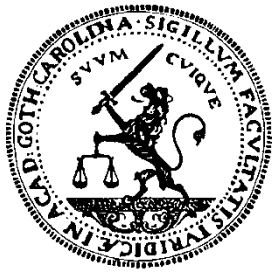


FACULTY OF LAW
Lund University

Seada Bogucanin Volic

Federated learning in autonomous vehicle setting- GDPR perspective

JAEM03 Master Thesis

European Business Law
30 higher education credits

Supervisor: Ana Nordberg

Term: Spring 2022

Contents

Acknowledgment

I hereby want to express my deepest gratitude to all the women in tech who shaped my influence to attain all of the aspects covering this paper; most notably, I would like to represent the support and mentoring given to me by a supervisor and Professor Ana Nordberg who strived to provide me with all the directions and affirm my thoughts in proper order in this period of my academic life.
Moreover, I would like to thank my family and friends, who never failed to show their love and understanding for my work during each writing stage.

Abbreviations

PET	Privacy Enhancing Technologies
FL	Federated Learning
ML	Machine Learning
GDPR	General Data Protection Regulation
CJEU	Court of Justice of the European Union
AI	Artificial Intelligence
IoT	Internet of Things
AV	Autonomous Vehicles
DPO	Data Protection Officer
DPIA	Data Protection Impact Assessment
RFID	Radio Frequency Identification Devices
PbD	Privacy by Design
DPA	Data Protection Authority
DPC	Data Protection Commissioner
EDPB	European Data Protection Board
DPS	Data Processing Systems

1 Introduction

From the very get-go, autonomous vehicles were the only reality in science fiction movies; today's autonomous vehicles never faced faster-emerging technologies. Even though the history of car making has been cautiously increasing the levels of automation, we still have only test models of fully unsupervised or level 5 autonomous vehicles.

The first-ever cruise control was introduced in 1958; later, in 1995, 'adaptive cruise control' was launched. In 2010, it functioned as a blind-spot intervention and active lane-keeping assist.[1] More recently, the testing of autonomous vehicles has become the focus of the substantial commercial investment.[2] For instance, Google presented driverless cars in 2009[3]; later, Nissan began testing driverless taxis in Japan. The UK government has invested millions in research and development, including various projects[4]. Many other countries as well. Uber itself started a company for driverless cars in 2015[5] and later sold its start-up in 2017[6] due to so-called allegations of technology theft.

Besides the fact that we had many broken promises about autonomous vehicles in the past few years, which are already supposed to be on the road[7], the technology is not subsided. In 2015 the Guardian predicted that we would be permanent backseat drivers until 2020[8]. Business Insider announced in 2016 that '10 million self-driving cars would be on the road by 2020[9]. Many car producers forecasted making self-driving cars, such as General Motors, Google's Waymo, Toyota, Honda, and Tesla. 2022 is here, and self-driving cars are not.

Regarding self-driving car technology, a fascinating fact is that law usually comes after the technology is implemented, or legislation is very much lately. However, in this case, we already have many national strategies for

---

[1] Channon M, McCormick L. and Noussia K., 'The Law and autonomous vehicles', (2019) at 1
[2] Ibid at 2
[3] https://www.investopedia.com/articles/investing/052014/how-googles-selfdriving-car-will-change-everything.asp#:~:text=In%202018%2C%20Waymo%20announced%20that,except%20in%20some%20trial%20programs.
[4] Ibid at 2 (UK autodrive, HumanDrive, Venturer, Gateway)
[5] https://www.nytimes.com/2020/12/07/technology/uber-self-driving-car-project.html
[6] https://www.bbc.com/news/business-55224462#:~:text=Uber%20is%20selling%20its%20driverless,self%2Ddriving%20cars%20a%20reality.
[7] https://www.vox.com/future-perfect/2020/2/14/21063487/self-driving-cars-autonomous-vehicles-waymo-cruise-uber
[8] https://www.theguardian.com/technology/2015/sep/13/self-driving-cars-bmw-google-2020-driving
[9] https://www.businessinsider.com/report-10-million-self-driving-cars-will-be-on-the-road-by-2020-2015-5-6?r=US&IR=T

autonomous vehicles[10]; for instance, the Swedish official Transport Agency already has an application form for testing autonomous cars and all the necessary guidelines. [11]Germany [12]considers itself a leader in this area and has introduced regulations accordingly[13]. France has had their national strategy[14] and law for regulating driverless vehicles since 2016[15]. Even Europe published an EU Strategy for the mobility of the future in 2018[16] and later in 2019 released Guidelines on the exemption procedure for the EU approval of automated vehicles.[17] Finally, in 2020, the EU published Guidelines on processing personal data in connected cars, as many of these questions were unanswered in the previous legislation. Anyhow, we have all of the upper mentioned legislation, but autonomous vehicles are not on the roads yet. The following sections of this paper will also touch upon this topic.[18]

All the hype about autonomous vehicles sounds very much attractive and so touchable. Still, from the lawyer's point of view, one big challenge (besides many others) is data usage in this setting. Connected autonomous vehicles are considered the Internet of things (more about this will follow in future sections). They require an enormous amount of data for their machine learning processes. Sweden sets Artificial intelligence (in further text Ai) and machine learning (in future text ML) as one of its main development goals[19]. Accordingly, in 2019, an organization named Ai Sweden was formed by the private sector and academia supported by the Swedish government[20]. '*The approach was based on assessing Sweden's Ai capabilities by innovation agency Vinnova, identifying several areas to address. During the summer of 2018, Lindholmen Science Park was assigned the task by nova to conceptualize and build a model for a national center for AI-related research, innovation, and education*[21]. AI Sweden has already set its highly ambitious agenda during its short existence. The AI Sweden's research and development 'kitchen' project is decentralized AI-

[10] https://leonard.vinci.com/en/the-national-strategy-for-automated-mobility-enshrines-cooperation-between-the-autonomous-vehicle-and-the-infrastructure/
[11] https://www.transportstyrelsen.se/en/road/Vehicles/self-driving-vehicles/
[12] https://www.bmvi.de/SharedDocs/EN/publications/strategy-for-automated-and-connected-driving.pdf?__blob=publicationFile
[13] https://www.jdsupra.com/legalnews/germany-takes-the-lead-with-a-new-law-7746782/#:~:text=German%20lawmakers%20have%20approved%20a,operation%20as%20soon%20as%202022.
[14] https://europe.autonews.com/article/20180808/ANE/180809840/france-pushes-for-highly-automated-vehicles-by-2022
[15] https://www.insidetechlaw.com/blog/france-new-legislative-developments-for-autonomous-vehicles
[16] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0283
[17] https://ec.europa.eu/docsroom/documents/34802
[18] https://edpb.europa.eu/our-work-tools/documents/public-consultations/2020/guidelines-12020-processing-personal-data_en
[19]https://www.government.se/4a7451/contentassets/fe2ba005fb49433587574c513a837fac/national-approach-to-artificial-intelligence.pdf
[20] https://www.ai.se/en/about-aisweden/our-story#:~:text=AI%20Sweden%20was%20launched%20in,the%20Swedish%20Innovation%20agency%2C%20Vinnova.
[21] Ibid

based learning, which aims to preserve and protect data. Ai Sweden defined this concept as follows: '*While centralized AI systems with access to all data and information in the cloud or a single device are easier to engineer and implement, decentralized systems are becoming increasingly important, not least a privacy restrictions and limited bandwidth. Decentralized AI will play a critical role in using AI in society and ensure that hospitals or autonomous cars can share and benefit from knowledge centrally while keeping the sensitive data safe and local*[22]*'.*
The largest company by turnover in Sweden, Volvo cars, and Volvo AB[23] is also setting the goal for fully automated vehicles. For Volvo, the future is 'happening now[24]. That being the case, Volvo releases the very first autonomous commercial solution in Norway, where the autonomous solution is transporting limestone from an open pit mine to a nearby port. '*The solution consists of limestone transported by six autonomous Volvo FH trucks on a five-kilometre stretch through tunnels between the mine and the crusher.*'[25]
Later on, autonomous vehicles, but this time based on federated learning, were presented at the Virtual ITS European Congress[26], the topic 'Federated learning to enable automotive collaborative ecosystem: opportunities and challenges presented by RISE Research Institutes of Sweden discussed the pros and cons of decentralized AI, its 'enormous potential of connected vehicle data.'[27][28]This has been conducted by illustrating the benefits of using federated learning (FL) driver's action classification by demonstrating the potential for collaborative machine learning without data sharing as a privacy preservation technique[29], such as live data sharing between Volvo cars and Volvo trucks for improving traffic safety.

As we face a new technological revolution and emerging development of Artificial Intelligence and Machine learning-based applications and services, data privacy and security of data started to become a critical component and more debated matter.
Conventionally, data is collected and aggregated in a data center where machine learning models are trained[30]. A segregated, centralized model has brought severe privacy risks to personal data leakage, misuse, and abuse.

---

[22] https://www.ai.se/en/projects-9/decentralized-ai
[23] https://www.largestcompanies.com/toplists/sweden/largest-companies-by-turnover/industry/manufacture-of-motor-vehicles-trailers-and-semitrailers
[24] https://www.volvogroup.com/en/future-of-transportation/innovation/automation.html
[25] https://www.volvogroup.com/en/future-of-transportation/innovation/automation.html
[26] https://www.diva-portal.org/smash/get/diva2:1590504/FULLTEXT01.pdf
[27] Ibid
[28] *'In view of the enormous potential of connected vehicle data and to create equal competition for innovative vehicle services and the data economy, the industry has been working on different initiatives to enable data sharing. One notable commercial solution is direct OEM to OEM data sharing such as the live data sharing between Volvo cars and Volvo trucks for improving traffic safety (2). Another solution is the on-going pilot project Data for Road Safety (3) for decentralized sharing of safety-related vehicle data between OEMs, service providers, as well as member countries' authorities'* Ibid
[29] Ibid
[30] Troung N, Sun K, Wang S and others- 'Privacy Preservation in Federated Learning: An insightful survey from the GDPR perspective' (2016)

These segregated models of data centres and their transfer to IoT devices are especially debated in the context of IoT (internet of things) as their security has been compromised often in previous years[31].
Besides the difficulties in transferring and sharing data across data sources, the challenge strains compliance with data protection regulations and administrative procedures such as the EU General Data Protection Regulation (GDPR)[32].
In this regard, Federated Learning (FL) sparked out as a potential solution that might facilitate distributed, collaborative learning without disclosing original training data while claiming that it still complies with GDPR[33].
This new method especially sparked the interest in the automotive industry in developing systems for autonomous vehicles.

Even though Google presented the FL technique in 2016[34] to overcome privacy challenges, it attracted enormous attention and was a promising innovative solution.
As autonomous vehicles represent IoT (internet of things), well-known information is that IoT carries a high cyber security risk of data breaches.[35]
Therefore, this system of machine learning sounds like a promising way to preserve data compliance and, at the same time, minimize the risk of data exposure.
A few federated learning engineers in the conducted interview claimed that different privacy preservation techniques might be used depending on the usage of the federated learning model.

1.1 Hypothesis

Therefore, this paper is dedicated to researching the question of privacy preservation, anonymization, and privacy risk assessment in the sample FL technique. This paper will try to manifest a bigger picture of privacy preservation techniques in FL as a 'good example' which is meant to comply with GDPR.[36]

1.2 Scope

This paper analyses the autonomous vehicle context's federated learning principles and technical advantages. The relevant legislation is General Data Protection Legislation. Moreover, even though many papers are respectively

[31] https://www.eurofins-cybersecurity.com/news/security-problems-iot-devices/
[32] Ibid
[33] https://www.ai.se/en/news/new-federated-learning-project-could-solve-gdpr-issues
[34] Mammen M. Priyanka, 'Federated learning: Opportunities and Challenges'. (2021) at 1
[35] Persson F. 'Information security risk review and analysis for the future autonomous vehicle', (2017) Luleå University of Technology
[36] Ibid at 1

discussing ethical matters, this paper will not touch upon this area of research.
Hereafter this work will focus on level 4,5 of automation, meaning- high or fully automated vehicles[37].

1.3 Research question

This paper will investigate the concept of federated learning as a newly developed machine learning technique. Furthermore, we will try to define personal data in this context. For instance, could a result from a local data station (autonomous vehicle) sent further to another station or a central server be regarded as non-personal data? And if, when exactly does it classify as personal data?
Furthermore, it will discuss the success/ unsuccess of anonymization and its relevance to the current technical developments.
Moreover, this paper will prove the benefit of federated learning as an anonymization technique and its supremacy compared to the pseudonymization technique and even other anonymization techniques.
Could Federated Learning as a machine learning technique be enough privacy preservation technique? Or should the federated model need to be empowered by efficient privacy-preserving methods to comply with GDPR? The following sections will briefly discuss this matter.

1.4 Research methods

An overview of the methodology undertaken is as follows:

- **Legal Review**: By a set hypothesis and for this work, a review of existing legal remedies under European data protection laws was considerable to conjoin new technological development regarding autonomous vehicles based on federated learning technique with the current legislation to give a critical overview.
- **Literature Review**: a proper review of all available sources in this area was conducted throughout this thesis where applicable. This covers all available legal papers, legislation, guidelines, and technology-related resources[38].
- **Inter-disciplinary Legal Research** was one of the essential research methods as there was an apparent deficit in current legislation regarding this type of technical development. This method was conducted to combine non-legal data with legal data. The intention was to prove the effectiveness of legal instruments in

[37] Version 4.1 The Guidelines hereafter have been supported by the Technical Committee on Motor Vehicles of 12 February 2019
[38] Tyler T, 'Methodology in Legal Research', 2017, accessed 23.03.2021.

this area and their compatibility with current technological developments[39].

- **Quantitative Empirical Research, together with the method of evaluation,** was used to analyse the current situation 'in the field' to understand the origin of legal deficiency within the process of the recent autonomous vehicle developments, as it was necessary to establish the level of effectiveness which might discourage new upgrades in this area.
- **Combination of structured and semi-structured interviews** was conducted among the limited number of experts in the industry in the Nordics. [40]This minor sample of interviewed experts did not aim to give statistical data which could define the picture in this industry. But rather to understand the processes they do while developing FL, their awareness of the legal background of their progress in FL processes, their understanding of current legislation, and finally, their stand on guarantees about all technological developments in this area. As engineers in question are involved in the projects in the R&D phase, their names, and the companies they work for will remain restricted[41].

1.5 Expected benefits of the research

As the first legal paper (*to our knowledge*) in this area, this paper aimed to summarize all the advantages of FL from a legal perspective as a privacy preservation technique compliant with GDPR. Furthermore, it aimed to contribute to the experts in the field of law responsible for implementing privacy measures within their organizations. Moreover, it helps those with a technical background relate their technology to the current legislation. Finally, this paper might be helpful to the relevant Data Protection Authority during the proposed (see conclusion remarks) prior consultation to ensure compliance with the current legislation.

[39] Aynale F & Vibhute K, 'Prepared under the Sponsorship of the Justice and Legal System Research Insitute', 2009

[40] For that purpose see Appendix 2

[41] Hoecke Van Mark, 'Legal Doctrine: Which Method (s) for What Kind of Discipline', 2021

2 Federated Learning Introductory notes

Federated learning (FL) is a privacy-preserving collaborative Machine Learning (ML). It promises a process that allows multiple stakeholders to share information through ML models without exposing raw data, thus protecting privacy.[42] As the automotive industry is motivated to use FL as it guarantees privacy in the autonomous vehicle setting, the following will discuss how this system works and whether could this be enough to preserve privacy as required by GDPR.

In the past few years, with AI and Machine learning development, data privacy and data security have become critical challenges usual process in ML is that data is collected and then aggregated in a data center on which machine learning models are trained. [43]However, this has been shown to have a high privacy risk of personal data leakage, misuse, or abuse. Moreover, in the era of the Internet of things and big data, in which data is generally distributed, transferring a considerable number of data-to-data centers for processing seems to be a risky solution. Besides the cyber security threats, sharing information needs to comply with GDPR, which does not make things any easier. In this regard. Federated Learning (FL) is a promising perspective solution that facilitates distributed learning without disclosing initial training while complying with GDPR.[44]

2.1 How does it work?

Symbol of the 20th-century economy, the vehicle is one of the mass consumer products that has impacted so many levels. As usually, the automobile is often associated with the realm of freedom where cars are often considered more than just a means of transportation. Indeed, they represent a private area in which people can enjoy a form of autonomous decisions without encountering any external interference[45].
The fact is that cars are becoming connected and infrastructure[46]. According to many predictions, 98% of the vehicles were supposed to be touched and generated 25GB of data per hour[47].

[42] Englund C, Torstensson M & Chen L- 'Federated Learning to enable automative collaborative ecosystem: opportunities and challenges' at 1
[43] Ibid
[44] Ibid
[45] Guidelines 01/2020 on processing personal data in the context of connected vehicles and mobility related applications (adopted on 09th March 2021) at 4
[46] Supra
[47] Frost & Sullivan, 'Otonomo, 2018 European Car Data Platform New Products Innovation Award.' At 3

In this ecosystem, not only vehicles but drivers and passengers become more connected. Accordingly, many launched models over the past few years on the market integrate sensors and connected onboard equipment, which may collect and record, among many other features, the engine performance, the driving habits, the locations visited, and potentially even the driver's eye movement, their pulse, biometric data to uniquely identify a natural person[48].

Therefore, the variety of data collected on this occasion requires the traditional players in the automotive industry to be shaped by the new players belonging to the digital economy. These new players may offer various services such as online music, road condition, and traffic information or provide driving assistance systems and services, such as autopilot software, vehicle condition updates, usage-based insurance, or dynamic mapping[49]. As those vehicles are interconnected via electronic communication networks, road infrastructure managers and telecommunication operators involved in this process also play a crucial role concerning the potential processing operations applied to the participants in the traffic (both drivers and passengers) and their data. Concrete about the type of data and their legal analysis will be conducted in the following sections.

Consequently, connected vehicles are generators of enormous amounts of data, most of which can be considered personal data since they will relate to drivers and passengers. Many of these technical data are still produced by a natural person and are still allowing their direct or indirect identification by the data controller or any other person.[50]
However, the model or use case presented in this paper is based on implemented federated learning into the system of autonomous vehicles where the model is trained in a decentralized manner by the clients, for instance, data curators, preventing the server from directly accessing those private data from the clients. This learning mechanism significantly challenges the attack from the server side as all the analytics is done locally on the IoT (in this case, autonomous vehicles[51]). The benefits of the shared model are trained from this rich data without the need to store it centrally.

Considering all this, the challenge is incorporating the 'protection of personal data' dimension from the product design phase and ensuring that car users enjoy transparency and control related to their data. This approach may help increase users' confidence and trust in this technology.[52]

---

[48] Ibid at 4 or for more information https://fpf.org/blog/future-privacy-forum-releases-infographic-mapping-data-connected-car-advance-ftc-nhtsa-workshop/
[49] Ibid Guidelines at 4
[50] Ibid Guidelines at 5
[51] Z. Wang, M. Song, Z. Zhang and others, 'Beyond inferring class representatives: User-Level privacy leakage from federated learning' (2018) at 1, accessed 18.04.2021.
[52] French Compliance Package-(October 2017 edition) Connected vehicles and personal data at 2

Today many countries in Europe (such as the UK, France[53], and Germany[54]) already have compliance packages and critical principles for data security and privacy guarantees for an autonomous vehicle. However, neither discusses the federated learning principle incorporated into the vehicle's system. Is federated learning a privacy preservation promise and a solution for the future? Federated learning is currently being developed by AI Sweden and its partners and will have many use cases[55].

2.2 What are the benefits of autonomous vehicles based on FL

According to World Health Organization data, almost 1,2 million people die each year because of road accidents[56]. Road traffic injuries are the leading cause of death among young people (aged 15-29)[57]. Moreover, half of those dying on the roads are 'vulnerable road users such as pedestrians, cyclists, and motorcyclists.[58]

Therefore, fully automated cars are defined as a car that can perceive their environment and decide which route to take to their destination or how to avoid traffic[59]. Those vehicles will primarily act as 'robocars'[60] as their performance will be based on the inputs from various sensors, computer processors, and road maps. Those 'supercars' have their aim to reduce car crashes, energy consumption, and pollution. [61]It is also claimed that autonomous vehicles will reduce driver stress and mobility of non-drivers, increases safety, increase fuel efficiency, provide more efficient parking, reduce costs, support shared vehicles,[62] etc.

For this purpose, they collect an enormous amount of data. More about collected data will be discussed in the following sections.

However, we should not forget about the socio-economic benefits of this technology, as it seems those developments will not be possible without the ability to process a vast amount of data. FL implementation in autonomous vehicles promises to be an 'along waited' solution for car producers as data processors which could guarantee compliance with GDPR.

---

[53] www.cnil.fr check under 'Connected vehicles and personal data-compliance package'
[54] https://www.roboticsbusinessreview.com/unmanned/germany-creates-ethics-rules-autonomous-vehicles/
[55] https://www.ai.se/en/node/81535/federated-learning
[56] https://www.who.int/news/item/23-04-2007-world-youth-assembly-meets-to-tackle-road-safety
[57] https://www.who.int/news/item/19-04-2007-road-traffic-crashes-leading-cause-of-death-among-young-people
[58] Ibid
[59] Jaswal A. Kumar & Rajasekhar MV, 'Autonomous vehicles: The future of automobiles' at 1
[60] Ibid
[61] Ibid
[62] Ibid

3 Personal data definition (concept of privacy vs. personal data)

The following two chapters will be dedicated to GDPR. Moreover, those two chapters will try to distinguish between personal data and privacy. Furthermore, this paper will attempt to illustrate the recommended approach from the legislator's perspective to buster the technological developments in this area. Additionally, it will describe and legally define all the data collected in the autonomous vehicle's ecosystem and present data controller obligations, particularly privacy by design measures.
As a first step in the legal analysis of autonomous vehicles based on FL, it is crucial to define whether the federation results might be considered personal data. While doing this, it is also necessary to draw a clear line between privacy as a concept[63] and personal data as a legal definition.
Privacy could result from a conceptual inversion that relates to how the purpose of privacy has been conceived. Some scholars claim privacy has a lousy reputation in privacy theory, as the right to privacy is usually not absolute. Even though in the EU, privacy represents a long-established notion even though the concept varies in the different member states. Other regulations ' perspectives and history will be mentioned below to discuss whether the type of data collected and processed for the purported model could be defined as 'personal.' Therefore, in the following text, this chapter will try to distinguish the concept of privacy compared to the personal data definition. For that reason, the next section will mention various legislation to demonstrate the importance and legal protection stipulated in the EU legal framework. However, this paper will assess the concept of federated learning from the GDPR perspective.

Personal data protection laws in Europe are not a new legal institute. GDPR Directive, which came into force in 2018, inherited the 1995 Data Protection Directive, which was adopted when the internet was in its infancy. According to General Data Protection Regulation (GDPR), the term 'personal data' is the 'entryway' to the application of this Regulation. This directive will only be applicable if the processing concerns personal data. The concept of personal data is defined in Art. 4 (1) as 'personal data are any information which is related to an identified or identifiable natural person.[64]'

'*The data subjects are identifiable if they can be directly or indirectly identified, concretely referencing to an identifier's name, an identification number, location data, an online identifier, or one of several special characteristics, which expresses the physical, physiological, genetic, mental, commercial, cultural or social identity of these natural persons. In practice, these also include all data that can be assigned to a person in any*

[63] Cohen J, 'How the privacy got a bad name for itself', 2012
[64] https://gdpr-info.eu/issues/personal-data/

way. For example, the telephone, credit card or a personal number of a person, account data, number plate, appearance, customer number or address are all personal data.'[65]

3.1.1 Eu Commission definition

Has a similar concept where '*personal data is any information related to an identified or identifiable living individual? Different pieces of information, which can lead to identifying a particular person, are constituted as personal data*. [66]According to EU Commission, the data which has been **de-identified**, **encrypted,** or **pseudonymized** but can be used to reidentify a person's identity is still considered personal data and falls under the scope of GDPR. On the other hand, personal data, which is anonymous so that individuals are no ger identifiable, is no longer considered personal data. For that purpose, the anonymization **must be irreversible**. The GDPR guarantees protection regardless of the technology used for processing data; it applies both to manual and automated processing, providing that data is previously organized by pre-defined criteria (for example, alphabetical order). The protection is also guaranteed regardless of how the data is stored in an IT through video surveillance or on paper; in all cases, personal data is subject to the protection requirements set out in the GDPR.[67]

3.1.2 United Nations

In their framework, the UN does not recognize personal data protection as such. However, the right to privacy is usually connected to this legal institute and is a long-established fundamental right in the international legal order. Notably, Article 12 of the Universal Declaration of Human Rights (UDHR) guaranteed for the first time an international instrument for the protection of an individual's private sphere against intrusion from others, especially from the state. In the following two resolutions (2013, 2016, and 2017), the United Nations has adopted a few resolutions on privacy issues entitled to 'the right to privacy in the digital age.'[68]

3.1.3 The European Convention on Human Rights

The contracting parties have an international obligation to comply with ECHR even though some of the Council of Europe member states have not incorporated or given effect to ECHR in their national law, which requires

[65] Ibid
[66] https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_en
[67] Ibid
[68] Handbook on EU data protection law at 22

them to act on the convention's provisions[69]. However, the Court affirmed that surveillance constitutes an interference with respect for private life in its practice. [70]
The Court extended the definition of personal data even though this right is not absolute. It might be limited, if necessary, for an objective of general interest or to protect the rights and freedoms of others. For instance, regarding modern scientific techniques in the criminal justice system, the Court ruled that the protection afforded by Article 8 of the Convention would be unacceptably weakened if such practices were allowed at any cost and without carefully balancing the potential benefits of the extensive use of such techniques against vital life interests.[71]

3.1.4 The European Union law

EU law comprises the primary and the secondary. For instance, all EU Member States ratified the Treaty on the European Union (TEU) and the Treaty on the Functioning of the European Union (TFEU) and formed 'primary EU law.' The original treaties of the EU do not contain any reference to human rights or their protection, and one of the main principles is the principle of conferral. According to this principle, the EU is entitled to intervene only within competencies conferred upon it by the Member States[72]. However, after numerous court decisions have been made alleging human violations, CJEU provided a crucial interpretation of the treaties. According to the CJEU, these general principles reflect the content of human rights protection found in national constitutions and human rights treaties, particularly ECHR. Therefore, to recognize that policies could

---

[69] *Moreover, in Convention for the Protection of Individual regarding Automatic Processing of Personal Data, personal data are defined almost identical to upper mentioned definitions. This Convention applies to all data processing carried out by both the private and public sector, including data processing by the judiciary and law enforcement authorities[69].*
As regard the processing of personal data, the principles laid down in the convention concern among other things, fair and lawful collection, and automatic processing data, for specified legitimate purposes. In fact, that means that data should not be used for ends incompatible with these purposes and should kept no longer than needed. This also concern the quality of the data, in particular- must be adequate, relevant, and not excessive (proportionality) as well as accurate.
All EU Member States have ratified Convention 108, but the EU never conduct an accession as a legal subject. However, Convention 108 is open to non- Contracting parties of the Council of Europe and to date 51 countries are parties to Convention 108, Ibid at 23

[70] See ECtHR Case *Klass and Others v Germany*, No5029/71 (1978), Case *Rotaru v Romania*, No28341/95 (2000) and Case *Szabo and Vissy v Hungary*, No37138/14 (2016)

[71] *Moreover, 'the protection of personal data is of fundamental importance to a person's enjoyment of his or her right to respect for private and family life as guaranteed by Article 8'[71]. Therefore, the use and release of information relating to an individual's private life which is stored in a secret register comes within the scope of Article 8 §*1.
Guide on Article 8 of the Convention- Right to respect for private and family life, more at Case *S. and Marper v the United Kingdom* No30562/04 and 30566/04 §112.

[72] Handbook on European data protection law, 2018 edition, at 27

impact human rights and make citizens feel more connected to the EU, the EU proclaimed the Charter of Fundamental Rights of the European Union (Charter). The Charter covers six different sections of fundamental rights- dignity, freedoms, equality, solidarity, citizen's right, and justice. The Charter became a legally binding primary EU law for the Member States. The Charter extended the right to private and family life (Article 7) to the right to protect personal data (Article 8). Not only does it explicitly mention a right to data protection in Article 8 (1), but it also covers the crucial principles of personal data protection in Article 8 (2).
This right is expressly provided in Article 16 of the TFEU, under the general principles of the EU. At the same time, Article 16 creates a legal basis, which grants the EU the competence to legislate on data protection matters. This Article served as a legal basis for adopting General Data Protection Regulation and forming the independent supervisory data protection authorities[73].

3.1.5 Personal data definition dispute

Therefore, considering a comprehensive set of legislation associated with privacy and the definition of personal data, could a result from a local data station (autonomous vehicle) sent further to another station or a central server be regarded as non-personal data? Is this current legal framework good enough to define all the current and future technical developments? It is essential to point out that the Guidelines[74] define data *associated with connected vehicles as personal data to the extent that it can link to one or more identifiable individuals*[75]. This covers technical data concerning the vehicle's movement and condition. Some of the data generated by connected cars may also warrant special attention given their sensitivity or potential impact on the rights and interests of data subjects[76]. But guidelines give recommendations over data collected in a 'regular' autonomous vehicle setting, not in federated learning, which already provides a high level of privacy to the individuals.
It cannot be denied that there is a specific dispute around the abstract definition of personal data. Accepting that even technical data can be categorized as personal data leads to when exactly data qualifies as personal data in each situation.[77]

[73] Handbook on European data protection law, 2018 edition, at 28
[74] Guidelines 01/2020 on processing personal data in the context of connected vehicles and mobility related applications
[75] Ibid at 15
[76] Ibid at 15
[77] Federation internationale de l' automobile region I- Europe, The Middle East and Africa, 'What EU legislation says about car data', Legal Memorandum on connected vehicles and data', 2017 at 8

3.2 Relative vs. absolute approach

In Article 4.1 GDPR data qualifies as personal and relates to an identifiable person, and such, *'An identifiable person can be identified, directly or indirectly.'*[78] According to the *Federation internationale de l' automobil* vital issue in this definition is triggering the legal dispute with implied reference to an unclear third party: The definition is not about the nature of data but rather about someone's capability to **identify a person behind the data**[79]. It concludes that the upper mentioned provision is entirely open regarding whose capabilities are relevant. As a result of this debate, the upper mentioned paper proposed two approaches to resolve the open question[80].
The so-called *'relative approach'* considers only the company controlling the data (the data controller) to be legally accountable; accordingly, the same data might be deemed personal data in the hand of one company[81] but not regarding companies that do not control such data (more about this will be in the following sections).
Contrary, the so-called 'absolute approach' considers the capability of virtually everyone relevant; *therefore, the absolute approach assumes an individual identifiable not only by the respective company itself but any third party that may identify this individual, even if this requires additional knowledge exclusively assigned to suche third party.* [82]Thus, the absolute approach regards almost all data for any party as personal data if only someone can identify the person behind that data. [83]This dispute might be highly relevant in assessing the legal nature of data as both theories result in different outcome.[84]

3.3 Interpretation by the CJEU in the case C-582/14

The European Court of Justice ruled in October 2016 that any information not directly identifying a person will be considered personal in the hand of

[78] Ibid at 8
[79] Simitis S & Dammann U, 'Federal Data Protection Act (BDSG), with national data protection laws and international regulations', 2014
[80] Federation internationale de l' automobile region I- Europe, The Middle East and Africa, 'What EU legislation says about car data', Legal Memorandum on connected vehicles and data', 2017 at 8
[81] Ibid at 8 also at Weichert, in: Däubler/Klebe/Wedde/Weichert, BDSG Kompaktkommentar
[82] Ibid at 8
[83] Ibid at 8
[84] Ibid at 8

any party that lawfully obtains sufficient additional data to link the data to a person and identify that person.[85]

'In so far as that recital refers to the means likely reasonably to be used by both the controller and by 'any other person,' its wording suggests that for information to be treated as 'personal data within the meaning of Article 2(a) of that directive, it is not required that all the information enabling the identification of the data subject must be in the hands of one person.'[86]

This led to the conclusion that for data to be treated as personal, the controller can or may employ legal means reasonably available to obtain corresponding additional knowledge from a third person through which the identification of the person in question is possible for the controller**.** Howbeit, the Court here ruled in favour of the relative approach but extended the scope by referring to legal means reasonably available to obtain corresponding additional knowledge[87]**.**

3.4 What does this mean for connected vehicles[88]

Nonetheless, data coming from connected vehicles are not automatically deemed personal data for everyone. Alternatively, it needs to be assessed whether a specific company controlling the data is in a position to identify the person behind that data. For instance, agreements about remote diagnostics or proactive maintenance will naturally result in information being collected in such a context that might be defined as personal data. As the customer's contractual partner, it is likely that the service provider is aware of its identity and can link personal data to individuals.[89] We cannot deny that this is almost similar to the current circumstances with regular vehicle maintenance.

These questions should be answered even more briefly as this machine learning technique is implemented in autonomous vehicles, where we speak about a high amount of personal data and the potential risks of its exposure. The crucial question in federated learning might be whether the data are anonymized in a way that is impossible to track the original data subject? This process needs to be irreversible, and the recipient of the anonymized data should not be able to be reversed by an engineer. In this case, anonymized data is no longer considered personal data, so federated learning will fall out of the scope of GDPR protection; however, more about that will be discussed in the following sections.

[85] Ibid at 9
[86] Case C-582/14 Patrick Breyer v Bundesrepublik Deutscgland §43
[87] Supra at 9
[88] Supra at 9
[89] Supra at 9

Furthermore, while doing the legal assessment of personal data in federated learning, it is essential to consider that the EU Courts usually have a broader approach to personal data and privacy[90].
On the other hand, the rules on protecting personal data go beyond preserving the broad concept of the right to respect for private and family life. Accordingly, they are stipulated in the Directive (GDPR) where a particular reference is made to the processing of personal data in a context outside of the home and family, like that provided for by labour law (Article 8.2 (b)), criminal convictions, administrativesanctions, or judgments in civil cases (Article 8.5) or direct marketing (Article 14 (b)). The European Court of Justice validated this broad approach[91].
Finally, the aim in Europe was to have a free flow of data among member states. For this purpose, there is Regulation 2018/1807 on the framework for the free flow of non-personal data in the European Union[92].

[90] Case of *Amann v Switzerland* (No27798/95) §65", *The term 'private life' must not be interpreted restrictively. In particular, respect for private life comprises the right to establish and develop relationships with other human beings; furthermore, there is no reason of principle to justify excluding activities of a professional or business nature from the notion of 'private life.'*
[91] Opinion 4/2007 on the concept of personal data at 7
[92] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018R1807&from=EN

4 Categorizing data in autonomous vehicles and its legal analysis

As a part of this research, in order as a lawyer to be able to distinguish and categorize collected data in autonomous vehicles, which will be using the federated learning concept, while writing this paper, I participated in numerous panels and debates where experts from different backgrounds were trying to define and categorize the type of data. One of the crucial conclusions is that people with diverse backgrounds and diverse scholars access data categorization differently. It is even more challenging than accessing these concepts from the GDPR perspective. As described in the section method used for this paper, one of the methods was o interviews with persons and entities developing federated learning concepts. Following will be just some of the results from my research and how some of them categorize personal data collected, processed, and further processed on this occasion. One of the conclusions was also that having a basic knowledge of the terminology used among persons with a technical background is crucial for the lawyer in charge of privacy matters to better legal advice.

For this purpose, any IoT uses various **inputs** collected via multiple sensors or other features. Inputs are generally defined as files containing data and serve as input to a device or a program.[93]

For the starting point, the lawyer needs to distinguish whether (inputs) data is personal or not according to upper mentioned criteria. If personal, data could be categorized as 'special categories data.' This means personal data about an individual

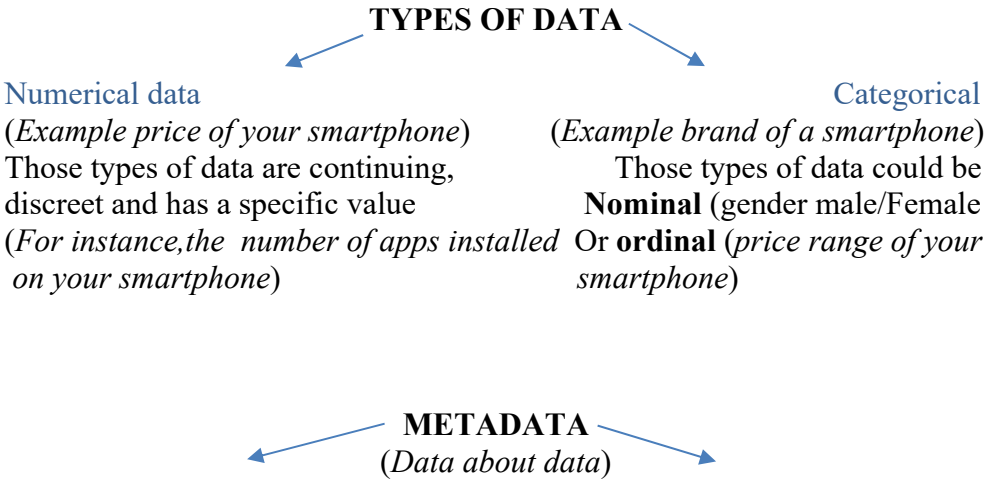
- Race
- Ethnic origin
- Political opinions
- Religious or philosophical beliefs
- Trade union memberships
- Genetic data
- Biometric data
- Health data
- Sex life or sexual orientation

Moreover, personal data can include information related to criminal convictions and offenses, which requires a higher level of protection[94]. Furthermore, according to Guidelines[95], the EDPB has identified three categories of personal data warranting special attention by vehicle and equipment manufacturers, service providers, and other data controllers:

[93] https://www.definitions.net/definition/input+data
[94] https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/what-is-personal-data/what-is-personal-data/
[95] Guidelines 01/2020 on processing personal data in the context of connected vehicles and mobility related applications (March 2021)

location data, biometric data (and any other particular category of data as defined above in Article 9 GDPR) and data which could reveal offenses or traffic violations.[96]

During one of the privacy experts panels, data were categorized as follows:



Descriptive data will be changed as the subject changes, for instance, different pictures, while structural data will give the information of all analyzed photos and their details.

But how to make a difference among them? For instance, descriptive data will cover information such as: who made the file, where it is made, when it is made, and how big the file is. In comparison, structural data are applied and analyzed in many files simultaneously.

Metadata, usually called the data about the data, will cover the author, publisher, and title. But it can also include a description of the object (weight, material). By analyzing those data types, we can learn a lot about someone's behavior and habits, such as who, how often, and how. Metadata is usually a statistical form of data. However, it can still include personal details depending on the anonymization technique used during the collection and processing (more about anonymization techniques will be discussed in the following sections).

[96] Ibid at 9

Moreover, personal data could be categorized as **commercial and transactional data and usage data**. The commercial will be data subject's identifying information, transaction-related data, data relating to means of payment, etc. Usage data will be personal data generated by vehicles, driving habits, location, etc.[97] Usage data could be classified as raw and aggregated data; however, data controllers should not process raw data. If necessary, raw data should be kept if they are required to elaborate on the aggregated data and check the aggregation process's validity. Aggregated data should be kept as long as necessary for the provision of the service or otherwise amended by the EU or Member State.[98]

Furthermore, according to the Guidelines, vehicles are equipped with image **recording devices** (e.g., car **parking camera systems or dashcams**). Since the deal with the issue of filming public places, which requires an assessment of the relevant legislative framework, is specific to each Member State, this data processing is out of the scope of these guidelines. This might be one of the crucial issues for autonomous vehicle implementation.

4.1 Guidelines' data categorization

Guidelines are referred to as Guidelines 01/2020 on processing personal data in connected vehicles mobility-related applications in the following text. In the interview conducted among the limited number of experts, one of the standard questions was What type of data do you collect in this process? The common answer was **EVERYTHING**. They claimed that this is necessary to make cars smarter and safer as machine learning needs enormous data. Moreover, some engineers claimed to process raw data as their quality is significantly higher than data from aggregated models. From the lawyer's perspective was a frightening fact due to the limited possibility of analyzing these data inputs due to many challenges. Firstly, to conduct a federated learning process, they have a few restrictions by GDPR or are limited by inadequately described requirements by the same regulation. It is important to stress that the limitation in this section is the fact that Guidelines are referring to data collected in a classical vehicle's setting, not in the federated learning process, making this qualification even more demanding, as it is clear that data in the federated learning process pass a considered level of privacy preservation techniques already on the very beginning of the machine learning process, more about the anonymization techniques will be discussed under the anonymization section.
Following is the intersection of our research and guidelines qualification. The main challenges for the engineers and, at the same time scope of this research are found to be:

[97] Guidelines 01/2020
[98] Ibid at 28

- **Firstly**, the engineers need all types of data collected via various input systems (such as vehicle sensors, cameras, thematic boxes, and mobile applications). Here all kinds of data that can make the car smarter will include
    1. **Mobility management**- functions that allow drivers to reach their destination quickly; this covers cost/ time-efficient management, traffic congestion, parking lot or garage assistance, optimization of fuel consumption, and road pricing[99].
    2. **Vehicle management**- functions that are aimed to aid drivers in reducing operating costs and improving the joy of use, such as notification of vehicle condition and service reminders, transfer of usage data (repair services, e.g.), insurance, remote operations, or profile configurations (e.g., seat position)[100]
    3. **Road safety**- functions that warn the driver of external hazards and internal responses, such as collision protection, hazard warnings, driver drowsiness detection, emergency call, or crash investigation[101]
    4. **Entertainment**- functions providing information to and involving the environment of the driver and passengers, such as smartphone interfaces, music, video, internet, social media, etc.[102]
    5. **Driver assistance**- functions covering partially or fully automated driving, such as operational assistance or autopilot in heavy traffic, in parking, or on highways[103]
    6. **Well-being**- functions monitoring the driver's comfort, ability, and driving fitness, such as fatigue detection or medical assistance.[104]

- **The second** challenge for them is to strip this data out of personal data but keep their quality. Some claimed that if the information is 'cleared' before the federated learning process, it will affect the data quality, affecting the machine learning process and its certainty. This claim is under question, as our research showed that anonymized data in the federated learning process could be without personal data and still provide good input for further federated learning. More about this process under the section 'anonymization'.
- **Thirdly, excessive data collection** could be one of the potential risks and challenges simultaneously. There is a high risk of excessive data collection compared to what is necessary to achieve the purpose. Machine earning processes require a large amount of data collected over a long period, directly diverging from GDPR principles. Secondly, one of the potential risks is an excessive amount of collected data, or more straightforward the ocean of data

[99] Ibid at 9
[100] Ibid at 9
[101] Ibid at 9
[102] Ibid at 9
[103] Ibid at 9
[104] Ibid at 9

that feeds the federated learning process, once it achieves the first federated learning model on the local level, which is aimed to be sent further into the federated learning pool will directly affect its transparency. This means it will be hard to understand why the machine made such a choice, and it will be almost impossible to do reverse engineering to find its root. This fact makes one of the significant oversights in GDPR as this regulation requires a transparent decision-making process but simultaneously requires anonymity which is in contradiction with each other. This could be solved by keeping initial inputs to provide transparency of the process in question. It is important to raise that this fact might also affect the legality of this process under Article 22 of GDPR.[105]

- **Fourthly**, a significant challenge in this data collection process is the security of personal data. This data, or federated models, must be processed locally; first-level federated modes are planned to be stored on vehicles or external locations (such as cloud infrastructures) but must be adequately secured against unauthorized access. For instance, a car must be handed to a technician who will require access to some of the vehicle's technical data during maintenance. It must be best, that a technician will be able to access only technical data but not all the data stored in the vehicles. [106]More about this matter will be discussed in the cyber security section.

4.2 Data guidance proposal and legal analysis

Guidelines[107] associate all data in connected vehicles as personal *to the extent that it is possible to link to one or more identifiable individuals*.[108] This will cover all types of data, including technical data provided by the vehicle itself. Following will be a legal analysis of upper mentioned categories of data.

4.2.1 Location data

Location data are among the most common and debated types of data collected in connected vehicles. An estimated 12-billion-dollar market of

[106] Ibid at 15
[107] Ibid
[108] Ibid

companies that buy and sell location data collected from our smartphones[109]. Moreover, location data are the critical element to support autopilot and make the performance and joy of autonomous vehicles more certain. But those data will indirectly or directly reveal crucial personal data about one person. For instance, place of work and residence, driver's most visited places (their interests), the site of worship (will reveal the person's religion), or even sexual orientation through the areas visited[110]. If a person has a child, where the child attends school, how often he goes to the doctor etc. More recommendations about children's data, other types of data, and how the illegal usage could be fined will be assessed more briefly under the section **privacy by design and default**. Consequently[111], the vehicle manufacturer, service provider, and other data controllers should be particularly vigilant not to collect location data except if doing so is necessary for processing as recommended by Guidelines. For instance, when processing detects a vehicle's movement, the gyroscope is sufficient to fulfill the function without collecting location data. [112]Here Guidelines also mention how collecting location data could be a subject of compliance with the GDPR principles[113].

Legal basis for processing location data[114] necessary for the full performance of any features for the performance of autonomous vehicles (parking assistance, renting a parking space, navigation, etc.)- Guidelines recommends if data is collected through a publicly available electronic communication, Article 5(3) of the ePrivacy directive applies. Those information does not require consent to access data stored in the vehicle when the subscriber explicitly requests such a service. Moreover, for the processing of personal data and only for data necessary for the performance of the contract to which the data subject is a party, Article 6 (1) (b) GDPR could be the legal basis[115]. However, as navigation and location

---

[109] https://themarkup.org/ask-the-markup/2022/02/24/who-is-policing-the-location-data-industry

[110] Ibid at 16

[111] https://themarkup.org/ask-the-markup/2022/02/24/who-is-policing-the-location-data-industry

[112] Ibid

[113] -adequate configuration of the frequency of access to, and of the level of detail of, location data collected relative to the purpose of processing. For example, a weather application should not be able to access the vehicle's location every second, even with the consent of the data subject;
– providing accurate information on the purpose of processing (e.g., is location history stored? If so, what is its purpose?);
– when the processing is based on consent, obtaining valid (free, specific and informed) consent that is distinct from the general conditions of sale or use, for example on the on-board computer;
– activating location only when the user launches a functionality that requires the vehicle's location to be known, and not by default and continuously when the car is started;
– informing the user that location has been activated, in particular by using icons (e.g., an arrow that moves across the screen);
– the option to deactivate location at any time; – defining a limited storage period

[114] When can personal data be processed at https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/grounds-processing/when-can-personal-data-be-processed_en

[115] Guidelines 01/2020 at 29

inputs are crucial for driver's safety, possibly legal basis as organization's legitimate interest could prevail, as aims to provide driver's security and override the person's rights, bearing in mind the principle of proportionality. Those fundamental rights and freedoms should not be seriously impacted in this case, considering the individual circumstances.
All of this could be disregarded if federated learning shows that even recorded data are encrypted in the federated learning model to the extent that it cannot link to the individual. This recommendation covers all the types of data below as well.

4.2.2 Biometric data

The GDPR defines biometric data broadly, in many cases requires privacy impact assessments for its processing, and, at the same time, empowers the Member States to seek divergent protections for biometric data. This means that biometric data are a subset of sensitive personal data, defined as the 'sensitive category of personal data.'[116] However, GDPR implicitly acknowledges that biometric technology is inceptive and will continue to evolve. [117]It seems that the usage of biometric data may arise through the development of new technologies.
Biometric data could be categorized as a person's physical or physiological traits. This will include facial information, fingerprints, iris scans, etc. The second group covers behavioral information that could be used for identification. Additionally, information about someone's habits, actions, or personality could be considered behavioral information within the scope of the definition. Biometric data have been discussed under Case C-434/16 (*Peter Nowak v Data Protection Commissioner*), where the court took a broad interpretation of the concept of personal data by extending the scope of biometric data, to the written form of an exam, as a warning to organizations of the velocity of the scope of data protection law[118].
Legal basis for processing: GDPR categorizes those types of data as 'sensitive' data and must proceed under the framework aimed at sensitive personal data generally. The GDPR typically prohibits the processing of biometric data to identify natural persons uniquely. The exemptions provided by GDPR are limited and very restrictive. [119]In the context of connected vehicles, biometric data are mainly used to uniquely identify a natural person within the remit of Article 9. Of GDPR and the national exceptions[120] to access the car, authenticate the owner, and enable access to

[116] https://iapp.org/news/a/processing-biometric-data-be-careful-under-the-gdpr/
[117] Ibid
[118] Brennan D. 'The expanding scope of 'personal data', CJEU delivers judgment Nowak', from January 2018
[119] https://www.dentons.com/en/insights/alerts/2020/december/22/gdpr-update-biometric-data
[120] For instance, in the Netherlands, article 29 of the UAVG provides such an exemption and states that the processing of biometric data for the purposes of uniquely identifying an individual is allowed if the processing is necessary for authentication or security purposes.

a driver's profile settings. Usually, data protection authorities have only two exceptional circumstances:

- If the data subject has given their explicit consent (an alternative must be provided)
- If it is necessary and proportionate for security or authorization purposes to serve a compelling public interest[121]

When biometric data is used, it is necessary to ensure that the subject has complete control over his data; on the other hand, it provides a non-biometric alternative (physical key or a code). While using biometric data, certain principles must be followed.[122]

4.2.3 Data revealing criminal offenses or other infractions

Criminal offense data is usually the data about a specific criminal conviction or trial and any other personal data 'relating to criminal convictions and offenses. That might include suspicion or allegations of illegal activity. Thus, according to UK Information Commissioner's office, one will enjoy extra protection in the UK, in line with Article 10 of GDPR[123]. [124]Moreover, in the *Case C-136/17 GC, AF, BH, ED v Commission Nationale de Informatique er des libertes (CNIL*), the court announced that conviction data relating to 'offenses and 'criminal convictions are considered special categories of personal data within the meaning of Article 8 (5) of Directive 95/46 and Article 10 GDPR.
Guidelines citing that EDPB recommends resorting to the local processing of the data where the data subject has complete control over the processing in question; however, external data processing revealing criminal offenses or other infractions is forbidden. [125]Therefore, advanced security measures

---

The necessity condition requires an assessment of whether the controller's interest in using biometric data is proportionate to the impact on an individual's privacy, and whether less privacy-intrusive measures are available to achieve the authentication or security purposes at Ibid

[121] Dutch Data Protection Authority

[122] – the adjustment of the biometric solution used (e.g., the rate of false positives and false negatives) is adapted to the security level of the required access control;
– the biometric solution used is based on a sensor that is resistant to attacks (such as the use of a flat-printed print for fingerprint recognition);
– the number of authentication attempts is limited;
– the biometric template/model is stored in the vehicle, in an encrypted form using a cryptographic algorithm and key management that comply with the state of the art;
– the raw data used to make up the biometric template and for user authentication are processed in real time without ever being stored, even locally - Guidelines at 17

[123] Processing of personal data relating to criminal convictions and offences or related security measures based on **Article 6**(1) shall be carried out only under the control of official authority or when the processing is authorised by Union or Member State law providing for appropriate safeguards for the rights and freedoms of data subjects. Any comprehensive register of criminal convictions shall be kept only under the control of official authority.

[124] https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/criminal-offence-data/what-is-criminal-offence-data/

[125] Guidelines at 17

need to be conducted against the illegitimate access, modification, and deletion of those data due to its sensitivity factor. On the other hand, some categories of personal data from connected vehicles could expose a criminal offense or other infraction; however, only competent national authority is entitled to pursue a criminal investigation, the safeguards provided in Article 10. GDPR.[126]
As mentioned above, EDPB recommended local processing, entirely in line with the federated learning process, as their data will be processed locally[127]. While ensuring high cyber security measures, no other specific obstacles to implementation have been found for this type of data usage in a federated learning setting, which will be discussed under the section Cyber Security.

4.3 Autonomous vehicles as the Internet of things

Before even going into a deeper legal analysis, it is essential to define autonomous vehicles based on federated learning (machine learning or AI intelligence). The way federated learning works or connected vehicles works sets them into a group of connected Internet of Things. Internet of things environment elaborates on three technologies in particular: (1) Radio Frequency Identification (RFID), which is typically used to identify objects and monitor their paths, (2) intelligent energy architectures, which measure and communicate energy data; and (3) intelligent wearable devices that are used to track health and fitness data of users. [128]As those devices trace many things in our surroundings, the IoTs create an environment of anticipation of every action and interaction between individuals and objects.

This concept describes the connection of different devices through the internet[129]. According to some scholars, users feel they are losing control of their privacy as the data is transferred between devices and remote data centers[130]. This is how the system in the connected autonomous vehicles setting is functioning indeed.

In the previous sections, the concept of personal data has been explained. It has been proven that the idea of privacy and personal data is, in fact, a long-established European institute. Not only that personal data protected in EU legislation, but the 'threat' of being identifiable will also set this concept

[126] Guidelines at 17
[127] Ibid at 17
[128] Aurelia Tamo Larrieux- 'Designing for Privacy and its Legal Framework- Data protection by desing and default for the internet of things' (2018) at 78
[129] Cosar Ahmet & Turk Ismail- 'Internet Connection Sharing Through NFC for Connection Loss Problem in Internet of Things Devices', (August 2015)
[130] Fahsi.M, Benslimane S., Rahmani A., 'A Framework for Homomorphic, Private Information Retrieval Protocols in the Cloud, (May 2015)

under the GDPR realm. How high is a threat in federated learning will be discussed in the following sections?
When looking closer at what type of data autonomous vehicles will be collecting, this comprehensive EU approach might question some of the core functions these IoTs will possess.
The applicable law is GDPR and will be applied in any case where data processing in the context of connected vehicles involves processing the personal data of individuals.[131]
Moreover, directive 2002/58/EC, as revised by 2009/135/EC (hereinafter 'ePrivacy directive'), sets concrete standards for all actors that wish to store or access information stored in the terminal equipment of a subscriber or user in the EU area.[132] Even though there is an inevitable confusion about the application of this directive, this directive is a general provision, and 'it does not cover only electronic communication services but also every entity, private or public, that places on or reads information from a terminal equipment without regard to the nature of data being stored or accessed.'
When it comes to the correlation between 2 the ePrivacy directive and GDPR[133], the ePrivacy directive provides that, as a rule, and subject to the exceptions to that rule mentioned in paragraph 17, prior consent is required for the storing of information or the gaining of access to information already stored, in the terminal equipment or a subscriber or user[134].

4.4 Autonomous vehicles based on machine learning and AI technology

As connected vehicles are based on machine learning algorithms, shortly Artificial Intelligence. European legislation published the white paper about Artificial intelligence overall[135], not mentioning connected vehicles particularly, except in part about safety, which this paper connects to the overall product liability Directive[136]. On the other hand, guidelines for connected cars do not define the AI system of the connected vehicles but instead mention machine learning in the context of a threat to extensive data collection[137].

[131] If collected data could be defined as personal data
[132] Guidelines 01/2020 on processing personal data in the context of connected vehicles and mobility related applications (March 2021) at 6
[133] European Data Protection Board- Opinion 5/2019 on the interplay between ePrivacy Directive and GDPR §40
[134] Guidelines 01/2020 on processing personal data in the context of connected vehicles and mobility related applications (March 2021) at 6
[135] White paper on Artificial Intelligence- A European approach to excellence and trust (19.02.2020)
[136] General Product Safety Directive (2001/95/EC)
[137] Guidelines 01/2020 on processing personal data in the context of connected vehicles and mobility related applications at 15

However, European legislation defines AI as a *collection of technologies that combine data, algorithms, and computing power.*[138] As Europe is heading to become a global leader in innovation in the data economy and its applications, this could be achieved by combining its technology and industrial strengths with high-quality digital infrastructure and regulatory framework based on its fundamental values[139].

4.5 Data Controller vs. Data Processor

As in the autonomous vehicles setting, or overall internet of things setting, there is usually a complex chain of different actors in the data processing. Other acts typically have quite different roles in delivering a quality service from the IoTs. Therefore, it is crucial to determine whether an organization has a data controller or a data processor role in the whole process.

The data controller determines the purpose for which and how personal data is processed. It can be done whether jointly or in common with other organizations.
This means that the data controller exercises overall control over the 'why' and the 'how' of a data processing activity[140]. The definition provides a certain level of flexibility; for instance, it can allow one data controller to mainly, but not exclusively, control the purpose of the processing with another data controller[141].

Firstly, the processing of personal data covers any operation that involves personal data as collection, recording, organization, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure, or destruction, etc.[142]

Secondly, the data subject is the natural person to whom the data covered by the processing relate. In the connected vehicles setting, it can, for instance, be the driver itself (both the main driver and the occasional one), the passenger, or the owner of the vehicle[143].

Thirdly, to determine whether the data processor is a data controller, it is crucial to ascertain which organization decides:

[138] White paper on Artificial Intelligence- A European approach to excellence and trust (19.02.2020) at 2
[139] Ibid
[140] Denley A, Foulsham M & Hitchen B,- 'How to Achieve and Maintain Compliance GDPR' (2019) at 22
[141] Ibid at 22
[142] GDPR, Article 4 (2)
[143] GDPR, Article 4 (1)

- To collect the personal data in the first place and the legal basis for doing so,
- Which items of personal data to collect, i.e,. the content of the data,
- The purpose data is aimed for
- Which individuals to collect data about,
- Whether to disclose the data and, if so, who to,
- Whether subject access and other individuals' rights might apply, i.e,. the application of exemptions,
- And for how long to retain the data or make non-routine amendments to the data[144]

The data controller can only make the decisions mentioned above in the data processing operation[145].

On the contrary, the GDPR states:

'Data processor means a natural or a legal person, public authority, agency, or other body which processes personal data on behalf of the controller.'

As stated by the Information Commissioner[146] :

"Data processor" about personal data means any person (other than an employee of the data controller) who processes the data on behalf of the data controller. [147]

"Processing" concerning information or data means obtaining, recording, or holding the information or data or carrying out any operation or set of functions on the information or data, including[148]-

a) *Organization, adaptation, or alteration of the information or data,*
b) *Retrieval, consultation, or use of the information data,*
c) *Disclosure of the information or data by transmission, dissemination, or otherwise making available, or*
d) *Alignment, combination, blocking, erasure, or destruction of the information or data[149].*

Consequently, it can be concluded that the definition of "processing" suggests that a data processor's activities must be limited to the more "technical" aspects of an operation, such as data storage, retrieval, or erasure. Furthermore, a data controller must carry out activities such as interpretation, professional judgment, or significant decision-making about personal data. Sometimes, the whole process can have actors; some actors

[144] Denley A, Foulsham M & Hitchen B,- 'How to Achieve and Maintain Compliance GDPR' (2019) at 22
[145] Ibid
[146] 'Data controllers and data processors 20140506'
[147] Denley A, Foulsham M & Hitchen B,- 'How to Achieve and Maintain Compliance GDPR' (2019) at 22
[148] Ibid
[149] Ibid

might even have sub-processors. Naturally, depend, the sub-processor could have a 'main' data controller depending on its role; it does not automatically mean that if an organization contracts or employs another organization to provide a service to it, it does not mean that a data controller or data processor will depend on their role and responsibilities in the process. [150]

So, looking at overall legal and another type of relationships among the actors in the whole process of autonomation of connected vehicles could be very challenging to distinguish relationships among them and access the data procession in the right direction. In this setting (autonomous cars), there can be many different actors who will have various functions in the whole process. Sometimes, it can be a lawyer providing legal advice, an accountant, a recruitment agency, an insurance agency, etc. In those cases, primarily, the client will not have sole data controller responsibility even though they initiated the work by asking for advice or commissioning a report. Lawyers, for instance, would have their professional responsibilities in terms of record-keeping, confidentiality, communications, and so forth. Guidelines on personal processing data[151] in connected vehicles also give recommendations[152].

[150] Ibid
[151] Guidelines 01/2020 on processing personal data in the context of connected vehicles and mobility-related applications.
[152] - Guidelines 01/2020 on processing personal data in the context of connected vehicles and mobility related applications at 11

5 Concept privacy by design and default

The General Data Protection Regulation addresses data protection by design as a legal obligation for data controllers and processors, making it an explicit reference to data minimization and the possible use of pseudonymization. Similarly, this provision introduces the burden of data privacy by default, which goes even further into stipulating personal data protection as the default option in the systems and services[153].
Regarding personal data protection and privacy in the concept of the Internet of things, in this case, autonomous vehicles (which we defined in the previous chapters), it is necessary to discuss the idea of privacy by design and default. The current trends toward increasing connectivity and the more remarkable ability to process many different data types are here to stay and become inevitably part of our everyday lives.
New connected devices can communicate across large spaces or register individuals' whereabouts, enabling big data and network analysis to an even greater extent[154]. Unfortunately, some predictions stress that the rise of intelligent things will increase security threats.[155]
The Privacy by Design (PbD) concept was developed by the Information and Privacy Commissioner of Ontario, Canada, Dr. Ann Cavoukian, back in the '90s.[156] Privacy by Design should provide proactive rather than reactive measures. It should anticipate and prevents privacy-invasive events before they happen[157].
Both privacy by design and default have deep roots in privacy technology[158]. This includes the very early work of David Chaum on anonymous communications, unlikable pseudonyms, and secure but untraceable payments. [159]
Somewhat, those concepts represent a clear intersection between two disciplines: law and technology. Instead, it could be defined as a concept now part of the EU data protection framework. Thus, it combines legal principle-based rationality to find technical mechanisms and organizational procedures to protect privacy[160]. For instance, *some scholars claim that*

[153] https://www.enisa.europa.eu/topics/data-protection/privacy-by-design
[154] Ibid at 146
[155] Ibid
[156] Privacy by Design- Information & Privacy Commissioner of Ontario
[157] Ibid
[158] Good N & Rubinstein I, 'The Trouble with Article 25 (How to Fix it): The Future of Data Protection by Design and Default', at 2
[159] See David Chaum, 'Security without Identification: Transaction Systems to Make Big Brother Obsolete' (1985) 28 Comm. of the ACM 1030. In the mid-1990s, Cavoukian and her colleagues in the Netherlands directly linked their work to Chaum's. See Ronald Hes and John J. Borking, *Privacy-Enhancing Technologies: The Path to Anonymity*
[160] Supra at 22

legal principles are reactive to past harms, while technological tools are proactive measures enacted to prevent infringements[161].
Privacy by design requires privacy to be a consideration from the very beginning. Privacy must be designed into the development phase and continue the data life cycle.
Moreover, Cavoukian, as a pioneer, advocated for the implementation of privacy by design and defined the seven principles that privacy by design entails. Therefore, PbD is:

1. Protects privacy in a proactive, not reactive way (*preventive, not remedial*)
2. Privacy must be the default setting (*it is built into the system*)
3. Must be embedded into the design (*privacy is integral to the system without diminishing functionality*)
4. It aims to achieve full functionality (*it is possible to have both- privacy and security*)
5. It aims to guarantee end-to-end security throughout the whole life cycle of personal data (*end-to-end secure lifecycle management*)
6. Visibility and transparency requirements on how personal data is processed and implemented (*its part and operations remain visible and transparent- Trust but verify principle*)
7. Overall respects user privacy (*keep it user-centric by providing firm privacy defaults, appropriate notice, and empowering user-friendly options*)[162]

5.1 GDPR and Privacy by design

EU legislators transferred these principles into Article 25. GDPR[163]. This Article requires data controllers to implement appropriate technical and organizational measures to ensure that, by default, only personal data necessary for each specific purpose of the processing is processed. [164]This obligation applies *to the amount of personal data collected, the extent of their processing, storage period, and accessibility*. [165]These measures shall ensure that personal data are not made accessible without an individual's intervention to an indefinite number of people.
The principle of privacy by default is primarily of importance for services and products where the data subject has the choice of sharing its personal data. [166]This stand is also taken in the official German '*Strategy for*

[161] Tamo-Larrieux Aurelia- '*Designing for privacy and its legal framework. Data protection by design and default for the internet of things'* (2018) at 22
[162] Ibid at 23
[163] https://gdpr.eu/article-25-data-protection-by-design/
[164] https://www.dentons.com/en/insights/alerts/2017/april/18/monthly-newsletter-gdpr-accountability-privacy-by-design-and-privacy-by-default
[165] Ibid
[166] Ibid

Automated and Connected Driving at variance with Spiekermann[167] and Windfield[168], who surveyed 124 engineers to understand ethical systems development issues concerning privacy and security engineering. The findings indicated that many engineers consider privacy important, but they do not enjoy working on them and struggle with their organizational environment[169].

5.2 Principles concerning the Design of Data Processing systems

Behind Article 25, the idea is that privacy is incorporated in the 'state of art' privacy technology and uses its enforcement powers to reward good examples of privacy engineering rather than *penalize failures*[170]. In other words, provisions target data controllers rather than engineers, developers, and manufacturers. The goal is to force them to pressure engineers and developers to develop an adequate solution. [171]The focus is on design mechanisms with the protection of certain principles, such as:

- ✓ **Data Minimization and Proportionality**

Article 5 GDPR limits the amount of personal data collected to the minimum necessary to achieve the purpose for which the data was gathered and processed. However, when it comes to this principle, sometimes, terms such as 'necessity' or 'proportionality' are used synonymously with the term 'minimality.'[172] Article 25 of GDPR implies the importance of establishing the data minimization principle in organizational measures.

- ✓ **Use, Disclosure, and Storage Limitations**

This principle is linked to the concept of purpose limitation. It prohibits the use or disclosure of personal data for purposes other than what the data controller had initially specified. The data subject had indicated their

[167] S. Spiekermann, J. Korunovska, and M. Langheinrich, "Inside the organization: Why privacy and security engineering is a challenge for engineers," Mar. 2019
[168] Winfield A & others, 'Machine Ethics: The Design and Governance of Ethical AI and Autonomous Systems', 2019 at 515
[169] Ibid at 515
[170] Good N & Rubinstein I, 'The Trouble with Article 25 (and How to Fix it): The Future of Data Protection by Design and Default' at 1
[171] Tamo-Larrieux Aurelia- '*Designing for privacy and its legal framework. Data protection by design and default for the internet of things'* (2018) at 86
[172] Ibid at 91, more see Bygrave, pp. 59-60; cf. also Bygrave, Data Privacy, pp. 151-152; The German Federal Data Protection Act employs the term "Datensparsamkeit" meaning data frugality. Cf. Simitis/Scholz, § 3a, marginal No. 31 in particular. Note that the criterion of "necessity" relates to the criteria of "proportionality", cf. Art. 7, 8, 13 Directive 95/46/EC

consent. [173]Furthermore, besides minimization and storage limitations are expressions of the principle of proportionality.

- ✓ **Data Security**

Data Protection law in Europe mandates the implementation of appropriate data security measures. Those measures should ensure data from accidental, unauthorized, or unintended use, modification, disclosure, dissemination, or destruction.[174] The GDPR requires data controllers to notify data protection authorities and data subjects of security breaches 'without undue delay.'[175]

- ✓ **Anonymity and Pseudonymity**

Are discussed in the previous sections

- ✓ **Data Quality and Accuracy**

Ensuring data quality and accuracy is an essential task for which data controllers are responsible. These steps should include regular security checks of data. Data quality, accuracy, and validity are also manifested in previous establishing data subjects' rights of rectification[176].

5.3 The trouble with Article 25 GDPR- a different perspective

Some scholars claim Article 25 is '*presently conceived and poorly aligned with privacy engineering methods and related privacy-enhancing technologies (PETs)*'[177]. Albeit Article 25 does a poor job in describing system designers and developers what requires or ensures the adoption of privacy engineering methodologies and rigorous PETs. Technology neutrality arguably favors broad statutory language to avoid discriminating

[173] Supra at 91 and 92, While this principle is neither directly expressed in Convention 108 nor in Directive 95/46/EC nor the GDPR, it is indirectly included in Articles 5(a), 5(b), and 6 of Convention 108, and in Article 6(1)(a) and (b) as well as Articles 7 and 8 of Directive 95/46/EC, and will probably be also read into Articles 5(1), 6, and 9 of the GDPR.130 Additionally, in Article 4(12), the GDPR defines the term "personal data breach" as a "a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to, personal data transmitted, stored or otherwise processed." Moreover, in Article 5(1)(e), the GDPR elaborates on the related principle of storage limitation.
[174] Zallone R, 'Connected Vehicles under the GDPR', 2019 at 4
[175] Ibid at 92
[176] Ibid at 93
[177] Good N & Rubinstein I, 'The Trouble with Article 25 (and How to Fix it): The Future of Data Protection by Design and Default', at 1

against certain technologies and discouraging innovation[178]. As a few academics noted, technology neutrality is a starting point for regulating technology, but more than one legislative approach may be required to achieve its objectives[179].
For instance, legislators might be unsuccessful in heeding the principle of technology neutrality with not-so-good results. In Europe, this could be regarded as the 'Cookie Directive' (2002/58/EC), which requires a consent requirement but at the same time ignores several alternative technologies that track online behavior.[180]
They claim that Article 25 has its broad view of 'technical and organizational 'measures' and 'design' and 'default' and yet significantly lacks clarity about the fitting technology for achieving required goals. Moreover, there should be a balance between technology neutrality and provision drafting. [181]*They also claim* that there is too much overlap between this article and other GDPR provisions, which causes confusion and uncertainty. *Secondly*, they claim that Article 25 is not having a clear scope. *Thirdly,* this article offers only one measure- pseudonymization. *Fourthly*, Article 25.1 states that controllers implement 'appropriate' technical and organizational measures and do so 'effectively. It is pretty unclear how. *Fifthly*, it fails to establish a clear baseline for what it means to design privacy controls.[182]

5.4 A good/ bad example of privacy by design

An excellent example of successful/ unsuccessful privacy was implemented by German retailer Real at the end of 2016. The Real initiated a digital out-of-home advertising program using AdPack, a system that relies on facial detection technology. AdPack works on software called SHORE and AVARD (Anonymous Video Analytics for Retail and Digital Signage).[183]AVARD is an e-privacy-certified technology that enables the analysis of faces about gender, estimated age, and emotional state. Similar features could be found in Microsoft's Face[184].
This technology implemented a privacy-by-design approach. The original pictures are never stored permanently or processed in the cloud; only the

[178] Ibid at 5
[179] Ibid at 5, more at Bert Jaap Koops, 'Should ICT Regulation be Technology-Neutral', in B.J. Koops and others (eds), Starting Points for ICT Regulation: deconstructing prevalent policy one-liners (Springer 2006), See also Hildebrandt and Tielemans (n 22); Christopher Reed, 'Taking Sides on Technology Neutrality' (2007)
[180] Ibid at 6
[181] Ibid at 6
[182] Ibid at 8
[183] Fraunhofer Institute for Integrated Circuits, 'AVARD – Anonymous Video Analytics for Retail and Digital Signage'
[184] George D, Reutimann K & Larrieux A, 'GDPR bypass by design? Transient processing of data under the GDPR' at 286

transient copy is kept in the memory of the camera system. The camera software detects the characteristics (metadata) and generates a value of the metadata. These values are used to choose an advertisement and are deleted after 150 milliseconds[185]. According to AVARD developers, it is impossible to re-identify a person[186]. Bavarian DPA[187] concluded that it does not collect any personal data; the Data Protection Commissioner of Ireland supported this stand. The Irish DPC made a difference between facial recognition and detection technology. [188]However, Real was forced to end AdPack's use due to a lack of customer acceptance and a criminal complaint filed by NGO Digital courage.[189] This might be clear evidence *that even 'an ideally constructed' certified privacy by design solution might not be a good solution in practice[190], as there are many papers regarding the lack of trust in the AI-based Internet of Things in Europe[191], which leads to the conclusion that the high level of awareness in Europe about their fundamental rights and privacy might prevail when it comes to the acceptance of new technological developments.*

5.5 Proposed privacy by design guidelines for autonomous vehicles based on federated learning

In light of all discussed above, how might controllers, processors, and software developers in autonomous vehicles based on FL translate GDPR core principles into concrete design requirements and methodologies? If the FL technique is considered a pseudonymization technique (Article 25 only refers to the pseudonymized data, but more about the anonymization assessment and FL will be discussed in the following chapters). This might be a task for privacy engineering and DPO.

- ✓ **Privacy engineering**

[185] Ibid at 286
[186] See also Fraunhofer Institute for Integrated Circuits, 'Eine Frage des Datenschutzes – die Bildanalysesoftware SHORE, Mit dem deutschen Datenschutz konform – Interview mit Jens Garbas'
[187] See Heike Anger, 'Gesichtsscan im Supermarkt ist unbedenklich' Handelsblatt (Dusseldorf, 12 June 2017)
[188] Ibid at 287- '*While facial detection technologies—such as those employed by AdPack—merely involve the detection of a human face and classification of its characteristics (e.g. gender, age range, and emotional status), facial recognition technology involves the storing of personal data in order to match it to a unique individual face (e.g. iris recognition*)'
[189] Ibid at 287, see also Real ends facial recognition in supermarkets | Car and technology | GQ (gq-magazin.de)
[190] https://www.statista.com/statistics/422787/europe-trust-in-the-internet-by-country/
[191] Fellander Anna, Teigland Robin & Holmberg Håkan, 'The importance of trust in digital Europe: Reflections on the sharing economy and blockchains'. 2018

Robinstein and Good, in the paper,[192] are supporting a proposal by a group of academics[193] regarding the first principle to have in mind when it comes to suitable privacy engineering. For the 'fixing' element for Article 25, they propose the principle of **data minimization.** They also support this approach by providing a few studies showing how a design process guided by the principle of data minimization can reduce privacy risks, avoid function creep, and limit the disclosure of sensitive information[194]. This stand might not be suitable in a federated learning setting, as a new ML technique requires more data (big data) to increase the process certainty. There is an apparent tension between Big Data and critical GDPR principles such as purpose limitation, data minimization, sensitive data, and automated decisions. [195]These questions the suitability of this provision for future technical developments.
In that event, all upper mentioned papers supported by a few studies claimed that Article 25 is not providing clear guidance for e engineers needed to decide between procedural solutions or privacy-preserving solutions. This *affects the constraints of the use of personal information, as companies will usually gravitate towards familiar procedural solutions that are less likely to disrupt existing business processes and revenue sources*[196].
Suppose the FL is a proven anonymization technique (see anonymization chapter). In that case, it does not need to incorporate privacy by design and default, as this provision only relates to pseudonymized data. However, specific risk assessments and privacy by design guidelines have been provided in Appendix 1, for the potentially '*unconvinced audience of readers.'*
Considering all presented above, it has been found that cyber security is more relevant in this regard. For this reason, in the next chapter, cyber security will be discussed briefly.

5.6 Data Protection Impact Assessment (DPIA)

A Data Protection Impact Assessment (DPIA) is a part of privacy by design and default concept; it is required under the GDPR any time a new system is implemented, likely to involve a 'high risk' to other people's personal information. The organizations must follow GDPR rules. If they fail to

[192] Good N & Rubinstein I, 'The Trouble with Article 25 (and How to Fix it): The Future of Data Protection by Design and Default' at 10
[193] Seda Gürses and Joris V. J. van Hoboken, 'Privacy After the Agile Turn' in Evan Selinger, Jules Polonetsky and Omer Tene (eds.), The Cambridge Handbook of Consumer Privacy (2018), also Gurses Seda, Kostova Blagovesta & Tronsoco Carmela, 'Privacy engineering meets software engineering' 2020. On the challenge of engineering privacy by design', at 7
[194] Good N & Rubinstein I, 'The Trouble with Article 25 (How to Fix it): The Future of Data Protection by Design and Default' at 11
[195] Ibid at 21
[196] Ibid at 20

comply, they are exposing themselves to severe penalties, including fines of up to 20 million dollars or 4% of annual revenue, whichever is higher[197]. The organization's way to demonstrate to the authorities that they comply with GDPR is to prepare DPIA for each of the high-risk data processing activities.

Article 35 of the GDPR covers Data Protection Impact Assessments. The DPIA is a new requirement under the GDPR as part of the 'protection by design principle. According to the law[198]:

Where a type of processing, in particular, using new technologies, and taking into account the nature, scope, context, and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, assess the impact of the envisaged processing operations on the protection of personal data.

To clarify, the following are examples of the type of conditions which would require a DPIA:

- ✓ If new technology is used
- ✓ If there has been location or behaviour tracking
- ✓ If data such as racial or ethnic origin, political opinions, religion, trade union membership, genetic data, biometric data to uniquely identify a natural person, data concerning health, a person's sex life, or sexual orientation
- ✓ If processed data is used to make automated decisions about people, that could have legal effects
- ✓ If children's data are processed
- ✓ If processed data could result in physical harm to the data subject if it is leaked.[199]

As most data breaches can cause specific regulatory requirements, it might still be prudent to conduct a DPIA to minimize liability and ensure best practices for data security. The official GDPR web page offers specific recommendations and templates regarding this matter.[200]

However, as proved above, Federated Learning could not be of high risk as soon as privacy by design and default is ensured. Therefore, we do not see a need for a DPIA assessment. Nevertheless, according to upper mentioned criteria to ensure GDPR compliance, the Data Protection Officer in charge could perform a thorough risk assessment or re-identification report. More about privacy by design will be discussed in the following section.

[197] https://gdpr.eu/data-protection-impact-assessment-template/
[198] Ibid
[199] Ibid
[200] https://gdpr.eu/wp-content/uploads/2019/03/dpia-template-v1.pdf

6 Privacy-preserving techniques in ML

FL is a machine learning technique; however, it differs from the conventional machine learning process. Considering that federated learning is not conventional machine learning, other privacy preservation techniques exist. These techniques could be incorporated into federated learning to enshrine the data set. For that purpose, it is necessary to distinguish those methods for a distributed learning system a) privacy of the training dataset and b) privacy of the local model parameters from an optimization algorithm such as a gradient descent variant which is exchanged with other nodes or a centralized server[201].

- Example for data preserving techniques:
- Data anonymization
- Differential Privacy
- Secure Multi-party computation
- Homomorphic Encryption

* **Data anonymization** or de-identification is a technique to hide or remove sensitive attributes, such as personally identifiable information, so that the data subject (irreversibly) cannot be identified within the modified dataset. [202]Consequently, data anonymization must balance the privacy guarantee and the dataset's utility. Unfortunately, it has been claimed that this privacy preservation technique cannot defend against so-called 'linkage attacks' whose adversaries possess some knowledge about the sensitive attributes[203].
* **Differential Privacy** is an advanced privacy-preserving technique that adds random noise to the actual outputs using rigorous mathematical measures. It is statistically indistinguishable between an original aggregate dataset and a differentially additive noise. That means that a single individual cannot be identified as the original dataset is practically the same regardless of the individual's existence. However, it is important to stress that there must be a balance between privacy guarantee and utility, as adding too much noise and improper randomness will significantly depreciate the reliability and usability of the dataset[204].
***Secure Multi-party Computation** or multi-party computation is a catalyst of functions that can collectively compute over a dataset owned by multiple parties using their inputs. SMC can be useful for data privacy preservation in distributed learning wherein compute nodes collaboratively perform model training on their local dataset without revealing such datasets to others.
* **Homomorphic Encryption** is a data privacy and security preservation technique, particularly in centralized systems, for instance, cloud servers,

[201] Truong N, Sun K, Wang S, Guitton F and Guo Y, 'Privacy Preservation in Federated Learning: An insightful survey from the GDPR Perspective' at 4
[202] Ibid At 4
[203] Ibid
[204] Ibid

wherein data is collected and trained at a server without disclosing the original information. The computation results are in encrypted form and can only be decrypted by the requester of the computation. Moreover, homomorphic encryption ensures that the decoded output is the same as the one computed on the original unencrypted dataset. As a result, employing homomorphic encryption in large-scale data training remains impractical[205].
It is necessary to raise those interviewed engineers who claimed that a combination of homomorphic encryption and federated learning is not possible to brake, representing the safest combination. Furthermore, some of them claimed that additional ML privacy preservation techniques would decrease the dataset's quality; others argued that they will not. However, they all agreed that homomorphic encryption is a perfect solution, but it consumes a significant amount of time and money. Therefore, the upper mentioned hypothesis was partially confirmed in our research.
This was shortly what could be done with a raw data set before further analytics. The following section will be more legal analysis of the given methods.

[205] Ibid

7 Federated learning as an anonymization technique in the autonomous vehicles

Anonymisation is one of the most debated criteria related to data, whether we talk about data protection, data transfer, or other GDPR criteria. However, even though it is debated, there are still many dilemmas regarding this term, both legal and tech.

Rooted from the Greek word "anonymia," the term anonymity/ anonymous stands for "namelessness," "not identified," or "unknown name" (Oxford Dictionaries) and usually bears on a person's appearance in public. Anonymity does not necessarily presuppose the complete anonymousness of a person's identity or the lack of a name; even the uncrowdedness of an individual's name could suffice. However, to distinguish anonymity from undetectability, it is necessary that one person vaguely knows about another person's existence without knowing their complete identity[206]. Some academics like to distinguish privacy from anonymity. They claim that *privacy allows for keeping certain information and data confidential and is supported by anonymity, describing a condition of being unknown or unacknowledged to others*.[207] Still, privacy is a fundamental but not an absolute right.[208]

According to GDPR regulation[209], fully 'anonymized' data does not meet the criteria necessary to qualify as personal data and is therefore not subject to the same restrictions placed on processing personal data under the GDPR. Data could be considered anonymized when individuals are no longer indefinable. This does not include only the names of the individuals but also any other information related to them, or more precisely, information that could be used to single them out. It is essential to say that this process needs to be irreversible. If this is not the case, data will be considered rather pseudonymized than anonymized and still defined as personal data[210].

However, data protection laws do not describe any technique for 'anonymization,' so it is up to data controllers to ensure that whatever anonymization technique they use is sufficiently robust[211]. More precisely, data must be used so that it can no longer be used to identify a natural person by using 'all the means likely reasonably to be used by either the controller or a third party. The focus is on the outcome: that data should be such as not to allow the data subject to be identified via 'all,' 'likely,' and 'reasonable' means[212].

---

[206] Heinrich U & Weber R (University of Zurich), '*Anonymization*' at 1 32
[207] Ibid at 36
[208] Ibid at 36, cited as well in European parliament
[209] Recital 26 excludes anonymized data from the scope of data protection legislation
[210] https://www.dataprotection.ie/en/dpc-guidance/anonymisation-pseudonymisation
[211] https://www.dataprotection.ie/en/dpc-guidance/anonymisation-pseudonymisation
[212] Article 29 data protection working party; Opinion 05/2014 on Anonymization Techniques adopted on 10 April 2014

Moreover, the e-Privacy Directive (Directive 2002/58/EC) also refers to 'anonymization' and 'anonymous data' very much in the same regard. Recital 26 defines that:

"*Traffic data used for marketing communications services, or the provision of value-added services should also be erased or made anonymous after the provision of the service*." [213]

Moreover, under Article 9 (1): "*Where location data other than traffic data, relating to users or subscribers of public communications networks or publicly available electronic communications services, can be processed, such data may only be processed when they are made anonymous, or with the consent of the users or subscribers to the extent and for the duration necessary for the provision of a value-added service."*

Shortly, as an introduction to this section, the outcome of anonymization as a technique applied to personal data should be, 'in the current state of technology, as a permanent as erasure, i.e., making it impossible to process personal data.'[214]

On the other hand, considering limited instructions about anonymization as a technique, re-identification incidents are always in jeopardy. Article 29[215]is describes a few anonymization techniques (some of them are mentioned above regarding privacy preservation techniques in machine learning). Still, a new generation of hackers every year is challenging those techniques.
However, much research showed that no proper database can ever be perfectly anonymous, and as the utility of data increases, privacy decreases. Therefore, cheap, powerful reidentification will cause significant harm that is difficult to avoid despite various anonymization techniques[216]. For a long time, it has been considered that technologists can robustly protect people's privacy by making small changes to their data. However, something significant has changed. The 'easy reidentification result' provides that the robust anonymization assumption is deeply flawed, not fundamentally incorrect, but deeply flawed[217]. That leads to the point that humanity needs new methods to protect its privacy. At the same time, this method will follow the latest technology developments, for instance, federated learning. Furthermore, it has been found that one of the challenges both from a technical and legal perspective is that described techniques vary in tech and legal literature, which makes things harder for accession from both sides. For instance, interviewed engineers usually argue to have difficulty implementing legal requirements. Data protection officers claim it is hard to

[213] Ibid at 6
[214] Article 29 data protection working party; Opinion 05/2014 on Anonymization Techniques adopted on 10 April 2014 at 6
[215] Ibid
[216] Ohm P, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization', UCLA Law Review (2010) at 1706
[217] Ibid at 1707

understand which applied anonymization technique matches legal literature and what their engineers did to the dataset.

7.1 Storage of Recorded Data

Below will be discussed whether federated learning provides enough anonymity, but there is another challenge besides anonymity. Data needs to be kept for federated learning purposes as long as possible. Although illegal according to most current national laws, many providers store recorded data over a long period. Therefore, data such as the time of visiting a webpage, the IP addresses, and the whole history of surfing are collected. The web page operators' prior intention is to collect all this data to conduct marketing habits to streamline their web pages and increase business opportunities[218].

Moreover, web page operators collect data to protect their web pages from potential misuse. This collection also threatens anonymity as the data storage period must be about the right to be forgotten encompassing, sing the right to have data deleted after a certain period[219]. This was one of the questions answered in C-136/17 GC and others (de-referencing of sensitive data), where the court took the shape of 'the right to be forgotten.' Here the court gave, among other things, a territorial scope. While according to Article 17 GDPR of, the right to be forgotten (right to erasure) is a general right, and it is of high importance in the context of search engines.[220]

7.2 Failures

Firstly, regarding Article 29[221], during this work, it was proved to be out of date as it does not follow the new technological developments in this area. It is vital to raise the fact that both from our perspective and the perspective of interviewed FL engineers, Article 29 is not helpful whatsoever, as all of us had the same impression that it is over eight years. It does not provide a clear guideline, and it does not fit the new trends in this area**.** As not suited for new technological developments, especially, in FL systems, many researchers have begun to develop suitable privacy techniques for use with ML techniques[222].

---

[218] Ibid
[219] Ibid
[220] Globocnik J. 'The Right to be Forgotten is Taking Shape: CJEU Judgments in GC and Others (C-136/17 and Google v CNIL (C-507/17
[221] Article 29 data protection working party; Opinion 05/2014 on Anonymization Techniques
[222] For instance, even before FL, in 2015*, University of Texas and Cornell University developed a privacy preservation technique which enabled multiple participants to learn*

Secondly, it is frightening to admit that anonymization could be overrated today. Still, many defend the privacy-protection power of anonymization and hold it out as a best practice despite evidence to the contrary[223]. For instance, Google announced, "*It is difficult to guarantee complete anonymization, but we believe Google's log file anonymization techniques will make it very unlikely that users could be identified*."[224]
For instance, famous anonymized data reidentification has been mentioned, in the '*Broken Promises of Privacy.* '[225][226]
Moreover, there is a long list of famous data breaches, leaks, and re-identification examples[227]. One of the well-known re-identification experiments was conducted by a Netflix provider. Researchers have analysed the geometric properties of that database consisting of more than 100 million ratings on a scale of 1-5 on over 18.000 movies, expressed by almost 500.000 users, publicly released by the company after being 'anonymized' according to an internal privacy policy with all customers identifying information removed except ratings and dates. It has been found that 99% of user records could be uniquely identified, allowing 68% of users to be identified.[228]

Furthermore, as one of the most critical inputs for autonomous vehicles is cameras, which record everything related to the driver and a random individual who are just transposing, this exposes even a higher risk regarding anonymity. This type of data breach exposes up to millions of personal data records, such as in the Verkada data breach incident. [229]But as mentioned above, most legal discussions are transferring this matter to Member states to regulate as a matter of national security.

7.3 Pseudonymisation

Pseudonymization replaces one attribute (typically a unique attribute) in a record with another. Therefore, the natural person is still likely to be identified indirectly; consequently, pseudonymization will not result in an

---

neural network on their own inputs, without sharing without sharing these inputs but benefitting from other participants who are concurrently learning similar models. Ibid at 23

[223] Ohm P, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization', UCLA Law Review (2010) at 1710

[224] Soghoian C. 'Debunking Google's log Anonymization Propaganda, Surveillance State CNET NEWS 2008

[225] Ohm P, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization', UCLA Law Review (2010)

[226] Ibid , AOL Data Release, ZIP, SEX and Birth Date, The Netflix Prize Data Study at 1720

[227] https://www.upguard.com/blog/biggest-data-breaches

[228] Supra

[229] Impact of this breach was accessing the feed of over 150.000 surveillance cameras placed in not only manufacturing companies such as Tesla, but hospitals, schools, prisons and police departments according to Bloomberg. https://www.wci360.com/hundreds-of-surveillance-cameras-hacked-in-data-breach/

anonymous dataset when used alone. However, it is claimed that pseudonymization reduces the linkability of a dataset with the original identity of a data subject; as such, it can be regarded more as a security measure than a method of anonymization. Therefore, the data used in this concept will still be considered personal data[230].
In Article 29, Opinion 05/14 in Anonymization Techniques, some pseudonymization techniques have been listed, such as Encryption with a secret, a key, hash function, keyed-hash function with the stored key, deterministic encryption, or tokenization.

7.4 General recommendations:

However, our analysis showed that federated learning does not fit in any of the pseudonymization techniques simply; it requires more work invested in the machine learning process while forming federated mode. According to our findings, those models as aggregated data are closer to the fully anonymized data set. Moreover, the upper mentioned Opinion[231]claims that there must be a specific misconception between pseudonymized and anonymized datasets. Still, a small sample of interviewed engineers claimed to be perfectly aware of those crucial differences in practice.
Some academic claims that EU lawmakers believed that the 'power of anonymization will avoid difficult balancing questions,' [232]The EU never intended that GDPR apply to all data (as it does not apply) as the very aim was to establish a free flow of data inside of the internal market.
Instead, it refers only to personal data, which is not 'directly or indirectly identifiable, such as anonymized data. With this, EU lawmakers desired to preserve space in society to store and transfer anonymized data, thereby providing room for unencumbered innovation and free expression. Later EU even sets the goal for new AI and robotics developments as one of the leading EU strategies in the new industrial revolution.[233]
This led to the conclusion that it might be necessary from **a legislator's point of view** to continuous legal guidelines updates regarding newly developed anonymization and pseudonymization techniques; as briefly described above, as current anonymization techniques seem to be quite old in compared to new trends in the industry (such as FL). Those techniques are proven not to be a good guarantee for personal data privacy protection.
From **a data processor point of view**, the legislation is behind technical developments in the industry. Therefore, until EDPB comes out with a newly updated guideline on FL/ML, we propose that companies engaged in FL/ML projects should carry out prior consultation with the DPA in charge. This kind of collaboration is also crucial so that the data protection

[230] In Article 29, Opinion 05/14 in Anonymization Techniques some of the pseudonymization techniques have been listed, such as: Encryption with secret key, has function, keyed- hash function with stored key, deterministic encryption or tokenization.
[231] Ibid
[232] Ohm P, 'Broken Promises of Privacy' at 1738
[233] https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

authorities can understand the technical protection mechanisms that FL affords. Prior consultation will also help speed up the technological development of a new guideline from the data protection authorities and/or EDPB.

In this regard, the next section will be more about privacy promises in federated learning.

7.5 Privacy promises in federated learning

'*Federated learning is a technique aimed to implement a Machine learning algorithm in decentralized collaborative learning settings wherein the algorithm is executed on multiple local datasets stored at isolated data sourced (i.e., local nodes) such as smartphones, tablets, PCs, and wearable devices without the need for collecting and processing the training data at a centralized data server. FL grants local nodes to collaboratively train a shared ML model while retaining both the training dataset and computation at internal sites. Only the training results are exchanged at a certain frequency, which appoints the local server to coordinate the training process, aggregate the training results, and calculate the global model*.'[234] Naturally, FL is in direct advantage compared to conventional Machine learning processes. Personal data in federated learning are stored and processed locally while only parameters are exchanged, making it presumably compliant with GDPR. [235]Approximately, federated learning could be defined as a privacy preservation technique for a distributed learning system that has two goals:

- Privacy of the training dataset
- Privacy of the local model parameters which are exchanged with other nodes or centralized server[236]

However, some scholars claim that, despite FL's distributed collaborative learning model empowered by additional privacy-preservation techniques, some personal information could covertly be extracted from the local training parameters. That concludes that service providers could still be liable within the regulatory personal data framework and are still accountable for implementing GDPR-compliant mechanisms when dealing with EU citizen's personal data.

7.5.1 Effectiveness of Deanonymization Attacks in Federated Learning as a Privacy Preservation Technique

[234] Troung N, Sun K, Wang S and others- 'Privacy Preservation in Federated Learning: An insightful survey from the GDPR perspective' at 2
[235] Ibid at 2
[236] Ibid at 4

Almost all published papers have been analysed to prove the effectiveness and privacy preservation in a federated learning setting. After 2016, when Google presented federated learning methods, many technical papers regarding this ML technique were published. There has not been until now (to our knowledge) any legal article assessing this matter.
Several predominantly technical papers aim to validate the effectiveness of deanonymization attacks or, better said, reverse engineering in federated learning in both federated models and datasets. Those papers tried systematically to study the influence of deanonymization attacks both by limiting the subset of users the adversary has prior knowledge of and by limiting the amount and quality of the prior knowledge. [237]In one of these papers, the researcher claimed that conjecture that the user bias holds rather well towards numerous simulated attacks and suggested various methods to enshrine the process and increase the privacy guarantee[238]. Moreover, they questioned whether devices could truly participate anonymously without compromising the identity of individuals. Their results indicated that '*devices can effectively be deanonymized using the transmitted model parameter updates and a reasonable amount of prior data. To mitigate such attacks, they proposed calibrated 'domain-specific data augmentation, which shows strong results in preventing de-anonymization with minimal impact on utility*.[239]

7.5.2 DTU Research on the FL

A fantastic paper from Danish Technical University raised a particular risk of the vulnerability of FL models (image reconstruction and others) while advising practitioners to wisely choose network architecture by using differential privacy mechanisms and 'mini-batch size and communication strategies' when designing Federated Learning[240]. In this paper, they claimed that while reconstructing input data within the FL environment by imitating an honest server or participant, data might be reconstructed with only knowledge of the gradient update and model parameters.

[237] Fritz M, Zhang Y and others, 'Understanding and Controlling Deanonymization in Federated Learning', 2020 at 7
[238] Ibid at 13
[239] Ibid at 15
[240] Hansen Kai Lars & others, 'On the limits to learning input data from gradients', Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2021 at 12

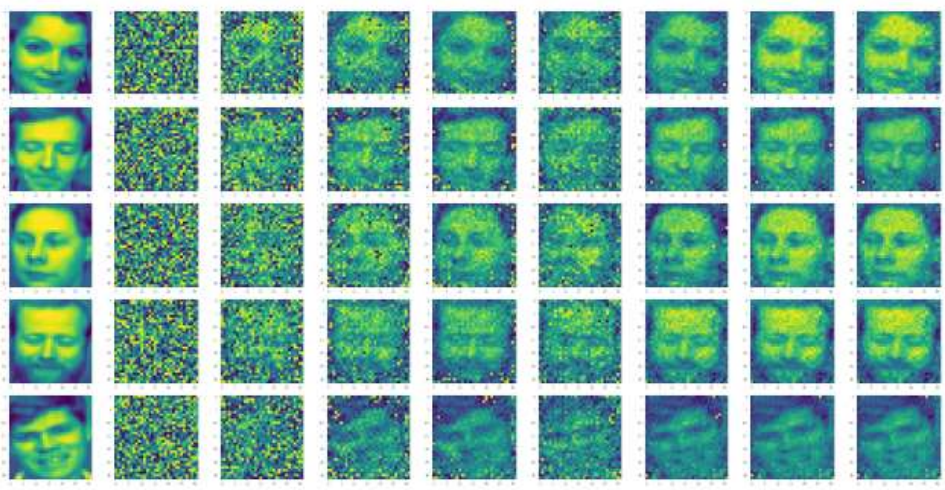


Figure 3: *Face* (batch reconstruction): Mini-batch contains 5 images, and we show partial reconstruction for 10, 50, 90, and final (400 iterations, with L-BFGS optimizer)iteration for [1] and our method accordingly.

Hansen Kai Lars & others, 'On the limits to learning input data from gradients,' Department of Applied Mathematics and Computer Science, Technical University of Denmark

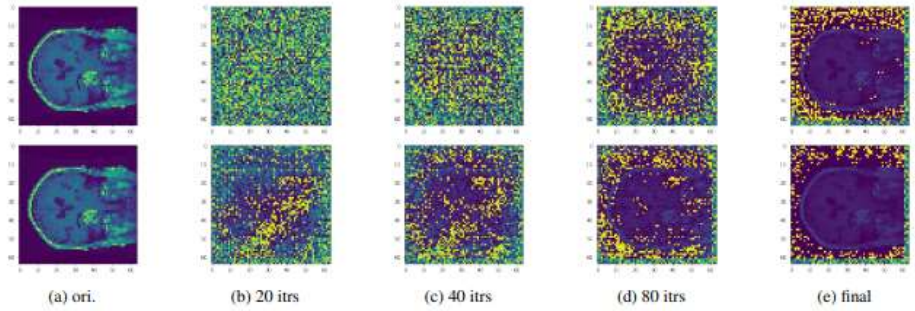


Figure 2: *FMRI*: The plots (first row) are produced by [1], and it shows the reconstructions after 20, 40, 80, and final iteration accordingly, whereas the second row is our reconstruction.

Hansen Kai Lars & others, 'On the limits to learning input data from gradients,' Department of Applied Mathematics and Computer Science, Technical University of Denmark

The upper presented pictures are datasets containing 3064 images of brain tumours from 233 patients. Face datasets contained 40 individuals. They also demonstrated the improvement of their method on the face dataset.[241] This clearly shows that FL with specific improvements guarantees a very high level of privacy protection.

7.5.3 Traditional ML vs. FL

In '*Privacy-Preserving Traffic Flow Prediction: A Federated Learning Approach*' Liu Yi and others claimed that data access and model

[241] Ibid at 8

performance in FL as a machine learning method demonstrates its privacy preservation superiority.[242]
Federated learning showed its dominance compared to conventional anonymization techniques and compared to other machine learning processes. *Traditional centralized machine learning cannot support ubiquitous deployments and applications due to infrastructure shortcomings such as limited communication bandwidth, intermittent network connectivity, and strict delay constraints*.[243] In this case, federated learning again pushed the training models to the devices from which they originated as a promising alternative to the ML paradigm. Federated learning advancingly; enables a multitude of participants to construct a joint ML model without exposing their private training data. FL can balance unbalanced and non-independent and identically distributed data which naturally arise in everyday life. Nowadays, FL benefits a wide range of applications such as next-word prediction, visual object detection for safety, etc.[244]

7.5.4 Decentralized ML vs. FL

On their web page, AI Sweden also categorizes federated learning as a decentralized AI[245]. However, in the paper published just a few days ago, '*On the privacy of Decentralized Machine Learning*[246], they clearly made a difference between decentralized machine learning and federated learning.

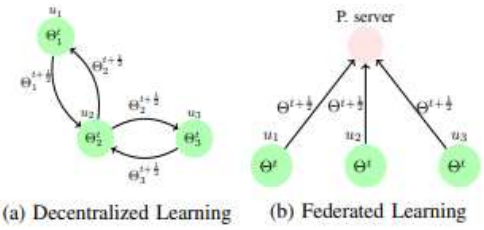


Fig. 1: Schematic representation of the decentralized learning and federated learning protocols.

[247]

They affirm that decentralized learning properties that affect users' privacy where they introduced a novel attack for both active and passive decentralized adversaries. Moreover, they demonstrated that contrary to what was claimed by decentralized learning proposers, decentralized

[242] Liu Yi, Kang J and others, 'Preserving Traffic Flow Prediction: A Federated Learning Approach', 2020, at 7761
[243] Yang Q, Fan L, Yu Han, 'Federated Learning- Privacy and incentive', Hong Kong University of Science and Technology, 2020 at 4
[244] Ibid at 4
[245] https://www.ai.se/en/projects-9/decentralized-ai
[246] Pasquini D, Raynal M, Tronsoco C, 'On the Privacy of Decentralized Machine Learning', 2022, at 1
[247] Ibid at 1

learning does not offer any security advantages over more 'practical' approaches such as federated learning. They reasoned this by clarifying that collaborative learning is gaining traction to train ML while respecting users' privacy. There are two approaches to collaborative machine learning: federated learning and decentralized learning. The main challenge is communication among users, which maintains the global state of the system [248] Furthermore, they reasoned this by claiming that using a unique central server brings limitations for both performance and privacy. In contrast, the server becomes a communication bottleneck on the communication side as the number of users in the system grows. The server becomes a single point of trust on the privacy side as it has complete control of the learning processes and can thus influence the used model. Finally, they concluded that users in decentralized learning could not reach the same privacy as users in federated learning, neither against passive nor active adversaries[249].

7.5.5 Anonymization techniques in the FL

Moreover, some researchers even proposed anonymization techniques inside the FL processes to enshrine the FL[250]. In '*Federated Learning with Blockchain for Autonomous Vehicles*[251], they suggest an *autonomous blockchain-based federated learning design for privacy-aware and efficient vehicular communication networking, where local on-vehicle machine learning model updates are exchanged and verified in a distributed fashion*. Here they proposed an enhanced FL technique for the performance and privacy of autonomous vehicles[252].

7.5.6 Local processing vs. FL

In '*A Hybrid Approach to Privacy-Preserving Federated Learning,'* they claim that simply maintaining data locally during the training process does not provide sufficient privacy guarantees. Therefore, this paper combines differential privacy and secure multiparty computation to reduce the growth of 'noise injection' as the number of parties increases without sacrificing privacy while maintaining a pre-defined trust rate[253]. Their suggestion is to implement FL systems with improved accuracy compared to existing approaches. Moreover, they propose including the 'tuneable trust' parameter while maintaining improved accuracy and privacy. They also claim that this

[248] Ibid at 1
[249] Ibid at 2
[250] Das A, Sylla I and others, 'Anonymizing Data for Privacy- Preserving Federated Learning'. 2020
[251] Pokhrel S & Choi J, 'Federated Learning with Blockchain for autonomous Vehicles: Analysis and Design Challenges'. 2020, at 4734
[252] Ibid at 4745
[253] Ibid at 1

provides end-to-end privacy guarantees where produced models can be safely deployed to production without infringing on privacy guarantees[254].

7.5.7 'KafkaFed' in the FL

Others propose the so-called 'KafkaFed' framework, which relies upon the Pub/Sub model based on FL's information-centric mode of communication.[255]
However, FL does not do personal data processing but only shares an FL model, which could be possibly considered as 'transient data.' [256]

7.5.8 Interim conclusions regarding the FL as an anonymization technique

Therefore, in light of all discussed above, about connected vehicles, legal scholars point out that transient data, i.e., data that is not stored in the 'long term', should be considered 'irrelevant' from a data protection standpoint[257][258].

Consequently, with complete confidence, we can conclude that federated learning showed its supremacy by administering a very high level of privacy protection comparing to classical anonymization techniques and other Machine Learning processes, even though FL is in its infancy and will be an active research area for the foreseeable future[259]. Moreover, it showed that it could be correlated with GDPR standards, with minimum risk for re-identifying original personal data inputs, with slight improvements to the current technology[260]. Finally, if constructed concerning privacy by design and default standards, it might be a future solution for all IoT technological improvements. Nonetheless, as mentioned above, for complete transparency of the implementation and decision-making process, a longer retention period could be a solution to ensure full transparency.

---

[254] Ibid at 10
[255] Bano S and others, 'KafkaFed: Two-Tier Federated Learning Communication Architecture for Internet of Vehicles', Department of Information Engineering, University of Pisa, Italy, 2022 at 6
[256] George D, Reutmann K & Aurelia Tamo-Larriex, 'GDPR bypass by design? Transient processing of data under GDPR', (2019) at 292, more at Benedikt Buchner, 'Datenschutz im vernetzten Automobil' (2015) 39 Datenschutz und Datensicherheit, 372, 374
[257] Ibid
[258] Concernedly, this view was supported by German case law handed down by Federal Administrative Court denying an intrusion into informational self- determination when data is immediately discarded after being recorded in an automated, anonymous, and untraceable fashion making it impossible to make any reference to a person at Ibid and more at BVerwG, 22 October 2014, Az.: BVerwG 6 C 7.13, para 27, with reference to BVerfGE 120, 378, 11 March 2008 – Automatisierte Kennzeichenerfassung; Buchner (n 64) 374.
[259] Ibid at 14
[260] Mammen P, 'Federated Learning: Opportunities and Challenges', 2021 at 4

8 Cyber security

To incentivize with this topic, I took an additional course to extend my cyber security knowledge. During this course and many published articles, the standard message is that we cannot secure 100% of the date 100% of the time. [261]The aim we should bear in mind while constructing a cyber security policy is to make sure how safe our data is **NOW**. To achieve 100% security, we should not have data at all[262]. But what can be done is to prioritize which security systems we can improve and how it can be done to mitigate the overall risk[263]. '*Cyber Security is big anxiety which needs to be chitchatted verbosely*.'[264]
Security has been well-defined as a progression to shield an object against physical destruction, unlicensed access, burglary, or loss, by preserving high confidentiality and truthfulness of information about the object and making information about the object presented whenever wanted. Therefore, certifying IoT security requires preserving the highest intrinsic value of both physical objects and immaterial ones.[265]

8.1 Cyber Security and GDPR

Cyber Security is a crosscut inside the GDPR rules as well.
Article 5 (principles relating to the processing of personal data).
'*Company must protect personal data to ensure appropriate security of the personal data, including protection against unauthorized or unlawful processing and accidental loss, destruction or damage, using appropriate technical or organizational measures*.'[266]

- **Article 28**- Data processors '*must only use processors providing sufficient guarantees to implement appropriate technical and organizational security and privacy measures*.'[267]

- **Article 32**- Or sometimes called '*state of the art provision.* '[268] Security of the processing- '*your organization must implement 'appropriate technical and organizational measures to ensure a level of security appropriate to the risk of data being processed*.'[269]

[261] https://hbr.org/2017/12/you-cant-secure-100-of-your-data-100-of-the-time
[262] https://www.functionize.com/blog/the-myth-of-100-code-coverage
[263] https://www.techrepublic.com/article/an-absolutely-secure-network-is-not-possible-but-the-risk-can-be-managed/
[264] Babu P and others, 'Cyber Security with IoT', 2019 at 1
[265] Ibid at 1
[266] https://fortifydata.com/gdpr-cyber-security/
[267] Ibid
[268] Guidelines at 8
[269] Ibid

- **Article 33**- Notification of a personal data breach to the supervisory authority- '*Within 72 hours after becoming aware of a breach, your company must notify the data breach of the supervisory authority. The Supervisory Authority is determined by a designated representative of the collector/processor (the company) in the EU*[270].

8.2 Prioritization as the first step

Thereupon, prioritizing is the first step in this assessment.
To conduct risk analysis, we first need to understand our assets and their value:

1. Which data is stored and processed?
2. What is its value?[271]

Secondly, we need to analyze threats and potential vulnerabilities:

1. How would a violation of confidentiality, availability, or integrity affect the value?
2. What is the probability of the threat?

Risk is calculated as follows:

VALUE X PROBABILITY= RISK

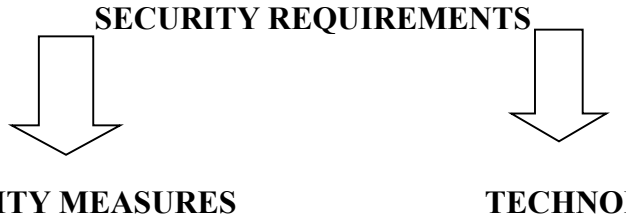
Firstly, high-risk areas need to be tackled to conduct a risk analysis as a basis for increasing security. [272]

Thirdly, after defining the value and threats, we need to formulate and specify what kind of security requirements we are dealing with. Accordingly, we need to ask a question: 'What are we wearisome to safeguard ourselves alongside?' there are three main threats:

- Unauthorized Access
- Unauthorized Deletion
- Unauthorized Modification[273]

[270] Ibid , more at https://www.bdo.dk/da-dk/services/advisory/cybersikkerhed?utm_source=bing&utm_medium=cpc&utm_campaign=Service%20-%20Cybersikkerhed%20(Dansk)&utm_term=Cyber%20Security&utm_content=Cyber%20Security
[271] Kun Wu- Chuan, '*Internet of Things Security- Architectures and Security Measures'* (2020) at 18
[272] https://www.educba.com/security-risk-analysis/
[273] Supra Babu P at 2

These three terms infer from their very universally known '**CIA**' triad, which stands for **C**onfidentiality[274], **I**ntegrity[275], and **A**vailability.[276][277]



Ergo, to prevent cyber-attacks, a continuous process of analysis must be performed:

1. Value
2. Threats
3. Requirements
4. New means

8.3 Cyber Security of Big data (potentially needed for FL)

Once organizations notice a new opportunity, they usually forget about the risks. But on the other hand, too rigorous or too early risk preventive measures could potentially decrease opportunities and suffocate the new technological developments. In an IoT system, the processing layer is specific as it may encounter an enormous, large scale of data, called *big data*. The concept of big data means that the data is too large or complicated to be adequately processed by traditional data processing techniques. Hence, the processing layer of IoT usually means cloud processing and is named 'cloud' for short[278]. Furthermore, big data are typically analyzed in real-time, meaning security network measures also need to be in real-time.

---

[274] *Confidentiality is the fortification of personal statistics. Confidentiality means possession of a client's data between you and the client, and not influential by others as well as colleagues, companions, bloodline, etc. Attacks are • Cracking encrypted data. • Man, in the mid attacks on plain text. • Data leakage / unauthorized copying of sensitive data. • Fixing spyware/malware on a server*. At Ibid

[275] *Integrity, in the perspective of computer systems, brings up techniques of making sure that data is tangible, precise, and fortified from unlicensed user amendment [5]. Attacks are • Web penetration for malware insertion. • Maliciously accessing servers and forgiving records. • Unauthorized Database scams. • Remotely controlling Zombie systems.* At Ibid

[276] *Availability, in the view of a computer system, mentions the knack of a handler to admittance statistics or assets in an indicated setting and in the spot-on layout. Attacks are • DOS / DDOS attacks. • Ransomware attacks forced encryption of key data. • Deliberately disrupting a server room power supply. • Flooding a server with so numerous appeals*. At Ibid

[277] Ibid at 2

[278] Kun Wu Chuan, '*Internet of things security'*, at 18

Therefore, specific security measures are necessary when dealing with an enormous amount of data; some of the challenges could be summarized as follows:

- We should identify if we deal with large-scale data, as more data could have more potential damage
- We need to identify whether we are dealing with heterogeneous data coming from different sources, bringing a unique challenge.
- And consequently, processing and storage in public clouds, security measures, or firewalls are not suitable for this purpose[279]

According to Harvard Business Review, most threats are coming from inside the organization. Most of the breaches are usually caused by an action or failure of someone inside the company. Wherefore, the organization fir needs to prevent unauthorized attacks. Thereupon, the organization should perform specific preventive measures such as:

- **Detective measures** (to discover unauthorized attacks)[280]
- **Administrative measures** (to clarify processes, rules, and standards)[281]
- And finally, **preventive measures** (such as encryption) to secure the data[282]

[279] More at Guidelines 02/2019 on Article 25 at 26

[280] **Detective measures** such as audits are to monitor if a system is actually secure and to detect attacks that cannot be prevented in order to identify whether measures are implemented as planned. Those measures can be active or passive. Active measures are improvement measures and passive measures are aimed to check if current measures work. A penetration test is a higher form of audit, where the organization can simulate a company's attack from a hacker's point of view to control if those measures work. The goal is to identify weaknesses and to improve the current measures that will make the system eventually more secure. Detective measures are aimed to increase security. Audits are aimed to correct planning configuration and a penetration test is aimed to remove weaknesses. DPIA as required by GDPR and could also check compliance with regulations. More at https://www.tutorialspoint.com/system_analysis_and_design/system_analysis_and_design_security_audit.htm

[281] Audit alone is not enough to identify risks, therefore, measures such as **monitoring** are necessary. Systems are dynamic and work by mathematical models and if everything goes well the data will behave according to the models. One technology that makes use of these models is known as anomaly or outlier detection. An outlier detection means to identify a specific anomaly in a world it actually should be consistent. More at https://www.wikiaccounting.com/what-is-audit-risks/

[282] While designing access control, measures for IT systems such as Big Data application, is crucial to distinguish between two separate processes: Authentication (means that the real person is the authentic and the real person have to use his user's ID) and authorization (is, for instance, I may be authentic, but I am not authorized to modify data). Authentication is a process by which you verify that someone is who they claim to be, while authorization is the process of establishing if the user is permitted to access a resource or perform a specific action. More at https://www.frescodata.com/blog/big-data-security-analysis-preventive-measure-security-threats/ or https://www.educba.com/authentication-vs-authorization/

Besides encryption as a security measure, some scholars are proposing the usage of blockchain technology for IoT. [283]They also raise the importance of this technology and how IoTs can operate without limitations enshrined with this technology.

The Guidelines are proposing similar construction and even going further as one of the priorities for connected autonomous vehicles[284]. In the Guidelines, EDPB proposes implementing additional security measures, which we could not cohere with our area of research[285].

Concerning the upper section regarding federated learning as an anonymization technique, where one of the proposals matches blockchain as a cyber security measure, this combination might occur as the best privacy engineering architecture for autonomous vehicles as IoTs, which could be compliant with GDPR but still provide privacy and cyber security guarantees.

[283] Won Lee Seok, Singh & Mohammadian Masoud, '*Blockchain Technology for IoT Applications'*
[284] Guidelines 01/2020 at 23
[285] Ibid at 23

9 Conclusion

Based on the preceding analysis and the hypothesis of this thesis, this paper managed to define the personal data in the autonomous vehicle setting based on federated learning. Moreover, this paper differenced all data types necessary for federated learning development.
Additionally, it defined the obligation of the data processor and proposed adequate privacy by design and default measures.
Finally, this paper analyzed all available technical articles to prove that federated learning could be a long-awaited 'green card' for the brand-new technological developments in the IoT world. FL showed that it could be correlated with GDPR standards, with minimum risk for re-identifying original personal data inputs, with slight improvements to the current technology. However, to ensure complete transparency of the implementation and decision-making process, a longer retention period could be a solution to provide full transparency.
Moreover, blockchain technology could guarantee a high level of cyber security.
Once again, we should raise the benefits in the socio-economic context of applying federated learning as it will bust the technological development and reach the European Union's goals of being a leader in AI.
Federated learning could even assure the free flow of data inside the internal market but potentially outside the EU; an initial goal of data privacy legislation was to establish the free flow of data as any other asset.
Federated learning might also solve various challenges regarding preventing cyber security incidents.
Our research undoubtedly showed that current legislation is vague. It does not follow the technical developments (as it usually does not), and it fails to guide users of FL/ ML with GDPR compliance. Therefore, once again, we propose that until EDPB comes out with a newly updated guideline on FL/ML, companies engaged in FL/ML projects should carry out prior consultation with the DPA in charge. This kind of collaboration is also crucial so that the data protection authorities can understand the technical protection mechanisms that FL affords. Prior consultation will also help speed up the technological development of a new guideline from the data protection authorities and/or EDPB.
Professor Felsberg from Linköping University believes that although many of the easiest problems for autonomous vehicles have been solved, there are still a lot of hard problems that are nowhere near resolution. Level 5 automation, in which vehicles do not require any human attention, is still a long way off.[286]

I consider myself a believer and optimist in technology, and we might be closer to fully automated cars than ever before.

[286] https://www.computerweekly.com/feature/Swedish-researcher-cuts-through-the-hype-around-autonomous-vehicles

10 Appendix 1- privacy by design guidelines

The privacy by design and default controls should be linked to the risk the new product/process presents. **High-risk processes require more management than low and medium-risk processes.** Below is a high-level overview of the risk associated with various types of processing. All kinds of data processing can result in high risk and administrative fines of up to 20 000 000 EUR (or, in the case of an undertaking, up to 4 % of the total worldwide annual turnover) if we fail to provide the data subject with a privacy notice, data subject rights or transfer data to a country outside of the EU/EEA

Consider if your new product/process will do the following:

Type of data	Example	Risk	
Fully anonymized data	The process by which Personal Data is **irreversibly** altered (no reverse engineering possible) in such a way that an individual can no longer be identified directly or indirectly, such as statistical variables that cannot identify an individual	Low	The anonymization process must be quality assured and tested. The risk of the administrative fine is low.
Personal Data	name, date of birth, contact details	Medium	Administrative fines up to 10 000 000 EUR (or in the case of an undertaking, up to 2 % of the total worldwide annual turnover)
Children data	Data relating to minors (the age of minors will vary between the EU Member States and there will also be differences in age of consent depending on purpose processing. For example, marketing and credit rating will have different ages of consent.[287]	Medium	Administrative fines up to 10 000 000 EUR (or in the case of an undertaking, up to 2 % of the total worldwide annual turnover)
Vulnerable individuals	Data related to, for example elderly. As with children, there may be differences between the EU Member States as to when an	Medium	Administrative fines up to 10 000 000 EUR (or in the case of an undertaking, up to 2 % of the total worldwide annual turnover)

[287] https://euconsent.eu/digital-age-of-consent-under-the-gdpr/

	individual is considered" elderly" or retired.[288]		
Other Sensitive Data	National ID number or other unique identifiers	Medium/High	Administrative fines up to 10 000 000 EUR (or in the case of an undertaking, up to 2 % of the total worldwide annual turnover)
Special Categories of Personal Data (also known as Sensitive Data)	data relating to health, race, religion, sexuality, political and philosophical[289]	High	Administrative fines up to 20 000 000 EUR (or, in the case of an undertaking, up to 4 % of the total worldwide annual turnover)
Combining and matching Personal Data	Taking data from various databases or registers and combining them with our data. Meets GDPR Article 5 b requirements[290]	High	Administrative fines up to 20 000 000 EUR (or in the case of an undertaking, up to 4 % of the total worldwide annual turnover)
Automated decision making	Automated decision-making is the process of deciding by automated means without any human involvement. These decisions can be based on factual data and digitally created profiles or inferred data. Examples of this include: - an online decision to award a loan Automated decision-making often involves profiling, but it does not have to.[291]	High	Administrative fines up to 20 000 000 EUR (or in the case of an undertaking, up to 4 % of the total worldwide annual turnover)
Profiling	Assessing or classifying individuals based on characteristics. More specifically defined as: "' *profiling' means any form of automated processing of Personal Data consisting of the use of Personal Data to evaluate certain personal*	High	Administrative fines up to 20 000 000 EUR (or in the case of an undertaking, up to 4 % of the total worldwide annual turnover)[293]

[288] https://ico.org.uk/media/for-organisations/guide-to-the-general-data-protection-regulation-gdpr-1-0.pdf - Guide to the General Data Protection Regulation

[289] https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/what-personal-data-considered-sensitive_en , Article 9 GDPR

[290] https://gdpr-info.eu/art-5-gdpr/ , Article 5 GDPR

[291] https://privacyinternational.org/sites/default/files/2018-04/Data%20Is%20Power-Profiling%20and%20Automated%20Decision-Making%20in%20GDPR.pdf , Privacy International- Data is Power: Profiling and Automated Decision Making in GDPR

[293] https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_gdpr_project_risk_white_paper_21_december_2016.pdf , 'Risk, High Risk, Risk Assessment and Data Protection Impact assessments under the GDPR' by CIPL, 2016

	aspects relating to a natural person, in particular to analyses or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behavior, location or movements." Profiling must have a significant/legal/economic impact for it to be considered high risk[292]		

[292] Ibid

11 Appendix 2 -FL engineers interview summary

Questions	Engineer 1	Engineer 2	Engineer 3	Engineer 4
What type of data do you collect Raw or metadata?	Everything	Everything	Everything	Everything
Do you find helpful Article 29?	No	No/yes	No	No
Do you understand how to implement article 25?	Not sure	Yes/no	No	No
What anonymization techniques do you use, and does it affect the data quality?	We do not use any as we work on a project out of national security importance.	We do not use any as we are in the R&D phase of the project	We can, but we do not use it now	We do not use any
Do you find that anonymization techniques can affect the quality of the dataset for FL?	I don't think it can affect the quality of the dataset for FL	We use raw data as it affects the quality of the dataset for FL, and we need as much data as possible	Nor one anonymization technique can affect the quality of the data for FL. Organizations usually choose cheaper versions of anonymization techniques even though homomorphic encryption in combination with FL is 'unbreakable' but it is time and money consuming	I do not believe that it can affect the quality of the data; sometimes it is just expensive to be done.
Do you understand the difference between pseudonymization and anonymization?	Yes	Yes	Yes	Yes
Do you find that FL could be closer to anonymization than pseudonymization?	Anonymization	Anonymization	Anonymization, adding some PET is making FL superior to other ML/ anonymization techniques.	Anonymization

12 Bibliography

Articles:

- Troung N, Sun K, Wang S and others- 'Privacy Preservation in Federated Learning: An insightful survey from the GDPR perspective' (2016) accessed 20.02.2021.
- Cosar Ahmet & Turk Ismail- 'Internet Connection Sharing Through NFC for Connection Loss Problem in Internet of Things Devices', (August 2015), accessed 14.03.2022.
- Fahsi.M, Benslimane S., Rahmani A., 'A Framework for Homomorphic, Private Information Retrieval Protocols in the Cloud, (May 2015), accessed 14.03.2022.
- Mammen M. Priyanka, 'Federated learning: Opportunities and Challenges'. (2021) accessed 23.03.2022.
- Persson F. 'Information security risk review and analysis for the future autonomous vehicle', (2017) Luleå University of Technology
- Tyler R. Tom. 'Methodology in Legal Research', (2017), Yale University, accessed 22.02.2022.
- Prof. Miodrag Jovanovic, 'Legal Methodology & Legal Research and Writing', accessed 23.03.2022
- Prof (Dr) Vibhute K & Aynalem F. 'Legal Research Methods, prepared under the sponsorship of the Justice and Legal System Research Institute', accessed 23.03.2022.
- Jaswal K. Anil & Rajasekhar M., 'Autonomus vehicles: The future of automobiles', accessed 29.03.2022.
- Englund C, Torstensson M & Chen L- 'Federated Learning to enable automative collaborative ecosystem: opportunities and challenges', Virtual ITS European Congress, 9-10 November 2020, accessed 20.03.2022.
- Frost & Sullivan, 'Otonomo, 2018 European Car Data Platform New Product Innovation Award', accessed 29.02.2022.
- Article 29 data protection working party; Opinion 05/2014 on Anonymisation Techniques adopted on 10 April 2014
- Globocnik J. 'The Right to be Forgotten is Taking Shape: CJEU Judgments in GC and Others (C-136/17 and Google v CNIL (C-507/17' accessed 18.05.2022.
- Ohm P, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization', UCLA Law Review (2010), accessed 10.11.2021.
- Fritz M, Zhang Y and others, 'Understanding and Controlling Deanonymization in Federated Learning' accessed 16.03.2021.
- Kohli N, Athavale S & Doomra S, 'Turn Signal Prediction: A Federated Learning Case Study', 2020, accessed 03.01.2021.
- Hansen Kai Lars & others, 'On the limits to learning input data from gradients', Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2021, accessed 01.03.2021.
- Wei K. and others, 'Federated Learning with Differential Privacy: Algorithms and Performance Analysis', 2020, accessed 20.10.2021.
- Lu Y and others, 'Federated Learning for Data Privacy Preservation in Vehicular Cyber-Physical Systems'.2020, accessed 26.02.2021.
- Berthold S, Fischer S,'Privacy-Enhencing Technologies', Karlstad University,2017, accessed 01.03.2021.
- Das A, Sylla I and others, 'Anonymizing Data for Privacy-Preserving Federated Learning'. 2020, accessed 21.02.2021.

- Song M and others, ‘Analysing User-Level Privacy Attack Against Federated Learning’, 2020, accessed 20.03.2021. ‘
- Peng W and others, ‘A Two-Stage Deanonymization Attack against Anonymized Social Networks.’, 2014, accessed 26.03.2021.
- Liu X and others, ‘Adaptive privacy- preserving federated learning’.2020, accessed 30.03.2021.
- Asad M and others, ‘A Critical Evaluation of Privacy and Security Threats in Federated Learning’, 2020, accessed 01.03.2021.
- Treux S and others, ‘A Hybrid Approach to Privacy Preserving Federated Learning’, 2019, accessed 15.03.2021.
- Parizi R and others, ‘A Survey on Security and Privacy of Federated Learning’, 2020, accessed 28.02.2021.
- Zhang C and others, ‘A Survey on Federated Learning’, 2020, accessed 26.02.2021.
- Mammen P, ‘Federated Learning: Opportunities and Challenges’, 2021, accessed 03.04.2021.
- Tal Zarsky, ‘Incompatible: The GDPR in the Age of Big Data’ (2017) Seton Hall Law Review 995
- George D, Reutmann K & Aurelia Tamo-Larriex, ‘GDPR bypass by design? Transient processing of data under GDPR’, International Data Privacy Law (2019), accessed 28.04.2021.
- Privacy by Design- Information & Privacy Commissioner of Ontario
- Good N & Rubinstein I, ‘The Trouble with Article 25 (How to Fix it): The Future of Data Protection by Design and Default’, accessed 20.03.2021.
- Pokhrel S & Choi J, ‘Federated Learning with Blockchain for autonomous Vehicles: Analysis and Design Challenges’. 2020, accessed 21.05.2022.
- Bano S and others, ‘KafkaFed: Two-Tier Federated Learning Communication Architecture for Internet of Vehicles’, Department of Information Engineering, University of Pisa, Italy, 2022, accessed 20.05.2022.
- Winfield A & others, ‘Machine Ethics: The Design and Governance of Ethical AI and Autonomous Systems’, 2019, accessed 19.04.2022.
- Spiekermann, J. Korunovska, and M. Langheinrich, “Inside the organization: Why privacy and security engineering is a challenge for engineers,”. 2019, accessed 21.05.2022.
- Privacy International- Data is Power: Profiling and Automated Decision Making in GDPR, 2017, accessed 22.05.2022.
- Fellander Anna, Teigland Robin & Holmberg Håkan, ‘The importance of trust in digital Europe: Reflections on the sharing economy and blockchains’. 2018, accessed 20.09.2021.
- Heintz F & Larsson, ‘Transparency in artificial intelligence’, 2020, accessed 20.10.2021.
- Fellander A, Heintz F and others, ‘Achieving a Data driven Risk Assessment Methodology for Ethical AI’, 2021, accessed 22.05.2021.
- ‘Risk, High Risk, Risk Assessment and Data Protection Impact assessments under the GDPR’ by CIPL, 2016, accessed 20.04.2021.
- Seda Gürses and Joris V. J. van Hoboken, ‘Privacy After the Agile Turn’ in Evan Selinger, Jules Polonetsky and Omer Tene (eds.), The Cambridge Handbook of Consumer Privacy (2018), accessed 20.05.2022.
- Gurses Seda, Kostova Blagovesta & Tronsoco Carmela, ‘Privacy engineering meets software engineering’ 2020, accessed 20.05.2020.
- Babu P, Pavani C, Naidu E,’Cyber Security with IOT’, 2019, accessed 20.05.2021.

- Sadek Islam, Momtaz A, Muhammed Ilyas, 'Securing IoT Devices using Blockchain Concept', 2021, accessed 15.01.2022.
- Pasquini D, Raynal M, Tronsoco C, 'On the Privacy of Decentralized Machine Learning', 2022, accessed 23.05.2022.
- Krontiris I & others, 'Buckle up: Autonomous Vehicles Could Face Privacy Bumbs in the Road Ahead,' accessed 28.04.2021.
- Zallone R, 'Connected Cars under the GDPR', 2019, accessed 28.05.2021.
- Veitas V & Delaere S, 'In-vehicle data recording, storage and access management in autonomous vehicles', 2018, accessed 28.04.2021.
- 'Innovation is great- connected and automated vehicles- UK model' at great.gov.uk
- PhD research, Montgomery David, 'Public and Private Benefits of Autonomous Vehicles', 2018, Securing America's Future Energy, accessed 28.04.2021.
- Truex S & others, 'A Hybrid Approach to Privacy-Preserving Federated Learning', 2019, accessed 20.04.2022.
- Zhang C & others, 'A survey on federated learning', 2020, accessed 26.02.2021.
- Mothukari V & others, 'A survey on security and privacy of federated learning', 2019, accessed 26.02.2021.
- Arachchige M & others, 'A trustworthy Privacy-Preserving framework for machine learning in industrial IoT systems', 2020, accessed 25.02.2021.
- Peng W & others, 'A two-stage deanonymization attack against anonymized social networks', 2014, accessed 26.02.2021.
- Liu X & others, 'Adaptive privacy-preserving federated learning', 2020, accessed 26.02.2021.
- Brasher E, 'Addressing the failure of anonymization: guidance from the European union's general data protection regulation,' 2021, accessed 26.02.2021.
- Gao J & other, 'Against signed graph deanonymization attacks on social networks,' 2017, accessed 26.02.2021.
- Song M & others, 'Analysing User-Level Privacy Attack Against Federated Learning,'2020, accessed 26.02.2021.
- Froomkin M, 'Anonymity and its Enmities (Article 4)', accessed 26.02.2021.
- Choudhury O & others, 'Anonymization Data for Privacy-Preserving Federated Learning', 2020, accessed 26.02.2021.
- Wang Z & others, 'Beyond Inferring Class Representatives: User- Level Privacy Leakage Federated Learning', 2018, accessed 20.04.2021.
- Palmbach J & others, 'Blockchain- orchestrated machine learning for privacy preserving federated learning in electronic health data', 2020, accessed 26.02.2021.
- Han W & Others, 'Darknet and Bitcoin De-anonymization: Emerging Development', 2020, accessed 26.02.2021.
- Qu Y & others, 'Decentralized Privacy Using Blockchain-Enabled Federated Learning in Fog Computing', 2020, accessed 26.02.2021.
- Wu C & others, 'Distributed modeling approaches for data privacy preserving', 2019, accessed 26.02.2021.
- Aynale F & Vibhute K, 'Prepared under the Sponsorship of the Justice and Legal System Research Insitute', 2009
- Bonatti P & Kirrane S, 'Big Data and Analytics in the Age of GDPR,' 2019, accessed 28.01.2021.
- Tene O & Polonetsky, 'Big Data and User Control in the Age of Analytics', accessed 03.02.2021.
- Yvonne D, 'Conceptualizing the right to data protection in an era

of Big Data', 2017, accessed 11.03.2021.
- Bordel B & others, 'Data Authentication and Anonymization in IoT Scenarios and Future 5G Networks Using Chaotic Digital Watermarking,' 2021, accessed 19.02.2021.
- Salgado A & others, 'Preliminary Tendences of Users' Expectations about Privacy on Connected-Autonomous Vehicles', 2020, accessed 24.05.2021.
- Ribeiro S, 'Privacy Protection with Pseudonymization and Anonymization In a Health IoT Systems,' 2019, accessed 19.02.2021.
- Acquisti A & others, 'The Economics of Privacy', 2019, accessed 03.02.2021.
- Churakov & Gubaydullina, 'Legal Regulation of Big Data in Industrial Systems: Problems and Development Prospects',2020, accessed 11.03.2021.
- Degree project, Persson F,' Information security risk review and analysis for the future autonomous vehicle', 2017, accessed 23.03.2022.
- Averin A & others, 'Review of Methods for Ensuring Anonymity and De-Anonymization in Blockchain', 2020, accessed 19.02.2021.
- Ekmefjord M & others, 'Scalable federated machine learning with FEDn', 2021, accessed 09.06.2021.
- Edwards L & Veale M, 'Slave to the algorithm? Why a 'Right to an explanation is probably not the remedy you are looking for', accessed 03.02.2021.
- Augusto C & others, 'Test-driven Anonymization for Artificial Intelligence, 2019, accessed 19.02.2021
- Augusto C & others, 'Test-driven Anonymization in Health Data: A Case Study on Assistive Reproduction,' 2020, accessed 19.02.2021.
- Lee Seok, Singh Iris, Mohammadian Masoud, '*Blockchain Technology for IoT Applications'*, 2021, accessed 23.05.2021.
- Patti F, 'The European Road to autonomous vehicles,' 2019, accessed 16.02.2021.
- Powell A & others, London School of Economics and Political Science- 'Understanding and Explaining Automated Decisions,' 2019, accessed 27.05.2021.
- Wachter S & others, 'Why a right to explanation of automated decision-making does not exist in the General Data protection Regulation,' accessed 03.02.2021.

Books:

- Lundgren Björn- '*Information, security, privacy, and anonymity: definitional and conceptual issues'* (2018) accessed 26.02.2021.
- Aurelia Tamo Larrieux- *'Designing for Privacy and its Legal Framework- Data protection by design and default for the internet of things'* (2018) accessed 27.02.2021.
- Lim Y. Hannah- '*Autonomous vehicles and the Law, Technology, Algorithms and Ethics'*(2018), accessed 21.03.2022.
- Channon M, McCormick L. and Noussia K., '*The Law and autonomous vehicles'*, 2019 accessed 20.03.2022.
- Hoecke Van M. '*Methodologies of Legal Research, What Kind of Method for What Kind of Discipline*?', (2011), accessed 23.02.2022
- Heinrich U & Weber R (University of Zurich), '*Anonymization*' (2012), accessed 09.05.2022.
- Davenport Thomas, '*Analytics at Work Smarter Decisions, Better Results*', 2010, accessed 10.02.2022.
- Yang Q, Fan L, Yu Han, '*Federated Learning- Privacy and incentive'*, Hong Kong University of Science and Technology, 2020, accessed 26.02.2021.

- Simitis S & Dammann U, '*Federal Data Protection Act (BDSG), with national data protection laws and international regulations'*, 2014, accessed 23.03.2022.
- Bakardjieva A and others, '*Trust in the European Union in challenging times: Interdisciplinary European studies*.'2019, accessed 22.05.2022.
- Butun Ismail, '*Industrial IoT- Challenges, Design Principles, Applications and Security'*, 2020, accessed 20.05.2020.
- Maanak Gupta and others, '*Access Control Models and Architectures for IoT and Cyber Physical Systems'*, 2022, accessed 23.05.2022.
- Kun C. Wu, '*Internet of Things Security'*, 2021, accessed 02.06.2021.
- Krontiris I & others, '*Autonomous Vehicles: Data Protection and Ethical Considerations*', 2020, accessed 04.06.2021.
- Fagnant D & Kockelman K, '*Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations*, 2015, accessed 28.04.2021.
- Ryan Mark, 'The Future of Transportation, Ethical, Legal, Social and Economic Impacts of Self-driving Vehicles in the Year 2025,' 2019, accessed 28.04.2021.
- Fischer Simone & Berthold Stefan, '*Computer and Information Security handbook', chapter 53, 'Privacy- Enhancing technologies'*, 2017, accessed 11.03.2021.
- Asmarina S & others, '*Engineering Economics: Decisions and Solutions from Eurasian Perspective'*, 2021, accessed 11.03.2021.
- Cohen J, 'How Privacy Got a Bad Name for Itself', 2012, accessed 03.02.2021.

Legislation and soft legislation:

- Regulation of the EU Parliament and of the Council on Union guidelines for the development of the trans-European transport network, amending Regulation (EU) 2021/1153 and Regulation (EU) No 913/2010 and repealing Regulation (EU) 1315/2013
- COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS Internet Policy and Governance Europe's role in shaping the future of Internet Governance
- Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions- A European Strategy for data
- COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE, THE COMMITTEE OF THE REGIONS On the road to automated mobility: An EU strategy for mobility of the future
- Decision of the Court of Justice of the European Union of 1 October 2019 on internal rules concerning restrictions of certain rights of data subjects in relation to the processing of personal data in the exercise of non- judicial functions of the Court of Justice of the European Union
- Federation internationale de l' automobile region I- Europe, The Middle East and Africa, 'What EU legislation says about car data', Legal Memorandum on connected vehicles and data', 2017, accessed 30.04.2021.
- EDPB-EDPS- Joint Opinion 05/2021 on the proposal for a

Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) 18.06.2021.
- National approach to Artificial Intelligence- Government Offices of Sweden https://www.government.se/4a7451/contentassets/fe2ba005fb49433587574c513a837fac/national-approach-to-artificial-intelligence.pdf
- Article 29 Data Protection Working Party- Opinion 03/2017 on Processing personal data in the context of Cooperative Intelligent Transport Systems
- Data Protection- Anonymisation: managing data protection risk, code of practice,
- Article 29 data protection working party, Opinion 05/2014 on Anonymisation Techniques
- Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data
- Guidelines 04/2019 on Article 25, Data Protection by Design and Default 2020
- White Paper on Artificial Intelligence- A European approach to excellence and trust, 2020
- Guide to Case-Law of the European Court of Human Rights, 2020
- Guidelines on the exemption procedure for the EU approval of automated vehicles, 2019
- Guidelines 01/2020 on processing personal data in the context of connected vehicles and mobility related applications, 2021
- Handbook on European data protection law, 2018
- Proposal for a Regulation of the European Parliament and of the Council, laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts, 2021
- Regulation (EU) 2018/1807 of the European Parliament and of the Council on a framework for the free flow of non-personal data in the European Union, 2018
- Report from the Commission to the European parliament, the Council and the European Economic and Social Committee- Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics
- The DPO Handbook, Guidance for data protection officers in the public and quasi-public sectors on how to ensure compliance with the European Union General Protection Regulation, 2016

Table of Cases

- Case of Amann v Switzerland (27798/95)
- Case *S. and Marper v the United Kingdom* No30562/04 and 30566/04
- Case *Satakunnan Markinaporssi Oy and Satamedia Oy v Finland* No931/13
- Case *Leander v Sweden* No9248/81, §48,
- Case *Rotaru v Romania* No28341/95
Case *C-434/16 Peter Nowak v Data Protection Commissioner*
- *Opinion of Advocate General Kokott- Case C-434/16 Peter Nowak v Data Protection Commissioner*
- Case C-582/14 *Patrick Breyer v Bundesrepublik Deutschland*
- *Case C-136/17 GC, AF, BH, ED v Commission nationale de Informatique er des libertes (CNIL)*

Links:

- https://inspec-analytics-app.theiet.org/#/landing
- https://gdpr.eu/checklist/
- https://gdpr.eu/what-is-gdpr/
- https://www.investopedia.com/articles/investing/052014/how-googles-selfdriving-car-will-change-everything.asp#:~:text=In%202018%2C%20Waymo%20announced%20that,except%20in%20some%20trial%20programs.
- https://www.vox.com/future-perfect/2020/2/14/21063487/self-driving-cars-autonomous-vehicles-waymo-cruise-uber
- https://www.theguardian.com/technology/2015/sep/13/self-driving-cars-bmw-google-2020-driving
- https://www.businessinsider.com/report-10-million-self-driving-cars-will-be-on-the-road-by-2020-2015-5-6?r=US&IR=T
- https://leonard.vinci.com/en/the-national-strategy-for-automated-mobility-enshrines-cooperation-between-the-autonomous-vehicle-and-the-infrastructure/
- https://www.nytimes.com/2020/12/07/technology/uber-self-driving-car-project.html
- https://www.bbc.com/news/business-55224462#:~:text=Uber%20is%20selling%20its%20driverless,self%2Ddriving%20cars%20a%20reality.
- https://www.transportstyrelsen.se/en/road/Vehicles/self-driving-vehicles/
- https://www.jdsupra.com/legalnews/germany-takes-the-lead-with-a-new-law-7746782/#:~:text=German%20lawmakers%20have%20approved%20a,operation%20as%20soon%20as%202022.
- https://www.bmvi.de/SharedDocs/EN/publications/strategy-for-automated-and-connected-driving.pdf?__blob=publicationFile
- https://europe.autonews.com/article/20180808/ANE/180809840/france-pushes-for-highly-automated-vehicles-by-2022
- https://www.insidetechlaw.com/blog/france-new-legislative-developments-for-autonomous-vehicles
- https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0283
- https://ec.europa.eu/docsroom/documents/34802
- https://edpb.europa.eu/our-work-tools/documents/public-consultations/2020/guidelines-12020-processing-personal-data_en
- https://www.government.se/4a7451/contentassets/fe2ba005fb49433587574c513a837fac/national-approach-to-artificial-intelligence.pdf
- https://www.ai.se/en/projects-9/decentralized-ai
- https://www.largestcompanies.com/toplists/sweden/largest-companies-by-turnover/industry/manufacture-of-motor-vehicles-trailers-and-semitrailers
- https://www.volvogroup.com/en/future-of-transportation/innovation/automation.html
- https://www.eurofins-cybersecurity.com/news/security-problems-iot-devices/
- https://www.who.int/news/item/19-04-2007-road-traffic-crashes-leading-cause-of-death-among-young-people
- https://www.dataprotection.ie/en/dpc-guidance/anonymisation-pseudonymisation
- https://www.upguard.com/blog/biggest-data-breaches
- https://iapp.org/news/a/processing-biometric-data-be-careful-under-the-gdpr/
- https://themarkup.org/ask-the-markup/2022/02/24/who-is-policing-the-location-data-industry
- https://www.definitions.net/definition/input+data
- https://www.dentons.com/en/insights/alerts/2020/december/22/gdpr-update-biometric-data

- https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/criminal-offence-data/what-is-criminal-offence-data/
- https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence
- https://gdpr.eu/data-protection-impact-assessment-template/
- https://hbr.org/2017/12/you-cant-secure-100-of-your-data-100-of-the-time
- https://hbr.org/2017/12/you-cant-secure-100-of-your-data-100-of-the-time
- https://www.functionize.com/blog/the-myth-of-100-code-coverage
- https://www.techrepublic.com/article/an-absolutely-secure-network-is-not-possible-but-the-risk-can-be-managed/
- https://www.leidenlawblog.nl/articles/traditional-legal-methodologywhat-if-you-have-never-seen-an-elephant-before