
# GDPR compliance of processings that embed Artificial Intelligence
# An introduction

(This document is a translation from the original in Spanish **Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial**)

February 2020

# SUMMARY

This guideline is aimed to be a first survey for the compliance to the GDPR of products and services that embed Artificial Intelligence components. The label “Artificial Intelligence” is used for many kinds of technical solutions, one of the most characteristic is Machine Learning or ML. Nowadays, AI is one more component that could be embedded in a processing carried out by a data controller. In many cases, such component has been developed by third parties. AI arise many concerns among users, researchers, professionals, authorities and industry regarding with compliance, human rights, and judicial certainty of the stakeholders. Such concerns and doubts difficult the right technological development.

The current guideline is not just a review of the GDPR, but to address some of such concerns regarding privacy compliance and to point of the more relevant aspects regarding with the design and implementation of IA based processings from the point of view of GDPR.

The guideline pays attention mainly to the legal basis for the processing, information and transparency, rights, automated decisions, profiling, privacy impact assessment and assessment of the proportionality.

This document is aimed to data controllers that embed AI in their processings, developers, data processors and any other one involved in the AI processing.

**Keywords**: Artificial Intelligence, AI, GDPR, LOPDGDD, rights, privacy, ethics, data protection, design, impact assessment, machine learning, ML, automated decisions, profiling, minimization, transparency, accuracy, bias.

# INDEX


I. INTRODUCTION TO THE AI FRAMEWORK AND DATA PROTECTION 5
A. AI Techniques 5
B. Data processing by means of AI solutions 6
C. Data protection and ethical dimension 7
D. GDPR definitions 8
Purpose of the GDPR 8
Personal data 8
Pseudonymisation and anonymisation 8
Special categories of personal data 9
Processing 9
Profiling 9
Automated decisions 10
Users of AI-based solutions 10
Controller 10
Joint controller 10
Processors 10
Exception concerning household activities 11
E. Life cycle of an AI solution 11
F. Personal data processing by means of AI 12
G. Assessment of AI-based solutions 14
H. Short summary of obligations lay down by the GDPR 14
II. ROLES, RELATIONSHIPS AND RESPONSABILITIES 16
III. COMPLIANCE 19
A. Lawfulness and limited purpose 19
Legitimate interest 20
Special categories 21
Processing for compatible purposes 21
B. Information 22
Relevant information on the implemented logic 23
C. General aspects related to the exercise of rights 23
D. Right to access 24
E. Right to erasure 24
Limitations to erasure 25
F. Blocking of data 25
G. Right to rectification 25
H. Portability 26
I. Decision-making based on automated processing 26
IV. PRIVACY RISKS MANAGEMENT 28
A. Risk assessment 28
B. Privacy Impact Assessment-(PIA) 29
C. Transparency 31
During training 31

Certification 31
Automated decisions and profiling 32
Data controller personnel 32
The Data Protection Officer as a tool for transparency 32
D. Accuracy 33
Factors affecting accuracy 33
Biometric information 34
Profiling combination 35
Verification vs. Validation 35
Accuracy assessment as continuous process 35
E. Minimisation 36
Training data 36
Minimisation techniques 37
Extent of the data categories in an AI-based solution 37
Extent of the training set 38
Personal data in the AI-based solution 38
F. Security 39
Specific threats in AI components 39
Logs or activity records 40
G. Assessment of the proportionality and the need for such processing 41
H. Audit 42
V. INTERNATIONAL TRANSFERS 44
VI. CONCLUSIONS 45
VII. REFERENCES 46
VIII. ANNEX: CURRENT AI-BASED SERVICES 48

# I. INTRODUCTION TO THE AI FRAMEWORK AND DATA PROTECTION

The term Artificial Intelligence or AI was used for the first time in 1956 by John McCarthy in reference to "the science and engineering behind the creation of intelligent machines, most especially intelligent computer programmes": The High Level Group for Artificial Intelligence (IA – HLEG) created by the European Commission in order to develop the European strategy on Artificial Intelligence applies this term to any "*systems that display an intelligent behaviour, understood as a capacity to analyse their environment and perform actions, with a certain degree of autonomy, for the purposes of achieving specific goals*"[1]. There are other definitions, although all may be summarized in the notion that AI is the capacity of a machine to act in a way that a human mind would, including the aspect of creativity and capacity to perform complex analyses and inferences[2] from complex and even incomplete information.

Depending on the scope and field of application of artificial intelligence, three different categories of AI may be distinguished: the so-called true AI, general AI and weak AI. General AI could solve any intellectual task which can be solved by a human being, while true AI or superintelligence would go beyond human capacity. The type of AI which has made the practical applications of these practices skyrocket is the discipline known as weak AI, which, unlike true and general AI, is characterised by the capacity to develop solutions for specific, well-defined problems[3]. These kinds of systems have multiple applications: from video games to defence systems, including healthcare, industrial monitoring, robotics, web search engines, natural language processing, marketing, personal assistance, human resources, optimization of public services, energy management, environmental management and any other activity we may imagine[4]. Therefore, the scope of application of AI solutions is extended to all sectors, each with their specific characteristics and the requirement to comply with regulations both at general and sector levels.

Implementation of Artificial Intelligence created doubts among users, researchers, experts, authorities and industries regarding compliance aspects, respect of the data subject rights and the legal certainty of all stakeholders. These doubts may be an obstacle for the appropriate technological development, and therefore it is necessary to develop guidelines and manuals which address such problems. In this document, we will focus in ensuring that any processing including weak AI components are compliant with the GDPR.

## A. AI TECHNIQUES

There are different approaches to AI-based solutions: neural networks, rule-based systems, fuzzy logic, machine learning, expert systems, adaptive systems, genetic algorithms, multi-agent systems, etc., terms which, usually, overlap[5]. All such techniques are designed to achieve models capable of addressing complex systems which cannot (or we do not know how to) address by means of sequential algorithms. Such difficulties could be due to the complexity to model a behaviour that is function of multiple variables, to deal with non-linear relationships (difficult to approximate by linear methods) and which even change over time.

One of the most successful branches of AI in commercial applications is Machine Learning (ML). ML designs predictive models that are capable to model the relationship between the

[1] Artificial Intelligence for Europe, European Commission
[2] Webster: Inference: (1) something that is inferred; (2) the act or process of inferring Infer: to derive as a conclusion from facts or premises
[3] In this case, inference may be used to solve a classification problem, such as identifying suspicious individuals; or a clustering problem, such as recommending a product to a customer based on their purchase history; or a regression programme (value estimation), such as detecting the voting intention of a community.
[4] In AI Index 2018 Annual Report http://cdn.aiindex.org/2018/AI%20Index%202018%20Annual%20Report.pdf a report on the scope and market of IA components can be found
[5] An expert system may be created by modelling knowledge of a series of experts by applying rules of fuzzy logic.

variables by means of analysing an original set of data, identifying patterns and setting classification criteria. Once the relevant criteria are set, the AI component is able of making an inference from a new data set. Machine learning is, therefore, related to data mining, optimization and big data techniques[6]. There are different types of machine learning (supervised learning[7], unsupervised learning[8], reinforced learning[9] and its variants), and each uses own techniques. Besides, there are specialist variants of ML as such as Deep Learning[10], and different learning models: centralized learning, decentralized learning or federated learning. Given an AI component, we may conclude that it is an adaptive system when the inference model is dynamically adjusted for each new dataset input, thus fine-tuning the already established relationship.

## B. DATA PROCESSING BY MEANS OF AI SOLUTIONS

The development and operation of an AI solution may involve processing data from natural persons, as may be the case on a marketing or vote profiling model, or it may involve processing data of a non-personal nature, as in a weather predictive model which collects data from geographically distributed weather stations.

Any automated process may affect natural persons (e.g. an authentication system) or not (e.g. an industrial monitoring system). When the relevant automated process makes decisions concerning natural persons, these decision may be related to the interaction of such person with its social context, such as their access to a certain product or service, or to the personalization of such service, as it is the case with AI applications for management of a vehicle or customize a television. Decisions may be of a predictive nature with regard the evolution of this subject, may assess the current status of such person or may decide whether a set of actions is to be carried out regarding him[11].

Besides, AI can perform two roles in the decision-making process:

- It can support the decision-making process by providing an inference or profile with regard to an individual or situation and leave the final decision to a human being.
- It may make and implement the relevant decision.

In any of the two scenarios above, the AI solution never acts in isolation; it is always a part of a wider processing. Besides, it shall be physically implemented in a system in which other applications are implemented, there will be data communications among processings, user interfaces, etc. The synergy which inherently occurs in the implementations including AI components shall relate those AI-based processing with big data, Internet of Things, 5G mobile systems Edge Computing[12] and Cloud Computing elements. Therefore, many technical components, but also human factors, take part either directly on in a collateral way, and they must be taken into account in order to determine the implications of using a processing which includes AI elements.

Finally, it must be taken into account that not all systems which make automatic decisions are AI systems; not all AI is machine leaning and not everything labelled as AI is actually AI[13],

[6] However, it is possible to perform ML on Small Data provided that it is performed on a dataset with high predictive power.
[7] The system is trained by means of a set of examples in which the output is known. Thus, algorithms work on labelled data to try to find a function which, provided the input variables, assigns such date the appropriate output label.
[8] Intends to extract significant information without any reference to known output variables by means of exploring the structure of such unlabelled data. They are based on unlabelled data for which there is no information regarding their classification or continuous dependent event. The role of the algorithm is to find out the internal structure of this dataset.
[9] There is human intervention in the learning process in order to reward or punish partial interventions.
[10] ML technique which uses multiple non-linear processing layers, and in which each layer is adapted for capture by means of learning a certain characteristic. These layers are organized in a hierarchy from the lowest to the highest abstraction level.
[11] For example, on a healthcare context, it may predict the evolution of a patient, perform an automatic diagnosis after an exploration or prescribe a specific treatment.
[12] Processing technique carried out at or near the physical location of the user device or the data source.
[13] https://www.elmundo.es/tecnologia/2019/03/12/5c829c0d21efa0760a8b45fa.html

but may be simply labelled as such for marketing purposes or to implement other types of business strategy.

## C. DATA PROTECTION AND ETHICAL DIMENSION

Protection of natural persons with regard to the processing of their personal data constitutes a fundamental right. Article 8, section 1, of the Charter of the Fundamental Rights of the European Union, and article 16, section 1, of the Treaty on the Functioning of the European Union establishes that everyone has the right to the protection of personal data concerning them.[14]. The fundamental right to the protection of personal data is developed in a regulatory framework which currently includes Regulation (EU) 2016/679 of the European Parliament and of the Council, of 27 April 2016, on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), complemented by Organic Act 3/2018, of 5 December, on Protection of Personal Data and Guarantee of Digital Rights (LOPDGDD), besides any sector-specific regulations entered into force before or after the enactment of the GDPR.

One of the most concerning aspects of the implementation of AI is its ethical dimension, as part of the "digital ethics".[15]. The ethical approach to AI pursues protecting values such as dignity, freedom, democracy, equality, autonomy for all individuals and justice regarding with machine-ruling. The European Commission is working on the definition of a trustworthy Artificial Intelligence and it has established that, to be considered as such, it must comply with seven key requirements: human action and supervision, technical security and robustness, data and privacy management, transparency, diversity, non-discrimination and equity, social and environmental well-being and accountability.

Those requirements must be assessed throughout the life cycle of an AI system in a continuous manner, and require constant surveillance both with regard to the ethical legitimation and any unexpected outcome of legitimate processing activities, and to the collateral impact of such processing activities in a social environment, beyond the intended purpose, duration and scope.

That is, the AI solutions must be analysed both by itself and into the framework of the processing where it is integrated, as well as the relationship of such processing with its context:

- Regarding with the ethical values and principles that rely in cultural differences.
- Regarding quality requirements due to the context in which the relevant service is deployed.
- Regarding with any other aspects arising from the massive interconnection of components and processings.

A critical aspect of AI systems is the possible existence of biases. A bias is an appropriate deviation in an inference process, and it is particularly concerning when, for example, it results in the discrimination a group in favour of another[16]. This problem is not exclusive to AI systems, but general to all decision-making processes, also those human-based or automated without AI.

However, there are other types of bias which may be even more concerning, and that is the bias when interpreting the AI outputs. The human bias means to assume AI results as

[14] Recital 1 of the GDPR.

[15] Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy, Cathy O'Neil, Broadway Books 2016

[16] *''systematically and unfairly discriminate against certain individuals or groups of individuals in favour of others. A system discriminates unfairly if it denies an opportunity or a good or if it assigns an undesirable outcome to an individual or group of individuals on grounds that are unreasonable or inappropriate''* (Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. ACM Transactions on Information Systems (TOIS), 14(3), 330-347).

true and certain without a critical analysis, and vesting with a infallibility principle which arises from the expectations created by such systems.

The scope and depth of ethical values is largely rooted in the cultural environment in which they are developed as persons. The private life of individuals constitutes an important aspect of ethical principles. In the same manner, the perception of such values held by each society can be vastly different[17], although certain worldwide accepted principles have been gradually agreed at international level[18].

## D. GDPR DEFINITIONS

The GDPR, mainly on its article 4, establishes a framework of concepts and definitions. Notwithstanding the reference to such article, some of them are developed below because they are of special interest with relationship to AI:

### Purpose of the GDPR

Article 1 of GDPR establishes that the goal of this regulation is to protect the fundamental rights and freedoms of natural persons, specifically in relation to the protection of their personal data[19].

### Personal data

Personal data are defined by article 4.1. of the GDPR 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".*

This definitions is developed in depth in Opinion 4/2007 on the concept of personal data of the Article 29 Working Party (WP29).

### Pseudonymisation and anonymisation

Pseudonymisation is defined in article 4.5 of the GDPR as *"the 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".*

Anonymisation is the process that allows to remove or minimize the risk of reidentification of an individual based to their personal data, by removing any direct or indirect reference to their identity but maintaining the accuracy of the results of the processing of such data. It

[17] In some countries, access to tax-related information is one of the most protected personal data, while in other countries it is considered public information. "Norway: The country where no salaries are secret" https://www.bbc.com/news/magazine-40669239

[18] Article 12 of the Universal Declaration of Human Rights. Convention for the Protection of Human Rights and Fundamental Freedoms (1950). International Covenant on Civil and Political Rights (1966) Council of Europe Resolution 509 on human rights and modern scientific and technological developments. OECD's Recommendation concerning guidelines governing the protection of privacy and transborder flows of personal data (1980; updated 2002). Council of Europe Convention 108 for the Protection of Individuals with regard to Automatic Processing of Personal Data (1981). Resolution of the General Assembly of United Nations, of 14 January 1990, concerning the Guidelines for the Regulation of Computerized Personal Data Files.

[19] Article 1 Subject-matter and objective 1. This Regulation lays 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 2. This Regulation protects fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data. …

would be ideal that, besides preventing identification of persons, processing of anonymised data keeps a suitable quality of the results[20].

Both definitions are complemented by the provisions of the WP2) Opinion 05/2014 on Anonymisation Techniques and the Guidelines and guarantees for the processes of anonymisation of personal data published by the AEPD.

**Special categories of personal data**

Article 9 of the GDPR lays down that special categories of personal data shall be those *"revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data[21], biometric data[22] for the purpose of uniquely identifying a natural person, data concerning health[23] or data concerning a natural person's sex life or sexual orientation."*

The aforementioned article lays down a generic prohibition to process such data. Such prohibition is extended in article 9 of the LOPDGDD, unless the requirements described in sections 2, 3 and 4 of article 9 of the GDPR and sections 1 and 2 of article 9 of the LOPDGDD.

**Processing**

Article 4.2 of the GDPR means 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.

Profiling (as defined below) and decision making regarding a natural person is considered processing according to Whereas 24[24] and 72[25] of the GDPR.

**Profiling**

Article 4.4. of the GDPR defines profiling as any form of automated processing of personal data involving the use of personal data to assess certain personal aspects referring a natural person, specially to analyse or predict aspects concerning that natural person[26].

Any processing involving profiling is characterised by three elements[27]:

- It must consist on an automated processing, including those processing activities where human beings are partially involved.
- It must be performed in reference to personal data;

[20] "Guidelines and guarantees for the processes of anonymisation of personal data" (AEPD 2016)

[21] Article 4.13: personal data relating to the inherited or acquired genetic characteristics of a natural person which give unique information about the physiology or the health of that natural person and which result, in particular, from an analysis of a biological sample from the natural person in question.

[22] Article 4.14: personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data.

[23] Article 4.15: 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.

[24] Recital 24 of the GDPR: " (...) 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 analysing or predicting her or his personal preferences, behaviours and attitudes."

[25] Recital 72 Profiling is subject to the rules of this Regulation governing the processing of personal data, such as the legal grounds for processing or data protection principles. The European Data Protection Board established by this Regulation (the 'Board') should be able to issue guidance in that context.

[26] Article 4.4 «profiling»: any form of 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 movements;.

[27] Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. Article 29 Working Party

- The goal of such profiling must be to assess personal aspects of a natural person.

**Automated decisions**

Decisions based solely on automated processing represent the ability to make decisions by means of technological solutions without human intervention[28]. Whereas 71 and 71 and article 22 of the GDPR limits and lays down the right of the data subject not being subdued to decisions based solely on automated processing [29] which have legal consequences of significantly affect the data subject. Automatic profiling is included in this framework of automated decisions. The Article 29 Working Party has assessed the implications of this right on the WP29 Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679.

However, it must be taking into account that automated decisions may involve profiling or not, and profiling may involve automated decisions or not.

**Users of AI-based solutions**

From the perspective of data protection, users of AI-based processing activities may be classified as follows:

- Organizations that use AI solutions on data from data subjects (employees, clients or others); for example, a company that uses AI in order to determine marketing policies based on its clients' preferences.
- Natural persons who purchase a product or subscribe to a service that includes an AI component for the purposes of processing their own personal data; for example, a person who purchases an activity bands in order to plan their training.

**Controller**

Data controller is defined by article 4.7[30] of the GDPR as the person who determines the purposes and means of the processing of personal data, and the scope of their obligations, as defined by article 24, include, among others, to "*implement appropriate technical and organisational measures to ensure and to be able to demonstrate that processing is performed in accordance with this Regulation*"

**Joint controller**

The GDPR includes, in its article 26, the figures of the processing controller as those (two or more) controllers that jointly determine the purposes and means of processing. Joint controllers shall mutually agree in a transparent manner their respective responsibilities in compliance of the obligations imposed by the GDPR.

**Processors**

Data processors are those persons who process personal data on behalf of the controller, as established by article 4.8[31] of the GDPR. The scope of their obligations is defined by article

[28] Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679

[29] In order to consider that there has been human intervention, the supervision of the decision must be carried out by an competent person authorised to modify the decision, and their intervention must be significant and not purely symbolic.

[30] Article 4.7 «data controller» or «controller»: means the natural or legal person, public authority, agency or other 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.

[31] Article 4.8 «data processor» or «processor»: means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller.

28, which states, among other aspects, that their relationship with the controller "*shall be governed by a contract or other legal act under Union or Member State law, that is binding on the processor with regard to the controller*".

**Exception concerning household activities**

Article 2.c. lays down, as further developed in Whereas 18, that the provisions of GDPR of not apply when the relevant processing is performed by a natural person in the course of a purely personal or household activity.

Such activities may be, for example "*correspondence and the holding of addresses, or social networking and online activity undertaken within the context of such activities*" and, in general, those "*with no connection to a professional or commercial activity*".

However, it must be considered that the exception concerning household activities does not apply to processing controllers or processor who provide the necessary means to process personal data with regard to such activities.

## E. LIFE CYCLE OF AN AI SOLUTION

Once some relevant framework concepts have been defined, we can proceed to analyse those processing activities which include an AI-based solution or an AI-component.

An AI-based solution means an element within a data processing which shall be included in at least one of the phases of a processing. In some cases, the AI component shall be specifically developed for such processing. In many other cases, such component shall be de developed by a third party other than the controller[32]. The AI component is not isolated but integrated in a specific processing with other components, such as data collecting, file systems, security modules, user interfaces etc. Indeed, once in the market, an AI component may be integrated in processing carried out by different controllers[33].

The life cycle of an AI system[34], from its creation to its removal, shall go through different stages, which are shared by all technological developments, but which, depending on the specific AI technology, could have certain specificities. Those stages are:

- Design and analysis, in which functional and non-functional requirements for the AI-based solution are set. Those shall be established by the business goals arising from the processing where the AI component is to be included or the market where it intends to be marketed. It shall include project planning, legal requirements, etc.
- Development, including research, prototyping, design, testing, training and validation. Not all these phases will be present in all cases, and the presence or not of a particular phase will depend on the specific AI-based solution adopted. For example, the training phase shall be included for ML-based AI components.
- Operation, including the implementation of certain actions, some of which could be performed in parallel: integration, production, deployment, inference, decision, maintenance and evolution[35].
- Final removal of processing or component.

[32] Such as the development of chat bots or customer care conversational platforms.

[33] Amazon Web Services offers AI modules to be included in processing procedures with, as displayed in their web page "No machine learning required" https://aws.amazon.com/es/machine-learning/ai-services/.

[34] The stages in the life cycle of an AI component, which is a component in the processing procedure, with the stages in which processing may be divided for analysing purposes. The AEPD's The Practical Guide for Data Protection Impact Assessments exposes the division of processing into phases from a data-based approach into data collection, classification/storage, assignment of data to third party and data destruction. This is a general division, an initial approach to analyse d a processing set up in operation. Each processing design must adapt thus general division to its specific conditions. For example, the stage of development of an ML-based AI component is a processing composed of different phases which are to be analysed.

[35] Evolutive maintenance

It must not be forgotten that, for the purposes of preparing this document, we have focused on the AI component, but when conducting an analysis of the processing, the AI component and all other component included in such processing must be assessed as a whole. Besides, depending on the type of applications, some of the stages described above might overlap. For example, validation may be overlapped during the development and operation stages, or evolution may be carried out simultaneously to inference.

## F. PERSONAL DATA PROCESSING BY MEANS OF AI

Considering a wide-ranging approach to possible processing activities by AI-based solutions[36], personal data may be found in the following stages of the AI-based solution life cycle:

- **Training**: considering an AI model based, for example, in ML, personal data may be used in the AI development. In other scenarios, as would be the case with a training system based on capturing knowledge from an expert, it may be considered that, in theory, no personal data processing activities are carried out. When the training stage involves personal data processing, it shall be considered, in itself, a processing of personal data. At its highest extent, it may include the following activities: defining, searching and obtaining the relevant dataset[37], data pre-processing (processing of unstructured data, data cleansing, balancing, selecting, transforming), splitting the dataset for verification purposes, information on traceability and audit.
- **Validation[38]**: this stage may include personal data processing when using data that correspond to the actual current situation of a processing activity in order to determine the eligibility of the experimental model. This dataset may be different from those used in the training stage (if present and if using personal data) and may even be carried out by a third party for model auditing or certification purposes.
- **Deployment**: in case that the AI-based solution is distributed to third parties as a component, it may be considered that there is data disclosure when the same model includes personal data or there is a way to obtain such personal data. Some AI-based solutions, such as Support Vector Machines (SVM[39]) may include examples of the training data within the model logic. In other cases, it may be found that the model has defined patterns that identify a natural person.
- **Operation**: it is possible that some of the operation activities of the AI-based solution includes personal data processing:
    - **Inference**: when the data subject's data are used to obtain certain results, or when third-party data are used for the same purposes, or when inferences or data subject's data are stored. If the data subject himself has an AI-component of their own, the exception concerning household activities would apply.
    - **Decision-making**: as seen before, any decision regarding a data subject is considered personal data processing.

[36] As stated before, not all will be present on the implementation of an AI-based solution.
[37] This is part of the data mining activities.
[38] Validation may be conducted or complemented by other methods, for example, analytic methods. The approach adopted by this text is focused on the use of personal data.
[39] Support Vector Machines or SVMs are a set of supervised learning algorithms. Those methods are related to classification and regression challenges. Given a set of examples for training purposes, labelled into classes, it is possible to train a SVM to build a model predicting the classes in which a new dataset will be divided. An intuitive description of an SVM is a model which represents sample data points in space, separating categories with two spaces, as far apart set as possible, by means of a separating hyperplane defined as the vector between the two closest data points for the two classes, which is called the support vector.

- **Evolution**: the AI-based solution may use data and results of the data subject to fine-tune the AI model. When it is found that this fine-tuning is performed in the AI component owned by the data subject themselves, in an isolated and autonomous way, the exception concerning household activities would apply. However, if such data are sent to a third party, it would be considered an event of data disclosure, possible storage processing, model-changing processing or could even give rise to new disclosures if such data are incorporated to the model and this model is available to third parties.

- **Removal**: service removal may include two different extensions: the AI component is removed when obsolete in all processing in which it is implemented, or a particular user decides not to use any longer the AI component. This user may be an entity or a natural person and may have effects on local, centralized or distributed data suppression, as well as service portability.

A reminder must be made that not all AI-based solutions process personal data at any stage of their life cycles, or make decisions solely based on automated processing affecting natural persons[40]. Some examples of such solutions are quality control systems for industrial products, or those algorithms used for decision making with regard to sale and high level manage of financial products.

If an AI processes personal data, conducts profiling on natural persons or makes decisions regarding such natural persons, their activities shall be subject to the provisions lay down by the GDPR. If it does not, it shall not be subject to such provisions. Concluding whether at a certain stage of the AI life cycle does not involve processing personal data may seem trivial, but, in some cases, it may not be that simple.

When personal data are somehow involved somehow during the life cycle of an AI-based solution (for example during their development) before stating that in further stages there is no personal data processing activities and, as a consequence, at least that processing stage is not subject to the obligations provided by the GDPR, evidence must be provided that the removal or anonymisation of such data is truly effective, and an assessment must be carried out in order to determine the potential risk of reidentification.

The operation stage of the system including an AI-based solution may present several scenarios:

- The owner of the AI-based system grants access directly to the data subject whose data are being processed by means of AI, for example, in case that the psychological evaluation of Facebook users carried out by Facebook in certain countries[41].

- The owner of the AI-based system decides to transfer the use rights of such AI solution as a component to a third party. The relevant AI component is thus treated as an off-the-shelf element by a controller. The owner is not involved in the system operation[42]. This could be the case with an AI-based driving assistant acquired as a component by a car manufacturer in order to have it included in their vehicles[43].

- The owner of the AI-based system is contracted by the user of the AI system in order to access a certain service, so as the owner of the AI effectively implements

[40] Using AI-based solutions for decision making in a financial environment may have serious legal consequences https://www.bloomberg.com/news/articles/2019-05-06/who-to-sue-when-a-robot-loses-your-fortune

[41] https://about.fb.com/news/2018/09/inside-feed-suicide-prevention-and-ai/

[42] Although they may be not disassociated from the system evolution.

[43] https://igniteoutsourcing.com/automotive/artificial-intelligence-in-automotive-industry/

one of the processings phases according to the order and instructions of the controller of the data subject's data[44].

## G. ASSESSMENT OF AI-BASED SOLUTIONS

Both the AI model including in a processing and the processing itself must have the purpose of providing a response to a real need of the entity and the industry. In this paper we are dealing with IA-solutions that exceed the experimental scope and are to be subject to the rules of the market, specifically a regulated market which is obliged to comply with certain quality standards and regulations[45].

With regard to their capacity to comply with processing requirements, all technical solutions share certain parameters which must be specified, such as, for example:

- Accuracy, precision or error rates required by the relevant processing[46].
- Data input quality requirements to the AI component.
- Precision, accuracy or effective error rates of the AI-based solution depending on the appropriate metrics to measure the eligibility of such AI-based solution[47].
- Convergence of the model when training or evolve the IA-solution.
- Consistency in the results of the inference process.
- Algorithm predictability[48] [49]
- And any other assessment parameters of the AI component.

Any technical solution which does not provide a certified answer to those issues cannot be considered a mature technology, but merely a technology without the ability to comply with basic requirements of accountability, transparency and legality. Besides, the answer to such performance issues will be the input to define some data protection requirements, for example, regarding with the implementation of the minimization principle.

## H. SHORT SUMMARY OF OBLIGATIONS LAY DOWN BY THE GDPR

The goal of this document is not to reproduce the contents of the GDPR, and therefore the text of the regulation must be used as a reference. Briefly, the GDPR is developed around six principles established in Chapter II:

1. Lawfulness, fairness and transparency
2. Purpose limitation (purpose specification)
3. Data minimisation
4. Accuracy
5. Storage limitation
6. Integrity and confidentiality

The same chapter establishes the conditions that make a personal data processing activity legitimate.

---

[44] https://es.wordpress.org/plugins/tags/captcha/

[45] What's your ML Test Score? A rubric for ML production systems, Eric Breck, 2018, Google Inc. https://www.eecs.tufts.edu/~dsculley/papers/ml_test_score.pdf

[46] These measures shall depend on whether the type of AI used is used for classification purposes, regression purposes or others. Precision refers to the dispersions of the set of values obtained by repeatedly measuring a magnitude. A common measure of variability is standard deviation of measurements. In classification models precision can be assessed as the proportion of correctly classified instances or instances with ROC curves. Accuracy is related to the bias of an estimation, and it refers to how close is the measured value to the actual value. Accuracy can be modelled with the parameters of total error, mean error, mean absolute error or mean squared error.

[47] Its precision and accuracy may or not exceed the processing requirements.

[48] Predictability allows to obtain an accurate estimation of the behaviour of the AI component under certain specified conditions. Although the halting problem states that there is no automatically computable manner of knowing whether all possible programmes will finish running, it is not denied that there are testing methods applicable to analysis of specific components.

[49] Factual unpredictability of an algorithm caused by a lack of analysis of such algorithm must not be confused with lack of knowledge of internal status, and lack of exhaustive testing procedures must not be confused with randomness.

Chapter III establishes the rights to which data subjects are entitled, as well as the obligation to guarantee such rights and implement effective mechanisms to exercise such rights, especially the rights to transparency, information, access, rectification, erasure, limitation, opposition, portability, and the rights of citizens with regard to automated decision-making, which is especially relevant when AI solutions are implemented.

Chapter IV establishes the model of responsibility and compliance based on accountability, and whose governing elements are:

- Identification of an element of responsibility in processing.
- Risk analysis with regard to rights and freedoms.
- Assessment of necessity and proportionality of processing activities considering their goal.
- Deployment of risk management measures, privacy by default and design privacy measures, security measures, incident management, etc.

And to conclude this brief summary, Chapter C establishes the conditions to carry out transfers of personal data to third countries or international organizations.

## II. ROLES, RELATIONSHIPS AND RESPONSABILITIES

One of the key aspects of the GDPR, which is central to determine whether accountability and transparency policies have been correctly implemented, is to duly identify the different roles and responsibilities within the processing.

In the different stages of the life cycle of an AI component, the personal data controllers shall be any natural or legal person, public authority or other organisation who makes the decision to conduct the aforementioned data processing as defined above[50]. Therefore, different responsibilities shall entail different obligations in the framework of data processing. It is possible that the aforementioned controller outsources different tasks to third parties, who shall carry out these tasks on the controller's behalf and according to their instructions. Such third parties shall be considered as data processors provided that they carry out any personal data processing activity under the instructions of the data controller. Any other additional processing of such data conducted for their own purposes shall make them the controller with regard to such processing.

The different stages of the life cycle of an AI system may involve different controllers and processors, as well as scenarios of data disclosure[51] between controllers:

| **Stage** | **Controller** | **Processors** |
|---|---|---|
| Development/Training | The organization that defines the purpose of the AI component and decides which data are to be used to train the system.<br>If development is outsourced to a third person, and it is this third person who makes any decision on the personal data used to train the AI component for their own purposes, they shall have the status of data controller.<br>If the person who defines the purposes of the AI component acquires a personal data set, this person shall have the status of data controller. | The organization to which development or training are outsourced, provided that the contracting party defines the terms that fix the purpose of the processing activity and the significant characteristics of data, regardless of whether the contracting party transfer to the contractor the relevant data or the contractor obtains such relevant data by their own means, and the data processor uses such data exclusively for the purposes established by the data controller. |
| Validation | Same as above | Same as above |
| Deployment/releasing | If the AI-based solution is a component [52] sold/rented to another entity (the IA could be part of an application) and this component includes personal data, both entities carried out a | Any entity that makes the access to AI-based model available to a controller so that it works in in the context of service provision, or any entity that does so because it |

[50] Please refer to the definition of "joint controller" as stated on this document and on article 26 of the GDPR in case that the relevant decisions are made by more than one person.
[51] Article 4.2 defines "…disclosure by transmission, dissemination or otherwise making available, alignment or combination…" also as data processing.
[52] Or are included in a more complex product, in which the AI component is just another element.

| | | |
|---|---|---|
| | data communication and both are controllers.<br>If the goal is to commercialise a product which includes an AI component to a natural person for private use, even if the model includes personal data, the exception concerning household activities shall apply, unless that they process the personal data included for their own purposes, in which case they shall also have the status of data controller. | is necessary for the appropriate execution of this service but does not use personal data for their own purposes. |
| Inference/profiling | The organization that decides to process the data subject's data by means of the AI system for their own purposes.<br>If the relevant processing is carried out by a natural person regarding their own personal data or to the personal data of their immediate environment for an exclusively personal or household activity, the exception concerning household activities shall apply. However, this exception shall not apply to those who provide the means to process personal data with regard to such personal or household activities[53] for their own means. | Same as above |
| Decision-making | The entity that carries out automated decision-making with regard to data subjects for their own purposes. | Same as above |
| Evolution | The entity that decides to process data from the data subject by means of the AI system, when it discloses user's data to a third organization, it shall be the controller with respect to the data disclosure | In case that the entity which decides to process personal data outsources AI-based processing to a third party, such third party shall have the status of data controller, provided that they do not |

[53] Recital 18 This Regulation does not apply to the processing of personal data by a natural person in the course of a purely personal or household activity and thus with no connection to a professional or commercial activity. Personal or household activities could include correspondence and the holding of addresses, or social networking and on-line activity undertaken within the context of such activities. However, this Regulation applies to controllers or processors which provide the means for processing personal data for such personal or household activities

| | | |
|---|---|---|
| | therefore, it is not a controller-processor relation.<br><br>The organization that determines evolution of the AI component based on user data shall be responsible for the corresponding evolution or retention processing, regardless of whether such data have been released directly by the subject or obtained through service provider. | process data for their own purposes. |

The controller-processor model may become more complicated when, for example, blockchain-type cooperative networks incorporating AI models could appear. The table above intends to address the most common cases and be used as guidelines to new situations which may arise in the market. Neither such table depicts the relationship between the stakeholders who may collect data in the framework of Big Data in order to put these data in the hands of developers. Particularly, joint controller models are not specifically analysed.

We must be aware that the issue of responsibility is being addressed from the data protection perspective, without further legal or ethical considerations which may be derived from using AI solutions.

The decision to adopt, in the frame of a processing activity, a technical solution based on AI or other technology, is made by the controller, since it is the controller who "*determines the means and purposes of processing*", and, therefore, in charge of making the decision of which technological or other solution is to be used. Such controller is obliged to act with due diligence when selecting the most correct solution, particularly when outsourcing or acquiring the relevant development[54]; they shall require and assess any quality specifications of the relevant solution, establish the scope of the processing and take the burden of facing the consequences of their decisions. The person in charge of conducting the processing shall be the person responsible, and they may not claim that they lacked information or technical knowledge to elude their responsibility to audit and decide whether a system is eligible.

Transferring the responsibility to the AI system itself is not acceptable in any case.

[54] Article 28.1 of the GDPR " Where processing is to be carried out on behalf of a controller, the controller shall use only processors providing sufficient guarantees to implement appropriate technical and organisational measures in such a manner that processing will meet the requirements of this Regulation and ensure the protection of the rights of the data subject."

# III. COMPLIANCE

The GDPR is a tool which may provide a high degree of flexibility in order to be able to guarantee and evidence the compliance of a certain processing with the law. However, it must be considered that there is a minimum set of terms and conditions that need to be fulfilled in order to guarantee that the processing is compliant, including:

- Grounds for a legitimate personal data processing (articles 6 to 11 of the GDPR).
- Obligation to inform data subjects and to be transparent in this respect (articles 12 to 14 of the GDPR).
- Obligation to provide data subjects with mechanisms to exercise their rights (articles 15 to 23 of the GDPR).
- Application of the principle of accountability (articles 24 to 43) that establish the need to include a series of additional guarantees, beyond a minimum, documented and oriented to manage the risks posed to the rights and freedoms of individuals); particularly, the obligation to keep a registry of processing activities (article 30 of the GDPR).
- Compliance with all terms and conditions to carry out international data transfers (articles 44 to 50 of the GDPR).

Referenced GDPR articles are further developed in the Regulatory Compliance Check-list guidelines published by AEPD. The key aspects to be considered when defining a processing that uses AI-based solutions in order to guarantee that it respects all principles established by the GDPR shall be described in detail in this chapter, and, in the following chapters, other aspects such as accountability and international transfers shall be covered.

## A. LAWFULNESS AND LIMITED PURPOSE

Establishing a legitimating legal base is the first step that needs to be taken to ensure that the AI-based solution is compliant with the GDPR. Lawfulness the personal data processing has to be considered before the design phase of the processing, regardless of whether such processing consists creating an AI component or to implement an AI component in the processing. From a Data Protection Approach, legitimacy is the first element that needs to be established within the stage of processing design. If no legitimate grounds can be found to proceed with the processing, such processing cannot be carried out.

In previous sections of this documents, several stages of the life cycle of an AI-based solution which may involve personal data processing have been listed. Each of such stages have different goals. They could process data subject's data to which the relevant service is being provided, and it is possible to also proceed third-subjects' data.

Due to the own nature of AI systems, in each stage of their life cycles different legal grounds could be used for:

- Training and/or model validation
- Use of third data subjects' data for inference purposes
- The disclosure of data implicit in this model
- The data subject's data processing in the framework of the service provided by the AI.
- Processing the data subject's data for the purposes of evolution of the model.

Article 6 of the GDPR establishes the six legal bases in virtue of which the processing of personal data may be considered lawful. The most frequent legal grounds claimed to consider that an activity involving data processing by means of an AI-based solution is lawful are:

- Processing the relevant data is necessary in order to execute a contract to which the data subject is a party, or in order to implement pre-contractual measures at the request of said data subject. This may be the case of developers who

outsource the system training stage to third parties which are allowed to use their personal data, or that the organization acting as data controller and which provides to third parties the relevant service including an AI-based solution uses the data of such third parties in the framework of such service agreement.

- Legitimate interest, provided that it does not prevail over fundamental rights or freedoms of the data subject which involve protection of their personal data, especially when such data subject is a child.
- Consent of the data subject, which, as established by article 4.11 of the GDPR, is 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.

And, in some specific cases, from the point of view of AI-based solution, the following can also constitute legal grounds.

- Protection of vital interest[55].
- Grounds of public interest or the exercise or public authority[56].
- Compliance with legal obligations.

It is very important to take into account that the two last legal grounds must be rooted in the legal framework of the European Union or of one of its Member States, which shall ley down the legal grounds for data processing. That is, a private entity may not claim public interest grounds if this is not laid down in a law.

Another important aspect to be considered is the principle of limited processing. This means that having legal grounds to process data does not authorise to process such data for any purpose and at any time whatsoever.[57] Contrary to that, the relevant processing must be limited to those express, specific and limited purposes that have been identified, and an effort shall be made to avoid any processing which is incompatible to such purposes. Besides, the data subject whose data are to be process must be aware of the uses of such data, which is closely related with the principle of information and transparency.

Extinction of a legitimating legal grounds, such as the removal of consent, does not have a retroactive effect regarding the results obtained in a processing activity already completed. For example, when personal data have been used to train an AI component, the applicable legal grounds do not invalidate the operation of this model, but the data controller must pay attention to any requests of exercises of rights with regard to data protection.

If the data controller uses third party data sets for training an AI component, it must act with due diligence for verifying that the original data source is legitimate, including the purchase or service agreement, or the corresponding contractual clauses that claim evidences and commitments of such legitimacy.

**Legitimate interest**

The Article 29 Working Party Opinion 06/2014 on the notion of legitimate interests of the data controller, developed in depth how should factors legitimating the interest of the data controller to process personal data be assessed and balanced with the also legitimate rights and interests of the data subjects. The legitimate interest is a legitimating alternative for a processing, like could be some cases ML, which could need access to some training data, and when they fulfil the conditions to be lawful.

[55] Article 9.2.c of the GDPR "processing is necessary to protect the vital interests of the data subject or of another natural person where the data subject is physically or legally incapable of giving consent".

[56] This is the case of Smartcities or border controls.

[57] Considering the exceptions provided by article 5.1b "(..) 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 purpose", and by 6.4 with regard to the processing with compatible purposes by the data collector.

It must be taking into accout that to ground the lawfulness of a processing in a legitimate interest requires from the organization acting as a controller a higher degree of commitment, formality and competence. It requires a careful assessment that their legitimate interests prevail over the possible impact on the rights, freedoms and interests of the data subjects, considering, among other things, eventual compensatory measures arising from the need to keep the processing activities continuously supervised, implement a high degree of accountability and stricter design default privacy measures or the adoption of best practices such as giving an opt-out option to data subjects[58]. In any case, the data controller must be able of proving that the relevance of the corresponding impact is not such that it does not allow to perform the intended processing on the corresponding grounds, and it must document the entire analysis and decision-making process in compliance with the accountability principle.

If the processing activities are based on the legitimate interest, it is not necessary to collect the consent of the data subject, but information obligations provided by articles 13 and 14 of the GDPR persist.

**Special categories**

In order to establish the legal grounds for the processing activity, it is important to bear in mind that special categories of data, as established by article 9 of the GDPR, include special processing requirements. In such cases, and before analysing the legal grounds legitimating the processing according to article 6 of the GDPR, the prohibition established in the aforementioned article 9 based on any of the circumstances included therein must be removed, albeit considering the additional limitations established by article 9 of the LOPDGDD[59], specifically the one stating that consent does not remove the prohibition to conduct any processing whose main purpose is to identify the ideology, union status, religion, sexual orientation, beliefs or racial or ethical origin of the data subject.

Besides, Whereas 71[60] of the GDPR establishes an additional restriction on the processing of special categories of data when they are intended to be used in automated decisions and for profiling purposes, establishing the limitation that such data can only be used under specific conditions. Particularly, article 22.4 establishes that decisions based solely on automated processing, including profiling, by reason of which the data subject may be subject to legal consequences or be similarly affected in a significant manner, shall not be based on special categories of personal data, unless the data subject has given their consent or such processing is required for reason of an essential public interest based on the legal frameworks of the European Union or of any Member State and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place.

Such limitations apply both to collected data regarding the data subjects for the purposes of development or operation of the AI component and to data pertaining to special categories which may be inferred in the course of the processing activities.

**Processing for compatible purposes**

As established by article 6, section 4[61], it is possible, under certain circumstances, the processing of personal data for purposes other than those for which they were collected in the first place. The legal grounds for this new purpose could be a new consent by the data subject, or on any legal grounds allowing it, but also the fact that both purposes, the original

[58] Opinion 7/2015 Meeting the challenges of big data. European Data Protection Supervisor. EDPS
[59] Organic Act 3/2018, of 5 December, on Personal Data Protection and the guarantee of digital rights
[60] Recital 71 Automated decision-making and profiling based on special categories of personal data should be allowed only under specific conditions.
[61] Recital 50 is related to article 6.4.

one and the new one, are compatible. Of course, the controller must comply with all requirements to ensure that the original processing was lawful and must assess said compatibility, considering the relationships between the purpose for which data were originally collected and the subsequently intended purposes, the context in which they were collected, the nature of data themselves, paying special attention, among other things, to special categories of personal data[62], the potential consequences for data subjects and the guarantees implemented in the processing in order to properly manage those risks to rights and freedoms.

Particularly, further processing activities for filling purposes for public interest, scientific or historical research purposes or statistical purposes must be considered compatible and lawful processing activities, with the guarantees and exceptions established by article 89 of the GDPR.

### B. INFORMATION

Information must be provided to data subjects by the data controller is established by articles 13 and 14 of the GDPR, and the specific contents must be adapted to the life cycle stage of the AI used for processing purposes. In order to support compliance with this obligation, AEPD published the Guide for compliance on the duty to inform, which has a general approach, as well as a more specific guidelines under The duty to inform and other accountability measures in apps for mobile devices. As a general approach, article 11 of the LOPDGDD provides the controller with the possibility of offering this information in a two-tiered or two-levelled scheme: a first, general layer, containing general processing information, and a second layer which completes the information of the first layer with a higher level of detail and which is accessible form it in an easy, intermediate manner, even by electronic means:

- The first layer should include:
    - The identity of the of the data controller or their representative.
    - The purpose of the processing.
    - The possibility of exerting the rights included in articles 15 to 22 of the GDPR.
    - If the processing involves profiling or automated decisions:
        - This circumstance must be clearly disclosed.
        - Informing of their right to oppose making any automated individual decisions pursuant to article 22 of the GDPR.
        - Relevant information on the implemented logic.
        - Relevance and foreseen consequences of the relevant processing for the data subject.
    - When the personal data which are subject to processing have not been obtained directly from the data subject, basic information shall also include:
        - The categories of data subject to processing.
        - The source from which the data originate.
- The second layer shall include all other information established by articles 13 and 14 of the GDPR.

If the AI component includes personal data which may be retrieved, this fact must be disclosed to the data subjects and, when appropriate, there must be legitimating legal grounds for any subsequent disclosure or processing of such data.

[62] Not only with the special categories defined in article 9, but also with personal data regarding criminal offences and penalties, pursuant to article 10.

**Relevant information on the implemented logic**

If the data subject is subject to profiling or automated decisions as provided by article 22 of the GDPR, an important aspect set forth by article 13.2.f of the GDPR is that the data subject must have "*meaningful information about the logic involved*" as well as about "*the significance and the envisaged consequences*".

The term "*meaningful*", defined by Merriam-Webster dictionary as "*having a meaning or purpose*" must be understood as information which, once provided to the data subject, serves the purpose of raising awareness of the type of processing undergone with his data, and providing certainty and trust on the associated results[63].

Complying with this obligation by making a technical reference to the algorithm implementation may be obscure, confuse or excessive and leading to information fatigue. However, sufficient information must be provided to understand the behaviour of the relevant processing. Although it shall depend on the type of AI component used, an example of the type of information which may be relevant for the data subject would be:

- Detailed information about the subject's data used for decision-making regardless of the category, especially regarding with how old the subject's data under processing are.
- The relative importance or weight of each data category in the decision making.
- The quality of training data and the type of patterns used.
- Profiling activities conducted and their implications.
- Error or precision values, according to the appropriate metrics used to measure the eligibility of the inference.
- Whether qualified human supervision is involved or not.
- Any reference to audits, especially on the possible deviation of inference results, as well as certification or certifications performed on the AI system. For adaptive systems or evolutive systems, the last audit conducted.
- If the AI system includes information referring to identifiable third-data subjects, prohibition of processing such information without legitimisation and of the consequences of doing so.

## C. GENERAL ASPECTS RELATED TO THE EXERCISE OF RIGHTS

Any data controllers that use AI-based solutions to process personal data, conduct profiling or make automated decisions must be aware that the data subject have rights related to the protection of their personal data which must be attended to.

Therefore, in the processing design stage, data controllers must be aware that they need to include appropriate mechanisms and procedures to attend any received claims, and that such procedures must be appropriately dimensioned to the scale of the processing that they are conducting.

For those personal data which are distributed across a set of controllers, for example, if an AI-based model includes personal data intended for either system training or system evolution, it is required, according to the principle of accountability[64], to include an effective information governance model which enables traceability of information for the purposes of being able to identify the relevant controller and empower data subjects to exercise their rights. This information governance model must also be provided when a processing involves a controller and including in the relevant agreement those tasks that are assigned to the controller in relation with the exercise of rights.

[63] This is related to the concept of processing "explicability" by means of AI.
[64] Particularly, the obligations established in articles 24, 25, 26 and Recital 66.

Finally, it must be considered that article 11 of the GDPR establishes that, when exercising the rights to access, erasure or limitation of processing, if it is not possible to identify the data subject, the controller shall not have the duty to keep, obtain or maintain additional information in order to identify the subject data for the sole purpose of complying with the GDPR provisions, although the relevant data subject shall be entitled to provide additional information which ensures their identification and thus enables them to exercise their rights.

### D. RIGHT TO ACCESS

The right of access must be executed by the controller of each life cycle stage of the AI-based solution involving personal data. This includes any training data which may be included in the AI components and which are susceptible to be retrieved by the controller who operates the relevant AI-based solution.

### E. RIGHT TO ERASURE

The right to erasure involves a proactive attitude by the data controller in order to, as established by Whereas 39[65], guarantee that data are erased when they are no longer necessary for the purpose of the processing, and, particularly, for including procedures for the period review of the relevant datasets, and terms for its erasure.

Article 17.1 establishes the obligation to erase without undue delay when the following situations concur:

- Personal data are no longer necessary with regard to the purposes for which they were otherwise collected or processed;
- the data subject removes their consent and such consent is the only grounds for such processing;
- the data subject opposes the processing and there are no other legitimate methods that prevail
- the data subject opposes that their data are processed for direct marketing purposes
- personal data have been unlawfully processed;
- personal data must be erased in order to comply with a certain legal obligation;
- Personal data have been obtained in regard with the provision of any information society service.

Data collected for the training stage, in compliance with the aspects set forth by article 11 of the GDPR and with the principle of data minimisation, must be cleansed or sanitized from all information which is not strictly necessary to train the model.

Once the training stage of an AI system is completed, the organization must execute its removal unless they are needed for the system fine-tuning or assessment, or they are used for other purposes which are compatible with the original purposes as per the provisions of article 6.4[66] of the GDPR. In any case, the data minimisation principles must be considered. If requests for erasure are placed by data subjects, the data controller should adopt a case-based approach, considering any possible limitations to this right provided in the relevant regulations.

[65] Recital 39 (...) ensuring that the period for which the personal data are stored is limited to a strict minimum. Personal data should be processed only if the purpose of the processing could not reasonably be fulfilled by other means. (...) Time limits should be established by the controller for erasure or for a periodic review.

[66] Article 6.4 Where the processing for a purpose other than that for which the personal data have been collected is not based on the data subject's consent or on a Union or Member State law which constitutes a necessary and proportionate measure in a democratic society to safeguard the objectives referred to in Article 23(1), the controller shall, in order to ascertain whether processing for another purpose is compatible with the purpose for which the personal data are initially collected, take into account, inter alia (...)

Particularly, when the AI-based solution is released to controllers and natural persons, if this includes data from data subjects:

- They must be either erased, or to assess that it is partially or entirely impossible because it would impair the systems,
- The relevant legal grounds to disclose data to a third party must be established, especially when data of special categories are included,
- The data subjects must be informed (as stated above),
- Proof must be offered that the relevant privacy by default and by design measures have been implemented (most specifically, data minimization) and,
- Depending on the risk involved for data subjects and the volume or categories of data, to carry out a privacy impact assessment.

In case that the data controller keeps the data subject's data for the purposes of personalizing the service which is offered by an AI-based solution, once that the service provision relationship finish, the relevant data must be erased.

**Limitations to erasure**

Article 17.3 of the GDPR establishes certain limitations to erasure. Besides, article 32 of the LOPDGDD establishes the obligation of the data controller to block the data which are to be rectified or erased.

## F. BLOCKING OF DATA

Blocking constitutes an obligation of the data controller whose sole purpose is to provide a response to any possible liabilities arising from the data processing, for the purposes of obtaining evidence of any possible non-compliance and exclusively during their validity term.

Article 32 of LOPDGDD establishes blocked data as the status of data kept outside the processing scope, applying any technical or organizational measures that prevent any type of process, including visualization, except for the purposes of making such data available to the relevant courts or judges, the public prosecutor or the competent Public Administration, especially the data protection authorities.

Therefore, the need to include the above measures to block any data related to the inference process (or at least inputs and results) which may be needed to respond a request or claim by the relevant data subject must be considered as a requirement when designing the processing. These mechanisms are also related to the log files[67] that shall be discussed below.

## G. RIGHT TO RECTIFICATION

The data controller shall be obliged to respond to the exercise by data subjects of their right to rectification, especially when this right arises from inferences and profiles created by the relevant AI-based solution.

On the other hand, inasmuch as the model includes inaccurate training data which do not have an effect on the AI component user and such inaccurate information is not possibly linked to any data subject when distributing the AI-based solution, feeding inaccurate training data may be advisable as part of the abstraction and obfuscation strategies intended to

[67] Files recording all events executed in a system.

guarantee that the principle of data minimisation is effectively applied[68]. If such strategies lead to prevent reidentification of an individual, referencing the above-mentioned article 11 of GDPR, the right to rectification would not apply.

However, if the model itself includes inaccurate personal data of third data subjects that are able to be reidentified and thus associated to them wrong information, the right to rectification must be attended.

## H. PORTABILITY

Article 20 of the GDPR establishes that, when the processing is carried out by automated means the data subject has the right to request any data that has provided to the relevant controller, get it in an structured, common-use, machine-readable format, and to transfer it to another data controller, when the legitimating grounds for the processing is either the data subject's consent or the processing is needed to execute an agreement.

The recognition of this right, as developed in of Article 29 Working Party's Guidelines on the right to data portability classifies personal data in: actively and willingly provided by the data subject; observed with regard of the data subject, by virtue of their use of the relevant service or device; or inferred, deducted or created by the data controller on the basis of the two former categories. The right to portability applies to the first two types of personal data.

Any data controller who includes an AI component must assess and document whether their processing must provide portability, considering the provisions of the aforementioned article 20. In such case, the requirement of portability must be considered from the earlier stages of conceptualization and design, as well as in the selection of the AI component by the AI component developers.

Article 20.2 of the GDPR establishes the right to have data directly transferred from one data controller to another, but only when it is technically feasible. In case that there are limitations to the right of portability, informing users beforehand of such limitations is considered an exercise of transparency.

## I. DECISION-MAKING BASED ON AUTOMATED PROCESSING

Applications offering or supporting a service based on AI solutions may make certain decisions which affect individuals, their private lives, their physical safety, their social position and their interaction with other persons.

The GDPR guarantees the right not to be submitted to automated decisions included profiling[69] when:

- There is no human intervention. In order to consider that there is human intervention, the relevant decision must be supervised by a competent person authorised to revert such decision by means of a significant action and not one purely symbolic.
- There are legal consequences derived from such decisions.
- Or the data subject is similarly and significantly affected[70].

An exception to the above can be made when processing:

[68] Such as application of differential privacy techniques.

[69] Automated decisions can be made with or without profiling; profiling can take place without making automated decisions. However, profiling and automated decision-making are not necessarily separate activities. Something that starts off as a simple automated decision-making process could become one based on profiling, depending upon how the data is used. (wp 251)

[70] Depending on the relevant case, the following aspects may be considered to have a significant impact: monitoring persons in different websites, devices and services, modifying the expectations and wishes of the involved persons, modifying the manner in which an advertisement is presented or using their knowledge of the vulnerabilities of the data subjects. (wp 251)

- Is based on explicit consent and guarantees to protect rights and freedoms are implemented.
- Is necessary to conclude or execute an agreement, does not affect special categories of data and besides, includes guarantees in order to protect the relevant rights and freedoms.
- It is based in the legal framework of the European Union or the Member State and does not involve special categories of personal data.
- It is based in the legal framework of the European Union or the Member State and it is required to protect a fundamental public interest.

The section "information" above, and, particularly, the subsection “Relevant information on the implemented logic", as well as the section on “Transparency” addresses the information requirements of such processing activities.

When the legal grounds for processing is explicit consent, the controller must design the processing in such a way that it protects the free choice of users. This is done, first, by providing viable and equivalent alternatives to automated solutions at the time when their consent is required. Besides, it is guaranteed that if the data subject chooses not to be subject to automated decisions, the decision concerning such data subject shall not be biased and goes against the data subject’s interests. If the above conditions are not met, consent may not be considered to have been freely given[71]. These alternatives must be implemented from the processing design stage.

As best practices, and beyond any requirement arising from data protection, human supervision may be an option which may be chosen within AI-based data processing, and in general with regard to automated decisions. The “dead man switch” approach must be avoided in system design[72]: human users must have the option to ignore the algorithm at a given time in all cases, and to document the situations in which this course is privileged. For this reason, it is recommended to document any incidence or any challenges to automated decisions by the relevant data subject, so that its analysis allows to detect situations which require human intervention, because the processing is not operating as expected.

With regard to processing data pertaining to minors, Whereas 71 establishes that decisions solely based on automated processing, including profiling[73], with legal or significantly similar consequences, should not be applied to minors. However, since this prohibition is not reflected in the relevant articles, it is not deemed absolute, and exceptions may be considered when they are unavoidable to protect the relevant minor’s well-being and appropriate and child-specific guarantees.

[71] “Automated Society Report 2019” describes an automatic candidate selection methods based on consent which requires the data subject access to their email address in order to have an algorithm assess their mail traffic and thus obtain a profile as a future employee. If there is no alternative, or the mere submission of a resume imposes a penalty in the selection process, the consent shall not be valid.

[72] Dead man switch

[73] WP29’s Opinion 02/2013 on apps on smart devices (WP 202), adopted on 27 February 2013, on its section 3.10 specifically on children, specifies , on page 26, that «Specifically, data controllers should not process children's data for behavioural advertising purposes, neither directly nor indirectly, as this will be outside the scope of a child's understanding and therefore exceed the boundaries of lawful processing».

# IV. PRIVACY RISKS MANAGEMENT

Management of the privacy risks is an continuous activity which belongs to the scope of accountability as established in GDPR.

When processing personal data and it is proportionally to the processing activities, the organization must adopt a Data Protection Policy as established by article 24 of the GDPR. This policy should be integrated in their quality policy, their secure information systems and their decision-making policy (among others).

Although the GDPR does not specifically establish which accountability elements need to be compulsorily implemented or how to implement them, but merely provides a highly flexible framework in order to allow very different processing approaches to be compliant with the relevant standard, GDPR lays down that such measures must be selected according a risk based approach (RBT[74]) which specifically considers the risks posed by personal data processing and profiling with regard to their rights and liberties[75].

RBT includes two fundamental stages: a first stage consisting on identifying threats and assessing the existing inherent risk level, and a second stage consisting in managing this risk by means of appropriate and proportional technical and organizational measures in order to remove or at least mitigate such risk, with the purposes of reducing the probability of impact of the identified threat. Once the chosen measures have been implemented, the remaining residual risk[76] must be assessed and controlled.

In order to determine the risk and to establish the appropriate measures to manage it, the processing must be assessed and divided into stages. Therefore, the special requirements and risks of each stage must be managed from the data protection point of view.

In order to determine the level of risk of a processing, which is based on or includes stages that feature an AI component, it must be taking into account:

- Any risks arising from the processing itself, in particular that derived from the bias present in decision-making systems and the natural persons discrimination (*algorithmic discrimination*[77]).
- Any risks arising for processing with regard to social context and collateral effects which may be derived from the processing, even those that are indirectly related to the purpose of the processing.

## A. RISK ASSESSMENT

Assessing the level of risk of a processing is the first step to establish what are the measures needed to remove or minimize this risk, especially when the conditions lay down in the regulation state that it is a high risk processing, especially:

- Article 35.3 of GPDR.
- Article 28.2 of the LOPDGDD.
- Lists of types of data processing requiring a PIA (Art. 35.4) issued by AEPD.

The above lists and articles lay down certain conditions which processings have to meet procedures in order to be considered high-risk, some of such conditions should be met simultaneously. The above lists are not to be considered exhaustive, but mere guidelines for

[74] Risk Based Thinking
[75] There are many types of risks: business continuity, fraud, financial, image-related, opportunity, technology reliability, safety, etc.
[76] ISO 31000 defines residual risk as the risk left over after you've implemented a risk treatment option. Treatment option are all preventive, detective or corrective measures intended to minimize risks.
[77] It allows to transpose those prejudices existing in society to the implementation of the AI component by means of spatial categories of data or other attributes, such as nationality. https://eticasfoundation.org/wp-content/uploads/2018/03/IG_AlgoritmicAFinal.pdf.

risk assessment. The controller must be able of identifying any additional risk arising from the new AI component and its implementation, as, for example, in case that the AI component is released to third parties including personal data. It must be stated that this analysis exceeds the limits of the controller organization, and it is designed to assess the impact that the performed activity may have in the social context in which the application is executed or deployed.

The controller of the development, maintenance and/or release of an AI components, as well as the data controller of any processing including AI components, must implement the appropriate measures to minimise or remove such risk factors for each respective stages and responsibilities.

## B. PRIVACY IMPACT ASSESSMENT-(PIA)

The PIA is an obligation laid down by the GDPR for high risk processing. This obligation requires to go further from the mere risk management related to the processing, and it requests an additional seriousness, accountability, at the time to manage such risk.

The need for each data controller to perform a data protection impact assessment is laid down in Article 35 of the GDPR when, as established in Paragraph 1, *"the processing, is likely to result in a high risk to the rights and freedoms of natural persons"*.

More precisely, but without limitation, as established in Article 35.3 a[78], it is necessary to perform a PIA whenever a profiling is being carried out that is based on automated processing (but not necessarily exclusively automated processing), and decisions are making that produce legal effects concerning the natural person or significantly affect the natural person.

The PIA is performed before the effective processing of the personal data, it means, before the actual operation of the processing. This means, inter alia, that the PIA must not be carried out during the processing validation phase, which includes the AI component, or simultaneously with the operation phase. Therefore, the validation must be performed before designing/selecting and implementing the AI solution for a processing so that the assessment of the privacy requirements can be made in advance and the privacy by design measures and the privacy by default measures may be effectively implemented.

Article 35.2 of the GDPR lays down the obligation of the controller to seek advice from the Data Protection Officer, when a Data Protection Officer has been appointed.

It is likewise important to take into account that Article 35.9 of the GDPR establishes that, when applicable, the controller shall seek the opinion of the data subjects.

In the case of AI solutions, this opinion is particularly important to understand not only the technology itself, but the precise context where this technology is to be used and the specific risks for the rights and freedoms of the data subjects. Then it is a good practice to extend such consultation to all the stakeholders, including human operators dealing or supervising the AI results and third-data subjects affected by the processing.

More precisely, when the processing that is carried out in an automated way creates profiles and makes decisions, all such decisions must be identified in all phases of the processing, and the operating parameters must be analysed, such as, for example, error rates, and the effects that such decisions have on the data subjects must be carefully analysed.

[78] Article 35.3.a systematic and extensive evaluation of personal aspects relating to natural persons which is 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.

The PIA must be documented and, when a high residual risk is detected, it must be subject to consultation before the supervisory authority under the conditions established in Article 36 of the GDPR.

A PIA must result in the adoption of a series of specific measures to manage risk, some of them aimed at reinforcing the obligations of fulfilment depending on such risk and affecting:

- The conception of the processing itslef, its division in phases, procedures, technologies and extent of the processing.
- The implementation of privacy by default measures and privacy by design measures in the processing pursuant to the principles of:
    - **Minimising** the amount of data that is being processed, both in terms of volume of information gathered and the size of the population that is the subject of the analysis, as well as throughout the different phases of the processing.
    - **Aggregating** the personal data to the extent possible, so as to reduce the level of detail that can be obtained as much as possible.
    - **Concealing** personal data and their interrelations to limit their exposure and to avoid their visibility before non-data subjects.
    - **Separating** the contexts of the processing to hinder the correlation of independent sources of information as well as the possibility to infer information.
    - Improving the **information** for the data subjects, in terms of time and the specific channels, about the characteristics and the legal grounds for the processing, thus enhancing transparency and allowing the data subjects to adopt informed decisions on the processing of their data.
    - Providing means for the data subjects so that they can **control** the way their data are collected, processed, used and released to third parties by implementing mechanisms that are adapted to the level of risk that enables them to exert their rights in terms of data protection.
    - **Complying** with a privacy policy that is consistent with the legal obligations and requirements imposed by the regulation.
    - **Proving,** as the accountability principle states, that the data protection policy that is applicable is being fulfilled, as well as all other legal requirements and obligations imposed by the Regulation, with regard to the data subjects and the supervisory authorities. This includes a dynamic audit of the output/results of the processing, assessing the accuracy divergences and deviations, including the algorithms that have been executed, in order to adopt corrective measures. Among such measures are to cancel data and to document in detail of the analysis performed and the measures adopted.
- The identification of the security requirements to reduce the risk with regard to privacy.
- The adoption of specific measures to implement a governance system of personal data enabling to provide evidence and traceability of the fulfilment of the principles, rights and guarantees in order to manage the risk.

The AEPD provides guidelines to address the principles listed above for generic processings: Guideline for PIA, la Guideline of Privacy by Design, Template of PIA assessment document for public administrations, and tools that help perform the EIPD, such as GESTIONA.

Before repeating what may be found in such guidelines, incorporated herein by reference, some specific aspects for IA solutions are going to be developed in the next sections.

## C. TRANSPARENCY

Pursuant to Recital 78, the principle of transparency is a privacy measure by default to enable, inter alia, that the data subjects supervise the processing to which they are being subject.

The principle of transparency is explained in Recital 39 and Recital 58. The obligation to inform data subjects is understood in such Recitals in a way that extends beyond the provisions in Articles 13 and 14 of the GDPR. More precisely, the Recitals state the obligation in the sense that *"any information and communication relating to the processing of those personal data be easily accessible and easy to understand" "concise", "that clear and plain language be used" "where appropriate, visualisation be used", that "it be provided electronically", "that further information to ensure fair and transparent processing"* be provided and that *"data subjects" "should be made aware of risks, rules, safeguards and rights in relation to the processing of personal data".*

In the case of AI-based processing, transparency may be considered a critical aspect, then, it must enable data subjects to become aware of the impact of such solutions. Transparency is addressed both to data subjects and to processing operators. More precisely, transparency is linked to an accurate information on the efficiency, the real capabilities and the real limitations of the AI systems to avoid false expectations among users and data subjects, which may give rise to an incorrect interpretation of the inferences made by the processing. Transparency is likewise linked to the information on the context of the processing, as well as the existence of third parties, the physical and virtual location of the AI solution, etc.

Transparency is not limited to a one moment or flahs, but rather it must be understood as a principle around which the processing to be carried out and affecting each and every element and stakeholder of the processing.

### During training

During training, if personal data are used, the data subject must be clearly informed on the possibility that such data be reidentifiable from the data of a model, or, as established in Article 11.2 of the GDPR, that such reidentification not be possible.

### Certification

One of the problems arising at the time to provide transparency with regard to AI algorithms is the question of the confidentiality needed to preserve industrial property.

This problem does not only affect AI, but it is a problem shared by other technological solutions, such as cryptographic modules in security systems or integrated circuits in credit cards. Notwithstanding, in sectors where the use of such solutions is mandatory, trust on sellers and developers is not enough, but rather compulsory mechanisms based on trusted third parties are established in order to guarantee the mandatory quality and reliability levels, either by granting the control authorities access to the internal aspects the productos or by establishing certification mechanisms through independent third parties[79].

Likewise, when the AI evolves from an experimental element to a product, it is necessary to add the same guarantees that are requested from any other technological service. The development and the application of certification schemes that establish reference frameworks (for manufacturers, data controllers, authorities and users) which enable an independent third

[79] Such is the case of the certification based on Common Criteria or CC, and where a list of certified products may be found https://www.commoncriteriaportal.org/

party to certify that both the AI in particular, and the implementation of the processing in general terms, are compliant with the GDPR and what was stated in this guidelines.

With regard to the certification schemes, which may cover several aspects regarding the use of AI and be tools to prove carefulness, the GDPR establishes in Article 42 the possibility to develop specific certification mechanisms in terms of data protection, seals and marks of data protection as tools to prove compliance with the GDPR, as explained in Recital 100, to increase the transparency of personal data processing.

**Automated decisions and profiling**

The data controllers must provide sufficient information to the data subjects with regard to the processing they are being subject to, as well as information on the mechanisms enabling a request of a human intervention in the assessment or questioning of the decision adopted by the system automatically, notwithstanding the provisions in Articles 13 and 14 of the GDPR and which have been described in the section “Information” of this document.

**Data controller personnel**

Data controllers adopting this type of solutions and systems must provide precise information and specific training to their personnel on the limitations of the AI system.

Whenever the processing is a tool that helps in decision-making, it is necessary to adopt measures to manage the risk that the humans behave like a mere link to the inferences made by the AI solution. Such measures include information on the operator (as indicated above), training and behaviour audits.

Errors on interpretability by the operators must be prevented. The inferred values must be presented in such a way that they reflect the reality of the inference and its limits. Such information must be given to the subsequent phases of the processing, carried out by humans or automatic. At the time to operate the system, it is necessary to offer real-time information to the operator on the accuracy values and/or quality values of the information inferred at each time. Furthermore, when the information inferred does not reach the minimum quality thresholds, it must be explicitly noticed that such information is not valid or that it has no value[80].

In the event that the solutions are released by a third party, the latter must provide sufficient information to the controller so that the controller may manage such risks as well as information on the best way to proceed in this regard.

**The Data Protection Officer as a tool for transparency**

The appointment of a DPO is one of the best measures that may be adopted by the controller to orientate the implementation of transparency policies and, more precisely, to manage an information channel for data subjects. It allows the data subject to get information about the AI solution directly from the controller (either on the development of the AI, the maintenance of the AI or the operation of the AI within the frame of a processing).

Both Article 37.1 of the GDPR and Article 34 of LOPDGDD la down the conditions that make compulsory for the controller/processor to appoint a Data Protection Officer (DPO) by virtue of the nature of the processing or by virtue of the type of activity.

The fact of using an AI solution does not imply, *per se*, that there is an obligation for a DPO, although two cases exist that appear detailed in lit. (b) and (c) of the referred Article

[80] For example, in the event of systems that may detect suspicious body behaviour, no subject should be pointed out only because no other subject exists with a higher risk assessed.

37.1 where the presence of a DPO within the organisation is compulsory: the regular and systematic monitoring of data subjects on a large scale or the processing on a large scale of special categories of data or personal data relating to criminal convictions and offences.

Pursuant to the Guidelines on Data Protection Officers issued by the WP29[81], the DPO is a key element to guarantee the fulfilment and the management of risk for the rights and freedoms of the data subjects.

Therefore, the DPO, even for cases where such an appointment is not requested, may be very useful in entities using AI-based solutions where personal data are processed, or where AI solutions that use personal data to train models are developed, thus becoming a key element to manage the risk and to effectively apply the proactive responsibility mechanisms. More precisely, Article 35 identifies de Data Protection Officer as a fundamental role in the performance of the PIA and one of the tools to implement transparency before the user.

## D. ACCURACY

The term "accuracy" is defined in Article 5.1 of the GDPR as one of the principles of the personal data protection principles. In the event of a processing that includes AI solutions, the existence of biases in the inference models is closely linked to the accuracy or the quality of the data[82].

Pursuant to the second paragraph of Recital 71 of the GDPR, the data linked to the data subjects,[83]be they directly collected or inferred, must be accurate. More precisely, it is specified that the data controller shall "*use appropriate mathematical or statistical procedures for the profiling*" which guarantee that the data linked to the data subject are accurate. That is to say, it is mandatory to prove and to document that the procedures used for the inference of the information on a data subject are accurate and, therefore, stable and predictable.

When a controller opts to use an AI component for a processing, it must ensure that the data that is being processed, generated and linked to the data subject fulfil the requirements above.

### Factors affecting accuracy

With regard to the need for inferred data on data subjects to be accurate, there are three factors that may affect such accuracy:

- The very implementation of the AI system. There are AIs, such as rule-based expert systems, where the conception of the system itself may introduce errors that may lead to erroneous inferences, or ML-based systems, which are unable to model the desired processing. As stated before, errors may be introduced because elements outside the AI, such as biometric readers, may introduce errors in input data. On the other hand, it could happen that programming errors or design errors exist that may erroneously transfer the model to its practical implementation. In such cases, it may be said that the bias is "hardwired" by the decisions adopted when an analysis model is built (assessment and aggregation biases).

[81] "The DPO plays a key role in fostering a data protection culture within the organisation and helps to implement essential elements of the GDPR, such as the principles of data processing, data subjects' rights, data protection by design and by default, records of processing activities, security of processing, and notification and communication of data breaches."
[82] The term "data quality" is contained in Article 47.2.d of the GDPR.
[83] All data subjects have the same rights with regard to data protection, not only groups classified as "risk groups", and no "positive discrimination" is possible.

- The training or validation dataset is corrupted by errors, information that is deliberately erroneous [84]or biases that make it impossible for the inferences to be accurate. Such biases may be inherent, such as bad quality of the data, absent data or selective sampling. They also could be errors of representation and measurement due to the way the dataset is shaped.
- Biased evolution of the AI model. For AI implementing adaptive techniques, it must be taken into account that the AI may be used mainly by a group of data subjects whose particular characteristics may introduce new feedback biases.

As for the required accuracy in the event of training data, sanitation and traceability metrics and techniques must be implemented in order to guarantee the faithfulness and the integrity of the datasets. Although it is not shared by all AI solutions, training and operation data in some models may be classified as “hard” data and “soft” data. The former are objective data, which are measurable in general terms, such as, grades, attendance percentage, analytic values, test results, etc. “Soft” data are data that are qualitative in general terms and contain a subjective component or a component of uncertainty. Some examples of “soft” data are data obtained through the processing of natural languages, opinions, personal assessments, surveys, etc. Although “hard” data are not free of errors or biases, the controller must carefully assess the accuracy problems that may arise from using or giving greater weight to “soft” data as a source of information.

Bias is not solely a problem for an AI system, but rather it may appear in any automated/non-automated processing system that adopts decisions or performs profiling. There are known techniques such as the Algorithm Impact Assessment (AIA) [85] that are oriented towards examining and detecting the existence of possible biases used in the AI solutions and towards guaranteeing equity in the implementation of the model. Such metrics must analyse the logic to be implemented so that such logic does not originate inaccuracies by design, and they must use mature test models and implementation checks to detect design errors[86].

**Biometric information**

The accuracy is particularly critical when the processing is based on biometric information, such as AI on face recognition, fingerprints, voice, etc. In such an event, performance factors need to be taken into account (false positives, false negatives, and other) as well as the impact on the personal data collection with regard to any disability or physical singularity. Especially, when an AI-based solution is unsable of identifying the subject as such,[87]an erroneous profiling occurs even before the performance of the processing itself.[88].

[84] Erroneous information may also appear either as a result of an attack performed over the dataset or as a result of the inclusion of backdoors through a conscious manipulation of the data.

[85] The Algorithmic Impact Assessments are tools that enable an assessment by the stakeholders to verify that the AI components meet certain quality parameters and that they are suitable for the endeavoured task. A reference thereto may be found, for example, at https://ainowinstitute.org/aiareport2018.pdf

[86] Recent papers have raised concerns on the risk of unintentional bias in these models originating a negative impact on certain individuals within several groups. Although many bias metrics and equity definitions have been suggested, no consensus exists on the definitions and metrics that need to be used in practice for the purposes of assessing and auditing such systems. Notwithstanding, several metrics have been developed that are oriented towards an assessment of algorithmic discrimination, such as Aequitas, a bias auditing tool kit with an open code developed by the Center for Data Science and Public Policy.

[87] In the study Discrimination, artificial intelligence, and algorithmic decision-making published by the European Council, the following examples are listed: Facial-tracking software by Hewlett Packard did not recognise dark-coloured faces as faces. And the Google Photos app labelled a picture of an African-American couple as "gorillas". A Nikon camera kept asking people from an Asian background: "Did someone blink?" An Asian man had his passport picture rejected, automatically, because "subject’s eyes are closed" – but his eyes were open. Buolamwini and Gebru found that "darker-skinned females are the most misclassified group (with error rates of up to 34.7%). The maximum error rate for lighter-skinned males is 0.8%.”

[88] There are cases of people whose fingerprints are faint and when subject to an access system through fingerprint identification they encounter access problems on an ongoing basis.

The controller must take into account that, even if such users may be a minority, alternative mechanisms must be put into place in order to avoid exclusion of a subject on the grounds that the AI solution is unable of capturing the biometric characteristics of the data subjects[89].

**Profiling combination**

Sometimes, an AI component may create profiles or adopt decisions on the same subject but under different circumstances and for different purposes. For example, some types of criminal processing use AI solutions which allow to analyse a data subject profile to fail to comply with a summons before a court and, simultaneously, to infer a profile on the risk for the same data subject to commit a criminal offense again. In such an event, unless supported by documentation, it is convenient for such profiles to be performed independently. The profile stated in one of the topics shouldn't be considered or used in the inference process of the other topic.

**Verification vs. Validation**

The tests and/or verification of a AI component is a fundamental part in the development of such component, as well as an important criterion for the controller to opt for one component or another from other developer. The controller must take into account that the inclusion of a verified AI component in a processing does not guarantee that the processing is validated and neither confirms the AI component's adequacy for such specific processing. The tests of the AI component guarantee that the design and development results comply with the requirements of the component. The scope of the validation of the processing extends even further. It should garantee that the resulting products and services meet the requirements regarding a specific application or envisaged use[90]. The validation guarantees that the processing, together with the AI solution it is based on, meet the results planned for a certain product or service.

That is to say, the validation of the processing including an AI component must be performed under conditions reflecting the real context where the processing is expected to be deployed[91]. At the same time, the validation process requires a periodic review, taking into account that such context or the processing itself could change and evolve.

**Accuracy assessment as continuous process**

In the event that the AI solutions evolve, namely when they are fed back both with their interaction with the data subject and with the interaction with third data subjects, it is necessary to carry out re-assessment on the model.

The drift in the accuracy of the profiling made by the AI-based solution due to the "bubble filter"[92], or other circumstances, may feedback biases that the user has about himself or other people[93].

[89] In summary, to avoid a discrimination for not being "biometrically suitable".
[90] ISO 9001:2015 Quality management systems. Requirements
[91] The "Practical Manual on Artificial Intelligence within a Healthcare Background", referred to in the Bibliography, describes the case of a paediatric diagnosis system for pneumonia that obtained a 97% precision. However, when such algorithm was applied to a Madrid population, the accuracy went down to 64%. The data analysis of the training evidenced that the population used was between 0 and 5 years old whereas the scope of the application of the paediatric treatment in Madrid included children of up to 14 years old.
[92] Bubble filter or Echo Chamber is a very common situation in content searchers, which learn the user's tastes and only offer the content they believe to be the user's taste and ignore all other information in such a way that the user is locked up in a "bubble" that prevents them from accessing all options possible.
[93] For example, an AI that assesses the psychological status of a subject may feedback the vision that the individual has about themselves, thus guiding them towards a drift of their own perception.

## E. MINIMISATION

Minimisation is implicitly defined in Article 5.1. C of the Regulation as the process that is aimed to guarantee that the personal data are accurate, relevant and limited to the extent necessary with regard to the purposes for which they are being processed. Pursuant to Recital 59, personal data must only be processed if the purpose of the processing may not reasonably be fulfilled by other means.

The data subjects, the data categories, and the retention period of the data that may be processed are linked to the legal grounds for such processing.

Minimisation is the process of optimising the processing from the point of view of data protection, analysing the needs of the data processing in the different phases of the process and in compliance with the requirements above for the purposes of:

- Limiting the extent of the data categories [94] that are used in each phase of the processing to categories that are strictly necessary and relevant.
- Limiting the level of detail or precision of the information[95], the granularity of the collection in terms of time and frequency and the collection date of the information used.
- Limiting the extent of the number of data subjects whose data are being processed.
- Limiting accessibility to the several data categories to the controller's/ processor's personnel or even to the final user (if there are data pertaining to third parties in the AI models) in all processing phases.

Therefore, it is necessary to assess how to implement such principles at the time to design the AI solution and the processing, and to describe and analyse the life cycle of the data throughout every phase of the processing. This analysis does not only focus on the AI solution from a technical point of view, but rather on the global processing where such solution is included. It must take into account both automated aspects and non-automated aspects of each and every phase of the processing.

### Training data

In case of ML, it is necessary to balance the need of the data to train the ML systems in regarding the risk for the rights and freedoms of the data subjects. The degree of quality of the training data is not only measured by the mere accumulation of data, but through relevance, actuality, reliability, soundness [96] and use the categories of data that are relevant for the intended processing. In order to apply such proportionality criterion, it is advisable to use professional profiles with expertise on data science[97], a discipline that goes beyond the specialisation in ML algorithms, but with skills on the principles of data protection, and in cooperation with the specialists in business logic and the data protection officer, if a data protection officer has been appointed[98].

[94] Extent of the data category regards with the number of data fields associated to a natural person refers to: name. physical and logical addresses, fields on their health, working status, social status, relations, tastes, beliefs, ideology...The more the fields, the more the precision and the more the diversity on an individual, the bigger the extent of the data categories that will be processed.

[95] Something as simple as a date of birth may be defined from specifying the decade up to specifying the year, the day and the hour.

[96] From the moment it is applicable to training for several processing.

[97] Data Science is a cross-disciplinary field that entails scientific methods, processes and systems in order to extract knowledge and a better understanding of data in their different shapes, be they structured or unstructured. This is a continuation of some fields in data analysis such as statistics, data-mining, machine learning and predictive analysis.

[98] By way of an example, in the book "Practical Manual on Artificial Intelligence within a Healthcare Background", it is established that 80% of the time of development of an algorithm has to do with the collection, the cleaning and the pre-processing of data and that only a 20% affects the training of the algorithm itself.

**Minimisation techniques**

There are several AI solutions data minimisation techniques[99], some of them specifically for ML, and they are in continuous development:

- Data assessment to check their high-quality and high predicting capacity for the relevant application.
- Critical analysis of the extent of the data categories used in each phase of the AI-based solution.
- Erasure of non-structured data or unneeded information that has been collected during the pre-process of the information.
- Identification and erasure, during the training process, of such data categories with no significant influence on the learning or the result of the inference.
- Suppression of non-relevant conclusions linked to a data subject during the training process, for example, in the event of non-supervised training.
- Use of verification techniques that require a lesser number of data, such as crossed validation[100].
- Analysis and configuration of hyperparameters [101]of the algorithm that may have an influence on the amount or extent of the data processed, with the aim to minimise them.
- Use of federated learning models instead of centralised learning models[102].
- Application of differential privacy strategies.
- Training with encrypted data using homomorphic techniques.
- Data aggregation.
- Anonymisation and pseudonymisation, not only in the communication of the data, but also in training data, possible personal data into the model [103]and during the processing of the inference.
- Etc.

**Extent of the data categories in an AI-based solution**

For each phase of a processing, regardless of whether the phase includes an AI component or not, a different extent of the total personal data regarding the same data subject needs to be processed. In general terms, it is not necessary to access the full extent of the available[104] personal data in every phase. Therefore, data minimisation strategies used for every phase of the processing should differ, and they should likewise differ in each of the phases of the life cycle of the AI-based solution: the training phase, the inference phase or the model evolution phase. It should be taken into account the limitations established by the legal basis itself[105].

One aspect that needs to be justified at the time to establish the extent of the data categories to be used in the AI component is the use of proxy variables. A proxy variable is

[99] See Privacy Design Guide of the Spanish Data Protection Agency [AEPD].

[100] Crossed validation entails de division of the database into training and testing several times and with a different data selection on each sub-group, comparing the global output of the training and the test of the algorithm on each iteration.

[101] Hyperparameters are parameters that allow to configure the operation of a specific AI model. For example, in neuronal networks, hyperparameters shall be the number of layers, the number of neurones per layer, the learning rate, etc.

[102] Federated Learning or FL is a distributed ML model that allows model training in final devices (process over the edge). It is an approximation to "bring the processing to the data instead of the data to the processing" and it permits to manage data privacy, data property and data localisation. https://arxiv.org/pdf/1902.01046.pdf

[103] Releasing the SVM Classifier with Privacy-Preservation https://ieeexplore.ieee.org/document/4781198

[104] This principle is related to the "need-to-know" principle that is used in security, but, in this case, it is understood from a point of view of privacy. For example, if the data set of the processing has a phase that submits a communication to the data subject, it shall be necessary to process the contact details during such phase, but it shall not be necessary to process all the information held on the data subject. Inclusive, for those performing the communication, it shall not be necessary to access the content of such communication.

[105] The legitimisation to process a datum within the frame of a processing, for example, the delivery address of a parcel, does not extend to the use of such an address as an input datum of the AI component to carry out a profiling of the client during the hiring phase.

a type of data that seems to not be in close relation with the aim of the processing but that could be strongly correlated to the inferred value[106]. When proxy variables are in use it is necessary to assess the validity of such correlation to legitimise the processing.

There are two important aspects that need to be analysed with regard to proxy variables. The first one is that their use must not be linked to any bias in the reasoning model, such as, for example, using the data subject's nationality as an element to grade their fraud profile or their social status to assess the possibility to reoffend. The second one[107], is to audit if the correlation between the proxy variables used and the purpose of the processing remains in time and within the scope of application, as such a relation may vary and be subject to fraud[108].

**Extent of the training set**

One of the specific aspects of the minimisation principle in AI-based solutions deals with the size of the population whose data are to be collected to train the models, mainly when such processing is based on the legitimate interest.

In such an event, the number of data subjects affected must be justified based on:

- The accuracy that the AI-based solution must reach in its inferences for the specific processing.
- The convergence factor of the training algorithm towards the value of desired accuracy required by the processing, and the relation with the amount of examples provided.
- The fact that the amount of data requested by the model may introduce biases in the inference, because lack of availability of data covering a balanced spectrum of population (per age, gender, status, race, culture...) or because due to the antiquity of the data collected, such data reflect obsolete social contexts[109].

The fact of feeding an AI learning model with data without a prior control or prior analysis, besides the fact that it is not justified, is factor that could lead to a lost of accuracy and become a bias multiplier. The dataset to be used must be carefully analysed in order to avoid such risks and the use of such dataset must be legitimated if, for example, the processing is based on the legitimate interest.

Furthermore, an excess of training data (overadjustment or overfitting) may result in a model that is very adjusted to cases easily reidentifiable,[110] therefore being more vulnerable in the event of a privacy attack (see below model inversion attacks).

**Personal data in the AI-based solution**

In the event that the AI is released as a processing component and that, as a part of such, third party data are included, a formal analysis must be carried out to know which personal data of the data subjects may be identifiable.

[106] Proxy variables such as pizza consumption at decision making-centres in the United States or the ocupation of the parking of such centres revealed the precise date of the operation Desert Storm in 1991, which led to a change in security procedureshttps://www.army.mil/article/2758/army_releases_new_opsec_regulation

[107] This is related to the accuracy principle.

[108] In the book "Weapons of Math Destruction"referred to in the Bibiliography, the case is described of the use of the proxy variable "Twitter followers" in order to assess the suitability of the candidates for a position as Social Media Expert instead of carefully assessing the marketing campaigns performed. Once the proxy variable is detected, it is easy to deceive the system through the hiring of a service for few money that increases the number of Twitter followers of the candidate.

[109] Such as using information on professional preferences of women including data from many years ago or outside the cultural context, thus creating an obsolete image of the current population.

[110] As well as other problems such as low performance for a population that is not included within the training set.

Furthermore, technical measures need to be adopted, as well as organisational or legal measures to minimise the identification or the extent of the identification to the minimum number of data possible.

Conversely, during the operation or inference phase, it is convenient to use early deletion strategies of the data subject's data in the AI-based solution, even when the exception concerning household activities is applied.

## F. SECURITY

The GDPR established in Article 32 that both the controller and the processor shall apply suitable technical and organisational measures to guarantee a suitable security level with regard to the data subjects' rights and freedoms. Such measures shall be adapted taking into account the costs of the implementation, the nature, the scope, the context and the purposes of the processing, as well as the variable risks of probability and severity. There is no standard solution for all processings and much less for those including an AI component. The solution must be assessed through a risk analysis that must be related to the risks for the rights and freedoms of the data subjects from the point of view of data protection.

### Specific threats in AI components

Apart from the analysis of the security measures that are common to any system there are specific guarantees for AI-based processing. Such guarantees should address specific threats derived from the fact that the IA components is developed by third parties or from the data disclosure to third parties.

There are attack and defence typologies with regard to AI components that have been analysed[111]. Among the different security measures, it is advisable to pay careful attention to those managing the following types of threats:

- Access and manipulation of the training dataset, for example, through poisoning techniques with adverse patterns.
- Inclusion of Trojans and [112]backdoors[113] during the development of the AI, either in the code itself or in the development tools[114].
- Manipulation of the user API that allow to access the model, both at the level of black box and white box, to manipulate model parameters, leaking of the model to third parties, integrity attacks or availability of the inferences[115].
- Attacks by "adversarial machine learning"[116] so that an analysis on the robustness [117]and control of the feed of data to the model should be necessary.
- Attacks through pattern imitation that are known to be admitted by the system[118].

[111] A Survey on Security Threats and Defensive Techniques of Machine, Qiang Liu et al., IEEE Access, ISSN:2169-3536, February 2018
[112] Trojans in IA, IARPA https://www.iarpa.gov/index.php?option=com_content&view=article&id=1150&Itemid=448
[113] Example of the manipulation of an image recognition algorithm https://www.bleepingcomputer.com/news/security/ai-training-algorithms-susceptible-to-backdoors-manipulation/ also Backdoor Embedding in Convolutional Neural Network Models via Invisible Perturbation https://arxiv.org/pdf/1808.10307.pdf
[114] Vulnerability in a library for the development of ML models: https://cyware.com/news/critical-vulnerability-in-numpy-could-allow-attackers-to-perform-remote-code-execution-33117832
[115] For the purpose of increasing the false positive rate and the false negative rate.
[116] Attack technique consisting in feeding the AI with example data, which may be indistinguishable from normal data from the point of view of human perception but may include small disturbances that force the AI to make erroneous inferences.
[117] Exploring the Landscape of Spatial Robustness, Logan Engstrom∗ MIT https://arxiv.org/pdf/1712.02779.pdf
[118] Oriented towards applications such as face-recognition or intrusion detection and often related to techniques of "adversarial machine learning".

- Reidentification of the personal data included within the model (belonging inference or [119]inversion of the model[120]) by internal and external [121]users.
- Fraud or deceive to the AI by data subjects, especially in such cases where in may entail a damage for other data subjects[122], which implies the necessity to perform an analysis of the robustness in the light of such actions and the performance of audits.
- Leak to third parties of the profiling results or the decisions inferred by the AI (also related to the user's APIs).
- Leak or access to the logs resulting from the inferences generated while interacting with the data subjects.

**Logs or activity records**

The existence of log files or activity records, the performance of audits (be they automated or manual) and the certification of the process are inherent to the "accountability" strategies or proactive responsibility strategies, but they also arise out of the legal requirements that are specifically established in the sectorial regulation.

The log files shall be necessary to support the audit processes and the security mechanisms, with regard to data protection, said log files shall provide evidence in order to:

- Establish who and under what circumstances accesses the personal data that may be included within the model.
- Provide traceability with regard to the update of the inference models, the communications of the user API with the model and the detection of abuse or intrusion attempts.
- Provide traceability to enable governance in data disclosure among all intervening parties in the AI-based solution with regard to the obligations arising out of Recital 66 of the GDPR.
- Provide a follow-up of the quality parameters of the inference when the AI is used for decision-making or in assistance processes to the decision-making.

The legitimate interest of the controller as a legitimate basis for the processing of the personal data of the files of the activity records for security purposes is explained in Recital 49 [123]of the GDPR, *"to the extent strictly necessary and proportionate for the purposes of ensuring network and information security".* In other cases, the sectorial regulation shall establish obligations on the preservation and the processing of activity records, such as Law 10/2010 of money-laundering prevention and counter terrorism financing[124], and therefore, in

[119] When it can be established if a certain individual, their data, are or not part of the training model.

[120]The attack by inversion of the ML model occurs when the attacker has access to certain personal data of the user included within the AI model and may infer the additional personal information of such individuals by analysing the inputs and outputs of the model.

[121] More precisely, when the AI models are acquired by third parties from the developer.

[122] A typical fraud example in CV analyses through AI is to write down unexisting merits in the same colour as the background of the document, which is impossible to read for a human but not for a machine, in such a way that a candidate selection system may be deceived to the detriment of honest candidates. This system was already used to deceive the Google searcher so that it would index pages through key words that were not visible for the user and had no relation with the real content of the page.

[123]Recital 49- The processing of personal data to the extent strictly necessary and proportionate for the purposes of ensuring network and information security shall be considered a legitimate interest of the data controller, i.e. the ability of a network or an information system to resist, at a given level of confidence, accidental events or unlawful or malicious actions that compromise the availability, authenticity, integrity and confidentiality of stored or transmitted personal data, and the security of the related services offered by, or accessible via, those networks and systems, by public authorities, by computer emergency response teams (CERTs), computer security incident response teams (CSIRTs), by providers of electronic communications networks and services and by providers of security technologies and services, constitutes a legitimate interest of the data controller concerned. This could, for example, include preventing unauthorised access to electronic communications networks and malicious code distribution and stopping 'denial of service' attacks and damage to computer and electronic communication systems.

[124] Article 25. Preservation of documents.

1. Obliged subjects shall store the documentation for a period of ten years since the fulfilment of the obligations established herein is executed, and they shall be erased afterwards. After five years pursuant to the termination of the business relation or the execution of the

such an event, the legal basis to perform the processing would be to comply with an applicable legal obligation for the data controller. Likewise, other legal basis may be employed to legitimate the processing of data in the log registers.

In either case, the controller shall be aware of the obligations and the limits established in the sectorial regulation, for such legal bases do not allow for the processing of personal data contained in the log file for different purposes such as the assessment of the performance or the evolution of the AI system. Therefore, the controller must ensure that guarantees are implemented to avoid access to and use of such record for purposes for which no legal grounds are available.

The developers of the AI component whenever they are using ML-based solutions should implement other registers for the purposes of documenting and meeting the principle of accountability to allow a traceability of the origin of the training data and the validation of such data, as well as records of analysis performed on the validity of such data and the results thereof.

## G. ASSESSMENT OF THE PROPORTIONALITY AND THE NEED FOR SUCH PROCESSING

Article 35.7 of the GDPR established that, at the time to perform a PIA, an assessment on the need and the proportionality of the processing with regard to the purpose of the processing needs to be carried out.

The use of AI-based solutions sets the controllers and the processor on a position where the processing, by virtue of its characteristics, and, more precisely, by virtue of the technology chosen, may entail a high level or risk. Therefore, an assessment should be made on whether the object of the processing may not be achieved through the use of another type of solution with the same functionality, with an acceptable performance and a lesser risk.

That is to say, the availability, or the novelty of a technology does not justify on itself the use of such a technology. Such technology must be subject to consideration, and an analysis must be carried out on whether the type of processing that is being proposed is a balanced solution. It should be assessed that the general interest and the society, as a whole, will get enough advantages and benefices as opposed to the disadvantages. Such assessment should be regarding the rights and freedoms of data subjects affected by the processing.

ISO-3100 “Risk Management. Principles and Guidelines” exposes in section 5.5 “Risk management” the general conditions needed to manage the risk in any type of scope:

> *“THE SELECTION OF THE MOST SUITABLE OPTION TO MANAGE THE RISK MEANS THAT A COMPENSATION NEEDS TO BE OBTAINED OF THE COSTS AND THE EFFORTS OF THE IMPLEMENTATION WITH REGARD TO THE ADVANTAGES THAT ARE BEING OBTAINED WHILE KEEPING INTO ACCOUNT THE LEGAL REQUIREMENTS AS WELL AS THE REGULATION REQUIREMENTS OR OTHERWISE, SUCH AS THE SOCIAL RESPONSIBILITY AND THE PROTECTION OF THE NATURAL BACKGROUND. THE DECISIONS SHALL ALSO BE ADOPTED WITH REGARD TO THE RISKS WHOSE PROCESSING IS NOT JUSTIFIABLE FROM AN ECONOMIC POINT OF VIEW, FOR EXAMPLE, SEVERE (HIGHLY NEGATIVE CONSEQUENCES) BUT RARE (LOW PROBABILITY) RISKS.*

occasional operation, the documentation stored shall only be accessed by internal control bodies of the subject obliged, with the inclusion of technical units of prevention and, as the case may be, those in charge of their legal defence.
More precisely, obliged subjects shall preserve them for their use in every research or investigation, in terms of possible money-laundering or terrorism financing by the Executive Service of the commission or any other legally entitled authority.
a) A copy of the documents that may be requested in application of the measures of due diligence for a period of ten years upon termination of the business relation or completion of the operation.
b) An original document or copy with probative force of the documents or records that duly credit the operations, the intervening parties to such operations and the business relations for a period of ten years upon completion of the operation or termination of the business relation.
2. Obliged subjects, with the exceptions established by regulation, shall store the copies of the identification documents referred to in Article 3.2 in optical, magnetic or electronic format so that the integrity of the data, an accurate reading of the data, the impossibility of manipulation and the suitable preservation and localisation of such data are ensured.
In any event, the filing system of obliged data subjects shall ensure a suitable management and availability of the documentation, both for internal control purposes and for due attention to the time and manner of the requirements by the authorities.

Whenever we are dealing with personal data protection, the risks to be taken into account are the risks regarding the rights and freedoms of the data subjects. If it is the society who is undertaking such risks, the advantages obtained should not only be assessed from the entity's perspective, but rather from the point of view of the society within the scope were the processing is deployed.

More precisely, in the event of processing using AI-based solutions in order to adopt decisions or to help adopt decisions, it is recommended that the PIA carries out a comparative analysis between the performance obtained by a qualified human operator and the output shown by models capable of forecasting scenarios or adopting decisions automatically[125]. In such an event, the real input conditions of the data need to be taken into account together with the context where the processing is deployed. Furthermore, such an analysis should cover both strict decision-making aspects and collateral effects for the data subjects.

## H. AUDIT

The audit process on a processing including an AI component may have a different extent: audit on the development process of the component, on the releasing of the model, on specific aspects of the processing, on the operation, the security, the robustness against model attacks, etc. Furthermore, the audit may be carried out externally or internally, and it may have a supervisory or a transparency purpose.

As establishes by the principle of accountability, the guarantees established to manage the risk must be documented and collect enough information to enable a satisfactory and verifiable accreditation of the actions taken. Such documentation must enable the traceability of the decisions and verifications made pursuant to the minimisation principles referred above. In summary, not only the processing needs to be audited, but the processing must be auditable throughout its life cycle, including upon withdrawal thereof.

More precisely, it is necessary to perform an audit to establish compliance with GDPR, as well as to verify the validity of the processing that is based on AI solutions. In order to be effective, the auditing process must be carried out under the same conditions as in a real operation context, more precisely, in order to assess:

- The existence of a documented process of analysis, development and/or implementation including, as the case may be, the relevant traceability evidence.
- The existence or absence of personal data, profiling or automated decisions on the data subjects without a human intervention, as well as the analysis of the efficiency of the anonymisation and the pseudonymisation methods.
- Analysis on the existence and legitimisation of the processing of special categories of data, more precisely with regard to inferred information.
- The legal grounds for the processing and identification of responsibilities.
- More precisely, when the legal grounds of the processing is the legitimate interest, an assessment of the balance between the different interests and impacts on the rights and freedoms with regard to the guarantees adopted.
- The information and the effectiveness of the implemented transparency mechanisms.
- The application of the principle of accountability and risk management for the rights and liberties of the data subjects and, more precisely, if the obligation or the need to carry out PIAs has been assessed and, if such was the case, the results of such PIAs.

[125]More precisely, in applications where the complexity of the processing has multiple aspects and collateral features that may not be simplified. A reflection on said aspect may be found at- Why we cannot trust artificial intelligence in medicine, Matthew DeCamp, Jon C Tilburt, The Lancet, correspondence, volume 1, issue 8, December 01, 2019

- With regard to the above, the application of data protection measures by default and by design, inter alia:
    - The prior analysis of the need to process personal data, in terms of quantity and extent, in the several phases pursuant to minimisation criteria.
    - The analysis of the accuracy, fidelity, quality and biases of the data used or gathered for the development or the operation of the AI component, as well as the data sanitation methods used with regard to the data.
    - The verification and performance of tests and validation of the precision, the accuracy, the convergence, the consistence, predictability and any other metrics on the eligibility of the algorithms used, profiling and inferences. Furthermore, the verification that such parameters meet the requirements needed for the processing.
- The suitability of the security measures in order to avoid privacy risks.
- The training and education of the data controller personnel involved in the development and operation of the AI component. In this last case, a special attention to the accurate interpretation of the inferences.
- The obligation, necessity and, as the case may be, the qualification of the DPO.
- The addition of mechanisms that guarantee the fulfilment of the rights of the data subjects, more precisely, the ex officio erasure of personal data, with a great attention to minors' rights.
- The fulfilment of the limitations on automated decisions without a human intervention, the assessment, as the case may be, of the quality of the human intervention and the supervisory mechanisms adopted. More precisely, when the legal grounds are the explicit consent, the identification of the guarantes adopted in order to verify that such a consent is free given.
- The application of some of the guarantees established in Chapter V of the GDPR in the event that international data transfers exist.
- In general terms, the fulfilment of the requirements and obligations of the GDPR and, more precisely, those stated in this document.

As previously established, the AI-based solution shall be integrated within a specific processing, with specific characteristics and a certain operational context. An audit of the isolated AI-based solution, without taking into account the context or the background, shall be incomplete and shall offer partial and unrealistic results[126].

A critical aspect of the audit is to guarantee that the AI-based solution is being used for the purpose for which it was designed, with special attention to its use by the operators of the system. Furthermore, when it is a component that has been acquired from a third party, the other collateral processing that could be performed by such component needs to assessed as well as whether legal or regulatory consequences that may arise out of such use[127].

The use of solutions or real-time automated audit tools is advisable in systems with automated decision-making in order to ensure that the output is consistent and precise, as well as in order to allow that erroneous decisions be aborted or cancelled before irreversible consequences arise.

[126]For example, the use of the ROC curves to prove that the performance of the face recognition systems must be performed in contexts reflecting real operating conditions.

[127]For example, the Google recaptcha component also performs analytic functions: https://developers.google.com/recaptcha/docs/analytics

## V. INTERNATIONAL TRANSFERS

The development or deployment of an AI component based on Cloud services, the disclosure of user data to third parties in order to evolve the AI model or the release of the AI component in the event that personal data exist that are inherent to the model, may imply cross-border data flows to third countries. Data flows of data occurring within the frame of the European Economic Area (EU Members plus Iceland, Norway and Liechtenstein) do not qualify as international transfers.

The guarantees appearing in Chapter V of the GDPR *"Transfers of personal data to third countries or international organisations"* need to be applied to such transfers. It is especially important to establish mechanisms in order to allow the contracts signed in this context of international transfers to be managed smoothly, ensuring at the same time that the client, as data controller, has enough information on the contractors or prospect contractors and keeps the power to adopt decisions. When international transfers exist, data subjects must be informed pursuant to the terms in Articles 13 and 14 of the GDPR and such international transfers must be included within the records of processing activities.

## VI. CONCLUSIONS

Commercialization, for entities and consumers in general, of processing including solutions based on disruptive technologies, such as technologies based on AI components, requests for the implementation of quality and security guarantees [128]like in every other product. The availability or novelty of a technology is not reason enough to commercialise products that do not meet a certain level of quality of service, especially when such requirements are laid down by regulation. Researchers and the AI-based industry need guidelines to help them to get compliance and legal certainty in their projects, products and services.

With regard to the protection of personal data, compliance with the provisions in the GDPR requests a certain level of maturity in the AI solutions that makes it possible to objectively assess the compliance of the processing and the implementation of guarantees and measures to manage the risks.

With regard to such risks, technology in general, [129]and AI technologies in particular, may have a multiplying effect of the ethical deficiencies that are already present in society or those that are perpetuated from the past and are registered in historical data. The application of the GDPR and the guarantees it incorporates allows to minimise such risks.

The use of transparency, risk management strategies and audit and certification mechanisms will allow for compliance with the provisions in the GDPR but this will also enhance trust by users in AI-based products and services and open a new market within this sector: privacy engineers, data scientist, auditors, certification schemes, authorised professionals, etc. These new development opportunities include the creation of portability schemes.

Even more, AI-applications may be a useful tool to implement guarantees to ensure data protection[130].

This document seeks to be a mere introduction to the compliance of the processing including AI components and does not cover all possibilities and risks arising out of the use of AI-based solutions in personal data processing[131].

Finally, it must be highlighted that one of the main problems of AI-based solutions is not the AI itself, but rather the way in which the AI technology and the new psychological biases arising out of its use are going to be used by individuals. More precisely, careful attention needs to be paid so as not to assign responsibilities to unsupervised AI components without adopting a rational, critical and discerning position. The delegation of decision-making on machines is not something new; it has already been used with determinist algorithms, but the bias of assigning a kind of authority or reputability to a result just inferred by an AI-based solution may increase the risks arising out of such a delegation of responsibility.

[128] Security understood as “safety”, that is to say, that it does not harm or cause damage.
[129] Such as with bullying, harassment or fake news situations.
[130] Such as AI applications to depurate databases or to analyse the security of the systems.
[131] There are other pending questions, such as the cases of human-machine hybridisation, group privacy, the demutualisation or even to what extent the decision adopted by an AI in military applications falls within the category of automated-decisions with legal consequences.

## VII. REFERENCES

The references hereunder have been organised pursuant to their relevance at the time to prepare this document:


Regulation (EU) 2016/679 of the European Parliament and of the Council, of 27 April 2016, relating to the protection of natural persons in relation to the processing of personal data and the free circulation of these data and repealing Directive 95/46/EC (General Data Protection Regulation).

Organic Law 3/2018 of 5 December on Personal Data Protection and Guarantee of Digital Rights.

Big data, artificial, intelligence, machine learning and data protection. ICO information commissioner's office, September 2017

How can humans keep the upper hand? The ethical matters raised by algorithms and artificial intelligence, CNIL, December 2017

Draft Ethics Guidelines for Trustworthy AI, High-Level Expert Group on Artificial Intelligence, European Commission, December 2018, Brussels

Artificial Intelligence for Europe. Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions -, European Commission, April 2018, Brussels.

Discrimination, artificial intelligence, and algorithmic decision-making, Prof. Frederik Zuiderveen Borgesius, Consejo de Europa 2018

Robustness and Explainability of Artificial Intelligence. From technical to policy solutions. European Commission. 2020

Spanish R+D+I in Artificial Intelligence, Ministry of Science, Innovation and Universities, 2019

Automating Society, Taking Stock of Automated Decision-Making in the EU, AW AlgorithmWatch gGmbH, January 2019, Berlin

Opinion 7/2015 Meeting the challenges of big data. European Data Protection Supervisor. EDPS, 19 November 2015.

Manual Práctico de Inteligencia Artificial en Entornos Sanitarios, Beunza Nuin J.J., Puertas Sanz E., Emilia Condés Moreno E., Elsevier, January 2020

Recommendations on the processing of personal data in Artificial Intelligence. Document to be subject to a public consultation. Text approved by the Member Entities of the Iberian-American Network on Data Protection during the meeting on 21 June 2019 in Naucalpan de Juarez, Mexico.

Artificial Intelligence and Data Protection in Tension, Centre for Information Policy Leadership, October 2018

AI and Data Protection – Balancing tensions – Slaughter and May, PLC Magazine, August 2019

Artificial Intelligence in Finance, Alan Turing Institute, Hanken School of Economics, April 2019, Finland

Report on Algorithmic Risk Assessment Tools in the U.S. Criminal Justice System, Partnership on AI (PAI), April 2019 https://www.partnershiponai.org/report-on-machine-learning-in-risk-assessment-tools-in-the-u-s-criminal-justice-system/

Human-AI Collaboration Framework & Case Studies, Partnership on AI (PAI), September 2019, https://www.partnershiponai.org/human-ai-collaboration-framework-case-studies/

AI Index 2018 Annual Report, Zoe Bauer et al., AI Index Steering Committee, Human-Centered AI Initiative, Stanford University, Stanford, CA, December 2018

Artificial Intelligence and the Future of Humans, J. Anderson et al., Pew Research Center, December 2018

Series regarding Artificial Intelligence from the ICO blog https://ico.org.uk/about-the-ico/what-we-do/tech-and-innovation/blogs/

Artificial Intelligence Safety and Cybersecurity: a Timeline of AI Failures Roman V. Yampolskiy, University of Louisville 2018

Why we cannot trust artificial intelligence in medicine, Matthew DeCamp, Jon C Tilburt, The Lancet, correspondence, volume 1, issue 8, December 01, 2019

Exploring the landscape of spatial robustness, Logan Engstrom, MIT https://arxiv.org/pdf/1712.02779.pdf

Towards federated learning at scale: system design Keith, Bonawitz et al. https://arxiv.org/pdf/1902.01046.pdf

What's your ML Test Score? A rubric for ML production systems, Eric Breck, 2018, Google Inc. https://www.eecs.tufts.edu/~dsculley/papers/ml_test_score.pdf

A Survey on Security Threats and Defensive Techniques of Machine, Qiang Liu et al., IEEE Access, ISSN: 2169-3536, February 2018

Opinion 05/2014 on anonymisation techniques, Article 29 Working Party, 2014

Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7of Directive 95/46/EC, Article 29 Working Party, April 2014.

Guidelines on the right to portability of the data Article 29 Working Party- Adopted on 13 December 2016, Last updated and adopted on 5 April 2017

Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. Article 29 Working Party. Adopted on 03 October 2017, Last updated and adopted on 06 February 2018

Guidelines on Data Protection Officers (DPO), Article 29 Working Party Adopted on 13 December 2016, Last updated and adopted on 05 April 2017

Guidelines on Privacy by Design, AEPD, October 2019.

Practical Guide for impact assessment with regard to personal data protection, AEPD, October 2018

"Guidelines and guarantees for the processes of anonymisation of personal data" (AEPD 2016)

Lists of types of data processing requiring a PIA (Art. 35.4), AEPD, 2019.

Report Model on the Assessment of the Impact of Data Protection for Public Administrations, AEPD, July 2019

ISO-3100 "Risk Management. Principles and Guidelines"

## VIII. ANNEX: CURRENT AI-BASED SERVICES

The list presented in this annex is without limitation and only seeks to be an example in order to illustrate the extent of the services that are currently being provided based on AI:

- Internet services
    - Captchas, chatbots, fraud detection, advertisement customisation…
- Human Resources
    - Selection of candidates
- Financial services
    - Forecasts about mortgages based on an analysis of the client's profile, monitoring of transactions to detect fraudulent activities based on consumption habits, automated financial investment...
- Health and Healthcare
    - Diagnosis based on the analysis of images, forecasts on readmission rates of patients based on the analysis of data, health maps, mental health analysis, suicide prevention, mental health chatbots, risk prediction based on analytical parameters, diagnosis through analysis of the pathological sample, natural languages processing of medical records, genetic analysis, electro-diagnosis, development of vaccines and medicines...
- Trade and communication:
    - Product recommendations based on the client's profile, and on the analysis of purchases, maximisation of the scope of the products and services for a group of clients, virtual travel agents, monitoring of social media...
- Public services and supplies:
    - Smart counters and forecast of the consumption demand of clients estimation of the cost of certain maintenance services, assignation of processing in the public healthcare system, automated processing of fines, support to a decision by the justice administration...
- Transport:
    - Self-driving cars, smart traffic lights, optimisation of routes and schedules for public transports...
- Education:
    - Customised content and training with regard to the needs of the students, exam marking, detection of plagiarism or fraud in papers, automatic mentoring, detection of abnormal students...
- Security:
    - Face recognition, fingerprints, behaviour detection, border patrolling, analysis of indication of fraud, analysis of activity records, intrusion detections, analysis of communications...
- Household:
    - Smart assistants, smart mirrors, household appliances, security...
- Other

- Drawing tools, assistance to artistic creation, optimisation of sport training programs...