AI in the GDPR - outline

- AI in the conceptual framework of the GDPR
- AI and the data protection principles
- AI and legal bases
- AI and transparency
- AI and data subjects' rights
- Automated decision making
- AI and privacy by design

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

AI in the conceptual framework of the GDPR

- Unlike the 1995 Data Protection Directive, the **GDPR contains** some terms referring to the **Internet** (Internet, social networks, website, links, etc.), but **it does not contain the term "Artificial Intelligence**", nor any terms expressing related concepts
- The GDPR is focussed on the **challenges emerging for the Internet** — which were not considered in the 1995 Data Protection Directive, but were well present at the time when GDPR was drafted— rather than on new issues pertaining to AI, which only acquired social significance in most recent years.
- **However, many AI provisions are relevant to GDPR**

Article 3
Territorial scope

1. This Regulation applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of whether the processing takes place in the Union or not.

2. This Regulation applies to the processing of personal data of data subjects who are in the Union by a controller or processor not established in the Union, where the processing activities are related to:

(a) the offering of goods or services, irrespective of whether a payment of the data subject is required, to such data subjects in the Union; or

(b) the monitoring of their behaviour as far as their behaviour takes place within the Union.

[…]

Article 4
Definitions

- **(1) 'personal data'** means any information relating to an identified or identifiable natural person **('data subject')**; an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;
- (**Data subject**: the natural person whom information relates to)

- **(2) 'processing'** 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,….
- **(7) '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…
- **(8) 'processor'** means a natural or legal person, public authority, agency or other body which **processes** personal data **on behalf of the controller;**

Article 5

Principles relating to processing of personal data

- Lawfulness, fairness and transparency
- Purpose limitation
- Data minimisation
- Data accuracy
- Storage limitation
- Integrity and confidentiality
- Accountability principle

Article 6
Lawfulness of processing

1. Processing shall be lawful only if and to the extent that at least one of the following applies:

(a) the data subject **has given consent** to the processing of his or her personal data for one or more specific purposes;

(b) processing is **necessary for the performance of a contract** to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;

(c) processing is **necessary for compliance with a legal obligation** to which the controller is subject;

(d) processing is **necessary in order to protect the vital interests** of the data subject or of another natural person;

(e) processing is **necessary for the performance of a task carried out in the public interest** or in the exercise of official authority vested in the controller;

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

2. […]

Article 4(1) GDPR: Personal data - identification

Here is how **personal data** are defined in **Article 4 (1) GDPR:**

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

- What about Natural phenomena?
- What about general medical information on human physiology or pathologies?

Article 4(1) GDPR: Personal data - identifiability

- **Recital (26)** addresses **identifiability**, namely, the conditions under which **a piece of data which is not explicitly linked to a person, still counts as personal data, since the possibility exists to identify the person concerned**.
- Identifiability depends on the availability of **"means reasonably likely to be used" for successful reidentification**, which in its turn, depends on the technological and sociotechnical state of the art:
  - *To determine whether a natural person is identifiable, account should be taken of **all the means reasonably likely to be used**, such as singling out, either by the controller or by another person **to identify the natural person directly or indirectly**. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all **objective factors, such as the costs of and the amount of time required for identification**, taking into consideration the available technology at the time of the processing and technological developments.*

Article 4(1) GDPR: Personal data - pseudonymisation

- **Pseudonymisation:** the data items that identify a person are substituted with a pseudonym, but **the link between the pseudonym and the identifying data items can be retraced by using separate info** (e.g., a table linking pseudonyms and real names, or through cryptography key to decode the encrypted names)
- Recital (26) specifies that **pseudonymised data still are personal data**.
  - *Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be* ***information on an identifiable natural person****.*

Article 4(1) GDPR: Personal data – connection with technological developments

- The connection between the personal nature of information and **technological development** is mentioned at **Recital (9) of Regulation 2018/1807*:**
  - If technological developments make it possible to turn anonymised data into personal data, such data are to be treated as personal data, and Regulation (EU) 2016/679 is to apply accordingly.

- The concept of **non-personal data** is **not positively defined in the EU legislation**
  - **Examples of non-personal data**: aggregated and anonymised datasets used for Big Data analytics, data on precision farming that can help to monitor and optimise the use of pesticides and water, or data on maintenance needs for industrial machines."

**(Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union)*

AI and GDPR definition of personal data: Reidentification & further inferences

In connection with the GDPR definition of personal data, AI raises in particular two key issues:

(1) the “re-personalisation” of anonymous data, namely the **reidentification** of the individuals to which such data are related;

(2) the **inference** of further personal information from personal data that are already available.

Reidentification

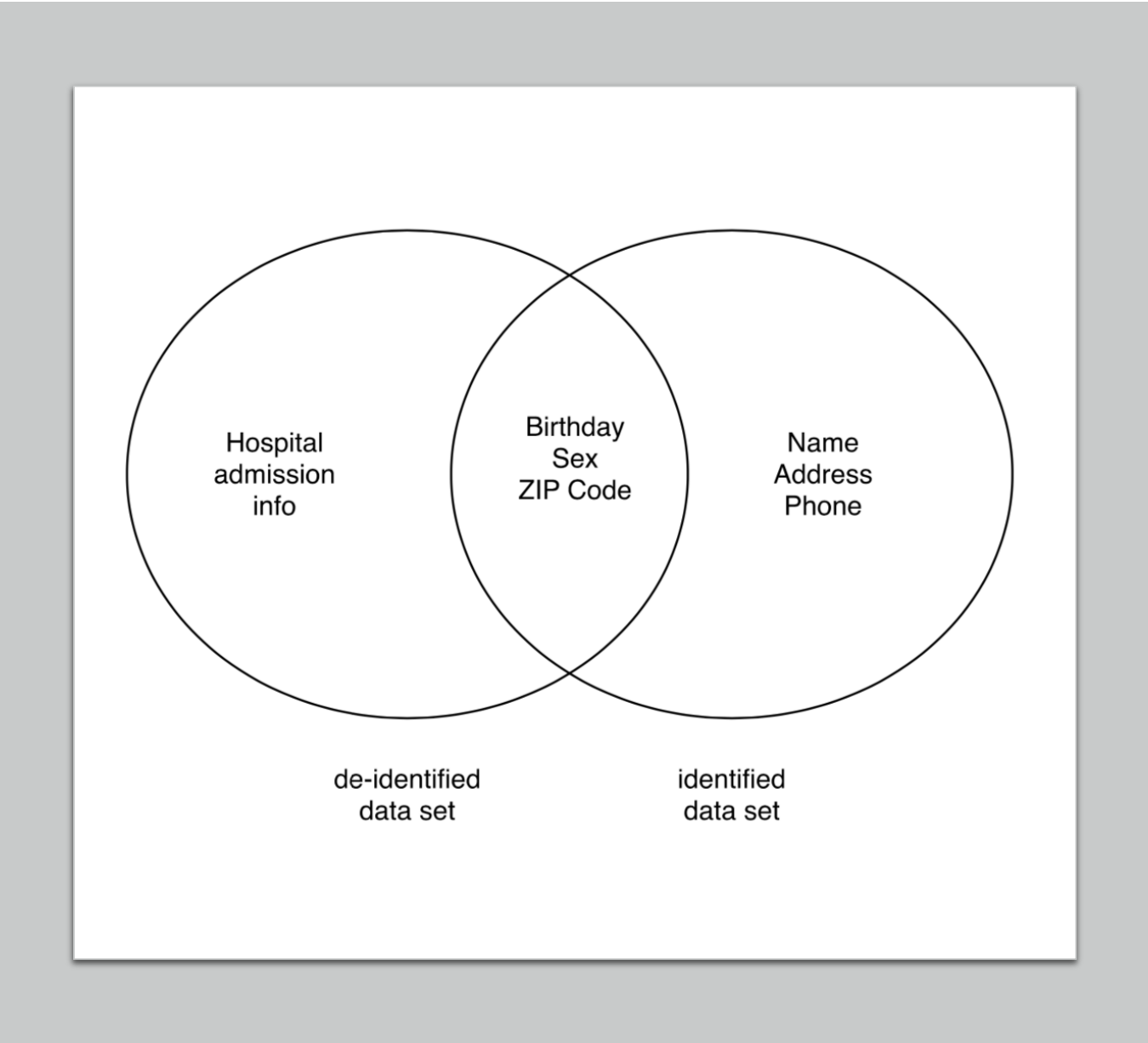
AI, and methods for **computational statistics**, increases the identifiability of apparently anonymous data, since they enable **nonidentified data** (including data having been anonymised or pseudonymised) to be **connected to the individuals concerned**

- **[N]umerous supposedly anonymous datasets have recently been released and reidentified**.
  - In 2016, journalists reidentified politicians in an anonymized browsing history dataset of 3 million German citizens, uncovering their medical information and their sexual preferences.
  - A few months before, the Australian Department of Health publicly released de-identified medical records for 10% of the population only for researchers to reidentify them 6 weeks later.
  - Before that, studies had shown that de-identified hospital discharge data could be reidentified using basic demographic attributes and that diagnostic codes, year of birth, gender, and ethnicity could uniquely identify patients in genomic studies data.
  - Finally, researchers were able to uniquely identify individuals in anonymized taxi trajectories in NYC27, bike sharing trips in London, subway data in Riga, and mobile phone and credit card datasets. (Rocher et al 2019).

The reidentification of data subjects is usually based on **statistical correlations between nonidentified data and personal data** concerning the same individuals.

The connection between identified and de-identified data

The figure illustrates the connection between an identified and a de-identified data set that enabled the reidentification of the health record of the governor of Massachusetts. This result was obtained by searching for de-identified data, such as the information on Hospital admission, that matched the Governor's date of birth, ZIP code and gender.



The connection between identified and de-identified data

The Netflix price database case, in which anonymised movie ratings could be re-identified by linking them to non-anonymous ratings in IMDb (Internet Movie Database). In fact, knowing only two non-anonymous reviews by an IMDb user, it was possible to identify the reviews by the same user in the anonymous database.

Reidentification

- Reidentification as **a specific kind of inference** of personal data. For an item to be linked to a person, **it is not necessary that the data subject is identified with absolute certainty; a degree of probability may be sufficient**

- Thanks to AI & Big Data the identifiability of the data subjects has vastly increased.
- As it has been argued, **"in any 'reasonable' setting there is a piece of information that is in itself innocent, yet in conjunction with even a modified (noisy) version of the data yields a privacy breach."**

This possibility can be addressed in two ways:

1. The first consists in ensuring that data is deidentified in ways that **make it more difficult to reidentify** the data subject;
2. The second consists in **implementing security processes and measures** for the release of data that contribute to this outcome.

Inferred personal data

- AI systems may **infer new information** about data subjects, by applying algorithmic models to their personal data.
- The key issue is **whether the inferred information should be considered as new personal data**, distinct from the data from which it has been inferred.
  - Assume for instance, that an individual's sexual orientation is inferred from his or her facial features or that an individual's personality type is inferred from his or her online activity. Is the inferred sexual orientation or personality type a new item of personal data? Even when the inference only is probabilistic?

- If the inferred information counts as new personal data, then automated inferences would trigger all the consequences that the processing of personal data entails according to the GDPR.

Legal status of automatically inferred information

- Some clues on the legal status of automatically inferred information can be obtained by considering the status of information inferred by humans: there is **uncertainty about whether assertions concerning individuals, resulting from human inferences and reasoning may be regarded as personal data.**

- This issue has been examined by the **ECJ** in **Joint Cases C-141 and 372/12,** where it was denied that the legal analysis, by the competent officer, on an **application for a residence permit** could be deemed personal data. **According to the ECJ, only the data on which the analysis was based** (the input data about the applicant) **as well as the final conclusion** of the analysis (the holding that the application was to be denied) **were to be regarded as personal data**.
- This qualification did not apply to the intermediate steps (the intermediate conclusions in the argument chain) leading to the final conclusion.

Legal status of automatically inferred information

- In the subsequent decision on **Case C-434/16,** concerning a candidate's request to exercise data protection rights relative to an **exam script and the examiners' comments**, **the ECJ apparently departed from the principle stated in Joint Cases C-141 and 372/12**, arguing that the examiner's comments, too, were personal data.
- However, **the Court held that data protection rights, and in particular the right to rectification, should be understood in connection with the purpose of the data at issue**. Thus, according to the Court, the right to rectification does not include a right to correct a candidate's answers or the examiner's comments (unless they were incorrectly recorded).
- In fact, according to the ECJ, data protection law is not intended to ensure the accuracy of decision-making processes or good administrative practices. Thus, an examinee has the right to access both to the exam data (the exam responses) and the reasoning based on such data (the comments), but **he or she does not have a right to correct the examiners' inferences (the reasoning)** or the final result.

The view that inferred data are personal data was endorsed by the Article 29 WP (Opinion 4/2007)

- *in case of automated inference (profiling) data subjects have the right to access both the input data and the (final or intermediate) conclusions automatically inferred from such data.*

Article 4(2) GDPR: Profiling

The definition of profiling, while not explicitly referring to AI, addresses processing that today is typically accomplished using AI technologies. This processing consists in using the data concerning person to infer information on further aspects of that person:

> *'**profiling**' means 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*

Article 4(2) GDPR: Profiling

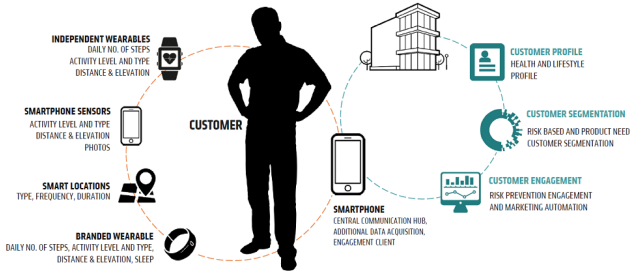
According to the **Article 29 WP,** profiling **aims at classifying persons into categories of groups sharing the features being inferred** (Opinion 216/679):

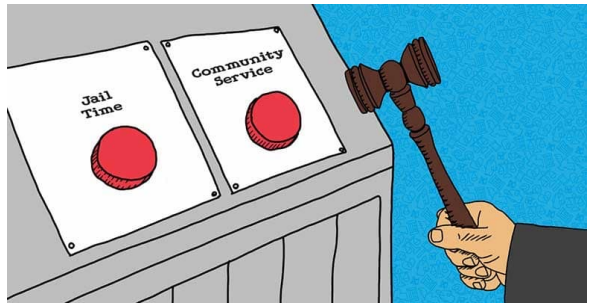
> "broadly speaking, profiling means ***gathering information about an individual*** (or group of individuals) and evaluating their characteristics or behaviour patterns in order to place them into a certain category or group, in particular to analyse and/or make predictions about, for example, their ability to perform a task, interests or likely behaviour."

AI and profiling

- **AI & Big Data have vastly increased the opportunities for profiling.**
- Assume that a classifier has been trained on a vast set of past examples, which **link certain features** of individuals (the **predictors**), to **another feature** of the same individuals (the **target**).
- Through the training, the system has learned an algorithmic model that can be applied to new cases: **if the model is given predictors-values concerning a new individual, it infers a corresponding target value for that individual, i.e., a new data item concerning him or her.**
  - the likelihood of heart disease of applicants for insurance on the basis of their health records, their habits or social conditions;
  - the creditworthiness of loan applicants on the basis of their financial history, their online activity and social condition;
  - the likelihood that convicted persons may reoffend on the basis their criminal history, their character (as identified by personality test) and personal background.

These predictions may trigger automated determinations concerning, respectively, the price of a health insurance, the granting of a loan, or the release on parole.





AI and profiling

A learned correlation may also concern a person's propensity to respond in certain ways to certain stimuli. This would enable the transition from prediction to behaviour modification (both legitimate influence and illegal or unethical manipulation).

- Examples: trigger the desired purchasing behaviour, or the desired voting behaviour.

Inferences as personal data

We need to **distinguish the general correlations that are captured by the learned algorithmic model, and the results of applying that model** to the description of a particular individual.

- *Consider for instance a machine learning system that has learned a model (e.g., a neural network or a decision tree) from a training set consisting of previous **loan applications and outcomes**.* The system's training set consists of personal data: e.g., for each borrower, his name, the data collected on him or her —age, economic condition, education, job, etc.— and the information on whether he or she defaulted on the loan.

- The learned algorithmic model no longer contains personal data, since it links any possible combinations of possible input values (predictors) to a corresponding likelihood of default (target). **The correlations embedded in the algorithmic model are not personal data, since they apply to all individuals sharing similar characteristics. We can possibly view them as group data, concerning the set of such individuals** (e.g., those who are assigned a higher likelihood of default, since they have a low revenue, live in a poor neighbourhood, etc.).

- **Assume that the algorithmic model is then applied** to the input data consisting in the description of a new applicant, in order to determine that applicant's risk of default. **In this case both the description of the applicant and the default risk attributed to him or her by the model represent personal data,** the first being collected data, and the second inferred data.

Rights over inferences: access

Since inferred data concerning individuals also are personal data under the GDPR —at least when they are used to derive conclusions that are or may be acted upon— **data protection rights should in principle also apply**, though concurrent remedies and interests have to be taken into account.

According to the Article 29 Working Party, in the case of automated inferences (profiling) **data subjects have a right to access** both the personal data used as **input** for the inference, and the personal data obtained as (final or intermediate**) inferred output**.

Rights over inferences: rectification

On the contrary, the right to rectification only applies to a limited extend. When the data are processed by a public authority, it should be considered whether review procedures already exist which provide for access and control. In the case of processing by private controllers, the right to rectify the data should be balanced with the respect for autonomy of private assessments and decisions.

According to the Article 29 Working Party data subjects **have a right to rectification** of inferred information not only when the inferred information is "verifiable" (its correctness can be objectively determined), but also when it is the outcome of unverifiable or probabilistic inferences (e.g., a the likelihood of developing heart disease in the future).

In the latter case, rectification may be needed not only when the statistical inference was mistaken, but also when the data subject provides specific additional data that support a different, more specific, statistical conclusion. This is linked to the fact that statistical inferences concerning a class may not apply to subclasses of it

A general right to “reasonable inference”?

Legal scholars have argued that data subjects should be granted a general right to “reasonable inference” i.e.,, the right that any assessment of decision affecting them is obtained through automated inferences that are reasonable, respecting both **ethical and epistemic standards.**

Data subject should be entitled to challenge the inferences (e.g. credit scores) made by an AI system, and not only the decisions based on such inferences (e.g., the granting of loans). **It has been argued that for an inference to be reasonable it should satisfy the following criteria:**

a) **Acceptability:** the input data (the predictors) for the inference should be normatively acceptable as a basis for inferences concerning individuals (e.g., to the exclusion of prohibited features, such as sexual orientation);
b) **Relevance:** the inferred information (the target) should be relevant to the purpose of the decision and normatively acceptable in that connection (e.g., ethnicity should not be inferred for the purpose of giving a loan).
c) **Reliability:** both input data, including the training set, and the methods to process them should be accurate and statistically reliable

A general right to “reasonable inference”?

Controllers, conversely, should be prohibited to base their assessment or decisions on unreasonable inferences, and they should also have the obligation to demonstrate the reasonableness of their inferences.

The idea that unreasonable automated inference should be prohibited only applies to inferences meant to lead to **assessments and decisions affecting the data subject**. They should not apply to inquiries that are motivated by merely cognitive purposes, such as those pertaining to scientific research.

Consent

Art 4(11)

'consent' of the data subject means any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her;

Art 7 (Conditions for consent)

1. Where processing is based on consent, the controller shall be able to demonstrate that the data subject has consented to processing of his or her personal data.

2. If the data subject's consent is given in the context of a written declaration which also concerns other matters, the request for consent shall be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language. Any part of such a declaration which constitutes an infringement of this Regulation shall not be binding.

3. The data subject shall have the right to withdraw his or her consent at any time. The withdrawal of consent shall not affect the lawfulness of processing based on consent before its withdrawal. Prior to giving consent, the data subject shall be informed thereof. It shall be as easy to withdraw as to give consent.

4. When assessing whether consent is freely given, utmost account shall be taken of whether, inter alia, the performance of a contract, including the provision of a service, is conditional on consent to the processing of personal data that is not necessary for the performance of that contract.

Information to be provided to the data subject (art 13-14, recital 42 GDPR, art29WP Guidelines on consent)

- **Identity of the controller** and (where applicable) controller's representative, + their contact details
- **Contact details** of the **data protection officer**
- **Purposes** of the processing for which the personal data are intended
- **Legal basis** for the processing
- **Categories of personal data concerned**
- **Recipients or categories of recipients** of the personal data
- **Period** for which the personal data will be stored, or if that is not possible, the criteria used to determine that period
- Existence of the right to request from the controller **access to and rectification** or **erasure** of personal data or restriction of processing concerning the data subject and to object to processing as well as **the right to data portability**
- **Right to lodge a complaint** with a supervisory authority
- **Source** from which personal data originate
- **Existence of automated decision-making**, including **profiling**

Article 17
Right to erasure (‘right to be forgotten’) (1/2)

1. The data subject shall have the **right to obtain** from the controller the **erasure** of personal data concerning him or her **without undue delay** and the controller shall have the obligation to erase personal data without undue delay where one of the following grounds applies:

- **(a) the personal data are no longer necessary in relation to the purposes for which they were collected or otherwise processed;**
- (b) the data subject **withdraws consent** on which the processing is based according to point (a) of Article 6(1), or point (a) of Article 9(2), and where there is **no other legal ground for the processing**;
- (c) the data subject **objects to the processing** pursuant to Article 21(1) and there are **no overriding legitimate grounds** for the processing, or the data subject objects to the processing pursuant to Article 21(2);
- (d) the personal data have been **unlawfully processed**;
- (e) the personal data have to be erased for **compliance with a legal obligation** in Union or Member State law to which the controller is subject;
- (f) the personal data have been collected in relation to the offer of information society services referred to in Article 8(1).
- […]

Article 17
Right to erasure (‘right to be forgotten’) (2/2)

2. Where the controller has made the personal data public and is obliged pursuant to paragraph 1 to erase the personal data, the controller, taking account of available technology and the cost of implementation, shall take reasonable steps, including technical measures, to inform controllers which are processing the personal data that the data subject has requested the erasure by such controllers of any links to, or copy or replication of, those personal data.

3. **Paragraphs 1 and 2 shall not apply to the extent that processing is necessary:**

(a) for exercising the right of freedom of expression and information;

(b) **for compliance with a legal obligation** which requires processing by Union or Member State law to which the controller is subject or for the performance of a task carried out in the **public interest** or in the exercise of official authority vested in the controller;

(c) for reasons of **public interest in the area of public health** in accordance with points (h) and (i) of Article 9(2) as well as Article 9(3);

(d) for archiving purposes i**n the public interest, scientific or historical research purposes or statistical purposes** in accordance with Article 89(1) in so far as the right referred to in paragraph 1 is likely to render impossible or seriously impair the achievement of the objectives of that processing; or

(e) for the establishment, exercise or defence of **legal claims**.

Article 9
Processing of special categories of personal data (1/2)

- 1. Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be **prohibited**.

Article 9
Processing of special categories of personal data (2/2)

2. Paragraph 1 shall not apply if one of the following applies:

(a) the data subject has given **explicit consent** to the processing of those personal data for one or more specified purposes…

(b) processing is necessary for the purposes of carrying out the obligations and exercising specific rights of the controller or of the data subject in the field of **employment** and **social security** and **social protection** law…

(c) 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…

(d) processing is carried out in the course of its **legitimate activities** with appropriate safeguards by a foundation, association or any other not-for-profit body with a political, philosophical, religious or trade union aim and on condition that the processing relates solely to the members…

(e) processing relates to personal data which are **manifestly made public by the data subject**;

(f) processing is necessary for **the establishment, exercise or defence of legal claims** or whenever courts are acting in their judicial capacity;

(g) processing is necessary for reasons of **substantial public interes**t…

(h) processing is necessary for the purposes of **preventive or occupational medicine**…

(i) processing is necessary for reasons of **public interest in the area of public health**…

(j) processing is necessary for **archiving purposes in the public interest**, **scientific or historical research purposes or statistical purposes**…

Article 22
Automated individual decision-making, including profiling

1. The data subject shall have the right **not to be subject** to a decision based **solely on automated processing**, **including profiling**, which **produces legal effects concerning him or her or similarly significantly affects** him or her.

2. Paragraph 1 shall not apply if the decision:

(a) is necessary for entering into, or performance of, a contract between the data subject and a data controller;

(b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or

(c) is based on the data subject's explicit consent.

Article 22(1) GDPR: The prohibition of automated decisions

- The first paragraph of Article 22 provides for a general right not to be subject to completely automated decisions significantly affecting the data subject:

  The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.

- According to the Article 29 Working Party:

  as a rule, there is a **general prohibition** on fully automated individual decision-making, including profiling that has a legal or similarly significant effect.

- For the application of the prohibition established by Article 22(1), four conditions are needed:

(1) a decision must be taken

(2) it must be solely based on automated processing

(3) it must include profiling

(4) it must have legal or anyway significant effect.

Article 22(1) GDPR: conditions for the prohibition of automated decisions

(1) **a decision must be taken:** requires that a stance be taken toward a person, and that this stance is likely to be acted upon (as when assigning a credit score).

(2) **it must be solely based on automated processing:** requires that humans do not exercise any real influence on the outcome of a decision-making process, even though the final decision is formally ascribed to a person. This condition is not satisfied when the system is only used as a decision-support tool for humans

(3) **it must include profiling:** requires that the automated processing determining the decision includes profiling. (A different interpretation of the condition may be suggested, but Recital (71) seems to confirm the first interpretation)

(4) **it must have legal or anyway significant effect:** Recital (71) mentions the following examples of decision having significant effects: the “automatic refusal of an online credit application or e-recruiting practices”. It has been argued that such effects cannot be merely emotional, and that usually they are not caused by targeted advertising, unless “advertising involves blatantly unfair discrimination in the form of web-lining and the discrimination has non-trivial economic consequences

Article 21 (1) and (2): Objecting to profiling and direct marketing

- Article 21 (1) specifies that the right to object also applies to profiling:
  - The data subject shall have the right to object, on grounds relating to his or her particular situation, at any time to processing of personal data concerning him or her which is based on point (e) or (f) of Article 6(1), including profiling based on those provisions.
- Profiling in the context of direct marketing is addressed in Article 21 (2), which recognises an unconditioned right to object:
  - Where personal data are processed for direct marketing purposes, the data subject shall have the right to object at any time to processing of personal data concerning him or her for such marketing, which includes profiling to the extent that it is related to such direct marketing.

- This means that the data subject does not need to invoke specific grounds when objecting to processing for direct marketing purposes, and that such purposes cannot be "compelling legitimate grounds for the processing which override the interests, rights and freedoms of the data subject".
- Given the importance of profiling for marketing purposes, the unconditional right to object to such processing is particularly significant for the self-protection of data subjects. Controllers should be required to provide easy, intuitive and standardised ways to facilitate the exercise of this right.

Information on automated decision making

Article 13(2)(f) and 14(2)(g) GDPR address a key aspect of AI applications, i.e. automated decision making. The controller has the obligation to provide:

(a) information on "**the existence of automated decision-making**, including profiling, referred to in Article 22(1)" and

(b) "at least in those cases meaningful information about **the logic involved**, as well as the significance and the envisaged **consequences** of such processing for the data subject."

Information on automated decision making

- **Computer scientists** have focused on the technological possibility of providing understandable models of opaque AI systems (and, in particular, of deep neural networks), i.e., model of the functioning of such systems that can be mastered by human experts. For instance, the following kinds of explanations are at the core of current research on explainable AI:

  - **Model explanation**, i.e., the global explanation of an opaque AI system through an interpretable and transparent model that fully captures the logic of the opaque system.

  This would be obtained for instance, if a decision tree or a set of rules was provided, whose activation exactly (or almost exactly) reproduces the functioning of a neural network.

  - **Model inspection**, i.e., a representation that makes it possible to understanding of some specific properties of an opaque model or of its predictions.

  It may concern the patterns of activation in the system's neural networks, or the system's sensitivity to changes in its input factors (e.g. how a change in the applicant's revenue or age makes a difference in the grant of a loan application).

  - **Outcome explanation**, i.e., an account of the outcome of an opaque AI in a particular instance.

  For instance, a special decision concerning an individual can be explained by listing the choices that lead to that conclusions in a decision tree (e.g., the loan was denied because of the applicant's income fell below a certain threshold)

- The explanatory techniques and models developed within computer science are intended for technological experts and assume ample access to the system being explained.

Information on automated decision making

Social scientists have focused on the objective of making explanations accessible to lay people, thus addressing the communicative and dialectical dimensions of explanations. For instance, it has been argued that the following approaches are needed (Miller 2019, Mittelstadt and Wachter 2019).

- **Contrastive explanation**: specifying what input values made a difference, determining the adoption of a certain decision (e.g., refusing a loan) rather than possible alternatives (granting the loan);
- **Selective explanation**: focusing on those factors that are most relevant according to human judgement;
- **Causal explanation**: focusing on causes, rather than on merely statistical correlations (e.g., a refusal of a loan can be causally explained by the financial situation of the applicant, not by the kind of Facebook activity that is common for unreliable borrowers);
- **Social explanation**: adopting an interactive and conversational approach in which information is tailored according to the recipient's beliefs and comprehension capacities.

While these suggestions are useful for the ex-post explanation of specific decisions by a system, they cannot be easily applied ex-ante, at the time of data collection (or repurposing).

Information on automated decision making

- Ex-ante the user should ideally be provided with the following information:

- The **input data** that the system takes into consideration (e.g., for a loan application, the applicant's income, gender, assets, job, etc.), and **whether different data items are favouring or rather disfavouring the outcome** that the applicant hopes for;
- **The target values** that the system is meant to compute (e.g., a level of creditworthiness, and possibly the threshold to be reached in order for the loan to be approved);
- **The envisaged consequence of the automated assessment/decision** (e.g., the approval or denial of the loan application).
- It may also be useful to specify what are **the overall purposes** that the system is aimed to achieve

A right to explanation?

According to Recital (71), the safeguards to be provided to data subjects in case of automated decisions include all of the following:

- specific information
- the right to obtain human intervention,
- the right to express his or her point of view,
- the right to obtain an explanation of the decision reached after such assessment
- the right to challenge the decision.

According to Article 22 the suitable safeguards to be provided include “at least”

- the right to obtain human intervention,
- the right to express his or her point of view,
- the right to challenge the decision.

Thus, two items are missing in article 22 relative to Recital (71): the provision of “specific information” and the right to *obtain an explanation of the decision reached after such assessment”*.

The second omission in particular raises the issue of whether controllers are really required by law to provide an individualised explanation

A right to explanation? Two possible interpretations

- **According to the first interpretation,** the European legislator, by only including the request for specific explanation in the recitals and omitting it from the articles of the GDPR, intended to convey a double message: **to exclude an enforceable legal obligation to provide individual explanations**, while recommending that data controllers provide such explanations when convenient, according to their discretionary determinations.
- **Following this interpretation, providing individualised explanation would only be a good practice, and not a legally enforceable requirement.**

A right to explanation? Two possible interpretations

- **According to the second interpretation**, the European legislator intended on the contrary **to establish an enforceable legal obligation to provide individual explanation,** though without unduly burdening controllers.
- This interpretation is hinted at by the qualifier “at least”, which precedes the reference made to a “right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.” The qualifier seems to suggest that some providers are legally required to adopt further safeguards, possibly including individualised explanations, as indicated in Recital 71.

- **On this second approach, an explanation would be legally needed,** whenever it is practically possible, i.e., whenever it is compatible with technologies, costs, and business practices.

- However, **we should be cautioned against overemphasising a right to individualised explanations as a general remedy to the biases, malfunctions, and inappropriate applications of AI & Big Data technologies** : the right to an explanation is likely to remain underused by the data subjects, given that they may lack a sufficient understanding of technologies and applicable normative standards. Moreover, even when an explanation elicits potential defects, the data subjects may be unable to obtain a new, more satisfactory decision.

Article 25
Data protection by design and by default

1. Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, **the controller shall**, both at the time of the determination of the means for processing and at the time of the processing itself**, implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation**, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects.

2. **The controller shall implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed.** That obligation applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility. In particular, such measures shall ensure that by default personal data are not made accessible without the individual's intervention to an indefinite number of natural persons.

[…]

Article 32
Security of processing

1. Taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk, including inter alia as appropriate:

(a) the pseudonymisation and encryption of personal data;

(b) the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services;

(c) the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident;

(d) a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing.

[…]

European Data Protection Board and European Data Protection Supervisor

Article 68
European Data Protection Board

1. The European Data Protection Board (the 'Board') is hereby established as a body of the Union and shall have legal personality.
2. The Board shall be represented by its Chair.
3. The Board shall be composed of the head of one supervisory authority of each Member State and of the European Data Protection Supervisor, or their respective representatives.
4. Where in a Member State more than one supervisory authority is responsible for monitoring the application of the provisions pursuant to this Regulation, a joint representative shall be appointed in accordance with that Member State's law.
5. The Commission shall have the right to participate in the activities and meetings of the Board without voting right. The Commission shall designate a representative. The Chair of the Board shall communicate to the Commission the activities of the Board.

[…]

Article 70
Tasks of the Board

1. The Board shall ensure the consistent application of this Regulation. To that end, the Board shall, on its own initiative or, where relevant, at the request of the Commission, in particular:

(a) monitor and ensure the correct application of this Regulation in the cases provided for in Articles 64 and 65 without prejudice to the tasks of national supervisory authorities;
(b) advise the Commission on any issue related to the protection of personal data in the Union, including on any proposed amendment of this Regulation;
[…]
(e) examine, on its own initiative, on request of one of its members or on request of the Commission, any question covering the application of this Regulation and issue guidelines, recommendations and best practices in order to encourage consistent application of this Regulation;
[…]

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423