



AI GOVERNANCE · EU AI ACT · WHITE PAPER

EU AI Act Data Governance: How to Turn Article 10 Requirements Into Enterprise Controls

A seven-control model — Discover, Understand, Qualify, Protect, Enforce, Trace, Prove — for turning Article 10's data governance language into repeatable enterprise controls.

"The practical challenge is not simply understanding the regulation. It's turning those expectations into controls that operate across the AI data lifecycle."

✉ info@solix.com

🌐 <https://www.solix.com>

☎ +1 408-654-6400



Article 10 of the EU AI Act makes data governance an operational requirement for relevant high-risk AI systems. Training, validation, and testing datasets must be governed around their origin, preparation, underlying assumptions, suitability, potential bias, data gaps, representativeness, quality, and intended context of use. The practical challenge is therefore not simply understanding the regulation. It is turning those expectations into controls that operate across the AI data lifecycle.

A useful way to do that is through seven controls: Discover → Understand → Qualify → Protect → Enforce → Trace → Prove. Together, they provide a practical model for moving from regulatory language to enterprise data governance.

IMPLEMENTATION RUNWAY

As of September 2026, the AI Act is generally applicable, although the Chapter III high-risk requirements apply from December 2, 2027, for Annex III high-risk systems and August 2, 2028, for high-risk AI embedded in regulated products under Annex I. That implementation runway gives enterprises time to establish the data foundations these requirements demand.

REGULATORY SCOPE

[Article 10](#) applies to high-risk AI systems under the conditions established by the Act. An organization's obligations depend on factors including its regulatory role and the classification of the AI system. The seven-control model below is an enterprise implementation framework, not terminology prescribed by the EU AI Act.

What does Article 10 actually require enterprises to govern?

Article 10 is titled “Data and data governance.” For high-risk AI systems that train AI models with data, it establishes requirements concerning the training, validation, and testing datasets used to develop those systems. Its scope goes well beyond generic data quality.

Article 10 requires appropriate data governance and management practices covering relevant design choices; data collection processes and origin; the original purpose of collection where personal data is involved; preparation activities such as annotation, labelling, cleaning, updating, enrichment, and aggregation; assumptions about what the data measures or represents; dataset availability, quantity, and suitability; possible bias and mitigation; and material data gaps or shortcomings.

The datasets must also be sufficiently relevant and representative and, to the best extent possible, free of errors and complete for their intended purpose. They should reflect appropriate statistical properties and, where necessary, the geographical, contextual, behavioral, or functional environment in which the high-risk AI system will operate. This changes the governance question. It is no longer enough to ask:

THE OLD QUESTION

“Do we have a data governance policy?”



WHAT AI TEAMS MUST NOW ANSWER

“Can we show how the data behind this AI system was selected, understood, assessed, controlled, and managed?”

That is where Article 10 moves from regulation into operating practice.

From Article 10 requirements to seven operational data controls

The following framework translates Article 10’s regulatory concerns into an enterprise control model. Not every control represents a literal requirement named in Article 10. Some are implementation mechanisms that help organizations operationalize those requirements or support related obligations under the AI Act.

THE SEVEN-CONTROL MODEL

An enterprise implementation framework for Article 10 data governance



1 Discover: Know which data enters the AI lifecycle

Governance begins with visibility.

Organizations need to know which datasets are being used for training, validation, and testing; where those datasets originate; whether multiple sources have been combined; and which systems, repositories, or third parties supplied them.

This connects directly to Article 10's requirements around data collection processes, data origin, availability, and quantity. The enterprise control is therefore more than a data inventory. The organization should be able to connect an AI use case to the datasets that support it and establish a reliable record of those sources. Without discovery, every downstream control starts with an incomplete picture.

2 Understand: Establish meaning, context, provenance, and ownership

Knowing that a dataset exists is not the same as understanding it.

Article 10 specifically requires attention to assumptions about what the data is supposed to measure and represent, as well as its origin and preparation. That makes semantic context important.

Organizations should establish what important fields, entities, relationships, metrics, and classifications mean; where definitions originated; which versions are authoritative; and who owns or stewards those definitions. Ownership and stewardship are enterprise governance practices rather than explicit Article 10 terminology, but they create accountability for maintaining that context.

This distinction matters because an organization cannot credibly determine whether a dataset is suitable for an AI use case if it cannot explain what that data actually represents.

Solix [Enterprise Data Governance](#) (EDG), for example, combines a Business Glossary, Data Catalog, Model Registry, and Semantic Store with ownership, versioning, approval workflows, and lineage. [Data Sense](#) extends this semantic layer through [Application Knowledge Graphs](#) that encode enterprise tables, relationships, terms, and governed business meaning.

3 Qualify: Determine whether the data is fit for the intended AI purpose

This is perhaps the most direct Article 10 control.

The regulation requires assessment of dataset availability, quantity, and suitability; examination for relevant bias; identification of gaps or shortcomings; and selection of datasets that are sufficiently relevant and representative, and, to the best extent possible, complete and free of errors. The characteristics of the intended environment also matter.

“High-quality data,” therefore, cannot be treated as a universal label. A dataset can be accurate yet unsuitable for a specific population. It may be complete for one business process but inadequate for another geography. It may also contain historical patterns that create bias when applied to a new decision context.

Qualification should therefore evaluate the dataset against the intended AI purpose, not only against generic dataquality thresholds. This is where profiling, quality rules, representativeness assessment, bias examination, exception handling, and formal suitability reviews become valuable controls.

[Solix AI Warehouse](#) provides automated profiling, quality rules and alerts, preparation workflows, and lineage tracking as part of the governed path from raw enterprise data to AI-ready data products.

4 Protect: Apply safeguards where AI data creates privacy or sensitivity risk

AI data governance also has to account for what information a dataset contains and what restrictions apply to it. This requires careful regulatory distinction. Article 10 does not impose a generic requirement to mask every AI dataset.

Following the 2026 Digital Omnibus amendments, Article 4a separately addresses exceptional processing of special categories of personal data for bias detection and correction. It requires strict necessity and safeguards, including access controls, documentation, privacy-preserving measures, and deletion conditions.

For enterprise governance teams, the operational principle is broader: sensitive data should be identified and handled according to the applicable privacy, security, access, and regulatory obligations. Controls may therefore include sensitive-data classification,

least-privilege access, masking or pseudonymization where appropriate, and restrictions on reuse.

Solix [Sensitive Data Discovery](#) identifies sensitive information across structured and unstructured environments, while Solix Data Masking supports techniques including masking, redaction, and format-preserving protection.

5 Enforce: Turn governance policies into operating controls

Policies establish what should happen. Controls determine what actually happens.

Article 10 requires organizations to properly manage and govern the data used by AI systems. It does not require them to use any specific automated governance tool. However, in large enterprises, relying only on policies and manual checks can be difficult. AI systems and data pipelines often access and process data automatically, so governance controls also need to work at that scale. In practice, this usually means combining automated controls with human oversight.

The implementation question becomes: “When data is accessed or used, do the applicable policies follow it?”

Solix EDG is designed around this distinction. It supports federated policy enforcement across data environments, permission-aware IAM, classification-driven controls, and auditable governance actions. This leads to a useful principle for AI governance:

Policy is not the control. The control is what actually happens to the data.

6 Trace: Preserve the data’s journey

Article 10 requires organizations to address data origin and relevant data-preparation operations, including annotation, labeling, cleaning, updating, enrichment, and aggregation. In practice, data lineage is one of the mechanisms enterprises can use to reconstruct that journey. For an AI dataset, traceability should help answer questions such as:

- Where did this data originate?
- What transformations occurred?
- Which version reached the AI workflow?
- What definitions or classifications applied?
- Where did the resulting dataset go next?

Lineage itself should not be misrepresented as the entirety of Article 10 compliance. Broader AI Act requirements also address technical documentation and record-keeping separately under Articles 11 and 12. But without traceability, explaining how a dataset moved from source to AI system becomes substantially harder.

7 Prove: Maintain evidence that governance actually happened

Trace and Prove are related, but they are not the same.

TRACE ASKS

What happened to the data?

PROVE ASKS

Can we demonstrate that it happened under appropriate governance?

Evidence may include dataset approvals, quality assessments, bias reviews, exceptions, access records, definition versions, transformation histories, policy decisions, and audit records.

Article 11 separately requires technical documentation for high-risk AI systems to demonstrate compliance with the requirements in Chapter III, Section 2. That makes it important not to attribute every documentation obligation specifically to Article 10. From an operational perspective, however, the principle is clear: controls that leave no durable evidence are difficult to defend later.

Solix EDG's Define → Bind → Enforce → Prove model supports this transition from governance definition to auditable execution. In the Prove stage, lineage and audit records can show the definition version, sources, policies, and approvals associated with the use of governed data.

Article 10 Enterprise Control Framework at a Glance

 CONTROL	 QUESTION ANSWERED	 ARTICLE 10 / REGULATORY CONCERN	 ENTERPRISE RESPONSE	 EXAMPLE EVIDENCE
1 Discover	What data do we have?	Collection, origin, availability, quantity	Inventory AI datasets and sources	Dataset and source inventory
2 Understand	What does it mean, and who governs it?	Origin, purpose, preparation, assumptions	Establish semantics, provenance, stewardship, and approved definitions	Glossary, provenance, ownership, and approvals
3 Qualify	Is it fit for this AI purpose?	Suitability, bias, gaps, relevance, representativeness, errors, completeness	Assess quality and fitness for intended use	Quality, bias, and suitability results
4 Protect	What restrictions apply?	Sensitive-data safeguards and related privacy obligations	Classify and protect restricted data	Classification and access records
5 Enforce	Are governance rules operating?	Appropriate governance and management practices	Apply policies during data preparation and use	Enforcement records
6 Trace	What happened to the data?	Origin and preparation	Maintain lineage and version history	Transformation and lineage records
7 Prove	Can we demonstrate governance?	Accountability plus related documentation requirements	Retain approvals and audit evidence	Audit and review records

Key Takeaway

The important point is that these controls are connected. You cannot **Qualify** data confidently if you do not **Understand** it. You cannot consistently **Protect** it if you have not **Discovered** and classified it. And you cannot **Prove** governance if you cannot **Trace** what happened.

Why data governance documentation alone is not enough for enterprise AI

Traditional governance programs often begin with policies, catalogs, ownership assignments, and documented standards. Those remain important. But AI changes the operating environment.

Documentation	Establishes intent and accountability
Operational controls	Apply those decisions
Traceability	Reconstructs the data journey
Evidence	Demonstrates what happened

Data moves through pipelines, models, applications, copilots, and automated workflows at a scale where retrospective review alone cannot reliably prevent governance drift. The stronger model is therefore not documentation versus enforcement. Enterprises need both.

Solix EDG was designed around this move from describing the data estate to enforcing governance across it, with federated controls that can extend across Solix and non-Solix environments. That is increasingly the difference between having an AI governance policy and having an AI governance system.

How Solix helps operationalize AI data governance controls

Solix Enterprise Data Governance

Provides the primary governance layer. Its Unified Asset Catalog connects business definitions, ownership, certified assets, policies, and lineage, while federated governance applies controls such as access, masking, retention, and other policies across distributed environments. The EDG model moves governance through **Define** → **Bind** → **Enforce** → **Prove** rather than stopping at cataloging.

Solix Data Sense

Adds semantic understanding and classification. Application Knowledge Graphs capture business objects and relationships in complex enterprise systems, while Intelligent Classification works against an organization's own taxonomy. Its workflow also incorporates human review before enriched definitions and intelligence are published.

Solix AI Warehouse

Provides a governed foundation for AI application development, with capabilities spanning data ingestion, preparation, profiling, quality rules, lineage, and governed AI-ready data products.

Sensitive Data Discovery & Data Masking

Complement these capabilities when enterprises need to identify and protect sensitive information in data used for analytics, development, or AI/ML workloads.

Where should enterprises start?

Start with one AI use case and its data, not an enterprise-wide compliance transformation. Determine whether the system could fall within the high-risk regime and establish the organization's regulatory role. Then identify the training, validation, and testing data associated with that use case and work through the seven controls:



The gaps will become visible quickly!

- The data is cataloged, but lacks semantic ownership.
- Quality is measured, but representativeness is not.
- Lineage exists, but approvals are not retained.
- Policies are documented, without an enforcement mechanism.

That gap analysis provides a much more useful starting point than simply asking whether an organization "has AI governance."

Conclusion: Move Article 10 from policy to operating practice

Article 10 makes the quality and governance of AI data part of the high-risk AI compliance architecture. The enterprise challenge is translating those expectations into repeatable controls.

Organizations that can discover, understand, qualify, protect, enforce, trace, and prove how AI data is governed will be better positioned not only for Article 10 readiness but for trustworthy enterprise AI more broadly.

Ready to operationalize AI data governance?

Request a demo of Solix Enterprise Data Governance to see how active policy enforcement, semantic lineage, permission-aware access, and auditable controls can support governed AI initiatives.

[Request a Demo »](#)

This article is intended for informational purposes and does not constitute legal advice.

FAQs

| Does Article 10 apply to every enterprise AI system?

No. Article 10 is part of the requirements for high-risk AI systems. Applicability depends on the AI system's classification and the organization's role under the Act. Most minimal- or no-risk AI systems are not subject to the high-risk regime.

| Who has the main Article 10 obligation: providers or deployers?

Article 10 primarily establishes requirements for the development of high-risk AI systems, making the provider role particularly important. Deployers have separate obligations elsewhere in the Act. Organizations should establish their regulatory role before translating requirements into controls.

| What data does Article 10 cover?

For high-risk systems that train models with data, Article 10 addresses training, validation, and testing datasets. For high-risk AI systems developed without techniques involving model training, the current Article 10(6) applies the relevant data quality provisions to testing datasets.

| Does Article 10 explicitly require data lineage?

Article 10 requires governance around data origin and data-preparation operations, but does not prescribe a specific "data lineage" technology. Lineage is an enterprise implementation mechanism that can help establish provenance and reconstruct transformations.

How are Article 10 and data privacy connected?

Article 10 addresses the original purpose of collection where personal data is involved and requires a bias examination. Article 4a separately establishes conditions for exceptional processing of special categories of personal data for bias detection and correction, alongside applicable EU data-protection laws.

What evidence should enterprises retain?

Evidence should reflect the controls that actually operate: dataset sources, definitions, preparation history, suitability and quality assessments, bias reviews where applicable, approvals, policy actions, lineage, exceptions, and access/audit records. Article 11 separately addresses technical documentation intended to demonstrate high-risk-system compliance.



About Solix Technologies

Solix Technologies, Inc. (Santa Clara, CA) is a leading Data + AI company helping Fortune 2000 enterprises transform enterprise data into trusted AI-powered business outcomes. The [Solix Common Data Platform \(CDP\)](#), [Enterprise Content Services \(ECS\)](#), and [Enterprise Data Governance \(EDG\)](#) provide a trusted foundation to unify, govern, secure, and activate enterprise data for AI. [Solix Enterprise AI \(EAI\)](#), a fourth-generation AI-native data platform, combines enterprise data warehousing, semantic intelligence, natural language interaction, and autonomous AI agents. With capabilities such as [Data Sense](#) and [Data Ask](#), organizations can interact with enterprise data in plain language, build AI-powered applications, automate business processes, and deploy intelligent agents on trusted enterprise knowledge. Supporting multi-cloud, hybrid, and sovereign AI deployments, Solix enables organizations to securely manage and activate data wherever it resides. Learn more at www.solix.com.

Stay Connected



www.solix.com



info@solix.com



+1.888.467.6549