



EU AI Act 2024 1689

Artikel 10

III - 2

Article 10

Data and data governance

1. High-risk AI systems which make use of techniques involving the training of AI 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 data sets are used.
2. Training, validation and testing data sets shall be subject to data governance and management practices appropriate for the intended purpose of the high-risk AI system. Those practices shall concern in particular:
 - (a)
the relevant design choices;
 - (b)
data collection processes and the origin of data, and in the case of personal data, the original purpose of the data collection;
 - (c)
relevant data-preparation processing operations, such as annotation, labelling, cleaning, updating, enrichment and aggregation;
 - (d)
the formulation of assumptions, in particular 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, have a negative impact on fundamental rights or lead to discrimination prohibited under Union law, especially where data outputs influence inputs for future operations;

(g)

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

(h)

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 data sets 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. Those characteristics of the data sets may be met at the level of individual data sets or at the level of a combination thereof.

4. Data sets 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 purpose of ensuring bias detection and correction in relation to the high-risk AI systems in accordance with paragraph (2), points (f) and (g) of this Article, the providers of such systems may exceptionally process special categories of personal data, subject to appropriate safeguards for the fundamental rights and freedoms of natural persons. In addition to the provisions set out in Regulations (EU) 2016/679 and (EU) 2018/1725 and Directive (EU) 2016/680, all the following conditions must be met 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 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 are subject to measures to ensure that the personal data processed are secured, protected, subject to suitable safeguards, including strict controls and documentation of the access, to avoid misuse and ensure that only authorised persons have access to those personal data with appropriate confidentiality obligations;

(d)

the special categories of personal data are not to be transmitted, transferred or otherwise accessed by other parties;

(e)

the special categories of personal data are deleted once the bias has been corrected or the personal data has reached the end of its retention period, whichever comes first;

(f)

the records of processing activities pursuant to Regulations (EU) 2016/679 and (EU) 2018/1725 and Directive (EU) 2016/680 include the reasons why the processing of special categories of

personal data was strictly necessary to detect and correct biases, and why that 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 AI models, paragraphs 2 to 5 apply only to the testing data sets.



[Verify AI Provenance](#)