of maintaining consumer trust. In today's increasingly digital world, consumers are more aware of their data privacy rights. Businesses that fail to comply with the CPRA risk not only financial penalties but also the erosion of consumer trust, which can have long-lasting effects on a business's reputation and customer relationships. (Ford, 2021 p. 12)

In terms of scope, the CPRA alters the threshold for what constitutes a "for-profit" business under the regulation. The CPRA raises the applicability criteria to entities catering to at least 100,000 consumers or households, thus modifying the landscape of businesses that need to comply with these regulations. (Mayfield, 2023 p. 9-10)

In conclusion, for businesses, adapting to the CPRA involves conducting thorough data assessments, updating privacy policies and contractual agreements, and ensuring an organizational culture that values and understands the importance of data privacy. The CPRA not only increases the responsibilities of businesses in terms of data protection but also offers an opportunity to strengthen consumer trust by demonstrating a commitment to safeguarding personal information in accordance with the evolving legal landscape.

2.2.3. Rights of Individuals

The CCPA, as the first comprehensive consumer privacy legislation in the U.S., laid the groundwork for these rights, which were further expanded by the CPRA.

Rights Under CCPA:

- **Right to Know**: Consumers can request information about the personal information (PII) a business collects and sells. (CCPA § 1798.110-115)
- **Right to Delete**: Individuals can request the deletion of personal information collected from them. (CCPA § 1798.105)
- **Right to Opt-Out**: Consumers have the right to opt out of the sale of their personal information. (CCPA § 1798.120)
- **Opt-in Rights for Minors**: Businesses must obtain opt-in consent to sell the personal information of consumers under 16 years of age. (CCPA § 1798.120)
- **Right to Nondiscriminatory Treatment**: This right ensures that consumers are not discriminated against for exercising their privacy rights. (CCPA § 1798.125)
- **Private Right of Action**: In cases of data breaches, consumers have the right to initiate legal action. (CCPA § 1798.150)
- **Expansion of the Right to Know**: The CPRA extends this right to include information about data that a business shares, and consumers can request information beyond the

standard 12-month period, provided it is feasible and not disproportionate. (CPRA § 1798.110)

- **Broader Opt-Out Rights**: The CPRA allows consumers to opt out of both the sale and sharing of their personal data. Data sharing includes transferring consumer information to third parties for advertising purposes, regardless of monetary exchange. (CPRA § 1798.110)
- **Enhanced Right to Delete**: Businesses are now required to instruct third parties to delete consumer data upon receiving a deletion request from a consumer. (CPRA § 1798.105)
- **Right to Correct**: Consumers have the right to correct inaccurate personal information held by a business. (CPRA § 1798.106)
- **Right to Limit Use and Disclosure of Sensitive Personal Information**: The CPRA introduces rights concerning sensitive personal information (SPI), which includes data such as social security numbers, geolocation data, racial or ethnic origin, religious beliefs, and more. (CPRA § 1798.121)
- **New Definitions and Protections for Sensitive Personal Information**: The CPRA provides more detailed definitions and protections for sensitive personal information, differentiating it from the broader category of personal information under the CCPA. (CPRA § 1798.121-135)
- **Discretionary 30-Day Cure Period**: The CPRA removes the automatic 30-day period for businesses to address violations, making it discretionary and subject to the judgment of the CPPA. (CPRA §1798.150

These changes enhance individuals' control over their personal data, offering more robust protection and options for managing their privacy.

In summary, both the CCPA and CPRA mark significant steps in US data protection laws, with the CPRA building upon the CCPA's foundation to offer greater protection and control to consumers over their personal data. These laws also pose new challenges and obligations for businesses, particularly in terms of compliance and adapting to the expanded rights of individuals. For a comprehensive examination of these laws, it is crucial to explore their specific provisions, impacts, and the evolving landscape of data privacy in the US.

2.3. Legal Status of Personal Data Use in Algorithm Training with GDPR, CCPA, and CPRA

The use of personal data in algorithm training is a significant concern under various data protection regulations like the GDPR, the CCPA, and the CPRA. Each of these regulations has its own set of rules and implications for organizations involved in such activities.

2.3.1. Assessment under GDPR

Under the GDPR, the use of personal data in training algorithms must adhere to several core principles to remind once more, including fairness, purpose limitation, data minimization, and transparency. The fairness principle requires personal data to be processed with respect for the data subject's interests, and measures must be taken to prevent discriminatory effects. Purpose limitation dictates that data subjects must be informed about the purpose of data collection and processing, and data minimization ensures that collected data is adequate, limited, and relevant to the purpose of the project. Transparency is critical, as data subjects have the right to know how their information is being used.

If we would need to assess, GDPR principles that are relevant to usage of personal data while training algorithm, the action could;

Comply with Fairness: The fairness of an AI system is closely tied to the data it's trained on. If the training data is biased, the AI's decisions may also be biased. For example, if an AI is trained predominantly on data from a certain demographic, it might not perform equally well for other demographics. Thus, the use of personal data for training algorithms can be fair if it is done in a way that does not lead to unjust discrimination between people. There are ways to mitigate the bias of machine learning algorithms, such as debiasing the information source. (Verma, Ernst & Just, 2021 p. 1)

Comply with Purpose Limitation: The purpose limitation principle under the GDPR mandates that personal data must be collected for specified, explicit, and legitimate purposes and not further processed in a manner that is incompatible with those purposes. Data subjects must be clearly informed about the purpose of data collection and processing at the time of data collection. Complying with this principle, will require careful management and clear communication to ensure adherence. (Giannopoulou, 2020 p. 8) Lawfulness of processing personal data is another key aspect under GDPR Article 5, which could be assessed with purpose limitation. This involves ensuring that one of the legal bases in Article 6 or Article 9

GDPR can be established for processing the data. This often depends on the specific use case, whether it's for internal purposes, customer-related activities, or involves sensitive data.

Comply with Data Minimisation: The usage of personal data while training an algorithm can comply with the data minimisation principle of the GDPR, but this requires careful consideration and planning. The data minimisation principle dictates that only data that is necessary for the purposes for which it is processed should be collected and used. This is particularly challenging in the context of AI and machine learning, where large datasets are often seen as beneficial for the accuracy and effectiveness of the algorithms. (Tschider, 2021 p. 174)

Comply with Transparency: To ensure GDPR transparency when using personal data for algorithm training, it is essential to clearly inform data subjects about the use of their data, ensure purpose specification, obtain consent, practice data minimisation, uphold data subject rights, consider data anonymisation, and conduct impact assessments. Nevertheless, the matter remains debatable, as certain companies may choose not to disclose every aspect of their processes in the interest of transparency, owing to the need to protect trade secrets. (Watson & Nations, 2019 p. 1)

Organizations using AI and personal data must also consider their role as either a 'controller' or 'processor' of data, as defined by GDPR. This distinction is important for determining responsibility and compliance requirements. In some cases, joint responsibility or 'joint controllership' may arise, especially when an AI user influences the AI training, for instance, by allowing the reuse of training data for general AI enhancement. In such scenarios, both the AI user and provider might be considered jointly responsible under GDPR, impacting their risk exposure significantly. (Colcelli, 2019 p. 1030-1034)

2.3.2. Assessment under CCPA-CPRA

Under the CPRA, which amends and extends the CCPA, the legal status of using personal data to train algorithms is subject to several key considerations:

Notification and Consent: Businesses are required to inform consumers about the collection of their personal information and the purpose behind this collection. This includes details on whether their personal information will be sold or shared, how it will be used, and how long it will be retained. (Jordan, 2022 p. 156)

Data Minimization Requirements: The CPRA enforces a "purpose limitation" provision. This mandates that the collection, use, retention, or sharing of a consumer's personal information must be "reasonably necessary and proportionate" to the purposes for which it was collected or processed. (Blanke, 2022 p. 70)

Sensitive Personal Information: The new category introduced under CPRA, "sensitive personal information," includes data types like biometric information, health information, and precise geolocation data. Consumers have the right to direct a business to limit the use and disclosure of such information. (Buresh, 2021 p. 67)

Consumer Opt-Out Rights: The CPRA provides consumers with the right to opt out of having their personal information sold or shared for targeted advertising. This extends to the requirement that businesses treat opt-out preference signals as valid requests to opt out of the sale or sharing of personal information for not only that browser or device but also for any consumer profile associated with that browser or device. (Zetoony, 2022 p. 14)

Compliance with Children's Privacy: The CPRA requires businesses to verify parental consent when selling or sharing personal information of consumers under the age of 13. (Arewa, 2023 p. 208)

Automated Decision-Making: The CPRA's stance on automated decision-making and profiling balances the need for privacy protection with the avoidance of overly restrictive regulations that could prevent innovation. This includes discussions on the type of automated decision-making activities that should be regulated, such as fully automated decision-making technology that produces legal or similarly significant effects. The part here requires attentionis regulating the rights of individuals to access and choose not to participate in companies' use of automated decision-making technologies, such as profiling. This includes the necessity for companies to provide substantial details in their responses to access requests. These details should encompass clear explanations of the reasoning behind these decision-making processes, along with a description of the probable consequences these processes may have on the consumer. (Weaver, 2022 p. 153)

2.3.3. Consequences of Non-Compliance

In terms of consequences, non-compliance with these regulations can result in substantial fines and legal actions.

Under GDPR, for instance, violations can lead to significant penalties, which can be as high as 4% of the annual global turnover or €20 million, whichever is higher.

Under the CCPA, businesses are given a 30-day notice period by the attorney general to rectify any compliance issues. If they fail to do so, they may face civil penalties of up to $2,500 per unintentional violation and $7,500 for intentional violations. Additionally, consumers have the right to private action, with statutory damages ranging from $1,000 to $3,000 or actual damages (whichever is greater) if their data is sold without consent. In cases of data breaches, affected consumers can claim damages between $100 to $750 per incident or actual damages, whichever is greater. (Hromisin, 2020 p. 60)

The CPRA, introduces further compliance obligations and reinforces the CCPA's provisions. It includes administrative fines for intentional violations involving sensitive personal information of individuals under 16 years of age, with fines of up to $7,500. (Buresh, 2021 p. 68)

The current legal status of using personal data in algorithm training is a dynamic and evolving area. Organizations must stay informed and adapt to the changing legal landscape to ensure compliance. This includes implementing Data Protection Laws-compliant AI development practices, such as using Generative Adversarial Networks (GANs) to use less data more efficiently, and continuously reassessing the type and quantity of training data required.

Overall, the key to compliance lies in understanding the specific requirements of each regulation and implementing robust data protection and privacy measures to safeguard personal data used in AI and algorithm training.

CHAPTER III – ETHICAL CONSIDERATIONS

3.1. Ethical Considerations

As a branch of philosophy, ethics refers to the effort of enlightening and substantiating what's been determined as moral, word that is dervied from Greek word "Ethos", sees ethics as a discipline concerned with what is right, wrong, good, or bad; and deals with the limits of the handled matter. (Çilingir, 2014 p. 711-713) To indicate such lines, limits; principles are used as a tool under the discipline. These principles, which are used to guide indiviaulds, societies; are dynamic and influenced by various cultural and philosophical beliefs, are essential in the context of technology and data usage, especially in the digital age. The ethical use of personal data in algorithm training, a key focus in the EU and US, requires balancing technological benefits with the protection of individual rights and societal values. This complex landscape intersects public attitudes, privacy concerns, and regulatory guidelines, highlighting the nuanced ethical challenges in online personalization and data privacy. These challenges go beyond legal compliance, encompassing diverse and often conflicting values and expectations.

With the continuous evolution of the digital environment, the distribution of personal data has become increasingly routine. This trend is fuelled by the widespread presence of online platforms and the simplicity with which individuals can share their information. Within the sphere of digital communications, personal data encompasses everything from elementary demographic details to intricate personal tastes and behaviours. Social media platforms such as Facebook, Twitter, and LinkedIn serve as depositories for enormous volumes of personal data, generously shared by their billions of global users. This increase in available data lays the groundwork for extensive data mining activities and the creation of advanced algorithms designed to improve user experiences, offering services like tailored recommendation systems and customised search functionalities. (Singh, 2016 p. 81)

However, this widespread sharing of personal data raises significant ethical issues, especially regarding privacy. The concern lies not only in the amount of data shared but also in the sensitivity of some of this data. The ethical handling of such data when training algorithms is a crucial matter, as it requires a careful balance between using this data for technological progress and ensuring the protection of individual privacy. Although social media platforms provide privacy options, many users either do not know about these options or opt not to use them. (Czerwiński, 2017 p. 8) Consequently, information that might have previously been shared in a more private context is now readily available to a broader audience, including those who may use this data for purposes beyond its original intention.

In response to these issues, there has been an emphasis on creating methods in data mining that protect privacy. These methods are designed to hide details that can identify users, while still keeping the data useful for analysis. Techniques like differential privacy, making data anonymous, and systems where users control their own identity information are leading this effort. However, despite these improvements, most of these methods are still mainly theoretical and haven't been fully applied in everyday systems. This situation is partly because users haven't shown enough concern or anger about privacy matters, which reduces the motivation for companies to make data privacy a priority. (Singh, 2016 p. 81)

Furthermore, the emergence of data ethics as a distinct discipline highlights the need for a fresh perspective on the moral implications of data sharing. Data ethics, divergent from traditional forms of ethics, calls for new strategies to empower users. These include establishing mechanisms for users to view and manage the data that companies hold about them and providing avenues for correcting inaccuracies in personal data. Such strategies would significantly improve the current situation by allowing users greater control over their personal behavioural data, thereby aligning technological innovation with ethical responsibility of tech companies, developers and designers, among others. (Lobschat et al., 2021 p. 876)

3.2. Consent

Consent in algorithmic contexts involves a complex interaction where a user (party A) consents to a company (party B) to process their personal data under specific explanations provided by the company. This process alters the moral relations between the user and the company, with informed consent being a key factor that legitimizes the company's actions. For consent to be meaningful, it must be voluntary, and the user must be competently informed about the data processing acts. This informed consent fundamentally includes the right to an ex-ante explanation, entailing that users should be made aware of how their data will be used in advance. (Giannopoulou, 2020 p. 4)

Achieving genuinely informed consent is challenging, especially given the complexities of data processing in machine learning and AI. The limitations of users' understanding of these processes and the quality of the explanations provided by companies can significantly impact the ethical standing of consent. The traditional view holds that consent must be of a high quality to be ethically valid, aligning with standards such as the Nuremberg Code of 1947, which emphasizes the responsibility for ascertaining the quality of consent. (Benedict, 2017 p. 90)

Another concept that should be pointed out in regards to "consent" is 'Notice and Choice', also known as 'notice and consent'. It is a key standard for obtaining online consent. It involves two main aspects: 'Notice', where terms are presented usually in a privacy policy or terms of use, and 'Choice', where the user shows agreement to these terms, often by clicking an 'I agree' button or simply by using the website. This approach is supported by organizations like the Federal Trade Commission, which offer guidelines for its application. The idea behind Notice and Choice is that if done correctly, it should ensure that people can freely and knowledgeably agree to how their data is collected and used. Additionally, it's believed that the collective result of individual consent decisions strikes a balance between privacy concerns and the advantages of data collection and use for consumers. The primary issue revolves around the complexity of these technologies, which often makes it challenging for users to fully understand what they are consenting to. While the notice provided is supposed to be clear and comprehensive, the reality is that the intricacies of data processing in AI are seldom conveyed in a manner that the average user can easily comprehend. (Sloan & Warner, 2014 p. 373-379)

Lastly, IoT is a modern concept increasingly used in daily life, especially noted for its role in producing large amounts of data, and is closely associated with consent. IoT refers to enabling various objects we utilise to have internet connectivity, transforming them into 'smart' items. In the realm of Internet of Things (IoT), anything that can be linked to a processing unit, such as a microcontroller, and connected to the internet, is regarded as a 'thing'. (Abed, 2016 p. 1) The IoT ecosystem, with its interconnected smart devices, generates vast amounts of data, presenting significant concerns about privacy, security, and informed consent. As these devices become more integrated into daily life, from smart homes to wearables, they blur the lines between public and private spaces, making it increasingly difficult to obtain informed consent. (Pathmabandu et al., 2023 p. 368-369)

3.3. Transparency

In this context, , it's essential to acknowledge how ethical principles are often interconnectedi with one principle can lead to the emergence of another. 3 key concepts in this dicsussion are, transparency, fairness and accountability. The link between transparency and accountability, is currently a subject of debate patricularly regarding the effectiveness of transparency.

To ciritically examine the effectiveness of transparency, we must first define two terms. Linear regression is a an approach used to model the relationship between two or more variables. (Rahman et al., 2018 p. 510) Meanwhile, machine learning, is a field of study that involves the development of algorithms that can learn from data and make predictions or decisions without being explicitly programmed. (Cecchetti, 2018 p. 1)

The challenge we face here is related to ‘black box nature of’ machine learning system. The term 'black-box',refers to lack of clarity in explainability and interpretability in AI systems. A problem arising mainly from the opacity of many modern AI models. As a result, although we can observe and have a basic understanding of the inputs and outputs, the exact internal processes remain unclear – this is what constitutes the 'black box' aspect. (Weber et al., 2023 p. 5) In instances, where linear regression is used, it is simpler to determine, explain and interpret the results compared to machine learning methods that utilise complex techniques like neural network.

Given the complexities highlighted above, it becomes clear that the achieving transparency in AI and machine learning is not just a technical challenge but also an ethical requirement. The 'black-box' nature of advanced machine learning models, while offering sophisticated analytical capabilities, presents significant challenges to ethical principles like transparency and informed consent. For further clarification, the challenge in the relationship between transparency and informed consent lies in the difficulty of explaining processes that are not fully acknowledged by the party responsible for providing the explanation, due to the nature of the process as mentioned above. It is crucial, therefore, to develop methods and tools that can reveal the inner workings of these models to some extent. This endeavour is not only crucial for maintaining public trust but also for ensuring that the algorithms we rely on are fair, unbiased, and accountable. Accountability, in this context, emerges as a complementary principle to transparency.

3.4. Accountability

First concept to address in this context is algorithmic accountability. The term refers to the obligation of organizations using algorithms to be answerable for the decisions these algorithms make. (Shin, 2022, p. 1172) In this process, there is a complicated interaction where the person making decisions needs to provide explanations about how the automated system is designed and works. The person affected by these decisions has the right to agree with or question these explanations. This could result in penalties or a need to change decisions if the explanations are not satisfactory. (Binns, 2018, p. 544)

Different strategies are used in Europe and the USA to regulate algorithmic accountability and transparency. In Europe, data protection laws govern audits of automated decision-making systems for instance by right to explanation ***. However, these laws currently don't explicitly require controllers to fully disclose the inner workings of their algorithms to data subjects. Future legislation may expand the GDPR to mandate providing data subjects with detailed insights into decision-making logic, prioritizing their interests over those benefiting from algorithm use. And The USA, lacks a comprehensive legal framework for algorithmic accountability and transparency. Regulatory requirements mainly exist within anti-discrimination laws, which are insufficient for addressing algorithm-related issues. Several US legislative initiatives are proposing mandatory impact assessments for automated decision-making systems, favouring entities that benefit from using algorithms. (Kuteynikov et al., 2020, p. 15)

A major challenge in this process comes from the difficulty in understanding in some algorithmic systems, especially those trained with machine learning as mentioned above. This lack of clarity often makes the decision-makers' ability to provide comprehensive accounts of their systems harder. Additionally, in addressing issues like algorithmic discrimination, decision-makers are required to explicitly include moral and political considerations into their models. This necessity underscores that the assumptions and values built into algorithms being a substantial part of any discussions about accountability. Therefore, the demand for algorithmic accountability is not just a call for transparency but also a deeper reflection of the underlying values and assumptions in these systems. (Binns, 2018, pp. 544-546)

3.5. Fairness

Fairness is defined as the practice of equally considering the moral interests of others. (Kwan et al., 2021, p. 5) The concept is assesed usually with transparency and accountability. Transparency is ensuring the standards applied by algorithms are clear and comprehensible to all involved parties, which would be vital for stakeholders to evaluate fairness of these algorithms. Moreover, the responsibility for the outcomes of these algorithms ultimately falls on their designers and operators. Collectively, these principles form a crucial connection for ethical decision making in algorithmic processes, ensuring technology is used in a way that is fair, transparent and responsible towards society.

Fairness in the context of AI and machine learning encompasses several dimensions. It's not merely about optimizing search engines or impartially ranking sevices. It encompasses the assurance that the decisions and predictions made by these systems do not perpetuate harmful human biases. This involves recognizing bias, as a manifestation of human actions that can impact individual rights, with algorithms having the potential to unintended bias and discrimination in algorithms. The fairness of the datasets used is also crucial; if the training data contains biased or discriminatory elements, outputs will likely present unfair action. (Varona and Suárez, 2022, p. 5)

Inclusion and diversity are integral aspects of fairness. Equal access and treatment through inclusive design are essential, and AI systems should ideally make the same recommendations for everyone with similar characteristics or qualifications. Regular testing of AI solutions in real-world applications is necessary to ensure they are free from biases related to gender, race, sexual orientation, age, religion, and other factors. (Varona and Suárez, 2022, p. 8)

For instance, use of advanced technologies like Human-Machine Learning Applications (HMLA) fairness would be compromised if certain groups of people are not allowed to question or challenge decisions made by these technologies. It's crucial that everyone has the opportunity to influence the development and application of these Technologies that impact their lives. It's also important for these Technologies to be designed in a way that people can understrand. However, companies creating them need to safeguard their unique designs and ideas, known as intellectual property. Policymakers, responsible for making laws and regulations, have a significant role in ensuring the fair use of these Technologies. They should carefully consider the appropriate use of these Technologies in critical sectors like healthcare. (Giovanola and Tiribelli, 2023, pp. 552-553)

Moreover, fairness in HMLA extends beyond mere non-discrimination and bias removal. Drawing from moral philosophy, it involves a more nuanced view that encompasses respect for individuals not just as equals but as specific persons with their own distinct attributes. This comprehensive view of fairness includes principles like equal opportunity and the right to justification, urging a shift in focus from solely anti-discrimination measures to developing technical and policy solutions that embrace the multifaceted nature of fairness. (Giovanola and Tiribelli, 2023, pp. 553-554)

In regards to fairness, "algorithmic neutrality" is as well an essential point. This concept adresses the need for algorithms to operate independently of external values, such as the financial interests or political views of its operators. Worth to note, algorithmic neutrality is a descriptive concept as opposed to the normative idea of fairness. (Phillips-Brown, 2023, p. 1-2)

Considering search engines as a case study, a search engine aiming for neutrality should ideally rank pages based on their relevance alone, excluding the influence of other values. However, this ideal is often not takes place in practice. For instance, the European Union's fine imposed on Google for prioritizing its own shopping service in search results over relevant ones by the reasonas financial interests, demonstrating a breach in neutrality. (Phillips-Brown, 2023)

CHAPTER IV – CASE LAW

4.1. 'Meaningful' Information Disclosure

The party involved in this preliminary ruling request is Dun & Bradstreet Austria, a firm specializing in business analytics and data provision. (CK v. Dun & Bradstreet Austria GmbH and Magistrat der Stadt Wien, No. C-203/22, 2022) Their case concerns whether, in cases of profiling, the controller must disclose information essential for making the result of the automated decision transparent in each individual case and whatwould be considered 'meaningful' as per Article 15(1)(h).

The case delves into connection between the right of access under Article 15(1)(h) of the GDPR is related to the rights guaranteed by Article 22(3) to express one's point of view and challenge an automated decision. It addresses the adequacy of the information provided on an access request is sufficiently 'meaningful' only if it enables the individual to exercise their rights under Article 22(3) effectively.

Moreover, the text examines whether Article 15(4) of the GDPR limits the scope of information to be disclosed under Article 15(1)(h) and how this limitation should be determined in each case. It also evaluates whether the provision of Article 4(6) of the Data Protection Law, which restricts access to information that would violate a business or trade secret, is compatible with the requirements of Article 15(1) in conjunction with Article 22(3) of the GDPR. Subsequently, questions whether this tension can be resolved by disclosing data required for accuracy checks to an authority or court instead of the data subject.

4.2. Extent of Data Access Rights

In another notable case concerning the interpretation of the GDPR, an individual named F.F. engaged in legal proceedings against the Austrian Data Protection Authority (DSB). (F.F. v. Österreichische Datenschutzbehörde and CRIF GmbH, No. C-487/21, 2021) Main dispute in the case was, the data processing company CRIF GmbH's refusal to provide F.F. copies of documents and database records that had his personal data. Important concerns were brought up by this circumstance about the applicability and reach of GDPR Article 15, which addresses people's right to access their personal data.

A number of essential GDPR elements were key to the argument. These included the regulation's recitals, which highlight the importance of protecting natural persons while processing personal data and ensuring transparency in such processes. Article 4 of the GDPR, defining 'personal data' and 'processing' in broad terms, and Article 12, which requires data controllers to deliver information in a clear, transparent, and accessible way, was also a key element.. However, the primary focus was on Article 15, which grants individuals the right to access their personal data and mandates data controllers to provide a copy of this data.

The court's responsibility was to interpret the meaning of 'copy' as used in Article 15(3) of the GDPR. The key questions included whether 'copy' meant a right to entire documents or database extracts containing personal data and the extent of information that should be provided electronically under this article. The court examined the GDPR' language, context, and aims, leading to several significant conclusions.

The concept 'copy' was interpreted as a faithful reproduction of personal data, covering a wide range of information processed in different ways. The access right was identified to include receiving copies of documents or database extracts if necessary for a full understanding of the data's context. The reproduction of personal data was required to be complete, intelligible, and respectful of the rights and freedoms of others. Notably, the term 'information' in Article 15(3) was clarified to refer specifically to personal data, excluding additional data like metadata.

4.3. Types of Automated Decision-Making Governed by GDPR

The case in question is a request for a preliminary ruling on the interpretation of the GDPR Article 6(1) and Article 22. The case arises from a dispute between an individual named OQ and the Land Hessen in Germany. (OQ v. Land Hesse, SCHUFA Holding AG, No. C-634/21, 2021) The main issue is centered around the refusal by the Hessischer Beauftragter für Datenschutz und Informationsfreiheit (HBDI) to mandate SCHUFA Holding AG to provide OQ access to, and erasure of her personal data.

In the case involving OQ, the issue arose when she was denied a loan based on a credit score calculated by SCHUFA. SCHUFA, known for its creditworthiness assessments using mathematical and statistical methods, partially complied with OQ's request to disclose information about her credit score and the data involved in its calculation. However, SCHUFA withheld certain details, claiming they were protected as trade secrets.

The main legal issue is whether the automated calculation of a credit score by SCHUFA falls under "automated individual decision-making" as defined in Article 22(1) of the GDPR. This is particularly relevant when such scores obviously influence third-party decisions, such as a bank's decision to grant or deny credit. For this reason the case was escalated to the Court of Justice for a preliminary ruling.

The Court's inquiry in this case focused on whether SCHUFA's probability value calculation constitutes "automated individual decision-making" under the GDPR. The Court took into account the broader intetnions and framework of the GDPR, which is primarily designed to protect individuals from the potential risks with automated data processing, including issues of transparency and fairness.

Ultimately, the Court decided that SCHUFA's method of calculating credit scores does indeed fall under the category of automated individual decision-making as per Article 22(1) of the GDPR. This decision was based on the substantial effect that these credit scores have on decisions made by third parties, such as banks, in relation to credit provision.

This case serves an example of the exten to which automated- decision-making processes, such as credit scoring, are governed by the GDPR. This highlights the GDPR's role in ensuring that automated systems perform transparently and fairly, particularly in situations where they have significant consequences for individuals.

4.4. Leaks out of Likes: The Cambridge Analytica Incident

The Cambridge Analytica, in 2018, centered on the unauthorized collection and use of personal data from millions of Facebook users. The core of the scandal involved a British political consulting firm, Cambridge Analytica, which had ties to the SCL Group. The company drew attention for its role in the 2016 U.S. presidential election and the UK's Brexit campaign. Cambridge Analytica gathered data from Facebook through an app “"This Is Your Digital Life" developed by Aleksandr Kogan, a data scientist at the University of Cambridge. The app, was presented as a research tool, and users who participated were paid for completing a survey. However, the app not only collected data from the survey participants but also from their Facebook friends, leading to the accumulation of a vast amount of personal information. (Bright, L., Wilcox, G., & Rodriguez, H., 2019)

The initial reporting of Cambridge Analytica's practices started in December 2015, however the full extent of the data breach became more widely known in March 2018, primarily thtough whistleblower Christopher Wylie, a former Cambridge Analytica employee. The scandal raised significant public concern over privacy and the impact of social media in politics, leading movements like #DeleteFacebook. The data collected included users' public profiles, page likes, birthdays, and current cities, and was detailed enough to construct psychographic profiles. These profiles were then used for politically tailored advertisement, influencing user opinions and behiavors. The data was notably used in campaigns such as those of Ted Cruz and Donald Trump in the U.S., and for the Leave.EU campaign in the UK demonstrating the significant impact of such targeted political advertising. (Rehman, I., 2019) Finally, Federal Trade Commission imposed a record $5 billion fine for violating consumer privacy. (Rachur, A., Putman, J., & Fisher, C., 2022) This incident underscored the significance of data protection laws like GDPR, which regulates personal data processing within the EU, and highlighted the need for comprehensive laws in the U.S., as exemplified by California's CCPA. The absence of a federal data privacy law in the U.S. creates inconsistencies and legal uncertainties for organizations. The Cambridge Analytica scandal influenced the development and emphasis on such data protection regulations. (Lee, C., 2020)

4.5. CCPA Opt-Out Requests

In a landmark CCPA case, the California Attorney General announced a settlement with Sephora, Inc., involving a $1.2 million fine. This was the first significant enforcement under the CCPA, focusing on Sephora's failure to disclose the sale of consumer data, not processing opt-out requests effectively, and not rectifying these issues within the allowed 30-day period under the CCPA.

Sephora's practices involved creating detailed consumer profiles based on their online activities, such as the type of device used, products added to the shopping cart, and location data. These practices, beneficial for targeted advertising, were deemed sales of personal information under CCPA. However, Sephora initially did not comply with the CCPA's requirements to inform consumers of this data sale and enable them to opt-out.

The settlement imposed several injunctive obligations on Sephora, including the requirement to provide clear mechanisms for consumers to opt out of the sale of their personal information and submit compliance reports to the Attorney General. This case underscores the importance of transparency in data practices and adherence to consumer data protection laws like the CCPA. (Linetzky, D.J., 2022)

CONCLUSIONS

1. There is a need for global consensus on data protection standards. Establishing such consensus would provide significant benefits, particularly for businesses operating in the international arena. A unified set of data protection guidelines would greatly simplify the complex landscape of compliance, reducing the burden on businesses to navigate varying regional and national regulations. This harmonization would not only support legal compliance but also promote a more consistent and transparent approach to personal data handling across borders, fostering trust and facilitating smoother international operations.

2. Businesses should be encouraged to develop more user-friendly and transparent consent mechanisms. This approach involves creating clearer, more understandable privacy policies, and providing users with easily accessible options to manage their data. Such mechanisms not only comply with legal requirements but also build trust with users by demonstrating respect for their privacy and autonomy in data usage. This balance ensures that businesses can leverage data effectively while upholding individuals' rights and choices regarding their personal information.

3. A collaborative approach is needed involving policymakers, technology companies, and privacy advocates to ensure that laws and regulations evolve in tandem with technological advancements. This collaboration is crucial for creating a regulatory environment that not only protects individual privacy rights but also supports innovation and growth in the tech sector. By working together, these diverse groups can develop comprehensive, forward-looking policies that address the rapid changes in technology, ensuring that regulations remain relevant, effective, and balanced in protecting personal data and encouraging technological progress.

4. A balance must be built between protecting business secrets and ensuring transparency. This becomes particularly delicate when disclosing requested information that may include trade secrets, which are of significant concern to businesses. One approach is to provide generalized information about the decision-making process without revealing specific algorithms or proprietary data. If necessary, detailed data can be disclosed to an independent authority or court for review, rather than directly to the individual. They would then verify the accuracy and fairness of the automated decision, providing an impartial assessment to the individual.

5. Businesses should be encouraged to invest in advanced data management systems tailored to comply with data protection laws. This includes ensuring that these systems adhere to principles like data minimization and purpose limitation, effectively reducing the amount of data collected

and used strictly for necessary purposes. Additionally, these systems should prioritize data security to protect against breaches and misuse, aligning with legal standards and boosting consumer confidence in how their data is managed and protected.

6. Finally, as technology becomes more integral to carious aspects of our lives, it is essential to manage it with a strong ethical compass and careful alignment with societal values. The advancements in technology, particularly in data management and algorithmic decision-making, offer tremendous benefits, but they also raise significant ethical and legal challenges. It is crucial to navigate these challenges by establishing global data protection standards, developing user-friendly consent mechanisms, collaborating across sectors to update laws, balancing transparency, and investing in robust data management systems. This approach ensures that while we leverage the power of technology, we also uphold the principles of privacy, fairness, and respect for individual autonomy, fostering a more harmonious integration of technology into our daily lives and societal framework.

LIST OF REFERENCES

1. Legal normative acts

California Consumer Privacy Act of 2018, Cal. Civ. Code § 1798.100 (West 2018) (amended 2020)

California Privacy Rights Act of 2020, Cal. Civ. Code § 1798.100 (West 2020)

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

2. Books, Articles and Reports

Abed, A. A. (2016) 'Internet of Things (IoT): architecture and design', in 2016 Al-Sadeq International Conference on Multidisciplinary in IT and Communication Science and Applications (AIC-MITCSA), IEEE, pp. 1-3.

Aksoy, H.C. (2022) 'Algorithmic Decision Making in terms of Personal Data Protection', Journal of Personal Data Protection.

Albrecht, J. P. (2016) 'How the GDPR will change the world', Eur. Data Prot. L. Rev., 2, pp. 288-289.

Anciaux, N., Frittella, S., Geoffroy, B., Nguyen, B., & Scerri, G. (2024, March) 'A new PET for data collection via forms with data minimization, full accuracy and informed consent', in EDBT 2024-International Conference on Extending Database Technology.

Asimov, I. (2004) I, robot, Vol. 1, Spectra.

Benedetta, G. and Tiribelli, S. (2023) 'Beyond bias and discrimination: redefining the AI ethics principle of fairness in healthcare machine-learning algorithms', AI & society, 38(2).

Berners-Lee, T. (2013) 'The power of the Web is in its universality', World Wide Web Consortium (W3C).

Blanke, J. M. (2022) 'The CCPA," Inferences Drawn," and Federal Preemption', Rich. JL & Tech., 29, pp. 18-19-70.

Borgesius, F.J.Z. (2020) 'Strengthening legal protection against discrimination by algorithms and artificial intelligence', The International Journal of Human Rights.

Brkan, M. and Bonnet, G. (2020) 'Legal and Technical Feasibility of the GDPR’s Quest for Explanation of Algorithmic Decisions: of Black Boxes, White Boxes and Fata Morganas', European Journal of Risk Regulation, 11(1).

Breen, S., Ouazzane K., and Patel, P. (2020) 'GDPR: Is your consent valid?', Business Information Review 37.1, p. 22.

Byun, D.Y. (2019) 'Privacy or Protection: The Catch-22 of the CCPA', Loy. Consumer L. Rev., 32.

Cara, C., & Dumitrașciuc, L. F. (2021) 'GDPR consent pop-ups. How are we thinking about them? An Elaboration Likelihood perspective', Journal of International Business and Management, 4(1), p. 2.

Cecchetti, A. A. (2018) 'Why Introduce Machine Learning To Rural Health Care?', Marshall Journal of Medicine, 4(2), p. 1.

Colcelli, V. (2019) 'Joint Controller Agreement Under GDPR', EU and comparative law issues and challenges series (ECLIC) 3, pp. 1030-1047.

Coleman, F. (2020) A human algorithm: How Artificial Intelligence is redefining who we are, Melville House UK.

Czerwiński, P. (2017) '“You Can’t Opt Out”: The Inescapability of Virtuality in Joshua Ferris’s To Rise Again at a Decent Hour', Roczniki Humanistyczne, 65(11), p. 8.

De Hert, P., Papakonstantinou, V., Malgieri, G., Beslay, L., & Sanchez, I. (2018) 'The right to data portability in the GDPR: Towards user-centric interoperability of digital services', Computer law & security review, 34(2), p. 194.

Finck, M. (2021) 'The Limits of the GDPR in the Personalisation Context', p. 5.

Ford, N. (2021) Data protection and privacy, p. 12.

Gamwell, B. (2022) 'California Consumer Protection Act (CCPA): Narrowing CCPA Exemptions Will Ensure Greater Privacy Protections', p. 3.

Garza, J. (2023) 'The Use of Artificial Intelligence (AI) in Medical Imaging', Across the Curriculum.

Giannopoulou, A. (2020) 'Algorithmic systems: the consent is in the detail?', Internet Policy Review, 9(1), pp. 4-8.

Gil González, E., & De Hert, P. (2019, April) 'Understanding the legal provisions that allow processing and profiling of personal data—an analysis of GDPR provisions and principles', in Era Forum (Vol. 19, No. 4, pp. 597-621). Berlin/Heidelberg: Springer Berlin Heidelberg, p. 612.

Goldman, E. (2021) 'An Introduction to California's Consumer Privacy Laws (CCPA and CPRA)', Santa Clara Univ. Legal Studies Research Paper, pp. 3-7.

Hill, R. K. (2016) 'What an algorithm is', Philosophy & Technology.

Hromisin, P. (2020) 'The CCPA and Law Practices: Figuring out Where You Stand', Prob. & Prop., 34, p. 60.

Jordan, S. (2022) 'Strengths and Weaknesses of Notice and Consent Requirements under the GDPR, the CCPA/CPRA, and the FCC Broadband Privacy Order', Cardozo Arts & Ent. LJ, 40, pp. 133-134-156-164-165.

Kahler, T. (2020, October) 'Data breach: 72 hours period extended on weekend?', in Turning Point in Data Protection Law (pp. 109-114). Nomos Verlagsgesellschaft mbH & Co. KG, p. 9.

Kearns, M., and Roth, A. (2019) The ethical algorithm: The science of socially aware algorithm design, Oxford University Press.

Kiesow Cortez, E. (2020) 'Data Breaches and GDPR', The Palgrave Handbook of International Cybercrime and Cyberdeviance.

Kuteynikov, D., Izhaev, O., Lebedev, V., & Zenin, S. (2020) 'Black box: transparency and accountability of automated decisionmaking systems', Revista Inclusiones, p. 15.

Kwan, D., Cysneiros, L. M., & Leite, J. C. S. D. P. (2021) 'Towards achieving trust through transparency and ethics', p.5.

Ledesma, S., et al. (2018) 'Analysis of data sets with learning conflicts for machine learning', IEEE Access, 6.

Lee, C. (2020) 'The aftermath of Cambridge Analytica: A primer on online consumer data privacy', AIPLA QJ, 48, p. 529.

Lindgren, P. (2016) 'GDPR regulation impact on different business models and businesses', Journal of Multi Business Model Innovation and Technology, 4(3), p. 250.

Lobschat, L., Mueller, B., Eggers, F., Brandimarte, L., Diefenbach, S., Kroschke, M., & Wirtz, J. (2021) 'Corporate digital responsibility', Journal of Business Research, 122, p. 876.

Mayfield, M. (2023) 'Talk Data to Me: Why Michigan Should Adopt a Comprehensive Data Protection Statute', Wayne St. UJ Bus. L., 6, pp. 9-10.

Minssen, T., Seitz, C., Aboy, M., & Compagnucci, M. C. (2020) 'The EU-US Privacy Shield Regime for Cross-Border Transfers of Personal Data under the GDPR: What are the legal challenges and how might these affect cloud-based technologies, big data, and AI in the medical sector?', EPLR, 4, 34, p. 47.

Moreira, H. (2023) 'Governing knowledge and technology: Technological pressure for convergence in EU, California, and China data protection regulation'.

Ong, R. (2023) 'Mandatory Data Breach Notification: Its Role in Protecting Personal Data', Journal of International and Comparative Law.

Partha Pratim, R. (2023) 'Benchmarking, ethical alignment, and evaluation framework for conversational AI: Advancing responsible development of ChatGPT', BenchCouncil Transactions on Benchmarks, Standards and Evaluations.

Pathmabandu, C., Grundy, J., Chhetri, M. B., & Baig, Z. (2023) 'Privacy for IoT: Informed consent management in Smart Buildings', _Future

Perumal, V. (2022) 'The Future of US Data Privacy: Lessons from the GDPR and State Legislation', Notre Dame J. Int'l Comp. L., 12, p. 114.

Phillips-Brown, M. (2023) 'Algorithmic neutrality', p. 1-2.

Politou, E., Alepis, E., & Patsakis, C. (2018) 'Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions', Journal of Cybersecurity, 4(1), p. 4.

Rachur, A., Putman, J., & Fisher, C. (2022) 'The effects of the digital age on privacy in the United States', The Business & Management Review, 13(2), p. 70.

Rahman, A.S., Ragu, V., Lee, M., Park, J., Cho, Y., Lee, M., & Shin, C. (2018) 'An Analysis Study Based on Linear Regression Model for Changes of Fruit Size over Plum Diseases', Journal of Knowledge Information Technology and Systems, 13(5), p. 510.

Reding, V. (2011) 'Your data, your rights: Safeguarding your privacy in a connected world', The Review of the EU Data Protection Framework, Brussels, 16.

Remenyi, D., and Griffiths, P. (2022) 'Data-its nature and management: A short note on some of the complexity behind the concept of data', Electronic Journal of Business Research Method.

Roberts, S. (2016) 'Claude Shannon, the Father of the Information Age, Turns 1100100', The New Yorker.

Rothstein, M.A., and Tovino, S.A. (2019) 'California takes the lead on data privacy law', Hastings Center Report.

Singh, L. (2016) 'Data Ethics—Attaining Personal Privacy on the Web', in Ethical Reasoning in Big Data: An Exploratory Analysis, pp. 81-90.

Singh, R., Amritha, P.P., & Sethumadhavan, M. (2022) 'Scoring Scheme to Determine the Sensitive Information Level in Surface Web and Dark Web', in International Conference on Advances in Computing and Data Sciences, Springer International Publishing, pp. 157-167.

Sloan, R.H., & Warner, R. (2014) 'Beyond notice and choice: Privacy, norms, and consent', J. High Tech. L., 14, pp. 373-379.

Tamò, A., & George, D. (2014) 'Oblivion, erasure and forgetting in the digital age', J. Intell. Prop. Info. Tech. & Elec. Com. L., 5, p. 71.

Thouvenin, F. (2021) 'Informational Self-Determination: A Convincing Rationale for Data Protection Law?', J. Intell. Prop. Info. Tech. & Elec. Com. L., 12, p. 246.

Tikkinen-Piri, C., Rohunen, A., & Markkula, J. (2018) 'EU General Data Protection Regulation: Changes and implications for personal data collecting companies', Computer Law & Security Review, 34(1), p. 135.

Tombal, T., & Graef, I. (2023) 'The Regulation of Access to Personal and Non-personal Data in the EU: From Bits and Pieces to a System?', p. 4.

Tschider, C.A. (2021) 'AI's Legitimate Interest: Towards a Public Benefit Privacy Model', Hous. J. Health L. & Pol'y, 21, p. 174.

ur Rehman, I. (2019) 'Facebook-Cambridge Analytica data harvesting: What you need to know', Library Philosophy and Practice, pp. 5-6.

Van Ooijen, I., and Vrabec, H.U. (2019) 'Does the GDPR enhance consumers' control over personal data? An analysis from a behavioural perspective', Journal of consumer policy.

Varona, D., & Suárez, J.L. (2022) 'Discrimination, bias, fairness, and trustworthy AI', Applied Sciences, 12(12), pp. 5-8.

Verma, S., Ernst, M., & Just, R. (2021) 'Removing biased data to improve fairness and accuracy', p. 1.

Vrabec, H.U. (2021) 'Data Subject Rights under the GDPR'.

Walker, L. (2020) 'California's Privacy Laws', Com. L. World, 34, p. 40.

Watson, H.J., & Nations, C. (2019) 'Addressing the growing need for algorithmic transparency', Communications of the Association for Information Systems, 45(1), p. 1.

Weber, A.L. (2021) 'Who Really Controls the Privacy Conversation? The Need for a Fundamental Right to Privacy in the United States', Corp. & Bus. LJ, 2, p. 198.

Weber, P., Carl, K.V., & Hinz, O. (2023) 'Applications of Explainable Artificial Intelligence in Finance—a systematic review of Finance, Information Systems, and Computer Science literature', Management Review Quarterly, p. 5.

Wong, M. (2021) 'Revising US Privacy Laws: New Laws Are Required to Fill in the Gaps of Current and Proposed Legislation to Account for New Technology and Future Emergencies', Brook. J. Corp. Fin. & Com. L., 16, p. 311.

Zhou, J., and Chen, F. (2022) 'AI ethics: From principles to practice', AI & Society, pp. 1-11.

Zetoony, D. (2022) 'Navigating the Chaos of the CCPA: The Most Frequently Asked Questions When Implementing Privacy Programs', Loy. U. Chi. J. Reg. Compl., 8, p.14.

3. Cases

CK v. Dun & Bradstreet Austria GmbH and Magistrat der Stadt Wien, CJEU, No. C-203/22, 2022.

F.F. v. Österreichische Datenschutzbehörde and CRIF GmbH, CJEU, No. C-487/21, 2023.

OQ v. Land Hesse, SCHUFA Holding AG, CJEU, No. C-634/21, 2021.

SUMMARY

Legal and Ethical Considerations in Personal Data Usage for Algorithmic Training

Anil Berk Gumus

This research delves into the current status of legal and ethical considerations in the use of personal data for algorithmic training. The introduction of the thesis aims an understanding the concepts in the era of big data and AI. It presents the terms from the intersection among law, technology and ethics. The introduction contextualizes the importance of the subject matter in the modern digital landscape, where personal data has become a cornerstone of technological advancement, particularly in algorithmic training and artificial intelligence (AI).

The thesis explores the complexities involved in ensuring data privacy and security, the rights of individuals in the digital sphere, and the responsibilities of organizations handling personal data. Furthermore, it discusses the global impact of these laws, considering the differences and similarities between the GDPR and CCPA, and how these regulations influence international data handling practices. The thesis primarily focuses on dissecting and analyzing major data protection regulations, notably the GDPR of the European Union and the CCPA of the United States. It aims to unravel how these legislations shape the practices of collecting, processing, and utilizing personal data, particularly in the context of training algorithms that increasingly permeate various sectors.

In addition to legal analysis, the thesis is adressing ethical considerations, highlighting the importance of ethical decision-making in the use of personal data for algorithmic purposes. This includes examining the implications of consent, data subject rights, and the societal impacts of data usage in algorithmic training. Moreover, thesis is navigating cases that serve as points of reference for understanding how legal theories, principles and regulations are interpreted and enforced in judicial settings. It also highlights the role of judicial decisions in shaping future policies and practices regarding personal data usage in technology.