Reliable Context

RELIABLE CONTEXT / RESEARCH & PRACTICAL GUIDANCE

EU AI Act Article 10 and Annex IV: Data Governance

> Reliable Context is a kernel of truth: a system of record for organizations that fuse private and public information and must defend every value they act on. One canonical record per real thing, a cited source and an observation date behind every value, unknowns declared rather than left blank, and any past state of the record reproducible.

Published by Reliable Context · 2026-10-29 · eu ai act article 10 · ai act data governance · ai act training data requirements · ai act annex iv technical documentation · ai act data quality

Original article: https://reliablecontext.com/articles/eu-ai-act-article-10-data-governance-and-technical-documentation



In this article

1. Why Article 10 is a data article
2. Who it applies to, and from when
3. Defined terms
4. Article 10(1) and 10(6): scope inside the system
5. Article 10(2): eight data governance practices
6. Article 10(3) and 10(4): quality criteria
7. Special-category data: from Article 10(5) to Article 4a
8. Article 11 and Annex IV: what has to be written down
9. Article 17: data management inside the quality management system
10. Worked example: an evidence pack for one data source
11. What cannot be reconstructed without provenance
12. Sources

Why Article 10 is a data article

Most of the EU Artificial Intelligence Act (Regulation (EU) 2024/1689, the "AI Act") is about systems: what they may do, who supervises them, how they are marked and registered. Article 10 is different. It regulates the data a high-risk AI system is built and tested on, and it does so in terms that a data governance team will recognise: origin, collection purpose, preparation steps, assumptions, gaps, errors, completeness and representativeness (AI Act, Article 10).

Article 11 requires technical documentation, Annex IV lists its contents, and Article 17 requires a quality management system with data management procedures. Together they require a provider to govern its data and to show, years later, how it did so.

This article breaks Article 10 into practices that can be tested, maps each to the Annex IV documentation that would evidence it, and states what changed with the Digital Omnibus on AI. It describes the text of an EU regulation. It is educational material, not legal advice.

Who it applies to, and from when

Article 10 sits in Chapter III, Section 2 of the AI Act, among the requirements for **high-risk AI systems**. A system is high-risk either because it is a safety component of, or is itself, a product covered by the EU harmonisation legislation listed in Annex I (Article 6(1)), or because it falls in one of the use areas listed in Annex III, such as employment, access to essential services or law enforcement (Article 6(2)) (AI Act). The obligations in Article 10 fall on the **provider**, the party that develops the system or has it developed and places it on the market under its own name.

The dates moved in 2026. Regulation (EU) 2026/1744 of 8 July 2026, the "Digital Omnibus on AI", was published in the Official Journal on 24 July 2026 and entered into force on the third day after publication, 27 July 2026 (Regulation (EU) 2026/1744, Article 4). It replaced Article 113, third paragraph, point (c) of the AI Act so that Chapter III, Sections 1, 2 and 3 apply from:

High-risk route	Classification basis	Application date of Chapter III, Sections 1-3
Stand-alone use cases	Article 6(2) and Annex III	2 December 2027
Products under EU harmonisation legislation	Article 6(1) and Annex I	2 August 2028

Source: Regulation (EU) 2026/1744, Article 1, point (40), amending Article 113. The Commission's AI Act Service Desk shows the consolidated wording. Before the amendment, the general application date for most high-risk obligations was 2 August 2026. Because Article 10 is in Section 2, the data governance requirements follow these later dates.

Under amended Article 111(2), high-risk systems already on the market before that date are caught only if their design changes significantly afterwards; systems intended for use by public authorities have a separate 2 August 2030 deadline (Regulation (EU) 2026/1744, Article 1, point (39)).

Defined terms

Article 3 of the AI Act defines the three data sets Article 10 regulates (AI Act, Article 3(29)-(32)):

- **Training data**: data used for training an AI system through fitting its learnable parameters.
- **Validation data**: data used for providing an evaluation of the trained AI system and for tuning its non-learnable parameters and its learning process, among other things to prevent underfitting or overfitting. A **validation data set** may be a separate data set or part of the training data set, either as a fixed or variable split.
- **Testing data**: data used for providing an independent evaluation of the AI system in order to confirm its expected performance before it is placed on the market or put into service.

Provenance, not an AI Act term, is used here for the documented origin and history of a data item (glossary: provenance). An **evidence pack** is the set of records a provider would show an authority to demonstrate a requirement was met.

Article 10(1) and 10(6): scope inside the system

Article 10(1) applies the quality criteria to high-risk systems "which make use of techniques involving the training of AI models with data". As amended by the Omnibus, it refers to the quality criteria in paragraphs 2, 3 and 4 of Article 10 and in the new Article 4a(1) (Regulation (EU) 2026/1744, Article 1, point (9); Service Desk, Article 10).

Article 10(6) covers systems that do not train models, for example a rules engine or a system that only uses a fixed third-party model. For those, the same criteria apply **only to the testing data sets**. A provider that buys in a model and does not train it is therefore not relieved of Article 10: its test data must still meet the governance and quality criteria.

Article 10(2): eight data governance practices

Article 10(2) requires training, validation and testing data sets to be "subject to data governance and management practices appropriate for the intended purpose of the high-risk AI system", and lists what those practices concern "in particular" (AI Act, Article 10(2)). The list is not exhaustive, but each item is concrete enough to test.

Point	What the practice concerns (text summarised)	A testable question
(a)	The relevant design choices	Is there a dated record of why these data, this split and these features were chosen?
(b)	Data collection processes and the origin of data; for personal data, the original purpose of collection	For each data source, can you name the supplier, the collection process, the date acquired and, for personal data, the purpose for which it was first collected?
(c)	Data-preparation operations such as annotation, labelling, cleaning, updating, enrichment and aggregation	For a given record in the training set, can you list every operation applied to it and by whom or what?
(d)	Assumptions, in particular about what the data are supposed to measure and represent	Is the construct each feature stands for written down, with its known limits?
(e)	Assessment of the availability, quantity and suitability of the data sets needed	Is there a pre-training assessment of whether the data are enough and fit for the purpose?
(f)	Examination for possible biases likely to affect health and safety, fundamental rights, or lead to prohibited discrimination, especially where outputs feed future inputs	Is there a documented bias examination, with method, scope and results?
(g)	Measures to detect, prevent and mitigate biases identified under (f)	For each bias found, is there a recorded measure and a re-test?
(h)	Identification of data gaps or shortcomings that prevent compliance, and how they can be addressed	Is there a gap register with owners and status?

Point (b) is the only place in the Act's high-risk requirements where the **origin of data** and the **original purpose of collection** are named as subjects of governance. It is a provenance requirement in all but name. It cannot be satisfied after the fact by inspecting the data set: a CSV file does not carry the purpose for which its rows were first collected. The information has to be captured when the data are acquired, or reconstructed from contracts and acquisition records, which is slower and less reliable.

Point (c) is the lineage requirement: an ordered record of operations, with code and reference-data versions, is what lets a reviewer reproduce a training set. The live article on minimum data provenance requirements sets out a field-level profile for this kind of record.

Article 10(3) and 10(4): quality criteria

Article 10(3) states that 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", and shall have appropriate statistical properties, including, where applicable, as regards the persons or groups on whom the

system is intended to be used. The characteristics may be met at the level of individual data sets or a combination of them (AI Act, Article 10(3)).

Three qualifiers matter. The criteria are judged **in view of the intended purpose**, so the purpose must be written down before the data can be assessed. Freedom from errors and completeness are required only **to the best extent possible**, which in practice means a documented error-detection process and measured error rates, not error-free data. And the criteria may be met by a **combination** of data sets, in which case the combination must be documented.

Article 10(4) adds that data sets shall take into account, to the extent required by the intended purpose, the characteristics or elements particular to the specific geographical, contextual, behavioural or functional setting in which the system is intended to be used. A model for triaging insurance claims trained on one Member State's claims data, then deployed in another, raises exactly this question.

Special-category data: from Article 10(5) to Article 4a

The original Article 10(5) allowed providers, "to the extent that it is strictly necessary" for bias detection and correction under points (f) and (g), to process special categories of personal data (the categories in Article 9 of the GDPR, such as racial or ethnic origin or health data), subject to six cumulative conditions (AI Act, Article 10(5)).

The Digital Omnibus **deleted Article 10(5)** and inserted a new **Article 4a**, "Processing of special categories of personal data for bias detection and correction" (Regulation (EU) 2026/1744, Article 1, points (6) and (9); Service Desk, Article 4a). The change has three parts:

- **Article 4a(1)** carries over the provider permission and the six cumulative conditions for high-risk systems, including that other data (synthetic or anonymised) would not be effective, pseudonymisation and strict, documented access controls, no access by other parties, deletion once the bias is corrected or the retention period ends, and reasons recorded in the GDPR record of processing activities.
- **Article 4a(2)** extends a permission to providers and deployers of other AI systems and models and to deployers of high-risk systems, where strictly necessary for bias detection and correction and with the same safeguards. It states that it does not create any obligation to conduct such bias detection.
- **Article 4a sits in Chapter I** of the AI Act. Recital (9) of the Omnibus explains that the legal basis should apply from the date of entry into application of the AI Act, so that providers can lawfully do bias work in preparation for the high-risk requirements (Regulation (EU) 2026/1744, recital (9)).

Amended Article 10(1) now refers to "Article 4a(1)" alongside paragraphs 2 to 4, so the conditions remain part of the data quality criteria for high-risk systems. For an evidence pack, condition (c) ("documentation of the access") and condition (f) (reasons recorded in the GDPR record of processing) are records that must exist; condition (e) means deletion events must be recorded too.

Article 11 and Annex IV: what has to be written down

Article 11(1) requires the technical documentation of a high-risk system to be drawn up **before** the system is placed on the market or put into service and to be kept up to date. It must be drawn up so as to demonstrate compliance with the Section 2 requirements, which include Article 10, and must contain at least the elements in Annex IV (AI Act, Article 11). The Omnibus replaced the second subparagraph to extend the simplified documentation form, previously offered to SMEs and start-ups, to small mid-cap enterprises (SMCs) (Regulation (EU) 2026/1744, Article 1, point (10)). The Omnibus does not amend Annex IV itself.

Annex IV has nine points. The ones that bear on data are:

- **Point 2(d)**: "where relevant, the data requirements in terms of datasheets describing the training methodologies and techniques and the training data sets used, including a general description of these data sets, information about their provenance, scope and main characteristics; how the data was obtained and selected; labelling procedures (e.g. for supervised learning), data cleaning methodologies (e.g. outliers detection)" (AI Act, Annex IV; Service Desk, Annex IV). The word *provenance* appears in the legal text here.
- **Point 2(b)**: design specifications, including key design choices with rationale and assumptions, "including with regard to persons or groups of persons" on whom the system is intended to be used.
- **Point 2(g)**: validation and testing procedures, "including information about the validation and testing data used and their main characteristics", the metrics used, and "test logs and all test reports dated and signed by the responsible persons".
- **Point 3**: capabilities and limitations in performance, including degrees of accuracy for specific persons or groups, and "specifications on input data, as appropriate".
- **Point 6**: a description of relevant changes made by the provider through the system's lifecycle.
- **Point 7**: "a list of the harmonised standards applied in full or in part" with references published in the Official Journal or, where none were applied, "a detailed description of the solutions adopted" to meet the Section 2 requirements (AI Act Service Desk, Annex IV). As of 28 October 2026 no harmonised standard covering Article 10 had been referenced; the first harmonised standard referenced in the Official Journal under the AI Act is EN 18286:2026, *Artificial intelligence - Quality management system for EU AI Act regulatory purposes*, published through Commission Implementing Decision (EU) 2026/2274 and announced on 19 October 2026; it gives a presumption of conformity with the Article 17 requirements it covers, and the Commission describes it as designed to be layered on top of standards such as ISO/IEC 42001 rather than replace them (European Commission), so for data governance point 7 currently means describing the solutions adopted.

The term "datasheets" in point 2(d) echoes a documentation practice proposed in the machine learning literature: Gebru and colleagues argued that every data set should be accompanied by a datasheet documenting "its motivation, composition, collection process, recommended uses, and so on" (Gebru et al., Datasheets for Datasets). The Act does not mandate that paper's template; it is one well-known structure for the content point 2(d) describes.

Article 18 then requires the provider to keep the technical documentation and the quality management system documentation at the disposal of national competent authorities for **10 years** after the system is placed on the market or put into service (AI Act, Article 18; Service Desk, Article 18). Data governance records written in 2027 may therefore have to be legible in 2037.

Article 17: data management inside the quality management system

Article 17(1) requires providers of high-risk systems to put in place a quality management system "documented in a systematic and orderly manner in the form of written policies, procedures and instructions", including at least thirteen aspects (AI Act, Article 17). Two concern data directly:

- **Point (f)**: "systems and procedures for data management, including data acquisition, data collection, data analysis, data labelling, data storage, data filtration, data mining, data aggregation, data retention and any other operation regarding the data that is performed before and for the purpose of the placing on the market or the putting into service".
- **Point (k)**: "systems and procedures for record-keeping of all relevant documentation and information".

Article 10 states what good data governance is; Article 17(1)(f) requires a documented procedure that produces it every time. The Omnibus amended Article 17(2) so that proportionality to the size of the provider refers expressly to SMEs, start-ups and SMCs, while keeping the requirement to "respect the degree of rigour and the level of protection required" (Regulation (EU) 2026/1744, Article 1, point (11)). Article 17(4), unchanged, deems financial institutions subject to EU financial services governance rules to meet most of the quality management system obligation through those rules, except points (g), (h) and (i) (Service Desk, Article 17).

Worked example: an evidence pack for one data source

Consider a fictional provider, Halden Therapeutics AG, building a system that ranks job applicants for laboratory roles. Recruitment and selection of natural persons is an Annex III use area, so this would be a high-risk system under Article 6(2), subject to Article 10 from 2 December 2027. (Halden Therapeutics AG and every detail below are fictional.)

One training source is a table of 18,400 historical applications from 2019 to 2025, exported from the provider's applicant tracking system, labelled with the hiring outcome. The evidence pack for that one source might contain:

Item	Content for this source	Article 10 point	Annex IV point
Source record	System of origin, export date, export query, row count 18,400, file hash	10(2)(b)	2(d) provenance
Collection purpose	Applications were collected to run recruitment; privacy notice version in force for each year	10(2)(b)	2(d)

Item	Content for this source	Article 10 point	Annex IV point
Preparation log	Ordered list: de-duplication (312 rows removed, rule stated), free-text CV fields dropped, outcome label mapped from five statuses to two	10(2)(c)	2(d) labelling, cleaning
Assumptions	"Hired" is used as a proxy for "suitable", with the known limitation that it reflects past hiring decisions	10(2)(d)	2(b)
Sufficiency assessment	Volume per role family; two role families under 200 examples flagged	10(2)(e)	2(d)
Bias examination	Method, groups examined, results; feedback-loop note because past decisions are the label	10(2)(f)	2(g), 3
Mitigation	Reweighting applied; re-test results dated and signed	10(2)(g)	2(g)
Gap register	Missing applications from one site in 2020 (system migration); status open	10(2)(h)	6
Representativeness note	Distribution of applicants by country of residence against intended deployment sites	10(3), 10(4)	3
Special-category handling	Whether any was processed under Article 4a(1); if so, necessity reasons, access log, deletion record	4a(1)	2(d), 5

The label "hired" was produced by Halden's own past decisions, which is the feedback loop Article 10(2)(f) singles out ("especially where data outputs influence inputs for future operations"). If the new system influences hiring, its outputs become future labels; the bias examination should say how later training sets will tell those records apart.

What cannot be reconstructed without provenance

Several Article 10 and Annex IV items are about the history of data rather than its current state. They are where documentation programmes most often fail, because the information was never captured.

Requirement	Why the data set alone cannot answer it
Origin of data (10(2)(b))	Files do not carry their supplier, licence or acquisition date
Original purpose of personal data collection (10(2)(b))	Purpose is a fact about the collection event, recorded elsewhere
Preparation operations (10(2)(c))	Cleaning and de-duplication remove the evidence of what was removed
Changes through the lifecycle (Annex IV point 6)	Requires dated versions of data sets and the reason for each change
Test reports "dated and signed by the responsible persons" (Annex IV 2(g))	Requires responsibility and time recorded with the result
Deletion of special-category data (4a(1)(e))	Requires a record that the deletion happened, and when

Each line asks what the item is, where it came from, and who did what to it when: the elements of an audit trail and a provenance record. Kept as a by-product of the pipelines, they make Annex IV a compilation exercise; missing, they make it archaeology.

A system of record for data facts should let a reviewer pick any training row and trace it to a dated, cited source; that is the property Reliable Context is built to make checkable.

Sources

1. Regulation (EU) 2024/1689 (Artificial Intelligence Act), EUR-Lex ELI: Articles 3(29)-(32), 6, 10, 11, 17, 18 and Annex IV (text read via the Publications Office CELEX 32024R1689).
2. Regulation (EU) 2026/1744 (Digital Omnibus on AI), EUR-Lex ELI: Article 4a insertion, deletion of Article 10(5), amendments to Articles 10(1), 10(6), 11(1), 17(2), 111 and 113; entry into force (Article 4); recital (9) (text read via the Publications Office, CELEX 32026R1744).
3. AI Act Service Desk, Article 10: consolidated text showing paragraph 5 deleted and Article 4a(1) references.
4. AI Act Service Desk, Article 4a: consolidated text of the new provision.
5. AI Act Service Desk, Annex IV: technical documentation contents.
6. AI Act Service Desk, Article 17: quality management system, including paragraph 4.
7. AI Act Service Desk, Article 18: 10-year documentation keeping.
8. AI Act Service Desk, Article 113: consolidated application dates.
9. Gebru et al., Datasheets for Datasets (arXiv:1803.09010; published in CACM, December 2021): datasheets as a dataset documentation practice.
10. European Commission, First harmonised standard supporting the AI Act published in the Official Journal, 19 October 2026: EN 18286:2026 (Article 17); no standard yet for Article 10.

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. https://eur-lex.europa.eu/eli/reg/2024/1689/oj
2. https://eur-lex.europa.eu/eli/reg/2026/1744/oj
3. https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-113
4. https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-10
5. https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-4a
6. https://ai-act-service-desk.ec.europa.eu/en/ai-act/annex-4
7. https://digital-strategy.ec.europa.eu/en/library/first-harmonised-standard-supporting-ai-act-published-official-journal
8. https://arxiv.org/abs/1803.09010
9. https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-18
10. https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-17
11. https://publications.europa.eu/resource/celex/32026R1744

THE PUBLISHER

About Reliable Context

Reliable Context is a system of record for organizations that fuse private and public information and must defend every value they act on.

The guarantees

One canonical record per real thing: duplicates resolve by rule, ambiguity is shown rather than hidden, and related organizations relate without ever being merged. A verbatim, cited source and an observation date sit behind every value. Unknowns are declared as not found, blocked or not applicable with a reason, never left as an empty cell. Any past state of the record is reproducible, together with what changed since.

Where it runs

It runs inside the customer's perimeter, one database per organization. Permitted sources, model routing, reader clearance and retention are set as the customer's own policies, and every value is exportable with its chain of custody for an auditor or regulator.

What it is not

It is not a vector database or a RAG framework, not a data catalog, MDM or data-quality suite, not a CRM or data-enrichment vendor, and not a chatbot or AI assistant. Where a reader needs one of those, the honest answer is to say so and point them there.

Visit

See Reliable Context.

Reliable Context is in production use in two sectors; more than six thousand of its identity decisions were made by rule and none has been reversed. Articles published here explain general practice, cite their primary and academic sources, and name the jurisdiction of any regulatory rule they describe. They are not legal, regulatory or compliance advice for a specific organization, and requirements change.

Website: https://reliablecontext.com

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.