Cost effective, Multilingual, Privacy-driven voice-enabled Services

www.compriseh2020.eu

Call: H2020-ICT-2018-2020
Topic: ICT-29-2018
Type of action: RIA
Grant agreement Nº: 825081

WP Nº5: Cloud-based platform for multilingual voice interaction

Deliverable Nº5.1: Data protection and GDPR requirements

Lead partner: ROOT

Version Nº: 1.0

Date: 30/05/2019

Document information	
Deliverable Nº and title:	D5.1 – Data protection and GDPR requirements
Version Nº:	V1.0
Lead beneficiary:	ROOT
Author(s):	Álvaro Moretón, Conrado Castillo, Laura Merlo
Reviewers:	Marc Tommasi (INRIA), Emmanuel Vincent (INRIA), Zaineb Chelly Dagdia (INRIA), Raivis Skadiņš (TILDE)
Submission date:	30/05/2019
Due date:	31/05/2019
Type[1]:	R
Dissemination level[2]:	PU

Document history			
Date	Version	Author(s)	Comments
10/05/2019	0.1	Álvaro Moretón, Conrado Castillo, Laura Merlo	First draft of the deliverable
17/05/2019	1.0	Álvaro Moretón, Conrado Castillo, Laura Merlo	Final version based on the reviewers' comments

[1] **R R**: Report, **DEC:** Websites, patent filling, videos; **DEM:** Demonstrator, pilot, prototype; **ORDP:** Open Research Data Pilot; **ETHICS:** Ethics requirement. **OTHER:** Software Tools

[2] **PU:** Public**; CO:** Confidential, only for members of the consortium (including the Commission Services)

Document summary

This deliverable (D5.1) is the output of Task 5.3 “Data protection and GDPR compliance” of WP5. The objective of this deliverable is to provide a comprehensible summary of the main aspects regarding the General Data Protection Regulation (GDPR), including an analysis of the type of data involved in voice interaction technologies, and to identify barriers and requirements to comply with such regulation. One of the main goals is to provide an analysis, when needed, regarding the particularities of the processing of personal data when technologies integrated in voice-enabled systems are involved, and also regarding the processing activities that will be carried out within COMPRISE.

The document contains recommendations to be compliant with the GDPR requirements as well as instructions and guidelines on how to carry out a Privacy Impact Assessment (PIA) and on Information Security management.

Finally, an overview of other possible legal aspects that may affect the activities related to voice-enabled technologies, such as Intellectual Property, Trade Secrets or even ethical aspects will be provided.

D5.1 is a living document and, since the project is in its early stages, the document will be refined, and new content will be included during the project’s lifecycle.

Table of contents

1 Introduction 7
2 Voice-enabled technologies overview 7
2.1 Machine Learning and Deep Learning 8
2.2 Voice-enabled technologies: Speech recognition and voice recognition. 8
3 Data protection legislation in Europe 9
3.1 European Convention on Human Rights 9
3.2 Council of Europe Convention No. 108 on data protection of 1981 10
3.3 General Data Protection Regulation (EU) 10
3.4 Directives (EU) affecting data protection 11
3.4.1 Directive (EU) 2016/680 11
3.4.2 Directive (EU) 2002/58 11
4 GDPR provisions 11
4.1 Objectives, material and territorial scope 11
4.2 Relevant concepts 13
4.2.1 Personal data 13
4.2.2 Special categories of personal data 15
4.2.3 Data processing 18
4.2.4 Profiling 18
4.2.5 Pseudonymisation and anonymization 19
4.2.6 Data Controller 21
4.2.7 Processor 21
4.2.8 Recipient 21
4.2.9 Third party 21
4.2.10 Consent 22
4.2.11 Supervisory authority 23
4.3 Principles 23
4.3.1 Principles relating to the processing of personal data 23
4.3.2 Processing of special categories of personal data 32
4.4 Data subject rights 34
4.4.1 Right to be informed 34
4.4.2 *Automated individual decision-making, including profiling and how it is connected to the right to information.* 37
4.4.3 Right of access 39
4.4.4 Right to rectification 40

4.4.5 Right to erasure (right to be forgotten) 40
4.4.6 Right to restriction of processing 42
4.4.7 Notification obligation regarding rectification or erasure of personal data or restriction of processing 42
4.4.8 Right to portability 42
4.4.9 Right to object 42
4.5 Data controller and data processor responsibilities 43
4.5.1 Data controller responsibilities 43
4.5.2 Data processor responsibilities 44
4.5.3 Data Protection Officer (DPO) 45
4.6 International transfers of personal data 46
5 Recommendations on how to implement a Privacy Impact Assessment (PIA) 48
5.1 Concept of the Privacy Impact Assessment 48
5.2 Analyse if the personal data processing operation is subject to the Privacy Impact Assessment 49
5.3 Description of the envisaged processing operation 51
5.4 Assess the necessity and proportionality of the processing 52
5.5 Risk management 53
5.5.1 Identification of the risks 53
5.5.2 Risk assessment 53
5.5.3 Mitigation measures 53
5.5.4 Action plan and conclusion 54
5.5.5 Supervision of the implementation of the actions 55
6 Information Security and cybersecurity 55
6.1 Information Security 55
6.2 Information security management 56
6.3 Audits 56
6.4 Personal data breach 57
6.5 Voice-enabled technologies and cybersecurity threats 57
7 Other legal issues affecting voice-enabled technologies 59
7.1 Intellectual property 59
7.2 Trade secrets 60
7.3 Ethics 61
8 Conclusions 62
9 Appendices 62

9.1 Appendix 1: Categories of personal information and examples list 62
9.2 Appendix 2: Recommendations to prepare an information sheet 66
9.3 Appendix 3: Threats and measures example list 68
10 Bibliography 71

1 Introduction

Current voice-enabled systems rely on the machine learning paradigm called deep learning which has led to major improvements in speech-to-text, spoken language understanding, and dialog management. These technologies typically operate as cloud-based services: the user's speech is sent to the cloud, where it is automatically transcribed and processed, and the system's reply is sent back to the user's device.

Speech-to-text, spoken language understanding, and dialog management require massive in-domain training data in the order of thousands of hours of speech or tens of thousands of dialogs from tens of thousands of users in every language to reach state-of-the-art performance and meet users' expectations.

Storing the users' speech raises serious privacy concerns. Each spoken message reveals information about the user's personality that he/she may not want to be centralised in a single place by the company providing the technology. On the other hand, during the course of COMPRISE, personal data (see Section 4.2.1) in the form of speech and text will be collected from the research participants and processed for research and demonstration purposes.

Deliverable D5.1 entitled "Data protection and GDPR requirements" aims to provide a comprehensible summary of the main aspects regarding the General Data Protection Regulation (GDPR) while considering and analysing the particularities of voice interaction technologies within COMPRISE and by identifying barriers and requirements to comply with such regulation.

D5.1 is the first deliverable of WP5 "Cloud-based platform for multilingual voice interaction" which will bring together the results of WP2, WP3 and WP4 to develop a cloud-based platform to collect neutral users' speech and text data and curate them. WP5 aims to provide access to the user-independent speech-to-text, spoken language understanding, and dialog management models trained on these data as a service via a web service API.

D5.1 will provide an overview on the main aspects of the GDPR and also recommendations for its compliance. This is essential considering the importance of the new regulation and the penalties that an organization that does not comply with the Regulation may face — there are two tiers of fines: the first is up to €10 million or 2% of annual global turnover of the previous year, and the second is up to €20 million or 4% of annual turnover of the previous year.

The rest of this document is structured as follows. Section 2 provides an overview of voice-enabled technologies. Section 3 presents the data protection legislation in Europe. Section 4 highlights the GDPR provisions. Section 5 elucidates the recommendations on how to implement a Privacy Impact Assessment. Section 6 describes the problem of information security and cybersecurity. Section 7 describes other legal issues affecting voice-enabled technologies. Finally, Section 8 presents a conclusion.

2 Voice-enabled technologies overview

In this section, an overview of the technologies integrated in voice-enabled systems is provided with the aim of better understanding how they work.

2.1 Machine Learning and Deep Learning

Currently, voice-enabled technologies rely on a machine learning paradigm called deep learning which has led to major improvements in speech-to-text, spoken language understanding, and dialog management.

Machine learning can be understood as a technique that helps Artificial Intelligence (AI) by training algorithms so that they can learn how to make decisions or predictions based on large amounts of inputs (data), acknowledging as much as possible about the processed information. In short, machine learning can be seen as a subset of AI.

Deep learning is a specific implementation of machine learning using artificial neural networks.

Just like humans, the algorithms used in deep learning try to compare new information to known items before making sense of it.[3] For example, without being technical, an algorithm can be taught to distinguish a soccer ball, a rugby ball or a tennis ball based on the known characteristics of each of them.

2.2 Voice-enabled technologies: Speech recognition and voice recognition

Speech recognition (a.k.a. speech-to-text) technologies recognise speech and convert it into readable form or text. It is often divided into three components[4]:

- **Signal level:** Which involves the speech signal from which speech segments and features are extracted.
- **Acoustic level**: In which features from the speech signal are classified into different phonetic classes.
- **Language level**: Where sequences of phonetic classes are combined into words and subsequently into sentences.

The algorithms used at all three levels mentioned above are based on deep learning techniques, which consist of "*learning multiple levels of representation and abstraction that help to make sense of data such as images, sound, and text*".[5] The term "speech recognition" is sometimes used in a broad sense as also encompassing natural language understanding and dialog, i.e., "*the ability of a machine or program to receive and interpret dictation, or to understand and carry out spoken commands*".[6] Natural language understanding and dialog algorithms are also often based on deep learning nowadays.

[3] Garbade, M. J. (2018, September 14). Clearing the Confusion: AI vs Machine Learning vs Deep Learning Differences. Retrieved, from https://towardsdatascience.com/clearing-the-confusion-ai-vs-machine-learning-vs-deep-learning-differences-fce69b21d5eb

[4] Lorica, B. (2016, July 14). Commercial speech recognition systems in the age of big data and deep learning. Retrieved from https://www.oreilly.com/ideas/commercial-speech-recognition-systems-in-the-age-of-big-data-and-deep-learning

[5] Chhavi Rana, R. (2015). A Review: Speech Recognition with Deep Learning Methods [Abstract]. International Journal of Computer Science and Mobile Computing,4(5), 1017. Retrieved from https://www.ijcsmc.com/docs/papers/May2015/V4I5201599a61.pdf.

[6] Chhavi Rana, R. (2015). A Review: Speech Recognition with Deep Learning Methods [Abstract]. International Journal of Computer Science and Mobile Computing,4(5), 1017. Retrieved from https://www.ijcsmc.com/docs/papers/May2015/V4I5201599a61.pdf

Voice recognition (a.k.a. speaker recognition) refers to a different range of technologies. Namely, speech recognition aims to "*extract the words that are spoken, voice recognition intends to identify the voice that is speaking*".[7] Voice recognition works by scanning the aspects of speech that differ from one individual to another, also known as voice physiology, i.e., accent, speaking style, voice pitch.[8]

While speech recognition programs are commonly present in mobile phones in the form of personal assistants, e.g., Siri, the triumph of voice recognition in the field of biometry has opened the possibility to identify or authenticate people using voiceprints.[9]

The large volume of data that is needed to train speech recognition and voice recognition algorithms can be considered as big data. This has important implications in relation to the enforcement of the GDPR among other EU laws.

According to Tech America Foundation, "*big data is a term that describes large volumes of high velocity, complex and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information*".[10]

In this respect, these large volumes of data may contain natural persons' personal and/or sensitive data, which fall under the scope of the GDPR and should be treated accordingly.

3 Data protection legislation in Europe

In this section, an overview on the main regulations in Europe that affect personal data protection is provided.

3.1 European Convention on Human Rights

The European Convention on Human Rights (ECHR), drafted in 1950, is an international convention that aims to protect human rights and political freedoms in Europe. All Council of Europe member states are party to the Convention.[11]

Art. 8 of the ECHR, which is tightly related to privacy and data protection, guarantees the right to respect for private land family life and home and correspondence.

Other situations involving data protection issues are examined in the ECHR, such as interception of communications, surveillance and protection against personal data storage by public authorities.[12]

[7] Coin, E. (2015, December 07). Introduction to Synthetic Agents: Speech Recognition - Part 1 - DZone Big Data. Retrieved from https://dzone.com/articles/introduction-to-synthetic-agents-speech-recognitio
[8] Kikel, C. (2019, April 6). Difference Between Voice Recognition and Speech Recognition. Retrieved from https://www.totalvoicetech.com/difference-between-voice-recognition-and-speech-recognition
[9] https://www.aware.com/voice-authentication/
[10] Gandomi, A., & Haider, M. (2014, December 03). Beyond the hype: Big data concepts, methods, and analytics. Retrieved from https://www.sciencedirect.com/science/article/pii/S0268401214001066
[11] https://en.wikipedia.org/wiki/European_Convention_on_Human_Rights
[12] Giakomopulus.C, Buttarelli. G, O'Flaherty M. (2018, April). Handbook on European Data Protection Law. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/5b0cfa83-63f3-11e8-ab9c-01aa75ed71a1

3.2 Council of Europe Convention No. 108 on data protection of 1981

Convention 108 was the first legal instrument created regarding personal data protection purposes. It applies to all data processing carried out by both private and public bodies and protects individuals against the abuses that may be committed while processing personal data. The Convention also seeks to regulate the transborder flow of personal data.

All the European Union (EU) Member States have ratified Convention 108, and, to date, 51 countries are part of the Convention, as the accession is open to non-contracting parties.

The Convention has been updated over the years to reinforce its potential as a universal instrument on data protection law.[13]

3.3 General Data Protection Regulation (EU)

The GDPR on the protection of natural persons with regard to the processing of personal data and on the free movement of such data was published on May 4, 2016, and came into force on May 25, 2018, repealing the previous 95/46 Directive.

The GDPR is a European Regulation, meaning that it is a binding legislative act that must be applied in its entirety across the EU. The previous central data protection regulation in the EU regulation was a Directive, a legislative act that sets out a goal that all EU countries must achieve but, in contrast with a European Regulation, the individual countries devise their own laws on how to reach these goals.[14]

One of the main objectives of the new GDPR is to harmonise and unify the applicable criteria in the effective guarantee of data protection and privacy rights.

With the publication of the GDPR, the EU is trying to find an equilibrium between the protection of personal data and the free movement of such data as well as to adapt its regulation to an increasingly technological society.

The structure of the GDPR is as follows:

- Chapter I: General Provisions
- Chapter II: Principles
- Chapter III: Rights of the data subjects
- Chapter IV: Controller and processor
- Chapter V: Transfer of personal data to third countries or international organizations
- Chapter VI: Independent supervisory authorities
- Chapter VII: Cooperation and consistency
- Chapter VIII: Remedies, liability and penalties
- Chapter IX: Provision relating to specific processing situations
- Chapter X: Delegated acts and implementing acts.

[13] Giakomopulus.C, Buttarelli. G, O'Flaherty M. (2018, April). Handbook on European Data Protection Law. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/5b0cfa83-63f3-11e8-ab9c-01aa75ed71a1

[14] https://europa.eu/european-union/eu-law/legal-acts_en

3.4 Directives (EU) affecting data protection

3.4.1 Directive (EU) 2016/680

"Directive 2016/680 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data" was published on April 27, 2017, with the purpose of complementing the GDPR, with a deadline for transposition of two years.

The Directive aims to regulate the processing of personal data by competent authorities for the purposes indicated in the previous paragraph, preserving data protection rights of natural persons and guaranteeing at the same time that the exchange of personal data by competent authorities of the EU is not restricted as a consequence of data protection issues.

3.4.2 Directive (EU) 2002/58

"Directive 2002/58/EC on Privacy and Electronic Communications", otherwise known as E-Privacy Directive, aims at ensuring data protection and privacy in the electronic communications sector in the EU, focusing specifically on the confidentiality of communications and the rules regarding tracking and monitoring.[15]

On January 10, 2017, a draft proposal for a new EU Regulation (EU E-Privacy Regulation) was published to repeal Directive 2002/58 and to adapt the new regulation to the digital age and to the GDPR.

4 GDPR provisions

This section provides a comprehensible summary of the GDPR, including an analysis on the main aspects of the Regulation and recommendations for its compliance. An analysis of the particularities of the processing of personal data involving integrated voice-enabled systems, as well as the processing activities that will be carried out within COMPRISE are also included.

4.1 Objectives, material and territorial scope

Art.1 of the GDPR establishes the objectives of the Regulation:

- Lay down rules relating to the protection of natural persons with regard to the processing of personal data and rules relating to the free movement of personal data.
- Protecting fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data.
- The free movement of personal data within the Union shall be neither restricted nor prohibited for reasons connected with the protection of natural persons with regard to the processing of personal data

The **material scope (Article 2)** of the GDPR includes the processing of personal data wholly or partly by automated means and the processing, by other means, of personal data which form part of a filing system or are intended to form part of a filing system.

[15] https://edps.europa.eu/data-protection/our-work/subjects/eprivacy-directive_en

However, there are some exceptions in which the Regulation does not apply to the processing of personal data:

- In the course of activities that fall outside the EU law.
- By Member States when carrying activities related to foreign policy and common security within the scope of Chapter 2 of Title V of the Treaty on European Union.
- By a natural person in the course of a purely personal or household activity. According to recital 18 of the GDPR, personal or household activities could include correspondence, and personal activity in the social media, for example.
- By competent authorities in the exercise of their duties.

Regarding the **territorial scope,** the GDPR "*applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of whether the processing takes place in the Union or not*" (Art.3 GDPR).

The GDPR also applies to the processing of personal data of data subjects who are in the Union by a controller or processor not established in the Union, where the processing activities are related to:

- The offering of goods or services, irrespective of whether a payment of the data subject is required, to such data subjects in the Union. Regarding this, Recital 23 indicates that "*factors such as the use of a language or a currency generally used in one or more Member States with the possibility of ordering goods and services in that other language, or the mentioning of customers or users who are in the Union, may make it apparent that the controller envisages offering goods or services to data subjects in the Union*".
- The monitoring of data subjects' behavior as far as their behavior takes place within the Union. Recital 24 of the GDPR states: "*In order to determine whether a processing activity can be considered to monitor the behavior of data subjects, it should be ascertained whether natural persons are tracked on the internet including potential subsequent use of personal data processing techniques which consist of profiling a natural person, particularly in order to take decisions concerning her or him or for analyzing or predicting her or his personal preferences, behaviors and attitudes*".

Finally, the Regulation can be applied to the processing of personal data by a controller not established in the Union, but in a place where Member State law applies by public international law (Art 3.3 GDPR).

COMPRISE
In the case of COMPRISE, all partners involved in the project are established in EU, and the research and demonstration activities that they will carry out within the project framework will take place within the EU as well as the processing of personal data performed in the context of these activities. On the other hand, the data subjects whose personal data will be processed during the demonstrations will be located in EU countries. This means that the GDPR will apply to the processing of personal data carried out within the project's activities framework, according to the material and territorial scope criteria set in the Regulation.

4.2 Relevant concepts

In this section, the main concepts tied to the GDPR management are explained.

4.2.1 Personal data

The GDPR defines personal data as "*any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person*" (Art.4.1 of the GDPR)

The definition provided in the GDPR reflects the intention of the European lawmaker for a wide notion of "personal data". The term "**any information**" also calls for a broad interpretation[16].

According to the definition provided in the GDPR, personal data is any sort of information related to an identified or identifiable individual. It is considered, in general terms, that information is related to an individual when it is about that individual; however, is not always easy to establish this relationship.[17]

One person can be considered **identified** when, within a group of persons, he or she can be distinguished from all the other members of the group. On the other hand, an **identifiable** person is a person who could be identified but has not been identified yet.[18] Recital 26 provides that "*to determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments*". According to this, the mere hypothetical possibility of identifying a natural person is not enough to consider that person as identifiable. As stated in the recital, several factors should be considered in each particular case.

Data can identify a person directly, e.g., the name of the person, but often the attributes are not unique (e.g., a large number of people may have the same first name) and must be combined with others in order to make it possible to identify a person. Other data, however, can make a person **indirectly identifiable**, e.g., telephone number, social security number, vehicle registration number.[19] In most cases, identification of a person

[16] The Art 29 Working Party (2007, June 20th). Opinion 4/2007 on the concept of personal data. Retrieved from https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf

[17] The Art 29 Working Party (2007, June 20th). Opinion 4/2007 on the concept of personal data. Retrieved from https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf

[18] The Art 29 Working Party (2007, June 20th). Opinion 4/2007 on the concept of personal data. Retrieved from https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf

[19] Giakomopulus.C, Buttarelli. G, O'Flaherty M. (2018, April). Handbook on European Data Protection Law. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/5b0cfa83-63f3-11e8-ab9c-01aa75ed71a1

is achieved through particular pieces of information (identifiers) which hold a particular relationship with a concrete natural person, i.e., signs of appearance like the hair colour, clothing, the quality of the person like the profession, function, etc. In the end, the extent to which specific identifiers are sufficient for identifying a person will depend on the context of the particular situation.[20]

From the nature point of view, any statement about a person could be personal data. The concept covers both objective information, e.g., health, physical attributes, etc., and subjective information, e.g., opinions, beliefs, assessments, etc.[21] Personal data can include any information related to the private life of the natural person as well as any information regarding whatever types of activity is undertaken by the individual, e.g., concerning working relations or the economic or social behaviour of the individual, regardless the position or capacity of these persons.[22]

Finally, it should be stressed that personal information includes information available in whatever form: alphabetical, numerical, graphical, photographic or acoustic.

Voice-enabled technologies
Voice-based technologies undoubtedly require the collection and processing of large amounts of data (including personal data) in order to achieve the execution of specific tasks demanded by users in the digital economy. Voice-enabled systems will collect and process mainly speech signals (voice) and the text transcription of these signals. Voice can be considered as an identifier. Indeed, the signal, its content, and all information we can derive from it could be considered as personal data if it is possible to identify a person by his/her voice. Personal information related to voice characteristics could include: • The user's identity • More general traits, e.g., gender, age, ethnic origin, nativeness, etc., and states, e.g., health condition, intoxication, sincerity, etc. Personal information related to the spoken message includes: • Words or utterances that explicitly mention the user's identity, general traits associated to the speaker's background, e.g., gender, age, ethnic origin, nativeness, etc., the information stated related to the user's health status, or otherwise critical information, e.g., credit card number, personal phone number, etc. User preferences not revealed by the dialogue outcome, e.g., in an e-commerce scenario, preferences revealed by asking the system about another product or a general category of products before settling on a specific

[20] The Art 29 Working Party (2007, June 20th). Opinion 4/2007 on the concept of personal data. Retrieved from https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf

[21] The Art 29 Working Party (2007, June 20th). Opinion 4/2007 on the concept of personal data. Retrieved from https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf

[22] The Art 29 Working Party (2007, June 20th). Opinion 4/2007 on the concept of personal data. Retrieved from https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf

product, health-related information revealed by the choice of products bought by the user.

COMPRISE

One of the main goals of COMPRISE is to design a privacy-driven voice interaction technology by:

- Designing privacy-driven speech transformations to delete individual voice characteristics from the users' speech, resulting in "neutral" speech data;
- Designing privacy-driven text transformations to delete private information relating to the spoken message, resulting in "neutral" text data;

To achieve this, it is imperative for the COMPRISE system to be able to identify personal information from what will be collected and processed. A list which includes different categories of personal information is attached in Annex 1.[23] It contains examples of words that could be identified as personal information and short sentences that can trigger the inclusion of personal information; for example: My account number is 0787698669.

The words included in the list can be considered as personal data only if they can be related to an identified or identifiable person.

One single word cannot necessarily be linked to an identified or identifiable person, but different pieces of information collected together can lead to the identification of a particular person, and in these cases, could be considered as personal data.

4.2.2 Special categories of personal data

Within the framework of the GDPR, the following information is considered as sensitive data (Art. 9 of the GDPR):

- Racial or ethnic origin
- Political opinions
- Religious or philosophical beliefs
- Trade union membership
- The processing of genetic data, biometric data to uniquely identify a natural person
- Data concerning health
- Data concerning a natural person's sex life or sexual orientation

Because of their nature, these data need enhanced protection as their processing may pose a higher risk for the data subject. According to the GDPR, the processing of special categories of personal data should be prohibited except under certain circumstances that will be analyzed later[24] (see Section 4.3.2).

[23] https://iapp.org/media/pdf/resource_center/Categories-of-personal-information.pdf
Based on infographic produced by Enterprivacy Consulting Group, offers an overview of types of data relating to an individual's public or private life.

[24] Giakomopulus.C, Buttarelli. G, O'Flaherty M. (2018, April). Handbook on European Data Protection Law. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/5b0cfa83-63f3-11e8-ab9c-01aa75ed71a1

Voice-enabled technologies
When a user is interacting with a voice-enabled system, sensitive data such as biometric data (see Section 4.2.2.1), health data (could be revealed when using an e-health app), sexual orientation (could be exposed depending on the products or services which a person is accessing online), political opinions, etc., may be collected. In many occasions, it is challenging to identify which data may be considered as sensitive information or may be used to make predictions regarding sensitive personal data, e.g., if a voice-enabled system is used to buy medicines, this could be considered as health-related information.
COMPRISE
In the specific case of the E-Health demonstrator developed by Tilde, it is possible that sensitive data will be collected and processed, as some of the research participants may be patients in partnering hospitals; and data concerning health is very sensitive. It is also possible that other sensitive data could be disclosed by research participants when interacting with the voice recognition application in other demonstrators, e.g., a product bought through the E-Commerce app could disclose sensitive information.

4.2.2.1 Biometric data

The processing of biometric data falls under the scope of sensitive data according to the GDPR (for the first time biometrics is included as a special category of personal data). The GDPR defines biometric data as "'*personal data resulting from specific technical processing relating to the physical, physiological or behavioral characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data*;" (Art.4.14). Typical examples of biometric data include fingerprints, retinal patterns, voices, but also hand geometry, vein patterns or behavioral characteristics.[25]

Voice-enabled technologies
Regarding the processing of biometric data involving voice-enabled technologies, a distinction should be made between speech recognition and voice recognition (see Section 2.2); the latter indicates biometric identification. Generally, speech-enabled devices in the market are not designed for the purpose of uniquely identifying a person through the biometric characteristics of his or her voice. The majority of these products use speech as a useful interface,[26] but it is expected that unique voice recognition as a consumer tool will spread over the coming years.
COMPRISE

[25] The Art 29 Working Party (2003, August 3rd). Working document on biometrics. Retrieved from https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2003/wp80_en.pdf

[26] Gray.S. (2016, Apriñ). Experts on the GDPR #3: What is personal data under the GDPR? Always On: Privacy Implications of Microphone Enabled Devices. Retrieved from https://fpf.org/wp-content/uploads/2016/04/FPF_Always_On_WP.pdf

To train the COMPRISE system, the user's speech will be transformed into "neutral" data before it is sent to the cloud. The transformation will result in a new "neutral" speech signal from which sensitive attributes related to the user's identity and the traits and states of his/her voice have been removed.

4.2.2.2 Health data

The GDPR defines data concerning health as personal data "*related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status*" (Art.4.15 GDPR). Health data is considered as sensitive data according to Art. 9 of the GDPR.

A wide range of personal data may fall into the category of health data. Article 29 of the Working Party[27] (currently the European Data Protection Board) identifies the main situations in which the data processed is considered as health data:

- Medical data, this is data about the physical or mental health status of the data subject generated in a medical context, e.g., data related to diagnosis or treatments, diseases, disabilities, etc.
- Other data from which is possible to conclude the health status of a data subject, including information such as the fact that someone uses glasses, alcoholic or smoking habits, allergies disclosed to a private entity, being a member of a patients support group, e.g., cancer, etc.[28]
- Information related to the purchase of medical products, devices, or to the request of health-related services when health status may be inferred from the data.

Voice-enabled technologies
The use of voice-enabled technologies in e-health applications, that will require the processing of sensitive health data, is expected to increase considerably over the coming years. Taking into account the wide range of personal data that may fall into the category of health data, there could be many other situations in which health data are collected and processed when a user is interacting with a voice-enabled system, for example if the person is using a voice-enabled app to buy medical products such as a pregnancy test, if some habits are revealed, etc.
COMPRISE
Concerning the e-health demonstrator, it is probable that health data will be collected and processed. In some of the other demonstrators that will be carried out, it is also

[27] The Art 29 Working. ANNEX - health data in apps and devices. Retrieved from https://ec.europa.eu/justice/article-29/documentation/other-document/files/2015/20150205_letter_art29wp_ec_health_data_after_plenary_annex_en.pdf

[28] 2015, February 9th) Article 29 Working Party Clarifies Scope of Health Data Processed by Lifestyle and Wellbeing Apps. Retrieved from https://www.huntonprivacyblog.com/2015/02/09/article-29-working-party-clarifies-scope-health-data-processed-lifestyle-wellbeing-apps/

possible that some health data are processed, e.g., it could be inferred that someone is allergic to some products from the data processed when using the cooking app.

4.2.3 Data processing

The GDPR defines data processing as "*any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction (Art.4.2 of the GDPR)*".

Processing of personal data covers a wide range of operations. Note that the mere storage of personal data is considered as data processing, even when such data will not be used in any other way.

COMPRISE
Within COMPRISE several processing operations will be performed on personal data such as recoding, storage, transformation, etc.

4.2.4 Profiling

According to the GDPR *"profiling means automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movement*s" (Art.4.4 of the GDPR).

Thanks to technological advances (data analytics, AI, machine learning) and the widespread availability of data on the internet and from Internet of Things (IoT) devices, it is possible to determine aspects of individuals' personality, preferences, interests or behaviours by finding correlations and creating links.[29]

Profiling is used in an increasing number of sectors as it has many commercial applications, e.g., aligning services or products with customer preferences and needs. However, this can pose a significant risk on individuals' rights and freedoms as this can perpetuate stereotypes and social segregation. Freedom of choice can also be undermined as profiling can lock a person in a specific category and restrict him/her to his/her suggested preferences.[30]

[29] The Art 29 Working Party (2017, October 3rd). Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053

[30] The Art 29 Working Party (2017, October 3rd). Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053

Voice-enabled technologies
Apps based on voice-enabled technology may apply profiling techniques in order to give personalised responses, e.g., calling and messaging, music, voice shopping, etc. The system recognises the user's voice (voice profile) and gives personalised answers (profile based on preferences and past interactions) to improve the quality of the service.
COMPRISE
One of the objectives of the project is to design a personalised machine learning methodology to adapt user-independent speech and dialogue models (that will be trained in the cloud) to each user by running additional computations **on the user's device** without disclosing his/her raw data.

4.2.5 Pseudonymisation and anonymization

The GDPR defines pseudonymization as "*processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person*" (Art.4.5 of the GDPR).

Pseudonymization is not an anonymization method, but a security measure that reduces risks for the data subjects. One of the most common pseudonymization methods consists of using data encryption techniques. After encryption, data is encoded in such a way that accessing the original data is possible only by using a decryption key.[31] However, for those entitled to use the decryption key, the re-identification is easy (it is also possible that someone unauthorised could use the decryption key). By using pseudonymization, it is possible to backtrack individuals, consequently individuals whose data has been pseudonymized are considered as indirectly identifiable; and pseudonymization is not exempt from the scope of the GDPR.[32]

On the other hand, according to Recital 26 of the GDPR "*the principles of data protection should therefore not apply to **anonymous information**, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not, therefore, concern the processing of such anonymous information, including for statistical or research purposes*."

Data anonymization consists in identifying and eliminating elements from a set of personal data so that the data subject is no longer identifiable.

[31] Giakomopulus.C, Buttarelli. G, O'Flaherty M. (2018, April). Handbook on European Data Protection Law. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/5b0cfa83-63f3-11e8-ab9c-01aa75ed71a1

[32] The Art 29 Working Party (2007, June 20th). Opinion 4/2007 on the concept of personal data. Retrieved from https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf

Anonymization can be an excellent strategy to mitigate risks. However, the creation of a genuinely anonymous dataset from a rich set of personal data is not easy and, in many cases, anonymization techniques are not completely effective.[33]

According to Recital 26, it will be considered that personal data has been anonymized when it is not possible to identify the data subject by using "*all the means likely reasonably to be used*" by either the controller or a third party. Consequently, the outcome of anonymization as a technique applied to personal data should be, in the current state of technology, as permanent as erasure (irreversible de-identification[34]). Every process of anonymization should be decided on a case-by-case basis.

When intending to anonymize personal data collected in a research project, it should be taken into account that the data collected will only be considered anonymized if the anonymization happens at the point and time at which the data is collected from the research subject. If the anonymization happens later, the raw data is considered as personal data and should be protected accordingly until the point at which they are anonymized or deleted[35].

COMPRISE
One of the main objectives of COMPRISE is to introduce a privacy-driven transformation to delete private information from the users' speech and the corresponding text data obtained by speech-to-text. As it is not possible to guarantee that personal data will be entirely anonymized, the GDPR will apply to personal data processed within the project activities. Additionally, pseudonymized data will be stored. Consequently, organisational and security measures will be taken in order to be compliant with this regulation. This approach is aligned with the following recommendation included in the "Ethics and Data Protection" document of the European Commission: "*As far as your research proposal is concerned, if there is a significant prospect of re-identification of persons whose data have been collected, the information should be treated as personal data. It is difficult to assess the risk of re-identification with absolute certainty and you should always err on the side of caution. A growing body of case studies and research publications in which individuals are identified from 'anonymous' datasets has demonstrated the fundamental constraints to anonymization as a technique to protect the privacy of individuals.*"[36]

[33] The Art 29 Working Party (2014, April 10th). Opinion 05/2014 on Anonymisation Techniques. Retrieved from https://www.pdpjournals.com/docs/88197.pdf

[34] The Art 29 Working Party (2014, April 10th). Opinion 05/2014 on Anonymisation Techniques. Retrieved from https://www.pdpjournals.com/docs/88197.pdf

[35] (2018, November 14th). Ethics and data protection. Retrieved from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/ethics/h2020_hi_ethics-data-protection_en.pdf

[36] (2018, November 14th). Ethics and data protection. Retrieved from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/ethics/h2020_hi_ethics-data-protection_en.pdf

4.2.6 Data Controller

According to the GDPR, the "controller" is "*the natural or legal person, public authority, agency or another body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law*" (Art. 4.7 of the GDPR).

The controller is responsible for compliance with the data protection rules and must exercise control over the processing, including legal liability.[37]

According to Art.26 of the GDPR "*where two or more controllers jointly determine the purposes and means of processing, they shall be joint controllers. They shall in a transparent manner determine their respective responsibilities for compliance with the obligations under this Regulation, in particular as regards the exercising of the rights of the data subject and their respective duties to provide the information."*

COMPRISE
Regarding the personal data processing activities that will be carried out within the research and demonstrations activities of the project, every consortium partner will be considered as a data controller (as legal persons) for each of the processing in which they determine purpose and means.

4.2.7 Processor

GDPR defines the "processor" as "*a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller*" (Art. 4.8 of the GDPR).

The existence of the processor depends on whether the controller decides to delegate all or part of the processing activities to an external organisation (a separate legal entity with respect to the controller), which will process the personal data on its behalf.[38] Examples of data processors can be found in companies that offer outsourcing services that require data processing such IT solutions, call centre activities, legal services, etc.

4.2.8 Recipient

The GDPR defines the "recipient" as "*a natural or legal person, public authority, agency or another body, to which the personal data are disclosed, whether a third party or not*" (Art. 4.9 of the GDPR).

4.2.9 Third party

The GDPR defines a "third party" as "*a natural or legal person, public authority, agency or body other than the data subject, controller, processor and persons who, under the*

[37] Giakomopulus.C, Buttarelli. G, O'Flaherty M. (2018, April). Handbook on European Data Protection Law. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/5b0cfa83-63f3-11e8-ab9c-01aa75ed71a1

[38] The Art 29 Working Party (2010, February 14th). Opinion 1/2010 on the concepts of "controller" and "processor". Retrieved from https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2010/wp169_en.pdf

direct authority of the controller or processor, are authorised to process personal data" (Art. 4.10 of the GDPR).

4.2.10 Consent

The GDPR defines "*consent of the data subject" as "any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her*;" (Art.4.11 of the GDPR).

The consent of the data subject is one of the legal bases for a lawful processing of personal data. According to this definition, five elements are needed for a valid consent:

- **Freely given:** This means that the data subject should have a real choice when he or she gives consent. If the data subject has no real choice (because he/she feels compelled, cannot refuse it, it will endure negative consequences or any other reason), the consent will not be considered as valid.[39]

  It is unlikely that consent is freely given when an imbalance of power situation exists, e.g., a public authority asking for consent to the citizens, an employer asking for consent to the employees, but it is not always the case.[40]

  It is also presumed that consent has not been given freely when the data subject is obliged to agree with the processing of personal data that is additional to what is strictly necessary for the performance of a contract or services.[41]

- **Specific:** The consent should be given for one or more specific purposes. If there are several purposes for the same personal data processing, the data subject should give consent for each of the purposes.[42]

- **Informed:** The consent should be informed. This means that the data subject should be informed about the processing of his/her data prior to obtaining his/her consent**.**

- **Unambiguous indication of the data subject's wishes**: The consent requires a statement of the data subject or a clear affirmative act (an active motion or declaration). Practices such as pre-ticked boxes or opt-out boxes are not allowed by the GDPR.

COMPRISE
Personal data collection and processing performed within COMPRISE activities will be based on valid consent. As will be later explained in this document, data subjects

[39] The Art 29 Working Party (2017, November 28th). Guidelines on consent under Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=51030
[40] The Art 29 Working Party (2017, November 28th). Guidelines on consent under Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=51030
[41] The Art 29 Working Party (2017, November 28th). Guidelines on consent under Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=51030
[42] The Art 29 Working Party (2017, November 28th). Guidelines on consent under Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=51030

will be duly informed prior to obtaining their consent and said consent would be asked for each of the specific purposes of the processing.

4.2.11 Supervisory authority

The GDPR defines the "supervisory authority" as "*an independent public authority which is established by a Member State*" (Art.4.21 of the GDPR).

The "supervisory authority concerned" is defined as the "*supervisory authority which is concerned by the processing of personal data because:*

- *the controller or processor is established on the territory of the Member State of that supervisory authority;*
- *data subjects residing in the Member State of that supervisory authority are substantially affected or likely to be substantially affected by the processing; or*
- *a complaint has been lodged with that supervisory authority*;" (Art.4.22 of the GDPR)

The powers of the supervisory authority are listed in Art. 58 of the GDPR. The most important powers are conducting investigatory audits; obtaining, access to premises and data; issuing warnings and reprimands, and imposing fines; ordering controllers and processors to comply with the GDPR and data subjects' rights; banning processing and trans-border data flows outside the EU; approving standard contractual clauses and binding corporate rules.[43]

COMPRISE
Partners that will process personal data during the project should cooperate with supervisory authorities, facilitating all the information required during an investigatory audit, always that the information requirement is according to the law. Partners should also comply with the instructions given by the supervisory authority without undue delay.

4.3 Principles

4.3.1 Principles relating to the processing of personal data

The principles governing the processing of personal data are set in Chapter II of the GDPR

4.3.1.1Lawfulness, fairness and transparency

"*Personal data shall be processed lawfully, fairly and in a transparent manner in relation to the data subject*" according to the GDPR (Art. 5.1.a).

[43] https://www.twobirds.com/~/media/pdfs/gdpr-pdfs/62--guide-to-the-gdpr--competence-tasks-and-powers.pdf?la=en

- **Lawfulness:** Personal data should be processed lawfully. This means that the processing should be based on the valid consent of the data subject or on other legitimate grounds provided by the Regulation.[44]

- **Fairness of processing:** This principle governs the relationship between the data subject and the controller. The data controller should inform and be able to demonstrate to the data subjects lawful processing of their personal data that is compliant with the GDPR. Additionally, data subjects should be aware of the potential risks associated with the processing of their personal data.

  Furthermore, the fairness principle requires that the processing of personal information be conducted with respect for the data subject's interest and used according to data subject's reasonable expectations. In the end, this principle seeks a fair treatment of the data subject when his/her personal data are processed, meaning that discriminatory and arbitrary treatment should be prevented.[45]

- **Transparency:** Data subjects should be informed about how their data are being used. Transparency may refer to the information given to the individual before the processing starts, which should be readily accessible to data subjects during the processing, but also to the information given to data subjects following a request of access to their own data.

In order to **lawfully** process the personal data**,** the processing must comply with one of the legitimate grounds provisioned in Art. 6 of the GDPR:

- The data subject has given specific consent for the processing of his/her personal data for one or more specific purposes.

  Note that the data controller should be able to demonstrate that the data subject has given consent for the processing of his/her personal data.

  The data subject should have the right to withdraw the consent given at any time. Withdrawing the consent should be as easy as giving it, free of charge and without another detriment for the data subject, i.e., the lowering of the service.[46] (Art. 7.3 GDPR).

- When the processing of personal data is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract. Consequently, the provision covers also pre-contractual relationships.

- When the processing of personal data is necessary for compliance with a legal obligation to which the controller is subject. The legal obligations referred to in

[44] Giakomopulus.C, Buttarelli. G, O'Flaherty M. (2018, April). Handbook on European Data Protection Law. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/5b0cfa83-63f3-11e8-ab9c-01aa75ed71a1

[45] The Norwegian Data Protection Authority. (2018, January). Artificial Intelligence and privacy. Retrieved from https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf

[46] The Art 29 Working Party (2017, November 28th). Guidelines on consent under Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=51030

this provision may affect to data controllers acting in the public sector as well as in the private sector[47] and should be laid down by Union law or Member State law to which the controller is subject.

- When the processing of personal data is necessary in order to protect the vital interests of the data subject or another natural person's vital interests.

  According to Recital 46 of the GDPR "*Processing of personal data based on the vital interest of another natural person should in principle take place only where the processing cannot be manifestly based on another legal basis."*

- When the processing of personal data is necessary for the performance of a task carried out in the public interest or the exercise of official authority vested in the controller.

- When the processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.

  If the personal data processing is based on the legitimate interest of the data controller or a third party, an assessment (which is usually complex) should be carried out in each particular case balancing the legitimate interest of the controller or third parties against the interests or fundamental rights or freedoms of the data subject.[48]

  Some examples of legitimate interest of the data controllers are provided in Recital 47 of the GDPR: "*Such legitimate interest could exist for example where there is a relevant and appropriate relationship between the data subject and the controller in situations such as where the data subject is a client or in the service of the controller [….]The processing of personal data strictly necessary for the purposes of preventing fraud also constitutes a legitimate interest of the data controller concerned. The processing of personal data for direct marketing purposes may be regarded as carried out for a legitimate* interest."

  The provision of the GDPR does not refer only to the legitimate interest of the data controller, but also the legitimate interest pursued by third parties or the third parties to whom the data are disclosed. Some examples where this provision would apply could be the publication of data for purposes of transparency and accountability, for historical or scientific research, or when there is a general public interest, amongst others.[49]

---

[47] Giakomopulus.C, Buttarelli. G, O'Flaherty M. (2018, April). Handbook on European Data Protection Law. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/5b0cfa83-63f3-11e8-ab9c-01aa75ed71a1

[48] The Art 29 Working Party (2014, April 9th). Opinion 06/2014 on the notion of legitimate interest of the data controller under article 7 of Directive 95/46/EC. Retrieved from https://fia.org/sites/default/files/uploaded/Excerpts%20-%20Opinion%2006-2014%20on%20the%20notion%20of%20legitimate%20interests%20of%20the%20....pdf.

[49] The Art 29 Working Party (2014, April 9th). Opinion 06/2014 on the notion of legitimate interest of the data controller under article 7 of Directive 95/46/EC. Retrieved from

Machine learning
Machine learning and "fairness principle" Machine learning plays a leading role in AI software applications and consequently also in speech recognition technologies. AI software algorithms are trained with large volumes of data. When an algorithm is trained with personal data, the model's result could be incorrect or discriminatory if biased data (data that renders a biased picture reality) or irrelevant data are used.[50] This would be contrary to the fairness principle stated in the GDPR. An example of a discriminatory output would be the case of a job searching app (which applies machine learning techniques to process data) that excludes job offers results depending on the gender of the user. Regarding voice recognition technologies, there have been cases in which the systems could not properly understand people with accents, or understanding was worse for female voices as a consequence of unbalanced training sets.[51] Models must be trained with correct and relevant data and must learn not to emphasis information related to gender, ethnic origin, beliefs, sexual orientation or other data that could lead to discriminatory treatments.[52] Actions may be taken in different stages of the processing to reduce and even eliminate biases: • Before the processing, by eliminating any sources of unfairness in the data before the algorithm is formulated. • During the processing, by making fairness adjustments as part of the process by which the algorithm is constructed. • After the algorithm is applied, its performance is adjusted to make it fairer.[53]
COMPRISE
Fairness principle

https://fia.org/sites/default/files/uploaded/Excerpts%20-%20Opinion%2006-2014%20on%20the%20notion%20of%20legitimate%20interests%20of%20the%20....pdf.

[50] The Norwegian Data Protection Authority. (2018, January). Artificial Intelligence and privacy. Retrieved from https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf

[51] Tatman.R.. (2016, July 12th). Google's speech recognition has a gender bias. Retrieved from https://makingnoiseandhearingthings.com/2016/07/12/googles-speech-recognition-has-a-gender-bias/

[52] The Norwegian Data Protection Authority. (2018, January). Artificial Intelligence and privacy. Retrieved from https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf

[53] Tiwari.A. (2017, July 4th). Bias and Fairness in Machine Learning. Retrieved from https://www.abhishek-tiwari.com/bias-and-fairness-in-machine-learning/

COMPRISE technology seeks to train the algorithms with anonymised sets of data, meaning that personal data will be removed from the data collected during the interactions of the user with the system as much as possible; this will help to eliminate biases in the data before using it to train the algorithms.

Lawfulness principle: Consent

Concerning the processing of personal data which is based on the consent of the data subject, the controller should be able to demonstrate by any means that the consent has been given by the data subject, e.g., by keeping the original signed consent form, by storing the acceptance log if the consent was provided via a web interface, by recording the consent if it was given orally, etc.

It is also recommended to set protocols to collect the consent and to make the withdrawal of the consent effective.

Finally, it is recommended to check that consent is given through an affirmative action.

4.3.1.2 Collected for specified, explicit and legitimate purposes

"*Personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes* ('purpose limitation'), according to GDPR (Art. 5.1.b)

From this article, we can conclude that this principle should gather four main requirements:

- **Specific:** Personal data should be collected for specific purposes that should be defined before processing has started.[54] Personal data that is not necessary, relevant or adequate for serving these specific purposes should not be collected.[55]

  On the other hand, purposes that are vague or too general will not meet the specification requirements.

  The controller should internally assess the specific purposes and be able to demonstrate that this assessment has been carried out.

- **Explicit:** The specific purpose that has been defined by the controller should be explained in an understandable way for data subjects, third-party processors or data protection authorities.

[54] Giakomopulus.C, Buttarelli. G, O'Flaherty M. (2018, April). Handbook on European Data Protection Law. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/5b0cfa83-63f3-11e8-ab9c-01aa75ed71a1

[55] The Art 29 Working Party (2013, April 2nd). Opinion 03/2013 on purpose limitation. Retrieved from https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

- **Legitimate:** Processing of personal data should be based at least on one of the grounds included in Art. 6 of GDPR (see Section 4.3.1.1).

- **Further processing should be compatible with the original purpose:** This provision sets a limitation on further processing of personal data making a distinction between further processing compatible and further processing incompatible with the original purpose.[56]

  Any processing following the collection of personal data, for the purposes initially specified or for any additional purposes, can be considered as "further processing". In order to assess the compatibility of the further processing with the original purposes the following factors should be considered (According to Art. 6.4 of the GDPR):

  - any link between the original and proposed new purposes;
  - the context in which data have been collected (in particular the relationship between subjects and the controller);
  - the nature of the data (particularly whether they are sensitive data or criminal offence data);
  - the possible consequences of the proposed processing; and the existence of safeguards (including encryption or pseudonymization).

  Any further processing of personal data for a new purpose that is not compatible with the original purpose must have its own particular legal basis (see Section 4.3.1).

  According to Recital 50 of the GDPR: "Further *processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes should be considered to be compatible lawful processing operations*". These provisions should not be interpreted as general authorisations for further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes. All relevant factors and circumstances should be taken in account in order to put in place adequate safeguards[57] such as encryption, pseudonymization, anonymization, etc. Besides, Recital 159 of the GDPR provides: "*For the purposes of this Regulation, the processing of personal data for scientific research purposes should be interpreted in a broad manner including, for example, technological development and demonstration, fundamental research, applied research and privately funded research*." This Recital provides an ambiguous wording of "scientific research".

Machine learning
Purpose definition and explanation

[56] The Art 29 Working Party (2013, April 2nd). Opinion 03/2013 on purpose limitation. Retrieved from https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf
[57] The Art 29 Working Party (2013, April 2nd). Opinion 03/2013 on purpose limitation. Retrieved from https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

When developing AI, it might be challenging to define the purpose of the processing of personal data, as it may not be possible to predict what the algorithm will learn.[58]

On the other hand, when personal data is used to train an algorithm, it may be difficult to explain the purpose as in many occasions the trained model cannot be understood by humans.

Development of AI as scientific research

It should be analysed if the development or application of AI may be considered as "scientific research ". The GDPR does not provide a definition of scientific research, but Recital 159 states that the concept of scientific research should be "*interpreted in a broad manner including, for example, technological development and demonstration".* It is possible, consequently, that the development and application of AI could be considered as scientific research.[59] However, every particular case should be analysed by considering if a scientific method is applied and if the activity aims to discover new knowledge.

In the case of models that develop and improve continuously (models that continue to learn when used), it is difficult to differentiate between the development and use stages. Consequently, it will not be clear if the development or application of AI will constitute scientific research in this case.

4.3.1.3 Minimisation

"*Personal data should be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation');"* according to GDPR (Art.5.1.c)

According to this principle, the processing of personal data must be limited to what is necessary to the purposes which are processed (legitimate purposes).[60] Besides, personal data only should be processed when the purpose of the processing cannot be reasonably fulfilled by other means (the information must be relevant for the legitimate purpose).

This principle should be applied to the quantity of the personal data collected, the length of time the data is stored, the extension of the processing as well as the number of data subjects from which personal data are collected.

[58] The Norwegian Data Protection Authority. (2018, January). Artificial Intelligence and privacy. Retrieved from https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf
[59] The Norwegian Data Protection Authority. (2018, January). Artificial Intelligence and privacy. Retrieved from https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf
[60] Giakomopulus.C, Buttarelli. G, O'Flaherty M. (2018, April). Handbook on European Data Protection Law. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/5b0cfa83-63f3-11e8-ab9c-01aa75ed71a1

Machine learning
Machine learning involves the collection of large amounts of data that in many cases remain unused.[61] This may be an issue when personal data are processed as it is contrary to the principle of data minimisation. As indicated before, in many occasions it is not possible to predict how the algorithm will learn, which makes the definition of the purpose difficult (that also may change as the machine learns). Consequently, this makes it unclear which data will be necessary or not to train the algorithm.[62] This challenges the principle of "data minimisation". In addition, the "data minimisation" principles seek to restrict the extent of the intervention in the privacy of the data subject avoiding disproportionate interferences. The data controller should examine the intended area of application of the model and consider how to achieve the objective in a way that is the least invasive for the data subject. The assessment should be continuous and needs to be documented.[63] Besides, the controller may apply technical measures in order to make difficult the identification of the data subject (such as pseudonymization or encryption).[64] ***Innovative tools and methods that may be used to fulfil the minimisation requirement.*** Some tools and methods that can help meet data protection challenges have been developed, but in many cases, they have not been evaluated in practice yet:[65] • **Generative Adversarial Networks:** This method may be used to generate huge volumes of high quality synthetic data. These data may in turn be used to train algorithms that require large volumes of training data. The need for real data that may contain personal information is reduced but not eliminated because Generative Adversarial Networks themselves require training data. • **Federated learning:** This method "*enables mobile phones to collaboratively learn a shared prediction model while keeping all the training data on device, decoupling the ability to do machine learning from the need to store the data in the cloud".*[66] The model is downloaded from the cloud to the device. Then, the model is improved by learning from data on the device.[67]

[61] Information Commissioner's Office. (2017, September). Big data, artificial intelligence, machine learning and data protection. Retrieved from https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf
[62] The Norwegian Data Protection Authority. (2018, January). Artificial Intelligence and privacy. Retrieved from https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf
[63] The Norwegian Data Protection Authority. (2018, January). Artificial Intelligence and privacy. Retrieved from https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf
[64] The Norwegian Data Protection Authority. (2018, January). Artificial Intelligence and privacy. Retrieved from https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf
[65] The Norwegian Data Protection Authority. (2018, January). Artificial Intelligence and privacy. Retrieved from https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf
[66] McMahan B., Ramage.D (2017, April 6th) Federated Learning: Collaborative Machine Learning without Centralized Training Data. Retrieved from https://ai.googleblog.com/2017/04/federated-learning-collaborative.html
[67] McMahan B., Ramage.D (2017, April 6th) Federated Learning: Collaborative Machine Learning without Centralized Training Data. Retrieved from https://ai.googleblog.com/2017/04/federated-learning-collaborative.html

- **Capsule networks**: This is a new variant of neural network which requires less data for learning than what is usually required in deep learning.[68]

4.3.1.4 Accuracy

"*Personal data should be accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay"* ('accuracy'), according to GDPR (Art. 5.1.d).

According to this principle, a data controller that holds personal information should check it and keep it up to date; if inaccurate data is detected, then this should be rectified or erased without any delay.[69]

Machine learning
When data are processed massively, a certain amount of inaccurate personal data may be tolerated when the model is trying to represent general trends. However, it is much more problematic when the personal data are processed with the purpose of profiling individuals,[70] as the use of inaccurate data may lead to a wrong prediction that in some contexts may bring adverse effects to them.

4.3.1.5 Storage limitation

"*Personal data kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject ('storage limitation')*" according to GDPR (Art. 5.1.d).

Recital 39 of the GDPR also provides: "*This requires, in particular, ensuring that the period for which the personal data are stored is limited to a strict minimum. […] In order to ensure that the personal data are not kept longer than necessary, time limits should be established by the controller for erasure or for a periodic review*."

The storage time limitation for personal data only applies if the identification of the data subject is possible. Consequently data could be stored lawfully, once they are no longer

[68] The Norwegian Data Protection Authority. (2018, January). Artificial Intelligence and privacy. Retrieved from https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf

[69] Giakomopulus.C, Buttarelli. G, O'Flaherty M. (2018, April). Handbook on European Data Protection Law. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/5b0cfa83-63f3-11e8-ab9c-01aa75ed71a1

[70] Information Commissioner's Office. (2017, September). Big data, artificial intelligence, machine learning and data protection. Retrieved from https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf

needed, if the data are anonymized (which means that the re-identification is not possible).

Machine learning
Machine learning needs and has the ability to process huge volumes of data, which may encourage data controllers to keep historical data beyond the period required for the original purpose. This is contrary to the "storage limitation" principle.
COMPRISE
If it is not possible to guarantee that personal data will be anonymized entirely. It is recommended to remove personal information once it is not needed anymore. It should be noted that archiving data for public interest, scientific or historical purposes, or for statistical use, may be for longer periods. Providing such data will be used solely for these purposes, and appropriate technical and organisational measures are implemented for the ongoing storage and use of personal data.[71]

4.3.1.6 Data security principle

"Personal data shall be processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('integrity and confidentiality'"), according to GDPR (Art. 5.1.d).

The application of this principle will be analysed in Section 6 of this deliverable.

4.3.2 Processing of special categories of personal data

According to the GDPR, the processing of special categories of personal data (see Section 4.2.2.) should be prohibited except under certain circumstances (Art. 9 of the GDPR):

- The data subject has given explicit consent to the processing of those personal data for one or more specified purposes.

  The term "explicit "means that the data subject must give an express statement of consent, which could be done in an obvious way by a written statement signed by the data subject.[72] However, there are other ways to obtain the explicit consent, such as filling an electronic form (in a digital or online context), sending an e-mail, uploading a scanned document, or even orally (it is recommended to record the statement for accountability purposes).[73]

[71] Giakomopulus.C, Buttarelli. G, O'Flaherty M. (2018, April). Handbook on European Data Protection Law. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/5b0cfa83-63f3-11e8-ab9c-01aa75ed71a1

[72] The Art 29 Working Party (2017, November 28th). Guidelines on consent under Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=51030

[73] The Art 29 Working Party (2017, November 28th). Guidelines on consent under Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=51030

Ways to ensure explicit consent include obtaining the (written or digital) signature of the data subject or through two-stage verification (i.e., sending an e-mail asking for consent first and an SMS with a verification code after[74]).

- Processing is necessary for the purpose of carrying out the obligations and exercising specific rights of the controller or of the data subject in the field of employment and social security and social protection law. The processing should be authorised by Union or Member State law or a collective agreement under to Member State law providing for appropriate safeguards.

- Processing is necessary to protect the vital interests of the data subject or of another natural person where the data subject is incapable of giving consent.

- Processing is carried out in the course of its legitimate activities with appropriate safeguards by a foundation, association or any other not-for-profit body with a political, philosophical, religious or trade union aim and on condition that the processing relates solely to the members or former members of the body or to persons who have regular contact with it.

- Processing related to personal data which are manifestly made public by the data subject.

  The personal data should be made public through a deliberate affirmative act of the data subject[75] (e.g., the data subject reveals personal information in an interview for a newspaper).

- Processing is necessary for the establishment, exercise or defence of legal claims or whenever courts are acting in their judicial capacity.

- Processing is necessary for reasons of substantial public interest, on the basis of Union or Member States law.

- Processing is necessary for the purposes of preventive or occupational medicine, for the assessment of the working capacity of the employee, medical diagnosis, the provision of health or social care or treatment or the management of health or social care systems and services on the basis of Union or Member State law or pursuant to contract with a health professional (data should be processed by or under the responsibility of a professional subject to the obligation of professional secrecy under Union or Member State law or rules established by competent national bodies) .

- Processing is necessary for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high standards of quality and safety of health care and of medicinal products or medical devices, on the basis of Union or Member State law.

[74] The Art 29 Working Party (2017, November 28th). Guidelines on consent under Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=51030
[75] Giakomopulus.C, Buttarelli. G, O'Flaherty M. (2018, April). Handbook on European Data Protection Law. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/5b0cfa83-63f3-11e8-ab9c-01aa75ed71a1

- Processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes based on Union or Member State law.

4.4 Data subject rights

The GDPR provides the following rights to the data subjects to assure the protection of their personal data.

Comprise
It is recommended that controllers implement protocols in order to fulfil the requirements of the data subjects, when exercising their rights, without undue delay.

4.4.1 Right to be informed

According to Art. 13 of the GDPR the controller shall inform the data subject about the intended processing when his or her personal data are collected.[76] The data subject should be provided with the following information:

- Controller's identity and contact details
- DPO's contact details (if applicable)
- The purposes of the processing as well as the legal basis for the processing
- When the processing is based on the legitimate interest of the controller (or a third party), the legitimate interests pursued
- The personal data's recipients or categories of recipients
- Where applicable the fact that the controller intends to transfer personal data to a third country or international organisation and the existence or absence of an adequacy decision by the Commission or whether relies on appropriate or suitable safeguards.
- The period for which the personal data will be stored or, the criteria used to determine that period, if it is not possible to establish the concrete period.
- The data's subject rights regarding the processing.
- The existence of the right to withdraw consent at any time.
- The right to lodge a complaint with a supervisory authority.
- Whether the provision of personal data is a statutory or contractual requirement, or a requirement necessary to enter into a contract, or if the data subject is

[76] Giakomopulus.C, Buttarelli. G, O'Flaherty M. (2018, April). Handbook on European Data Protection Law. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/5b0cfa83-63f3-11e8-ab9c-01aa75ed71a1

obliged to provide the personal data and the possible consequences of not providing the data.

- The existence of automated decision making, including profiling, and meaningful information about the logic involved (see Section 4.4.2).

- If the personal information has not been obtained from the data subject, from which source the personal data originates.

The quality, accessibility and comprehensibility of the information are as important as the content.[77] Art. 12 of the GDPR (Transparency) establishes that the information regarding the processing of the personal data should be provided to the data subject "*in a concise, transparent, intelligible and easily accessible form, using clear and plain language, in particular for any information addressed specifically to a child".* Recital 39 of the GDPR provides: "*It should be transparent to natural persons that personal data concerning them are collected, used, consulted or otherwise processed and to what extent the personal data are or will be processed*".

When the GDPR establishes that information should be provided in a "concise" form, it means that it should be presented efficiently and succinctly by the data controller, avoiding information fatigue.[78]

Additionally, the information should be "intelligible", meaning that it should be understood by an average member of the intended audience. This is directly related to the "clear and plain language" requirement which seeks to avoid the use of complex sentences and structures.[79]

The controller should be especially careful when the data subject is a child, in which case, an appropriate language and style should be used when the information about the personal data processing is provided. An example of child-centred language can be found in "UN Convention on the Rights of the Child in Child Friendly Language".[80] It is important to point out that children do not lose their right to be informed according to the transparency requirements just because the consent has been given by the holder of parental responsibility.[81] According to Art. 8 of the GDPR "*the processing of the personal data of a child shall be lawful where the child is at least 16 years old. Where the child is below the age of 16 years, such processing shall be lawful only if and to the extent that consent is given or authorised by the holder of parental responsibility over the child. Member States may provide by law for a lower age for those purposes provided that such lower age is not below 13 years".*

[77] The Art 29 Working Party (2017, November 29th). Guidelines on transparency under Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=622227
[78] The Art 29 Working Party (2017, November 29th). Guidelines on transparency under Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=622227
[79] The Art 29 Working Party (2017, November 29th). Guidelines on transparency under Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=622227
[80] The Art 29 Working Party (2017, November 29th). Guidelines on transparency under Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=622227
[81] The Art 29 Working Party (2017, November 29th). Guidelines on transparency under Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=622227

Regarding the form of providing information, the use of layered privacy notices is recommended. A first layer will contain the most critical aspects of the information for the data subject, however, the data subject should be able to access the complete information (that should be available in a one single place) from first layer (e.g., through a link for example if the privacy statement is provided through a website, or in the backside of the paper if the information is provided in a paper information sheet[82]).

Other methods to provide information such as videos, infographic, cartoons or other means can be used by the data controllers as far as the chosen method is appropriate to the particular circumstances (e.g., in the case of a device that captures personal data but does not have a screen, the privacy notice could be included in a website which its URL is provided in the instructions manual of the device).

When "*the controller intends to further process the personal data for a purpose other than that for which the personal data were collected, the controller shall provide the data subject prior to that further processing with information on that other purpose*" (Art. 13.3 GDPR).

According to Art. 14.3. of the GDPR, if the information is not collected directly from the data subject the controller should provide the information:

- Within a reasonable period after obtaining the personal data, but at the latest within one month.
- If the controller uses the data for communication with the data subject, at the latest, the information should be provided at the first communication.
- If the controller discloses the personal data to a third party, at the latest when the personal data are disclosed.

Machine learning
Explaining machine learning processes
One problem related to machine learning processes is that the model in many occasions produces a result that cannot be explained. This makes it very difficult to know how the result is produced. Machine learning systems work as what is called a "black box".[83]
Artificial intelligence is an advanced technology difficult to understand and the "black box" concept makes it almost impossible to explain how the information is correlated and weighed in a particular process. This may raise challenges on how to satisfy the transparency principle and to properly inform the data subjects about the processing of their personal data.[84]

[82] The Art 29 Working Party (2017, November 29th). Guidelines on transparency under Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=622227
[83] The Norwegian Data Protection Authority. (2018, January). Artificial Intelligence and privacy. Retrieved from https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf
[84] The Norwegian Data Protection Authority. (2018, January). Artificial Intelligence and privacy. Retrieved from https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf

Given the propensity of finding unexpected correlations between data, it may be difficult for organizations to foresee all the uses they may make of the data collected, and therefore, to define the purposes of the processing. The data controller should identify the purposes of the processing in an early stage and inform the data subjects in comprehensible way about these purposes.[85]

When there is an automated decision making, referred to in Art. 22, the GDPR provides that the logic involved in the processing of the personal data should be explained (this particular case will be analysed in Section 4.4.2).

Innovative tools and methods that may be useful to fulfil the minimization requirement.

Some tools and methods that can help meet data protection challenges have been developed, but in many cases, they have not been evaluated in practice yet:[86]

- **Explainable Artificial Intelligence:** The Explainable AI (XAI) programme aims to create machine learning techniques that produce more explainable models (maintaining at the same time a high level of learning performance) and that enable human users to understand and trust the emerging generation of AI partners.[87] The produced models will be combined with human-computer interfaces capable of translating models into understandable and useful explanation dialogues for the users.[88]

- **LIME:** This is an approach to XAI that consists in a model-agnostic solution that produces explanations for ordinary people.

COMPRISE

For COMPRISE, some recommendations were provided to the partners in order to prepare the information sheet and consent form (see Annex 2).

4.4.2 *Automated individual decision-making, including profiling and how it is connected to the right to information.*

According to Article 22 of the GDPR, there is a general prohibition on fully automated individual decision-making, including profiling (see Section 4.2.2) that produces legal effects concerning the data subject or similarly significant effects, with some exceptions that will be mentioned later.[89]

[85] Information Commissioner's Office. (2017, September). Big data, artificial intelligence, machine learning and data protection. Retrieved from https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf

[86] The Norwegian Data Protection Authority. (2018, January). Artificial Intelligence and privacy. Retrieved from https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf

[87] Gunning D., Ramage.D. Explainable Artificial Intelligence (XAI). Retrieved from https://www.darpa.mil/program/explainable-artificial-intelligence.

[88] Gunning D., Ramage.D. Explainable Artificial Intelligence (XAI). Retrieved from https://www.darpa.mil/program/explainable-artificial-intelligence.

[89] The Art 29 Working Party (2017, October 3rd). Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053

The decision to which Art. 22 refers to should be solely based on automated processing, therefore it should not involve a human at all.[90] The decision will not be solely based on automated processing if a human, besides the results of the automated processing, takes account of other factors and makes a final decision.[91]

Additionally, the decision based on automated processing should produce legal effects concerning the data subject or similarly significant effects. That the automated decision produces legal effects means that it affects someone's legal rights, legal status or rights under a contract.[92]

Even if the decision has no effect on someone's legal rights but produces similarly significant effects on someone, it will fall within the scope of Art. 22. Recital 71 of the GDPR provides some examples of similarly significant effects *"such as automatic refusal of an online credit application or e-recruiting practices".*

Art. 22 sets out some exception to the general prohibition if the automated decision:

- is necessary for entering into, or performance of, a contract between the data subject and a data controller;
- is authorised by the Union or Member State law to which the controller is subject, and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or
- is based on the data subject's explicit consent.

If sensitive data is involved in the processing, the exemption to the general prohibition only applies if the data subject has given explicit consent to the processing or if the processing is necessary for reasons of substantial public interest, on the basis of Union or Member State law. In both cases suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place (Art. 22.4 GDPR) — It should be taken into account that the joint processing of several types of personal data may reveal sensitive data of the data subject.[93]

When there is an automated decision with the characteristics described in Art. 22 the data subject will have the following rights:

- The right to be informed when his/her personal data are collected, and the right to obtain information from the data controller about the automated decision making, that should include meaningful information about the logic involved and the envisaged consequences of such processing (Art. 13.2.f and 15.h of the GDPR).

[90] The Art 29 Working Party (2017, October 3rd). Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053
[91] The Art 29 Working Party (2017, October 3rd). Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053
[92] The Art 29 Working Party (2017, October 3rd). Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053
[93] The Norwegian Data Protection Authority. (2018, January). Artificial Intelligence and privacy. Retrieved from https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf

- The right to obtain human intervention on the part of the controller, to express his/her point of view and to contest the decision. However, the individual cannot exercise this right if the automated decision is authorised by Union or Member State law to which the controller is subject, and which also lays down suitable measures to safeguard the data subject's rights and freedoms (Art 22.3 of the GDPR).

Machine learning
It is not expected that the automated decision resulting from the training of algorithms in a voice-based technology produces legal effects concerning the data subject or similarly significant effects. However, these voice-based technologies are usually used in combination with other software or applications that may make fully automated individual decisions including profiling (see Section 4.2.4) that produce legal effects concerning the data subject or similarly significant effects. ***Right to information when there are automated decisions referred to in Art. 22.*** When an automated decision falls under Art. 22, the data subject should be informed about the logic involved in the processing and the envisaged consequences of such processing. As explained in previous sections, the "black box" concept may make it difficult to explain how an algorithm reaches a decision. It is recommended to find a simple way to explain the data subject about the rationale behind or the criteria relied on to reach the decision. Besides, the GDPR does not require to provide a complex explanation of the algorithms (the "black box" should not necessarily be open[94]), but the data controller should provide as much information as necessary in such a way that the data subject is able to understand the result and to exercise his or her rights.[95] The data subject should be able to understand why a particular decision is reached, how it affects him or her, or what would need to be changed to reach a different output.[96] It is recommended to exercise caution before relying on machine learning decisions that cannot be rationalised in human understandable terms.[97]

4.4.3 Right of access

According to Art. 15 of the GDPR, the data subject shall have the right to obtain from the controller confirmation as to whether or not personal data concerning him/her are

[94] The Art 29 Working Party (2017, October 3rd). Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053

[95] The Norwegian Data Protection Authority. (2018, January). Artificial Intelligence and privacy. Retrieved from https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf

[96] The Norwegian Data Protection Authority. (2018, January). Artificial Intelligence and privacy. Retrieved from https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf

[97] Information Commissioner's Office. (2017, September). Big data, artificial intelligence, machine learning and data protection. Retrieved from https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf

being processed and, when that is the case, access to the personal data as well as to the information related to the processing of the personal data.

The GDPR has enhanced this right by broadening the information to which the data subject should have access to, including among others the purpose of processing, the categories of personal data concerned, the recipients to whom the personal data have been or will be disclosed, and the period for which the personal data will be stored or the existence of automated decision-making.

The transparency requirement (see Section 4.4.1) established in Art. 12 also applies to the information obtained by the data subject when exercising his or her right to access.

Machine learning
High volumes of personal data For organisations that collect large volumes of data to train algorithms it may be difficult to provide the data subject with the information required in the exercise of his/her right to access. The data controller should practice a good data management, meaning that he/she will need good metadata, the ability to use this data to find all the information about an individual and tracking of the data that has been anonymized and the data that can still be linked to an individual.[98] ***Online access to personal data*** Currently, many organisations are making available to their customers their personal information on request or proactively online through a secure log-in.[99] This practice is aligned with Recital 63 of the GDPR:" […] *Where possible, the controller should be able to provide remote access to a secure system which would provide the data subject with direct access to his or her personal data*. […]".

4.4.4 Right to rectification

According to Art. 16 of the GDPR, data subjects have the right to request the controller for their personal information to be rectified if inaccurate without undue delay.

4.4.5 Right to erasure (right to be forgotten)

According to Art. 17 of the GDPR, the data subject shall have the right to obtain from the controller the erasure of his or her personal data with undue delay when:

- Personal data are no longer necessary regarding the purposes for which they were collected or processed.

[98] Information Commissioner's Office. (2017, September). Big data, artificial intelligence, machine learning and data protection. Retrieved from https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf

[99] Information Commissioner's Office. (2017, September). Big data, artificial intelligence, machine learning and data protection. Retrieved from https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf

- The data subject withdraws consent on which the processing is based and there is no other legal ground for the processing.
- The data subject objects to the processing and there are no overriding legitimate grounds for the processing.
- The personal data have been unlawfully processed.
- The personal data have to be erased for compliance with a legal obligation in Union or Member State law to which the controller is subject.
- The personal data have been collected in relation to the offer of information society service to children.

The right to be forgotten is the personal data erasure obligation applied to internet search engines. Data subjects shall obtain from the data controller (the internet search engine) the erasure of their personal information anytime. If any controller has made the data subjects' personal data public, he must take reasonable measures to inform other controllers that all links to this personal data, as well as copies or replicates of the personal data, must be erased.[100]

There are some exceptions to the erasure right, that should not apply when the processing is necessary (Art. 17.3 of the GDPR):

- For exercising the right of freedom of expression.
- For compliance of a legal obligation.
- For reasons of public interest in the area of public health.
- For archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in so far as the right to obtain the erasure of personal data is likely to render impossible or seriously impair the achievement of the objectives of that processing.

Machine learning
In those cases in which systems operate machine learning, the right to erasure could damage the system. A machine learning system must learn relationships between elements in a dataset.[101] For that purpose, it will need to remember all the data used for training in order to sustain the rules derived from that data. Erasing data may damage the system and make it less accurate.[102]

[100] https://gdpr-info.eu/issues/right-to-be-forgotten/
[101] https://en.wikipedia.org/wiki/Unsupervised_learning
[102] Wallace.N, Castro D. (2018, March 17th) The Impact of the EU's New Data Protection Regulation on AI. Retrieved from http://www2.datainnovation.org/2018-impact-gdpr-ai.pdf

4.4.6 Right to restriction of processing

According to Art. 18 of the GDPR, the data subject shall have the right to obtain from the controller temporal restriction of processing of his/her personal data where:

- The data subject contests the accuracy of the personal data
- The processing is unlawful, and the data subject opposes the erasure of the personal data and requests the restriction of their use instead.
- The data are not needed for the purposes of the processing but are kept for the exercise or defence of legal claims.
- The data subject has objected pending the verification whether the legitimate grounds of the controller override those of the data subject.

4.4.7 Notification obligation regarding rectification or erasure of personal data or restriction of processing

According to Art. 19 of the GDPR, the controller shall communicate any rectification or erasure of personal data or restriction of processing carried out to each recipient to whom the personal data have been disclosed, unless this proves impossible or involves disproportionate effort. Additionally, the controller should inform the data subject about the recipients of his/her personal data if he/she requests it.

4.4.8 Right to portability

According to Art. 20 of the GDPR, the data subject shall have the right to receive his/her personal data, which he/she has provided to a controller in a structured, commonly used and machine-readable format and has the right to transmit those data to another controller. This means that the personal data should be transmitted directly from one controller to another, where technically feasible. The right to data portability only applies in situations where the personal data processing by automatic means is based on consent or a contract.[103]

The controller, when responding to a data portability request, is not responsible for the recipient's compliance with the data protection law.[104]

4.4.9 Right to object

According to Art. 21 of the GDPR, the data subject shall have the right to object to the processing of his/her personal data:

- On grounds to his/her particular situation, at any time, when the processing is based on public interest or legitimate interest pursued by the controller or a third-party ground. The controller shall no longer process the personal data unless the controller demonstrates compelling legitimate grounds for the processing which

[103] Giakomopulus.C, Buttarelli. G, O'Flaherty M. (2018, April). Handbook on European Data Protection Law. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/5b0cfa83-63f3-11e8-ab9c-01aa75ed71a1

[104] Giakomopulus.C, Buttarelli. G, O'Flaherty M. (2018, April). Handbook on European Data Protection Law. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/5b0cfa83-63f3-11e8-ab9c-01aa75ed71a1

override the interests, rights and freedoms of the data subject or for the establishment, exercise or defence of legal claims.

- Where the personal data are processed for direct marketing purposes.
- In the context of the use of information society services, the data subject may exercise his/her right to object by automated means.
- Where personal data are processed for scientific or historical research purposes or statistical purposes unless the processing is necessary for the performance of a task carried out for reasons of public interest.

4.5 Data controller and data processor responsibilities

4.5.1 Data controller responsibilities

Data controllers should implement appropriate technical and organisational measures to ensure and be able to demonstrate that personal data processing is performed in accordance with GDPR (Art. 24 of the GDPR). To do this, data controllers need to take into account the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for the rights and freedoms of natural persons.

These measures should be applied both when determining the means of the processing and during the processing of the personal data.[105] This is identified in the GDPR as "**privacy by design"** which is conceived as a prevention model that demands a proactive attitude to the controller. Data protection should be considered from the beginning, when a service, an application or a product that involves the processing of personal data is designed.

The data controller should implement appropriate technical and organisational measures to ensure that only personal data which are necessary for the purposes will be processed **by default**.[106] This obligation will be extended to:

- The volume of personal data collected
- The extension of the data
- The storage periods
- The accessibility to the data

The data controller should maintain a record of processing activities (Art. 30.1 of the GDPR) that should contain amongst other information the contact details of the controller, the purposes of the processing, the categories of data subjects and of the recipients and information about international transfers if applicable, technical and organisational security measures. This obligation does not applies if *"an enterprise or an*

[105] Giakomopulus.C, Buttarelli. G, O'Flaherty M. (2018, April). Handbook on European Data Protection Law. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/5b0cfa83-63f3-11e8-ab9c-01aa75ed71a1

[106] Giakomopulus.C, Buttarelli. G, O'Flaherty M. (2018, April). Handbook on European Data Protection Law. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/5b0cfa83-63f3-11e8-ab9c-01aa75ed71a1

organisation employing fewer than 250 persons unless the processing it carries out is likely to result in a risk to the rights and freedoms of data subjects, the processing is not occasional, or the processing includes special categories of data".

Data controllers should inform the supervisory authority in case of personal data breach without undue delay, and no later than 72 hours after becoming aware of it (Art. 33 of the GDPR). If the personal data is likely to result in high risks to the rights and freedoms of natural persons, the controller shall communicate the personal data breach to the data subject without undue delay too (Art. 34 of the GDPR).

COMPRISE
Controllers should assure that privacy by design and privacy by default are applied to every new personal data processing activity initiated during the project, as well as control that all the ongoing processing activities are applying it. Controllers should also monitor periodically the adequacy of the organisational and technical measures applied.

4.5.2 Data processor responsibilities

When a data processor is carrying out a processing on behalf of the controller, the processor should provide sufficient guarantees to implement appropriate technical and organisational measures to ensure the compliance with the GDPR (Art. 28 GDPR).

If the processor wants to engage another processor, the written authorisation of the controller, that should be informed about any intended changes concerning the addition or replacement of other processors, will be needed.

The processing of the processor shall be governed by a contract or other legal act under the EU law, that is binding on the processor with regard to the controller that sets out the subject-matter and duration of the processing, the nature and purpose of the processing, the type of personal data and categories of data subjects and the obligations and rights of the controller (Art. 28.2).

The processor should maintain a record of processing activities (Art. 30.2) that should contain amongst other information the contact details of the processor and the controller, the categories of the processing, information about international transfers if applicable, technical and organisational security measures. This obligation does not apply to "*an enterprise or an organisation employing fewer than 250 persons unless the processing it carries out is likely to result in a risk to the rights and freedoms of data subjects, the processing is not occasional, or the processing includes special categories of data*".

COMPRISE
If a data processor will carry out a processing on behalf of any of the partners (controllers) a contracting protocol should be defined. All the processors should be identified and should be assured that the relation with each of them is governed by a contract or another legal act under EU law.

4.5.3 Data Protection Officer (DPO)

The Data Protection Officer (DPO) figure is a cornerstone of accountability that will facilitate compliance with GDPR.[107] According to Art. 37 of the GDPR the data controller and the processor shall designate a Data Protection Officer in any case where:

- The processing is carried out by a public authority or body.
- The core activities of the controller or the processor consist of processing operations which, by virtue of their nature, their scope and/or their purposes, require regular and systematic monitoring of data subjects on a large scale.
- The core activities of the controller or the processor consist of processing on a large scale of special categories of data or personal data relating to criminal convictions and offences.

The Work Group of Art. 29 provides some guidelines to interpret some concepts such as "core activities", "large scale" or systematic monitoring:

- **Core activities**: Can be defined as the key operations to achieve the controller's or processor's objectives.
- **Large scale**: To determine whether a processing is carried out on a large scale, factors such as the number of data subjects concerned, the volume of data processed, the duration of the data processing activity and the geographical extent of the processing activity should be considered.
- **Regular and systematic monitoring**: This notion includes all forms of tracking and profiling in internet, but the notion is not restricted to the online environment. Other activities in which processing requires regular and systematic monitoring include providing telecommunication services, email retargeting, profiling and risk assessment, location tracking, etc.

The main tasks of the DPO are defined in Art. 39 of the GDPR, which provides that the DPO shall:

- Inform and advise the controller or the processor (and their employees) about their obligations regarding data protection regulations.
- Monitor compliance with GDPR. However, the GDPR makes it clear that it is the data controller and not the DPO who is required to implement the technical and organisational measures to ensure the compliance with the Regulation.[108]
- Assist the data controller when carrying out a Privacy Impact Assessment (PIA).
- Cooperate with the Supervisory Authority and act as a contact point (facilitate access by the Supervisory Authority to documents and information as well as for

[107] The Art 29 Working Party (2016, December 13th). Guidelines on Data Protection Officers ('DPOs'). Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612048
[108] The Art 29 Working Party (2016, December 13th). Guidelines on Data Protection Officers ('DPOs'). Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612048

the exercise of its investigative, corrective, authorisation, and advisory powers).[109]

- Prioritise the activities on issues that present higher data protection risks without neglecting the compliance of operations with a lower risk level.

Art. 38 of the GDPR establishes some guarantees in order to ensure that the DPO acts in an independent manner:

- The DPO should not receive instructions by the controller or the processor regarding its tasks.

- The DPO should not be penalised or dismissed by the controller for the performance of its tasks.

- The DPO's tasks should not result in a conflict of interest. This means that the DPO cannot hold a position within the organisation that leads him or her to determine the purposes and means of the personal data processing (such as senior management positions or other roles in the organisation structure that also could lead to the determination of purposes and means of the personal data processing[110]).

4.6 International transfers of personal data

Personal data may be transferred to third countries (outside of the European Economic Area) in which personal data is not adequately protected or in which data subject rights could be undermined. It should be noted that if a third party (i.e., a partner or a provider) that is located outside of the EEA is able to access the personal data collected by the data controllers this amounts to a 'data transfer' in the context of the GDPR.[111]

According to the GDPR, there are two ways of allowing the transfer of personal data to third countries or international organisations, on the basis of an adequacy decision by the European Commission or when the controller or processor has provided appropriate safeguards.

Transfers of personal data may take place on the basis of an adequacy decision by the European Commission, this happens when the Commission has decided that the third country, a territory or one or more specified sectors within that third country, or the international organisation in question ensures an adequate level of protection (Art. 45 of the GDPR). This adequacy decision will be reviewed at least every four years. The website of the Commission or the Official Diary of the EU should be consulted on a regular basis to know in relation to which countries an adequacy decision has been taken.

[109] The Art 29 Working Party (2016, December 13th). Guidelines on Data Protection Officers ('DPOs'). Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612048
[110] The Art 29 Working Party (2016, December 13th). Guidelines on Data Protection Officers ('DPOs'). Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612048
[111] (2018, November 14th). Ethics and data protection. Retrieved from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/ethics/h2020_hi_ethics-data-protection_en.pdf

If there is no adequacy decision by the Commission regarding the third country to which the personal data will be transferred, the controller or processor may also transfer personal data to a third country or an international organisation when the controller or processor has provided appropriate safeguards. The appropriate safeguards according to Art. 46 of the GDPR are:

- A legally binding and enforceable instrument between public authorities or bodies.

- Binding Corporate Rules (BCRs), which are internal rules for data transfers within multinational companies,[112] that should be approved by the competent supervisory authority.

  This applies for international transfers that take place within the same group of enterprises or undertakings that are part of a joint economic activity.[113] The BCRs should be legally binding, applied to and be enforced by every member of the group or undertaking.

- Standard data protection clauses (between the data-exporting controller and the recipient in the third country) adopted by the Commission.

- Standard data protection clauses adopted by a supervisory authority and approved by the Commission.

- An approved code of conduct or an appropriate certification mechanism together with binding and enforceable commitments of the controller or processor in the third country to apply the appropriate safeguards.

Where international transfer of personal data is carried out through ad hoc contracts between the controller or processor and the controller, processor or recipient of the personal data, which includes contractual clauses not adopted by the Commission, it will be necessary to obtain the authorisation from the competent supervisory authority.

In the absence of an adequacy decision or of appropriate safeguards, a transfer or a set of transfers of personal data to a third country shall take place on one of the following conditions (Art. 49 GDPR):

- The data subject has explicitly consented to the proposed transfer, after being informed of the possible risks of such transfers for the data subject.

- The transfer is necessary for the performance of a contract between the data subject and the controller or the implementation of pre-contractual measures taken at the data subject's request.

[112] https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/binding-corporate-rules-bcr_en

[113] Giakomopulus.C, Buttarelli. G, O'Flaherty M. (2018, April). Handbook on European Data Protection Law. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/5b0cfa83-63f3-11e8-ab9c-01aa75ed71a1

- The transfer is necessary for the conclusion or performance of a contract concluded in the interest of the data subject between the controller and another natural or legal person.
- The transfer is necessary for important reasons of public interest.
- The transfer is necessary for the establishment, exercise or defence of legal claims.
- The transfer is necessary in order to protect the vital interests of the data subject or of other persons, where the data subject is physically or legally incapable of giving consent.
- For the transfer of data from public registers.

5 Recommendations on how to implement a Privacy Impact Assessment (PIA)

5.1 Concept of the Privacy Impact Assessment

The Privacy Impact Assessment(PIA) is a preventive instrument introduced by the GDPR (Art. 35) which should be used by the controller to identify, evaluate and manage the risks concerning the personal data processing activities carried out, with the objective of guaranteeing the rights and freedoms of the data subjects.[114]

According to Art. 35 of the GDPR, the PIA is only mandatory when the processing is "*likely to result in a high risk to the rights and freedoms of natural persons".* In the cases where it is not clear whether a PIA is mandatory or not, it is recommended that the PIA is carried out, as it is a very useful tool that helps controllers to comply with the GDPR and to demonstrate that appropriate measures have been taken to comply with the Regulation.[115]

It is an obligation of the data controller to conduct the PIA when required. The DPO (if it has been appointed) must provide assessment and support to the data controller when performing this task. However, other professional profiles such as information security specialist or legal specialists[116] may be involved (which is recommended). It is also recommended to gather feedback from the data subject.

The PIA concept is aligned with the privacy by design principle, so it should be carried out before starting the processing of personal data in order to take appropriate decisions. However, the PIA should be understood as a permanent improvement process that must

[114] Agencia Española de Protección de Datos. Guia práctica para las evaluaciones de impacto sujetas a la RGPD.Retrieved from: https://www.aepd.es/media/guias/guia-evaluaciones-de-impacto-rgpd.pdf
[115] Agencia Española de Protección de Datos. Guia práctica para las evaluaciones de impacto sujetas a la RGPD.Retrieved from: https://www.aepd.es/media/guias/guia-evaluaciones-de-impacto-rgpd.pdf
[116] Agencia Española de Protección de Datos. Guia práctica para las evaluaciones de impacto sujetas a la RGPD.Retrieved from: https://www.aepd.es/media/guias/guia-evaluaciones-de-impacto-rgpd.pdf

be reviewed periodically, especially when there are updates or important modifications in the processing activities.[117]

A PIA may concern a single data processing operation or "*a set of similar processing operations that present similar high risks"* (Art. 35.1 GDPR). Regarding this, Recital 92 of the GDPR provides :""*There are circumstances under which it may be reasonable and economical for the subject of a data protection impact assessment to be broader than a single project, for example where public authorities or bodies intend to establish a common application or processing platform or where several controllers plan to introduce a common application or processing environment across an industry sector or segment or for a widely used horizontal activity*".

The steps that should be taken to carry out the PIA are to:

- Analyse whether a PIA is required
- Assess the necessity and proportionality of the processing
- Detect and manage risks
- Define an action plan and conclusions
- Supervise the implementation of the actions

All the tasks, analysis, assessments carried out, as well as the conclusions drawn within the framework of the project, should be documented to keep track of the actions performed and be able to justify the decisions taken.[118]

5.2 Analyse if the personal data processing operation is subject to the Privacy Impact Assessment

According to Art. 35 of the GDPR, carrying out a PIA is mandatory when the processing of personal data "*is likely to result in a high risk to the rights and freedoms of natural persons".* Art. 35.3 provides a non-exhaustive list of situations in which conducting a PIA is mandatory (but this does not mean that a PIA will not be required in other situations):

- A systematic and extensive evaluation of personal aspects relating to natural persons which are based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person.
- Processing on a large scale of special categories of data, or of personal data relating to criminal convictions and offences.
- A systematic monitoring of a publicly accessible area on a large scale.

[117] Agencia Española de Protección de Datos. Guia práctica para las evaluaciones de impacto sujetas a la RGPD.Retrieved from: https://www.aepd.es/media/guias/guia-evaluaciones-de-impacto-rgpd.pdf
[118] Agencia Española de Protección de Datos. Guia práctica para las evaluaciones de impacto sujetas a la RGPD.Retrieved from: https://www.aepd.es/media/guias/guia-evaluaciones-de-impacto-rgpd.pdf

Besides, the "*supervisory authority shall establish and make public a list of the kind of processing operations which are subject to the requirement for a data protection impact assessment*" (Art. 35.4 GDPR), as well as *"establish and make public a list of the kind of processing operations for which no data protection impact assessment is required"* (Art. 35.5 GDPR).

Art. 29 of the Data Protection Working Party provides additional criteria to be considered to identify processing operations that require a PIA:[119]

- **Evaluation or scoring:** Including profiling and predicting (aspects related to health, work, economic situation, behaviour, etc.)
- **Automated-decision making with legal or similar significant effect** (see Section 4.4.2)
- **Systematic monitoring:** Is the process used to observe, monitor or control data subjects
- **Sensitive data** (see Section 4.2.2)
- **Data processed on a large scale** (see Section 4.5.3)
- **Datasets that have been matched or combined:** Could be data originated in two different processing operations for different purposes.
- **Data concerning vulnerable data subjects:** When there is a power imbalance between the controller and the data subject (e.g., employees that cannot oppose to the processing carried out by the employer). Children would also be considered vulnerable data subjects, as well as segments of the society such as elderly, patients or mentally ill people.
- **Innovative use or applying technological or organisational solutions:** The use of new technologies can trigger the need to conduct a PIA, as the use of such technologies involves novel forms of data collection and processing that may entail a high risk for data subject's rights and freedoms.
- **Data transfer across borders outside the EU** (see Section 4.6)
- **When the processing itself prevents data subjects from exercising a right or use a service or a contract**

The more of these criteria are met by the processing, the more likely it is to present a high risk to the rights and freedoms of the data subject and therefore to require a PIA.[120]

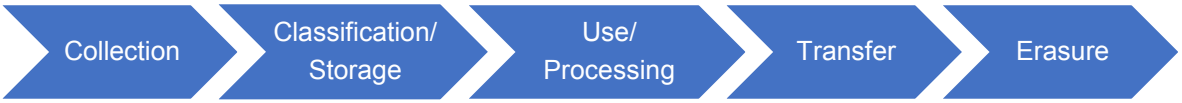
[119] The Art 29 Working Party (2017, April 4th). Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236

[120] The Art 29 Working Party (2017, April 4th). Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236

5.3 Description of the envisaged processing operation

Art. 35.7 provides that the PIA shall contain **"***a description of the envisaged processing operations and the purposes of the processing, including, where applicable, the legitimate interest pursued by the controller***".** To do this, it is necessary to know the personal data's lifecycle and flow as well as the elements and actors involved in the processing.

The personal data's lifecycle can be divided into the following stages:[121]



- **Collection:** Collection of data may be carried out in different ways and using different techniques such as paper form, web form, audio or video recording, etc.
- **Classification/Storage:** Set categories to classify personal data and store them in the systems or archives.
- **Use/Processing:** Any operation or set of operations which is performed on personal data or sets of personal data, whether or not by automated means (such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure, etc.).
- **Transfer of data to a third party**: "Transfer" is a broad concept that includes transfer, communication, consultation, interconnection or any other means that enable access to the data**.**
- **Erasure:** The erasure of data stored in systems or archives.

Additionally, for each of these stages the elements involved in each of them should be identified:[122]

- **Activities of the processing:** The activities and operations carried out in relation to personal data should be identified. Every processing will include a set of operations performed with a purpose based in the same legal ground**.**
- **Personal data:** All personal data processed should be identified and categorised. Their importance in the processing should be assessed and a decision should be made whether they can be excluded or not from that processing.
- **People involved in the processing activities:** People that participate in different processing activities should be identified. Their functions and responsibilities should be defined and limited as well.

[121] Agencia Española de Protección de Datos. Guia práctica para las evaluaciones de impacto sujetas a la RGPD.Retrieved from: https://www.aepd.es/media/guias/guia-evaluaciones-de-impacto-rgpd.pdf
[122] Agencia Española de Protección de Datos. Guia práctica para las evaluaciones de impacto sujetas a la RGPD.Retrieved from: https://www.aepd.es/media/guias/guia-evaluaciones-de-impacto-rgpd.pdf

- **Technology:** Technological elements (software and hardware) involved in all processing activities should be identified. A detailed analysis not needed.

Below is an example of a table that could be used to identify all elements involved in each stage of the personal data lifecycle:[123]

	Personal data lifecycle stages				
	Collection	**Classification /Storage**	**Use/ Processing**	**Transfer**	**Erasure**
Activities					
Data					
People involved					
Technologies involved					

5.4 Assess the necessity and proportionality of the processing

To analyse the necessity and proportionality of the personal data processing that will be carried out, the purpose and means of the processing as well as the identification of the data that will be used for the processing should be both analysed first.[124]

Once the means and the purposes are identified, the legal grounds on which the personal data processing will rely (see Section 4.3.1.2) should be analysed. When the processing relies on the legitimate interest of the data controller, each particular case should be carefully evaluated, balancing the interest of the controller and the rights and interests of the data subjects.

Additionally, the minimisation principle should be considered (see Section 4.3.1.3) when assessing the necessity and proportionality of the personal data processing. It should be considered which personal data are strictly necessary to carry out the processing activities for the purposes decided as well as if all the actions that will be performed during the processing are necessary and proportionate to these purposes.

[123] Agencia Española de Protección de Datos. Guia práctica para las evaluaciones de impacto sujetas a la RGPD.Retrieved from: https://www.aepd.es/media/guias/guia-evaluaciones-de-impacto-rgpd.pdf
[124] Agencia Española de Protección de Datos. Guia práctica para las evaluaciones de impacto sujetas a la RGPD.Retrieved from: https://www.aepd.es/media/guias/guia-evaluaciones-de-impacto-rgpd.pdf

5.5 Risk management

5.5.1 Identification of the risks

A "risk" is a scenario describing an event and its negative consequences, estimated in terms of severity and likelihood.[125] Art. 35 of the GDPR refers to risks to the rights and freedoms of the data subject, which primarily concern the right to privacy but may also involve other rights such a freedom of speech and freedom of thought, amongst other fundamental rights.[126]

The personal data lifecycle can be useful to determine or identify the different scenarios in which an infringement of the rights and freedoms of the data subject may occur.[127]

5.5.2 Risk assessment

When assessing the risks, the probability and the impact of the risk materialising should be considered, for which it is necessary to define a criterion. This criterion may be based on a standard (such as ISO 29134) or defined by the controller.[128]

When assessing the risks, two concepts need to be acknowledged:[129]

- **The inherent risk:** This is the raw, untreated risk associated with the concrete activity.
- **Residual risk**: This is the resulting risk after applying to the inherent risk the mitigation measures selected.

5.5.3 Mitigation measures

Once the possible risks are assessed, the next step will be to take the appropriate measures to mitigate them. Four different kinds of measures can be taken:[130]

- **Risk reduction measures:** Measures to reduce the probability and impact level associated with the risk.
- **Risk retention measures**: If the inherent risk is low and within what is considered an acceptable risk, additional measures are not required.
- **Risk transference:** Consist of sharing the risk with an external organisation.
- **Risk avoidance:** When the risk is very high it may be decided not to carry out the processing, and consequently the risk is not assumed.

---

[125] Agencia Española de Protección de Datos. Guia práctica para las evaluaciones de impacto sujetas a la RGPD.Retrieved from: https://www.aepd.es/media/guias/guia-evaluaciones-de-impacto-rgpd.pdf
[126] The Art 29 Working Party (2017, April 4th). Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236
[127] Agencia Española de Protección de Datos. Guia práctica para las evaluaciones de impacto sujetas a la RGPD.Retrieved from: https://www.aepd.es/media/guias/guia-evaluaciones-de-impacto-rgpd.pdf
[128] Agencia Española de Protección de Datos. Guia práctica para las evaluaciones de impacto sujetas a la RGPD.Retrieved from: https://www.aepd.es/media/guias/guia-evaluaciones-de-impacto-rgpd.pdf
[129] Agencia Española de Protección de Datos. Guia práctica para las evaluaciones de impacto sujetas a la RGPD.Retrieved from: https://www.aepd.es/media/guias/guia-evaluaciones-de-impacto-rgpd.pdf
[130] Agencia Española de Protección de Datos. Guia práctica para las evaluaciones de impacto sujetas a la RGPD.Retrieved from: https://www.aepd.es/media/guias/guia-evaluaciones-de-impacto-rgpd.pdf

A different classification of the measures that may be applied would be:

- **Organisational measures:** Measures associated with the procedures, organisation and governance of the entity.
- **Legal:** Measures associated to legal compliance.
- **Technical:** Measures to protect the logic and physical security of the information assets.

Once the measures are applied, the residual risk should be evaluated by considering the probability and impact of the risk after applying the measures. It should be assessed if the residual risk is within what is considered an acceptable risk.

In Annex 3, a catalogue of examples regarding possible risks and mitigation measures is provided.

5.5.4 Action plan and conclusion

Once the measures that will be applied to control and mitigate the risks are defined, an action plan to reduce the risk of the processing activity to an acceptable level should be incorporated. It is recommended that the action plan includes:[131]

- Control
- Description of the control
- Implementation responsible
- Implementation period

The conclusion of the PIA should be based on the resulting residual risk. If the result is not acceptable (where the data subjects may encounter significant, or even irreversible, consequences, which they may not overcome, and/or when it seems evident that the risk will occur)[132] additional measures may be applied or the supervisory authority should be consulted (especially whenever the controller cannot find sufficient measures[133]).

If the result is within what is considered acceptable, the process can be carried out as far as the measures included in the action plan are applied.[134]

---

[131] Agencia Española de Protección de Datos. Guia práctica para las evaluaciones de impacto sujetas a la RGPD.Retrieved from: https://www.aepd.es/media/guias/guia-evaluaciones-de-impacto-rgpd.pdf

[132] The Art 29 Working Party (2017, April 4th). Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236

[133] The Art 29 Working Party (2017, April 4th). Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236

[134] The Art 29 Working Party (2017, April 4th). Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236

5.5.5 Supervision of the implementation of the actions

The PIA is a theoretical exercise that must be put into practice. The implementation of the measures defined in the PIA should be supervised and controlled to reduce an inherent risk to an acceptable residual risk. To assess the effectiveness of the measures implemented, it is recommended to carry out periodic audits prioritising the risks with the highest likelihood and severity. If the result of the audit is considered not acceptable, corrective actions should be implemented.

6 Information Security and cybersecurity

In this section, Information Security principles and some recommendations are analysed and provided to manage Information Security in an organisation framework.

6.1 Information Security

According to Art. 32 of the GDPR "*taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk*". These measures include among others:

- Pseudonymizing and encrypting personal data
- Ensuring the ongoing confidentiality, integrity, availability and resilience of processing systems and services
- Restoring the availability and access to personal data in a timely manner in the event of a physical or technical incident
- Regularly testing, assessing and evaluating the effectiveness of technical and organisational measures taken

To implement adequate measures that ensure appropriate levels of security, it is important to know the main information security principles:

- **Confidentiality principle**: This principle aims to ensure that information should solely be accessed by people with correct and proper privileges and consequently, should be hidden from people unauthorised to access it.[135]
- **Integrity principle:** This principle aims to ensure that data are protected from unauthorised modifications (modifications by unauthorised people or processes).
- **Availability principle**: This principle aims to ensure the continuity of the processes and the business guaranteeing that the information is accessible and usable.

[135] https://resources.infosecinstitute.com/guiding-principles-in-information-security/#gref

- **Resilience principle:** This principle aims to ensure the information system recovery after an incident.

6.2 Information security management

Data security in an organisation is not achieved just by implementing technical measures; it also requires proper organisational rules and measures as well as raising awareness of information security among organisation workers or employees. Internal organisational measures to carry out a successful information security management include:[136]

- Implementation of Information Security policies in the organisation.
- Informing the employees about data security rules and arise awareness on data security threats and consequences of security breaches.
- Defining a precise distribution of responsibilities and competencies in matters of data processing, especially regarding the decisions on processing of personal data or transmission of personal data to third parties.
- Use of personal data according to the instructions of the competent person only.
- Protecting the access to locations and to hardware and software of the controller or processor
- Ensuring that authorisations to access personal data have been assigned by the competent person
- Automated protocols on electronic access to personal data
- Recording of all data processing activities

To implement a proper Information Management system, it is recommended to adopt reference standards such as ISO 27001. The ISO 27001 is an international standard that specifies "*the requirements for establishing, implementing, maintaining and continually improving an information security management system within the context of the organization*[137]". This standard "*also includes requirements for the assessment and treatment of information security risks tailored to the needs of the organization*[138]*".*

6.3 Audits

To assess the effectiveness of the measures implemented, it is recommended to carry out periodic information security audits prioritising the risks with the highest likelihood and severity. If the result of the audit is considered not acceptable, corrective actions

[136] Giakomopulus.C, Buttarelli. G, O'Flaherty M. (2018, April). Handbook on European Data Protection Law. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/5b0cfa83-63f3-11e8-ab9c-01aa75ed71a1

[137] https://www.iso.org/obp/ui/#iso:std:iso-iec:27001:ed-2:v1:en

[138] https://www.iso.org/obp/ui/#iso:std:iso-iec:27001:ed-2:v1:en

should be implemented, and the effectiveness of these actions should be evaluated in the future audits.

6.4 Personal data breach

According to Art. 4.12 of the GDPR, data breach *"means a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed."*

In case of a personal data breach, the controller should inform the supervisory authority without delay and no later than 72 hours after becoming aware of it, unless the personal data breach is unlikely to result in a risk to the rights and freedoms of the data subject (Art. 33 of the GDPR). If the notification is not within 72 hours, it shall be accompanied by reasons for the delay.

If the personal data breach is likely to result in a high risk to the rights and freedoms of the data subject, the controller should communicate the data breach to the data subject without undue delay (Art. 34 of the GDPR). There are some exceptions when the communication to the data subject is not necessary:

- When the controller has applied appropriate technical and organisational measures to the personal data affected by the personal data breach (such as encryption)
- When the controller has taken subsequent measures, which ensures that the high risk to the rights and freedoms of the data subjects referred to are no longer likely to materialise;
- When this would involve a disproportionate effort (in that case it should be a public communication or a similar measure)

COMPRISE
It is recommended that the partners, as controllers, document any personal data breach, indicating the event, the impact and the measures taken. It is also recommended to register all communications to the data subjects regarding data breaches and keep them as an evidence.

6.5 Voice-enabled technologies and cybersecurity threats

Voice-enabled technologies are taking over different sectors. The possibility of executing remote commands at home or on automobiles, phone searching through voice assistants, accessing account information, to mention a few, are making voice-enabled systems popular among users around the world.

Voice-enabled systems rely on deep learning which has led to major improvements in speech-to-text, spoken language understanding, and dialogue management. These technologies typically operate as cloud-based services: the user's speech is sent to the cloud, where it is automatically transcribed and processed, and the system's reply is sent back to the user's device. Consequently each spoken message is centralised in a single place by the company providing the technology.

The process of personal data in the cloud implies exposure to risks that may involve loss of confidentiality and availability, and potential unauthorised access (in the case of a cyber-attack to the data or the loss of the data).

Consequently, it is expected to witness a greater market emphasis of local processing and storage.[139] A few companies such as Mycroft and Snips advertise "private-by-design" voice-controlled assistants that run locally on the user's device. The dialogue outcome (e.g., the desired temperature to be achieved, the food item to be added to the cart, the TV channel to be displayed, etc.) does not leave the user's home (e.g., it is sent to the heating controller) or, when it does, it is sent only to the company that requires it to deliver the expected service (e.g., the online food retailer or the TV access/contents provider).

The input speech signal and the intermediate dialogue steps do not leave the user's home and, crucially, neither these nor the outcome are centralised in a single place by the third-party company providing the voice interaction technology. This approach addresses privacy concerns during the operating phase. However, the technology must still be trained on external data, whose collection still raises privacy concerns.

Also new techniques such as Federated Learning are being applied to reduce the training of personal data in the cloud (see Section 4.3.1.3).

The use of anonymization, pseudonymization or other disassociation mechanism and encryption methods is recommended, so that personal data cannot be easily accessed by cybercriminals. Establishing properly designed dissociation procedures, and allowing proper management of stored data, will be a key element in improving the protection of privacy.

Other recommended measures to be taken in order to improve the control that users have over the personal information stored in the cloud include:

- Preventing the device's microphone to start recording (and consequently sending voice samples to the cloud) until it listens to a "wake phrase". The microphone will passively listen until it listens to the "wake phrase[140]".
- Providing a way to manually disable the microphone in the voice-enabled device.
- Including visual cues in the voice-enabled device that indicate when it is recording or transmitting the information.[141]
- Facilitating the erasure of the information stored in the cloud by the user.

[139] Gray.S. (2016, Apriñ). Experts on the GDPR #3: What is personal data under the GDPR? Always On: Privacy Implications of Microphone Enabled Devices. Retrieved from https://fpf.org/wp-content/uploads/2016/04/FPF_Always_On_WP.pdf

[140] Gray.S. (2016, Apriñ). Experts on the GDPR #3: What is personal data under the GDPR? Always On: Privacy Implications of Microphone Enabled Devices. Retrieved from https://fpf.org/wp-content/uploads/2016/04/FPF_Always_On_WP.pdf

[141] Gray.S. (2016, Apriñ). Experts on the GDPR #3: What is personal data under the GDPR? Always On: Privacy Implications of Microphone Enabled Devices. Retrieved from https://fpf.org/wp-content/uploads/2016/04/FPF_Always_On_WP.pdf

There are other common cyber-threats associated with voice-enabled technologies such as:

- **Voice squatting**. Consists of the use of third-party applications commonly referred to as "skills", that provide additional functions to voice-enabled devices. Malicious users may create skills that give them the ability to record audio or order items without the owner's permission. This can lead to getting access to personal data (including sensitive data).[142]

- **Software attack:** Consists of the possibility to gain access to the operating system of a voice recognition-based device and control it remotely through unsecured internet connections. This can lead to the stealing of personal data (and sensitive data) of the owner as well as to the control and use of the device's microphone for unknown purposes.[143]

- **Audio adversarial attacks:** Machine learning and deep learning models, are particularly vulnerable to adversarial attacks, whose goal is to manipulate the behaviour of AI models while remaining unnoticed to humans. It is possible to create audio that sounds normal to humans, but an automated speech recognition system will pick as a command (e.g., opening a door). These audio adversarial attacks are performed by adding a layer of noise imperceptible to the human ears that creates a different output when it is processed by AI.[144]

Companies like Google or Amazon are implementing new features to fight back these threats such as requiring users to confirm commands before completing any task.[145]

7 Other legal issues affecting voice-enabled technologies

In this section, an overview regarding other legal aspects that may affect activities that involves the use of voice-enabled technologies is provided.

7.1 Intellectual property

Machine learning methods are exposed to intellectual property infringement. Machine learning training requires very large datasets (that may have different origins) to improve its decision-making abilities. When a dataset comes from or has been generated by third parties, things may get complicated as the data can be protected by different Intellectual Property rights or its use can be limited by contractual restrictions.[146]

---

[142] Voice Assistants and Cyber Security. (2018, December 14). Retrieved from https://www.micomlabs.com/2018/12/14/voice-assistants-and-cyber-security/
[143] Voice Assistants and Cyber Security. (2018, December 14). Retrieved from https://www.micomlabs.com/2018/12/14/voice-assistants-and-cyber-security/
[144] Dickson.B. (2019, April 29th) Protecting AI models against audio adversarial attacks. Retrieved from https://bdtechtalks.com/2019/04/29/ai-audio-adversarial-examples/
[145] Voice Assistants and Cyber Security. (2018, December 14). Retrieved from https://www.micomlabs.com/2018/12/14/voice-assistants-and-cyber-security/
[146] Bond.T. (2017, Junel 16th) Artificial Intelligence and IP – Part 1: Developing AI systems. Retrieved from https://digitalbusiness.law/2017/06/artificial-intelligence-and-ip-part-1-developing-ai-systems/

Directive 96/9/EC protect databases by copyright if they are original, while non-original databases can also be protected if the investment in obtaining, verifying and presenting the data is substantial (this protection is known as "sui generis" protection). Consequently, the use of a database without the permission of the rights holder may represent an infringement of his/her Intellectual Property rights.

The use of datasets that rely on copyrighted works for training purposes may also lead to Intellectual Property infringement of the said works.

Some voice-enabled assistants are incorporating song-searching functions. The song recognition technology works by giving a unique fingerprint to each audio piece that is compared against a database.[147] Fingerprinting consists of extracting essential information from digitised audio recordings so they can be searched in a database.[148]

Some fingerprinting applications include[149]

- Song identification (e.g., Shazam). Users can record small samples of a song using a device, for example, a smartphone, which is then fingerprinted. The server is going to match the fingerprint against a database of fingerprints of known songs. Then it will provide the artist and song title to the user.

- Collection of royalties. Songs played in radio stations, clubs and similar venues are fingerprinted and identified automatically.

Many issues can arrive from fingerprinting in relation to intellectual property. First of all, it is essential to recall that samples in the databases are subject to copyright, as well as databases itself.

Another problem lies in whether fingerprints can be considered as a derivative work (a work, protected by copyright, based upon one or more pre-existing works for which permission is required,) or not, and this will depend on the technical details of the fingerprinting process itself.

7.2 Trade secrets

Article 2 (1) of the Directive 2016/943 of 8 June 2016 (Trade Secrets Directive) defines trade secrets as information which:

- Is secret in the sense that it is not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question;

- Has commercial value because it is secret; and

147 https://www.cnet.com/news/google-assistant-just-got-way-better-at-recognizing-which-songs-are-playing/

148 Benschop, L. (n.d.). Could music fingerprinting be copyright infringement? Retrieved from https://lennartb.home.xs4all.nl/fingerprint.html

149 Benschop, L. (n.d.). Could music fingerprinting be copyright infringement? Retrieved from https://lennartb.home.xs4all.nl/fingerprint.html

- Has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret.

When developing new technologies or trying to improve existing ones, results can be subject to trade secrets as long as they meet directive requirements; however, the scope of protection given by trade secrets can cover a lot more than that.

When an organisation provides information about the model involved in the processing of personal data, in compliance with the transparency principle, this may disclose trade secrets. However, the result of those considerations should not be a refusal to provide all information to the data subject, according to Recital 63 of the GDPR. The controller should find a solution to provide the data subject with the information that he or she may need without disclosing trade secrets.[150]

7.3 Ethics

A report from the recent conference on Computers, Privacy and Data Protection suggests that the European Commission may be considering the possibility of legislating in AI.[151]

For the moment, a first draft of the AI Ethics Guidelines was published. It was presented by the European Commission's High-Level Expert Group on Artificial Intelligence (AI HLEG), an independent group of 52 experts coming from academia, business and civil society.[152] The aim of the document is to set out how developers and users can ensure that AI respects fundamental rights, applicable regulations and core principles, and how the technology can be made technically robust and reliable.[153]

The guidelines set out a framework for trustworthy AI and are structured as follows:

- *"**Chapter I** deals with ensuring AI's ethical purpose, by setting out the fundamental rights, principles and values that it should comply with.*

- *From those principles, **Chapter II** derives guidance on the realisation of Trustworthy AI, tackling both ethical purpose and technical robustness. This is done by listing the requirements for Trustworthy AI and offering an overview of technical and non-technical methods that can be used for its implementation.*

- ***Chapter III** subsequently operationalises the requirements by providing a concrete but non exhaustive assessment list for Trustworthy AI. This list is then adapted to specific use cases[154]."*

[150] The Norwegian Data Protection Authority. (2018, January). Artificial Intelligence and privacy. Retrieved from https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf

[151] https://www.forbes.com/sites/washingtonbytes/2019/02/08/the-eu-should-not-regulate-artificial-intelligence-as-a-separate-technology/#4920b85552c9

[152] https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai

[153] https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai

[154] https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=57112

8 Conclusions

This deliverable provides a comprehensible summary (which will be further refined if needed during time) of the main aspects regarding the GDPR, including an analysis of the type of data involved in voice interaction technologies and an identification of the barriers and requirements to comply with such regulation.

It is essential to know the GDPR requirements before starting a research activity that involves the processing of personal data. Providing an overview of the main aspects of this Regulation and recommendations on how to comply with such requirements is essential.

This document will be useful not only to the COMPRISE project partners, but also to anyone that is interested in starting development or exploitation activities relating to voice-enabled technologies as the document includes the analysis, when needed, regarding the particularities of the processing of personal data when such technologies are involved.

Based on this deliverable, it can be concluded that when techniques such as machine learning (and deep learning) are used to process personal data, fulfilling the requirements and principles of a complex regulation such as the GDPR may become challenging. However, this cannot be an excuse for not complying with the regulation, as organisations may face very high fines if infringements are committed. Several techniques and processes are pointed in this document that may help to be compliant with the data protection regulations.

9 Appendices

9.1 Appendix 1: Categories of personal information and examples list

Categories of personal information	Examples
Knowledge and Belief	Communist, fascist, liberal, leftist
	Christian, Muslim, Jewish, atheist,
	Aesthetic, nihilist, pagan
	I think that_____
	I know that_____
	I believe in___
	My philosophy is_____
	My religion is_____
Authenticating	Card number_____
	Access word____
	Secret word___
	Password_____
	PIN _____
Preference	I like____

	In my opinion____
	I prefer_____
	I'm interested in_____
	I always buy ____
	I want.___
	My favourite_____
	When my (relative/ friend...) died____
	_____happened to me
	When I was in the university_____
	When I was a kid_____
Life History	During my last holidays_____
	PIN __
	Card Holder ___
	Short code ___
	CVV____
Account	Credit card____
	Car brand____
	My house is in___
	I have rented____
	I borrowed_____
	I have a___
Ownership	I own_____
	I've received___euro/dollars/.....
	I've paid_____euros/ dollars/.....
	I've charged_____euros/dollars/....
	I bought it for_____euros/dollars/....
Transactional	I was paid_____euros/dollars/...
	Balance
	Loan
	Red numbers
	Payroll
	Mortgage
	I have ___euros
Credit	Broke
Identifying	Identification document, passport, driving license

	Name, surname, family name
	Nickname
	My name is____
	I'm known as_____
	They call me _____
	Indo-European
	Turkic
	Caucasic
	Austronesian
	Basque
	English
	German
Ethnicity	Spanish
	Heterosexual, straight, homosexual, gay, lesbian
	Vondage, sado, porn, fetiche
	Travestite, transgender
	Asexual
Sexual	Hermaphrodite
	I visit_____website
	I buy through_____
	I practice_____
	I usually____
Behavioural	I always ____
	My age is ____
	Single, married, divorced
	Middle class, high class, working class
	Young, mature, old
Demographic	Employed, unemployed, retired
	Alcoholic, addict
	Disease
	Deaf, blind, physical disability, mental disability.
	Bipolar, depression, schizophrenia
	Pills, tablets, antibiotic, treatment
	The test results show____
Medical and Health	I've been prescribed ____

	I suffer from_____
	Blonde, brown, redhead
	I'm ___ years old
	Man, woman, male, female
Physical Characteristic	Tall, short, skinny, fat
	Lawyer, police, doctor, commercial
	Directive, manager, senior, trainee
	Certificate, Master, speciality
	I was dismissed from____
Professional	I work in_____
	Robbery, burglary, heist.
Criminal	Parole, sentence, conviction, sentence,
	Wife, husband, son, daughters, girlfriend
	Married, divorce,
	My wife____
	My boyfriend_____
	I'm married with___
Family	I'm going out with____
	My friends are___
	My contacts are ___
	I'm a member of_____
	Association
Social Network	I'm a member of the____group
	IP address
Computer Device	Device ID
	My e-mail address is___.
	Telephone number. _____
	Social network_____
	Pager number. _____
Contact	Address (home and work) ___
	Club, restaurant, hospital, shop, cafe
	Altitude and latitude
	My address is___
	I live in___
Location	I am in ____

	I'm close to_____
	I'm visiting_____

9.2 Appendix 2: Recommendations to prepare an information sheet

DATA CONTROLLER
• **Data controller (organization):** • **Address:** • **Phone:** • **E-mail:** • **DPO (e-mail):**
PURPOSES
• **Short explanation of the COMPRISE project (Contextualization**) • **Type of data that will be processed:** For example, speech signal**,** text transcription of the signal, traits, states or other personal data included in the speech signal or text transcription (e.g., age, gender, location, etc.), maybe other data needed for registration, etc. <u>If any sensitive data will be processed, then it should be mentioned and specifically explained why these are needed.</u> **Explanation about the purpose of each processing operation and about how personal data will be processed**: Description of the purpose: In most of the cases the main purposes will be for research or/and demonstration. The research action or the concrete demonstration should be explained as well as the concrete processing operations (what for and how). This should be explained in a comprehensible manner, so the data subject will be able to understand it (transparent, intelligible and easily accessible information). **Inform about automated decision-making or profiling if applies:** Automated processing, including profiling, which produces legal effects concerning the data subject or similarly significantly affects him or her. • **Period:** Indicate for how long the personal data will be stored, or which will be the criteria to determine that period. • **Indicate the security measures that will be implemented and risks**: Explain the dissociation techniques that will be applied, which data will be stored in the device and which in the cloud and main risks (data in the cloud is not totally anonymized). Also, explain additional security measures that will be taken, and other possible risks associated to the data processing.
LEGAL BASES FOR PROCESSING
Inform about the legal basis of the processing. The legal basis for the processing activities carried out within COMPRISE will be the consent given by the data subject to the processing of his or her personal data for one or more specific purposes.

RECIPIENTS

- **The recipients or categories of recipients of personal data, if any:** Inform about data controllers, joint controllers, processors or third parties to whom data is transferred to or disclosed.

  According to the information provided in the questionnaires, the main recipients of the personal data in most cases will be the project partners and worldwide research community. Maybe in other particular cases, such as app demonstrators, there will be other recipients (e.g., the retailer in the case of the e-commerce app).

- **Intention of transferring personal data to a third country or international organisation** and the existence or absence of an adequacy decision by the Commission. In these cases, it is crucial to refer to the appropriate or suitable safeguards and the means by which to obtain a copy of them or where they have been made available.

  If the data will be transferred to a third country for which does not exist an adequacy decision of the Commission and the transfer is not subject to specific safeguards, the data subjects should be informed about the risks associated with the transfers made to these countries.

RIGHTS

- **The rights of the data subject** (GDPR: Chapter III) should be explained including a summary of what the rights involve and how the data subject can exercise them. The rights mentioned in the GDPR are the following:

  - access;
  - rectification;
  - erasure;
  - restriction on processing;
  - objection to processing
  - portability

- **The right to lodge a complaint** with a supervisory authority (in particular in the Member State of his or her habitual residence, place of work or of an alleged infringement of the GDPR) should be explained

- **Explain that the data subject can withdraw the consent at any moment** and how the consent may be withdrawn. Withdrawing the consent should be as easy as giving it for the data subject.

CONSENT

- The legal bases for processing personal data in COMPRISE will be the consent provided by the data subject (consent should be valid, freely given, specific informed and active).
- **Explicit consent should be provided by the data subject when sensitive data will be processed.** In this case, it is recommended to obtain the manuscript or digital signature of the data subject.
- **Explicit consent of the data subject is also needed for making transfers to third countries or international organizations**. This will be mandatory when there is no adequacy decision (Article 45) or appropriate safeguard (Article 46).

9.3 Appendix 3: Threats and measures example list

Note: This is not a closed list. The list includes examples of possible threats (once analysed the probability and impact will become risks) and solutions. However, every specific case should be analysed when trying to identify risks, and the effectiveness of the measures should be assessed to find the ones that should be applied in each particular case.

Some of the measures including examples are inspired in the ones proposed in the ISO 27001 and in the "Guia práctica para las evaluaciones de impacto sujetas a la RGPD (AEPD)"

Threats (Risks)	Measures
Breach of the law as a consequence of the lack of specialised knowledge on data protection.	Provide training to the staff on GDPR and other sectorial regulation (which also regulates data protection aspects).
	Appoint a DPO.
	Appoint a person or a department responsible of data protection issues (if it is not mandatory to appoint a DPO).
Process personal data that is not needed for the purpose for which it is processed (breach of the minimisation principle).	Use anonymization or pseudonymization techniques.
	Examine the intended area of application of the model an consider how to achieve the objective in a way that is the least invasive for the data subject.
	Determine a data collection policy
The consent obtained is not clear or is invalid.	Analyse the possibility of relying in a different legal ground.
	Avoid relying on consent ground if exist an imbalance situation.

Data subjects cannot exercise their rights (access, rectification, erasure, portability, objection, revoke consent)	Implement protocols and policies that should be followed in order to attend data subject requests without undue delay.
	Appoint a point of contact person to attend the requests of the data subject (this person could be the DPO if there is one appointed).
Process sensitive data without the required safeguards.	Analyse whether the processing of sensitive data is absolutely necessary for the purpose intended.
	Assess whether the processing may be justified on any of the legal grounds provided by the GDPR.
	Implement a protocol to ensure that the explicit consent is obtained and to enable to demonstrate it.
Unclear information forms that may confuse the data subjects regarding the purposes and other elements associated to the processing of their personal data.	Implement clear and accessible privacy policies using standardized formats and templates.
Breach of the storage limitation principle because personal data are not erased once they are no longer needed.	Define clear cancellation periods for all personal data storage.
	If possible, use an automatic warning system that is triggered when the defined storage period is coming to an end.
Confidentiality infringements of personal data by the staff.	Provide training to the staff on confidentiality and raise awareness regarding their responsibilities
	Set internal penalties for the employees that infringe confidentiality.
Lack of accountability (not being able to demonstrate compliance with the GDPR)	Implement a documentation policy and protocols to register processing of the operations and document the evidences of compliance with data protection regulations.
Lack of transparency in the processing of personal data which involves models based on deep learning that may make it difficult to provide information to the data subjects.	Have the support of experts (internal or external) with a strong knowledge of AI based systems, that know what to look for and what questions to ask for.
	Try to select models that can explain how they reached a specific result when possible.

	Use notifications, icons, videos, and other visual representations if these can help to explain complex concepts in an easy way.
Machine learning involves the collection of large volumes of data and in many cases some of the collected data is not subsequently used.	Research can be carried out on solutions that use less training data.
	Use anonymization techniques.
When an algorithm is been trained with personal data, the model results could be incorrect or discriminatory if biased data or irrelevant data are used.	Conduct system audits to ensure the system isn't biased, especially audits by third parties.
	Implement, if possible, discrimination detection into the machine learning system to prevent such decisions from being made in the first place.
Humans participating in the machine learning process (e.g., data annotators) disclose confidential information.	A Nnon-Diclosure Agreement should be signed with all the persons that will have access to personal data or other confidential information.
Information Security threats and measures	
Unauthorised access to personal data.	Implement an access policy.
	Implement a user registration and de-registration process.
	Assign passwords to the authorised users and change the passwords periodically.
	Review access rights periodically.
	Use encryption mechanisms.
	Use anonymization and pseudonymization mechanisms.
	Store data locally (less risky than cloud storage).
Damage caused to the information facilities of the organisation as a consequence of an unauthorised physical access	Define security areas
	Protect secure areas with entry controls.
Loss of data that are accidentally deleted	Make backup copies of the information
Data breach caused by a supplier.	Data security should be addressed within the agreements with each supplier that may access store or process personal data.

	Carry out audits of the security systems of the supplier.
Information security cannot be ensured after a cyber-attack.	Implement an Information Security Continuity Plan.
	Implement periodical Information Security continuity controls.

10 Bibliography

Giakomopulus.C, Buttarelli. G, O'Flaherty M. (2018, April). Handbook on European Data Protection Law. Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/5b0cfa83-63f3-11e8-ab9c-01aa75ed71a1

(2018, November 14th). Ethics and data protection. Retrieved from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/ethics/h2020_hi_ethics-data-protection_en.pdf

The Norwegian Data Protection Authority. (2018, January). Artificial Intelligence and privacy. Retrieved from https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf

Information Commissioner's Office. (2017, September). Big data, artificial intelligence, machine learning and data protection. Retrieved from https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf

Agencia Española de Protección de Datos. Guia práctica para las evaluaciones de impacto sujetas a la RGPD.Retrieved from: https://www.aepd.es/media/guias/guia-evaluaciones-de-impacto-rgpd.pdf

The Art 29 Working Party (2007, June 20th). Opinion 4/2007 on the concept of personal data. Retrieved from https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf

The Art 29 Working Party (2003, August 3rd). Working document on biometrics. Retrieved from https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2003/wp80_en.pdf

The Art 29 Working Party (2017, October 3rd). Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053

The Art 29 Working Party (2014, April 10th). Opinion 05/2014 on Anonymisation Techniques. Retrieved from https://www.pdpjournals.com/docs/88197.pdf.

The Art 29 Working Party (2010, February 14th). Opinion 1/2010 on the concepts of "controller" and "processor". Retrieved from https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2010/wp169_en.pdf

The Art 29 Working Party (2017, November 28th). Guidelines on consent under Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=51030

The Art 29 Working Party (2014, April 9th). Opinion 06/2014 on the notion of legitimate interest of the data controller under article 7 of Directive 95/46/EC. Retrieved from https://fia.org/sites/default/files/uploaded/Excerpts%20-%20Opinion%2006-2014%20on%20the%20notion%20of%20legitimate%20interests%20of%20the%20....pdf.

The Art 29 Working Party (2013, April 2nd). Opinion 03/2013 on purpose limitation. Retrieved from https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

The Art 29 Working Party (2017, November 29th). Guidelines on transparency under Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=622227

The Art 29 Working Party (2016, December 13th). Guidelines on Data Protection Officers ('DPOs'). Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612048

The Art 29 Working Party (2017, April 4th). Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679. Retrieved from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236

The Art 29 Working. ANNEX - health data in apps and devices. Retrieved from https://ec.europa.eu/justice/article-29/documentation/other-document/files/2015/20150205_letter_art29wp_ec_health_data_after_plenary_annex_en.pdf

Media. (2018, May 8). GDPR: New rights that we will all enjoy. Retrieved from https://blog.signaturit.com/en/gdpr-new-rights-that-we-will-all-enjoy

Chhavi Rana, R. (2015). A Review: Speech Recognition with Deep Learning Methods [Abstract]. International Journal of Computer Science and Mobile Computing,4(5), 1017. Retrieved from https://www.ijcsmc.com/docs/papers/May2015/V4I5201599a61.pdf.

Lorica, B. (2016, July 14). Commercial speech recognition systems in the age of big data and deep learning. Retrieved from https://www.oreilly.com/ideas/commercial-speech-recognition-systems-in-the-age-of-big-data-and-deep-learning

Gandomi, A., & Haider, M. (2014, December 03). Beyond the hype: Big data concepts, methods, and analytics. Retrieved from https://www.sciencedirect.com/science/article/pii/S0268401214001066

Meyers, J. (2019, January/February). Artificial Intelligence and Trade Secrets. Retrieved from https://www.americanbar.org/groups/intellectual_property_law/publications/landslide/2018-19/january-february/artificial-intelligence-trade-secrets-webinar/

Protection of Databases | Digital Single Market. (n.d.). Retrieved from https://ec.europa.eu/digital-single-market/en/protection-databases

Benschop, L. (n.d.). Could music fingerprinting be copyright infringement? Retrieved from https://lennartb.home.xs4all.nl/fingerprint.html

Cooper, C. (n.d.). How secure is voice recognition technology?, from https://www.business.att.com/learn/tech-advice/how-secure-is-voice-recognition-technology.html#

Voice Assistants and Cyber Security. (2018, December 14). Retrieved from https://www.micomlabs.com/2018/12/14/voice-assistants-and-cyber-security/

Copeland, M. (2016, July 29). The Difference Between AI, Machine Learning, and Deep Learning? | NVIDIA Blog. Retrieved from https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/

Garbade, M. J. (2018, September 14). Clearing the Confusion: AI vs Machine Learning vs Deep Learning Differences. Retrieved, from https://towardsdatascience.com/clearing-the-confusion-ai-vs-machine-learning-vs-deep-learning-differences-fce69b21d5eb

Kikel, C. (2019, April 6). Difference Between Voice Recognition and Speech Recognition. Retrieved from https://www.totalvoicetech.com/difference-between-voice-recognition-and-speech-recognition

Coin, E. (2015, December 07). Introduction to Synthetic Agents: Speech Recognition - Part 1 - DZone Big Data. Retrieved from https://dzone.com/articles/introduction-to-synthetic-agents-speech-recognitio

Acs.G. (2017, November 22). Experts on the GDPR #3: What is personal data under the GDPR? Retrieved from https://tresorit.com/blog/personal-data-under-the-gdpr/

Gray.S. (2016, Apriñ). Experts on the GDPR #3: What is personal data under the GDPR? Always On: Privacy Implications of Microphone Enabled Devices. Retrieved from https://fpf.org/wp-content/uploads/2016/04/FPF_Always_On_WP.pdf

(2015, February 9th) Article 29 Working Party Clarifies Scope of Health Data Processed by Lifestyle and Wellbeing Apps. Retrieved from https://www.huntonprivacyblog.com/2015/02/09/article-29-working-party-clarifies-scope-health-data-processed-lifestyle-wellbeing-apps/

Tatman.R.. (2016, July 12th). Google's speech recognition has a gender bias. Retrieved from https://makingnoiseandhearingthings.com/2016/07/12/googles-speech-recognition-has-a-gender-bias/

Tiwari.A. (2017, July 4th). Bias and Fairness in Machine Learning. Retrieved from https://www.abhishek-tiwari.com/bias-and-fairness-in-machine-learning/

McMahan B., Ramage.D (2017, April 6th) Federated Learning: Collaborative Machine Learning without Centralized Training Data. Retrieved from https://ai.googleblog.com/2017/04/federated-learning-collaborative.html

Gunning D., Ramage.D. Explainable Artificial Intelligence (XAI). Retrieved from https://www.darpa.mil/program/explainable-artificial-intelligence.

Wallace.N, Castro D. (2018, March 17th) The Impact of the EU's New Data Protection Regulation on AI. Retrieved from http://www2.datainnovation.org/2018-impact-gdpr-ai.pdf

Dickson.B. (2019, April 29th) Protecting AI models against audio adversarial attacks. Retrieved from https://bdtechtalks.com/2019/04/29/ai-audio-adversarial-examples/

Bond.T. (2017, June 16th) Artificial Intelligence and IP – Part 1: Developing AI systems. Retrieved from https://digitalbusiness.law/2017/06/artificial-intelligence-and-ip-part-1-developing-ai-systems/

https://www.aware.com/voice-authentication/

https://en.wikipedia.org/wiki/European_Convention_on_Human_Rights

https://europa.eu/european-union/eu-law/legal-acts_en

https://edps.europa.eu/data-protection/our-work/subjects/eprivacy-directive_en

https://iapp.org/media/pdf/resource_center/Categories-of-personal-information.pdf

https://www.twobirds.com/~/media/pdfs/gdpr-pdfs/62--guide-to-the-gdpr--competence-tasks-and-powers.pdf?la=en

https://en.wikipedia.org/wiki/Unsupervised_learning

https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/binding-corporate-rules-bcr_en

https://resources.infosecinstitute.com/guiding-principles-in-information-security/#gref

https://www.iso.org/obp/ui/#iso:std:iso-iec:27001:ed-2:v1:en

https://www.cnet.com/news/google-assistant-just-got-way-better-at-recognizing-which-songs-are-playing/