Operationalizing AI: Managing the input of sensitive data and validating the output from AI

Eugenio Fontenla Suárez

What personal data is considered sensitive?

The following personal data is considered ‘sensitive’ and is subject to specific processing conditions:

- personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs;
- trade-union membership;
- genetic data, biometric data processed solely to identify a human being;
- health-related data;
- data concerning a person’s sex life or sexual orientation.

GDPR

Articles:

- Art. 4 (13,14 and 15)

  (13) Genetic data

  ‘genetic data’ means 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;

  (14) Biometric data

  ‘biometric data’ means 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;

  (15) Health data

‘data concerning health’ means personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status;

- Art. 9

  Processing of special categories of personal data

  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.**

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

- Consent (1.a.)

  the data subject has given <u>explicit consent</u> to the processing of those personal data for one or more specified purposes, except where Union or Member State law provide that the prohibition referred to in paragraph 1 may not be lifted by the data subject;

- Public interest (1.g.)

  processing is necessary for reasons of substantial <u>public interest</u>, on the basis of Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject;

Automated decision-making and sensitive data

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:

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

3. In the cases referred to in points (a) and (c) of paragraph 2, the **data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests**, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.

4. Decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1), unless point (a) or (g) of Article 9(2) applies and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place.

EU AI Act

Article 10
Data and Data Governance

1. High-risk AI systems which make use of techniques involving the training of models with data shall be developed on the basis of training, validation and testing data sets that meet the quality criteria referred to in paragraphs 2 to 5 whenever such datasets are used.

2. Training, validation and testing data sets shall be subject to appropriate data governance and management practices appropriate for the intended purpose of the AI system. Those practices shall concern in particular:

(a) the relevant design choices;

(b) data collection processes and origin of data, and in the case of personal data, the original purpose of data collection;

(c) relevant data preparation processing operations, such as annotation, labelling, cleaning, updating, enrichment and aggregation;

(d) the formulation of assumptions, notably with respect to the information that the data are supposed to measure and represent;

(e) an assessment of the availability, quantity and suitability of the data sets that are needed;

(f) examination in view of possible biases that are likely to affect the health and safety of persons, negatively impact fundamental rights or lead to discrimination prohibited under Union law, especially where data outputs influence inputs for future operations;

(fa) appropriate measures to detect, prevent and mitigate possible biases identified according to point (f);

(g) the identification of relevant data gaps or shortcomings that prevent compliance with this Regulation, and how those gaps and shortcomings can be addressed.

3. Training, validation and testing datasets shall be relevant, sufficiently representative, and to the best extent possible, free of errors and complete in view of the intended purpose. They shall have the appropriate statistical properties, including, where applicable, as regards the persons or groups of persons in relation to whom the high-risk AI system is intended to be used. These characteristics of the data sets may be met at the level of individual data sets or a combination thereof.

4. Datasets shall take into account, to the extent required by the intended purpose, the characteristics or elements that are particular to the specific geographical, contextual, behavioural or functional setting within which the high-risk AI system is intended to be used.

5. To the extent that it is strictly necessary for the purposes of ensuring bias detection and correction in relation to the high-risk AI systems in accordance with the second paragraph, point f and fa, the providers of such systems may exceptionally process special categories of personal data referred to in Article 9(1) of Regulation (EU) 2016/679, Article 10 of Directive (EU) 2016/680 and Article 10(1) of Regulation (EU) 2018/1725, subject to appropriate safeguards for the fundamental rights and freedoms of natural persons. In addition to provisions set out in the Regulation (EU) 2016/679, Directive (EU) 2016/680 and Regulation (EU) 2018/1725, all the following conditions shall apply in order for such processing to occur:

(a) the bias detection and correction cannot be effectively fulfilled by processing other data, including synthetic or anonymised data;

(b) the special categories of personal data processed for the purpose of this paragraph are subject to technical limitations on the re-use of the personal data and state of the art security and privacy-preserving measures, including pseudonymisation;

(c) the special categories of personal data processed for the purpose of this paragraph are subject to measures to ensure that the personal data processed are subject to suitable safeguards, including strict controls and documentation of the access, to avoid misuse and ensure only authorised persons have access to those personal data with appropriate confidentiality obligations;

(d) the special categories of personal data processed for the purpose of this paragraph are not to be transmitted, transferred or otherwise accessed by other parties;

(e) the special categories of personal data processed for the purpose of this paragraph are deleted once the bias has been corrected or the personal data has reached the end of its retention period, whatever comes first;

(f) the records of processing activities pursuant to Regulation (EU) 2016/679, Directive (EU) 2016/680 and Regulation (EU) 2018/1725 includes justification why

the processing of special categories of personal data was strictly necessary to detect and correct biases and this objective could not be achieved by processing other data.

6. For the development of high-risk AI systems not using techniques involving the training of models, paragraphs 2 to 5 shall apply only to the testing data sets.

Insights

Article 22(4) introduces a prohibition, limited by an exception, to ground automated decisions on sensitive data, i.e., the special categories set out in Article 9(1):

Decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1), unless point (a) or (g) of Article 9(2) applies and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place.

The exception concerns the cases in which the data subject has given explicit consent (Article 9(2)(a)) or processing is necessary for reason of public interest (Article 9(2)(g)). The role of the data subject's consent needs to be clarified since consent does not exclude that the method used for the decision is inacceptable (as when it is discriminatory).

AI challenges the prohibition of processing sensitive data. First of all, sensitive data can be (probabilistically) inferred from non-sensitive data.

For instance, sex orientation can be inferred from a data subject's Internet activity, likes or even facial features. In this case. the inference of sensitive data should count as a processing of sensitive data, and therefore would have to be considered unlawful unless the conditions under Article 9 are met.

Secondly, non-sensitive data can work as proxies for sensitive data correlated to them, even though the latter are not inferred by the system. For instance, the place of residence can act as a proxy for ethnicity. In this case, unlawful discrimination may take place.

Interesting read

https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf