

KEEP IT SECRET, KEEP IT (IN A) SAFE, THEN CAST IT INTO THE FIRE

Engineering AI Act Data Governance:
A Practical Guide for Data Teams

Lorenzo Colombani

French-qualified lawyer (CAPA, the French professional qualification for lawyers), LL.M.

(University of Pennsylvania), Data Vault practitioner

Independent researcher · lorenzo.colombani@live.fr · ORCID 0009-0009-4207-4300

Preprint, 2 September 2026 · DOI 10.5281/zenodo.22255574

Abstract

Data Vault preserves history; data protection can require parts of that history to be erased. The EU AI Act adds bias-detection and correction duties that may depend on sensitive attributes whose processing the GDPR restricts. Reconciling these demands requires decisions about data collection, storage, transformation, release, use and deletion. This paper maps selected AI Act data obligations to those decisions, using Data Vault 2.x and a recruitment-screening example. Article 4a receives particular attention: its permission to process sensitive data carries six cumulative safeguards, including deletion at the earlier of bias correction or retention expiry. The implementation patterns connect each relevant requirement to evidence, an accountable owner and a test. They support lawful processing without supplying a missing legal basis, resolving contested interpretations or replacing the Act's wider organisational requirements.

This paper provides general legal and architectural analysis. Specific systems require advice from qualified counsel. The architectures, owner assignments and tests below are suggested engineering practices; statutory duties and conditions are identified by their provisions.

Keywords: EU AI Act; Article 4a; GDPR; special categories of personal data; Data Vault 2.0; bias testing; erasure; data governance

Introduction

One Ring to rule them all, One Ring to find them, One Ring to bring them all and in the darkness bind them.

– *Inscription on the One Ring, J. R. R. Tolkien, The Lord of the Rings*

In *The Fellowship of the Ring*, Gandalf's advice to Frodo on handling the One Ring comes in two parts: "keep it secret, keep it safe". But Elrond, recalling what Isildur refused to do when he had the chance to destroy the Ring, was more drastic: *cast it into the fire*. Data teams have heard all three.

Keep it secret. Under Article 9 of the GDPR, ethnicity, religion, health and the other special categories of personal data may not be collected or processed except by narrow exception.

Keep it (in a) safe. In data warehousing, and in the Data Vault methodology in particular, the phrase is true twice over: the data is locked away, and it does not change. Every record is kept, every version, for good; hubs, links and satellites are append-only. That is what gives the business analysts who rely on the warehouse auditability, traceability and trust in the numbers. And because Data Vault allows satellite splitting, so that personal data can sit in a PII satellite of its own, apart from purely business data, the method lends itself naturally to GDPR compliance. "Customer 33 bought item 8 at store 45 on 19 December 2022" is useful business information and belongs in an ordinary satellite. "Jean Dupont bought a crucifix at the abbey shop of Mont-Saint-Michel on 19 December" invites the inference that he is a Christian, true or not, and belongs in the sensitive one. Better safe than sorry. On sensitive data the regulation and the warehouse have always agreed.

Cast it into the fire. Then came the AI Act, and with it the fire. Before a high-risk model can be trusted, its makers must test it for bias: show, for instance, that no ethnic or religious group is over- or under-represented in the training data, or that the model draws no false conclusions about one of them, such as that Jean is a Christian. Yet such checks cannot be run without knowing each person's ethnicity or religion: the data team must collect and process exactly the data the GDPR tells it to keep secret. At face value, the Act throws the GDPR's protection into the fire.

This paper shows how Data Vault practitioners can adapt when two binding rules collide: Article 9 of the GDPR and Article 10 of the AI Act.

1. Establish the system, the actor and the purpose

One warehouse can serve several AI systems with different purposes, users and legal duties. Its controls must distinguish those uses. Start with the system or model, the organisation's role, the data involved and the decision they will support. A platform-wide label such as 'AI Act compliant' leaves these questions unanswered.

1.1 Establish which obligations apply

A duty to examine bias can create a need for group information without authorising access to it. Article 10 imposes bias-related data duties for relevant high-risk systems; Article 4a governs the exceptional use of special-category personal data for that work. Regulation (EU) 2026/1744 moved the former Article 10(5) derogation into Article 4a and extended its reach. The amendment entered into force on 27 July 2026.^{1,2,3}

Article 4a(2) also reaches deployers of high-risk systems and providers and deployers of other AI systems and models, subject to its conditions. It permits qualifying processing but creates no duty to conduct that bias testing. The processing record must therefore distinguish an applicable obligation from an optional audit: both need lawful authority, but they begin from different legal positions.^{4,5,6,7}

Enactment and application dates can differ. Article 4a is already available, subject to its conditions; the Article 10 duties follow the applicable high-risk-system timetable: 2 December 2027 for systems classified under Article 6(2) and 2 August 2028 for those under Article 6(1), subject to exclusions and transitional provisions. Article 2(2) limits the regime for Annex I Section B product systems. Provisions elsewhere in the Act have their own timetable.^{8,9,10}

Have the product or compliance lead, with legal review, record the actor, system, applicable provision and date, dataset or processing purpose, and owner of the resulting control. Leave unresolved questions open in that record. A delivery deadline cannot settle them.

Test. Trace one data flow to its system, purpose, accountable owner and applicability decision. If the only authority recorded is a wish for a fairer model, stop: that aim alone does not permit special-category processing.

The analysis below assumes GDPR-governed processing. EU institutions and law-enforcement bodies may fall under different data-protection instruments. The wider system obligations in the map also retain their own actor, scope and transition conditions. Coverage is limited to the selected data-related requirements and their organisational interfaces.

1.2 Follow the data through a recruitment model

Consider a provider preparing a recruitment-screening model and an audit of selection outcomes across groups. Before collecting or reading a protected attribute, it must resolve the system's classification, applicable dates, legal grounds and safeguards. The example assumes those questions receive a separate assessment; it assigns no legal status to a real system. The employer's responsibilities enter where it controls input data, system use or retained logs.

In Data Vault 2.x, a **hub** holds business keys, a **link** represents relationships, and a **satellite** holds time-stamped descriptive history. The **Raw Vault** preserves received data; the **Business Vault** adds derived rules and query aids; the **AI-Mart** supplies curated model data. These structures provide a worked architecture. Other warehouses and lakehouses can implement the same controls through different components.

The work proceeds from permission and collection through quality checks, release, evidence and erasure. This build order follows technical dependencies; the Act does not prescribe it. The appendix supplies legislative history and interpretive disputes where readers need the fuller legal argument.

2. Map obligations to controls and evidence

Each row connects a selected legal requirement to a possible control and the evidence needed to inspect it. Some controls govern the data directly; others supply material for documentation, oversight or monitoring elsewhere in the organisation. Apply the actor, scope and timing conditions before adopting a row. The named components are design choices, not statutory specifications.

Table 1. Selected obligations, possible controls and evidence, subject to the relevant actor, system and timing conditions.

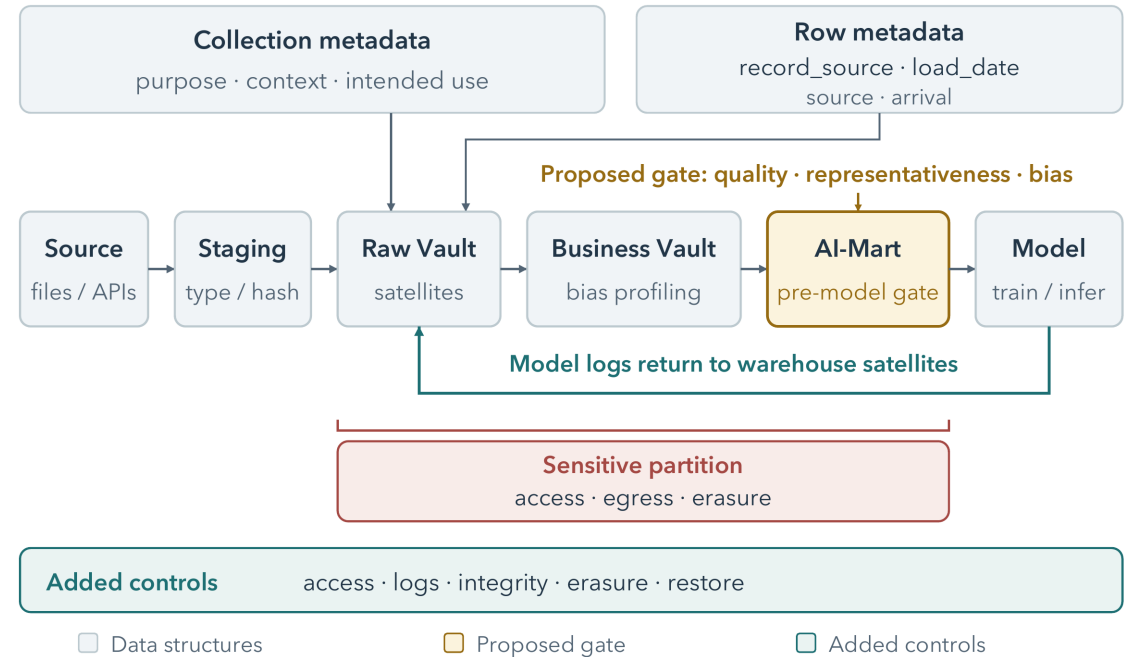
Duty or condition	Proposed design response	Evidence to produce	Guide
Lawful use and necessity Art. 4a; GDPR Arts. 6, 9	Permission boundary; assess alternatives before protected reads	Applicability, lawful-basis and necessity records	§3
Purpose limits, access and egress Art. 4a(1)(b)–(d)	Sensitive partition; controlled joins, reads and exports	Permission tests, access records, transfer boundaries	§3
Collection and origin Art. 10(2)(b)	Link row provenance to collection purpose and context	Dataset inventory and collection metadata	§4
Preparation and assumptions Art. 10(2)(a), (c), (d)	Version transformations, design choices and semantic definitions	Pipeline versions and transformation records	§4
Availability and data gaps Art. 10(2)(e), (h)	Assess coverage; track remediation and restrictions	Gap register, actions and closure evidence	§5
Bias examination, prevention and mitigation Art. 10(2)(f), (g)	Measure distributions and errors; validate corrective action	Assessment results and mitigation checks	§5
Quality and context Art. 10(3), (4)	Release checks linked to intended use; proposed AI-Mart gate	Versioned results and release decisions	§5
Deployer-controlled inputs Art. 26(4)	Validate relevant input data within the deployer's control	Input checks and recorded limitations	§5
Documentation and logging Arts. 11, 12, 19; Art. 26(6)	Link datasets, model versions, events and protected records	Traceable releases, logs and retention rules	§6
Erasure triggers and scope Art. 4a(1)(e); GDPR Art. 17	Apply Article 4a's earlier trigger; assess Article 17 separately	Trigger record, deletion results and residual-data assessment	§7
Necessity documentation Art. 4a(1)(f)	Link the processing record to the alternatives assessment	Contemporaneous reasons for protected processing	§3, §8
Wider system governance Arts. 13–15, 17, 27, 72, 73	Supply evidence to system owners and review processes	Instructions, monitoring and assessment inputs	§8

Article 10 reaches well beyond the bias measurement itself. It covers design choices, collection, preparation, assumptions, data availability and suitability, mitigation, and gaps and their treatment. For high-risk systems not developed using model-training techniques, Article 10(6) applies the specified data requirements to testing datasets. Record the dataset type so the correct requirements follow it.¹¹

The provider’s training dataset and the employer’s live applicant records sit under different responsibilities. Article 26(4) requires deployers to ensure relevant and sufficiently representative inputs insofar as they control those data. The employer therefore needs input checks within its control; the provider remains responsible for its own Article 10 duties.¹²

2.1 Place the controls in the architecture

Collection metadata records what a dataset represents and why it was obtained. Row metadata records origin and arrival. In Figure 1, the Business Vault brings the data together for assessment; a gate at the AI-Mart controls which dataset version reaches a model. Model logs return to the warehouse. Sensitive payloads and derivatives remain within their assessed access, transfer and erasure boundaries throughout.



Capability and evidence alone do not establish compliance.

Figure 1. A possible control architecture. Data Vault structures preserve provenance and history; the AI-Mart gate links assessment results to release. Model logs return to warehouse satellites. Define the sensitive-data boundary and erasure scope for each processing activity.

A quality result needs a dataset version, rule version and recorded release decision. An erasure event needs a scope, dependency inventory and response to partial failure. Assign an owner who can produce and explain that evidence. A catalog or orchestration

service may hold it; a new hub or satellite is useful only where the modelling need warrants one.

3. Establish permission and define the sensitive boundary

A recruitment audit comparing selection outcomes may need data revealing ethnicity. Those data fall within Article 9 GDPR: collecting or using them is processing even if the report contains only statistics.¹³ Before choosing tables or running queries, establish the permitted purpose, records and users. An instruction to “check fairness” leaves those decisions unresolved.

3.1 Establish the legal basis

Special-category processing needs both an Article 6 lawful basis and an Article 9 exception.¹⁴ Article 4a operates alongside the GDPR and does not identify the organisation’s Article 6 ground.^{15,16,17} The controller, supported by counsel, should explain why each route covers the actor and activity. If the necessary grounds or Article 4a conditions cannot be established, sensitive-data processing must not proceed; technical safeguards cannot cure missing authority.¹⁸

Article 6(1)(c) may provide a basis for processing necessary to meet an applicable Article 10 bias-related duty.¹⁹ An eligible private-sector controller considering Article 6(1)(f) must establish that the interest is lawful, that processing is necessary, and that the individual’s interests or fundamental rights and freedoms do not override it.²⁰ Public authorities need another applicable Article 6 basis for processing in the performance of their public tasks.¹⁴ Record whether testing fulfils a duty or is voluntary: Article 4a(2) does not itself create a duty to test.⁵

3.2 Assess alternatives before reading protected records

Condition (a) requires an evidenced conclusion that other data, including synthetic or anonymised data, cannot effectively fulfil the objective; condition (f) requires the strict-necessity explanation in the records of processing activities.^{21,22} Define the intended comparison and assess already anonymous results, suitable synthetic data and other separately lawful sources. Record what each can establish and where it falls short. A controlled comparison may help, but the law prescribes neither one experiment nor an exhaustive sequence of techniques. Protected records cannot be read for this assessment before permission and safeguards are established.

Overall selection rates may conceal differences between groups. A postcode proxy may be unreliable, may itself reveal ethnicity, and may require the protected attribute for validation.^{23,24} Computing an anonymous aggregate from identifiable protected records also remains record-level processing. A third party holding those records relocates the Article 9 question and may conflict with Article 4a’s restriction on other parties’ access.^{25,26}

Use sufficient alternatives within their lawful boundary; otherwise document their limits before authorising sensitive processing.

3.3 Define what may enter the warehouse

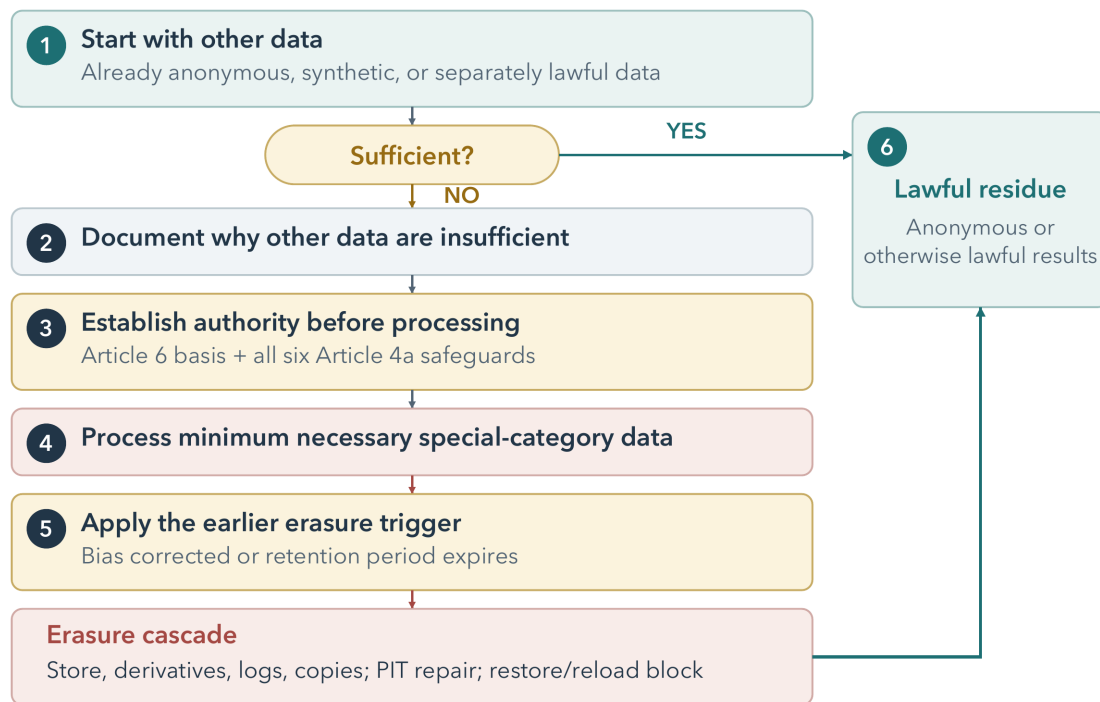
Ethnicity can sit in a dedicated sensitive satellite* linked to the audit's subject records, with access to identifying information controlled separately. A campaign-specific partition is another option. Choose boundaries according to permitted use and what must be erased together. The controller and data engineer should record the purpose, dataset and model versions, necessary attributes, permitted users, destinations, retention deadline and deletion unit. Hashing the business key does not make associated records anonymous when identification remains possible.²⁷

All six Article 4a conditions must be met, with safeguards in place, before processing begins. Condition (b) requires technical limits on reuse and state-of-the-art security and privacy-preserving measures, including pseudonymisation; condition (c) requires strict access controls, confidentiality obligations and access documentation; condition (d) prohibits transmission, transfer or access by other parties. "Other parties" needs assessment for the arrangement; it cannot be reduced to "no unauthorised sharing".^{28,29,26} The platform owner should test satellite permissions and outbound paths; the data owner should check fields and joins against the approved purpose. Use actual queries to establish that the controls hold.

Condition (e) requires deletion at the earlier of bias correction and retention expiry.³⁰ Before the first sensitive load, define the correction event and an erasure path covering dependent data and recovery processes. The controller owns the permission and necessity record; engineering owners supply evidence that the controls work.

Test. Using non-personal fixtures, attempt separate loads with a missing permission reference, an unapproved attribute and an undefined erasure scope. Each should be refused and recorded. Run a permitted load and confirm that only authorised roles can query it. Record the configuration and owners; successful fixture tests do not establish lawful processing.

* 'Sensitive satellite' is descriptive, not a term of art. The published practitioner vocabulary (Linstedt; Scalefree, n 54) is a 'satellite split' into a personal-data satellite and a non-personal satellite. 'PII satellite' is also common, but 'PII' is American vocabulary (NIST) and means all personal data, names and addresses included. 'Sensitive' here is narrower: it means the special categories of Article 9 GDPR only, such as ethnicity or religion. In EU terms the safest phrasing is 'special-category satellite'.



Anonymous output does not make record-level input processing anonymous.

Figure 2. Assess other data first. Establish the Article 6 basis and all six Article 4a conditions before processing protected records; erase at the earlier trigger and retain only anonymous or otherwise lawfully held results.

4. Capture provenance, purpose and transformation history

For providers whose training, validation and testing data fall under Article 10, Article 10(2)(b) covers collection processes and origin, including the original purpose of collecting personal data.³¹ A reviewer needs to follow the dataset from collection to model use and understand what the records represent. Data Vault can hold this evidence; the team must supply it.

4.1 Record why and where data were collected

The row fields *record_source* and *load_date* identify source and arrival, not purpose or context. "Applicant system" and a timestamp do not identify the recruitment campaign, original collection purpose, geography or population covered. Link the load to a collection record containing those facts and their supporting evidence. The data owner should verify the collection description; the engineer should preserve the connection from loaded rows to the correct version of that record.

Collection information can live in catalog metadata, a reference structure or a satellite that records changes to the documented purpose. Preserve earlier versions so a new description does not rewrite the explanation attached to an old extract. Article 10(2)(d) also requires assumptions about what the data measure and represent.³² Define whether a missing outcome means "not yet decided" or "not selected"; the column name alone cannot distinguish them.

Record relevant design choices alongside source facts.¹¹ Explain why the audit counts applications rather than people or vacancies, and why the selected outcome represents the model's intended task. Link those decisions to the collection and semantic records. A reproducible query can still measure the wrong thing if its unit of analysis is poorly chosen or undocumented.

4.2 Trace each transformation

Document how applications were matched to outcomes, duplicates treated and outcome labels derived. Article 10(2)(c) covers preparation operations including annotation, labelling, cleaning, updating, enrichment and aggregation.³³ The Raw Vault can preserve received history, the Business Vault hold derived rules, and the mart deliver model extracts. To trace a result through those layers, the pipeline must record which rule and execution produced it.

A released extract should point to its source collections, transformation versions, execution records and semantic definitions. The data engineering owner should maintain that history through orchestration logs, version-controlled code and catalog entries. References must identify the versions actually used. For manual changes to labels or mappings, record the reason and accountable owner.

A transformation log need not copy protected values to prove that a rule ran. Diagnostic payloads that contain them need access controls and erasure coverage. Preserve the method, decisions and permitted results, while acknowledging that they cannot support indefinite recomputation once the underlying protected records have been lawfully deleted.

Test. Trace a transformed field in a released extract to its collection record and exact transformation version. Ask a reviewer other than the pipeline author to explain its purpose, missing-value meaning and derivation. After changing a mapping rule, verify that the earlier extract still points to the earlier rule and the new extract identifies the change.

5. Measure data quality and bias, then control release

For providers subject to Article 10, datasets must be relevant, sufficiently representative, have appropriate statistical properties, and be as complete and error-free as possible. The Act also requires examination of specified biases and appropriate measures to detect, prevent and mitigate them.^{34,6} Set and justify assessment criteria for the intended use, calculate the evidence and withhold unsuitable extracts. These duties supply no universal acceptance threshold.

5.1 Define the population and measures

A correctly loaded dataset may cover the wrong region, period or recruitment process. The model and data owners should define the intended population, setting and outcomes. Article 10(4) addresses the geographic, contextual, behavioural and functional setting of use.³⁵ Checks can cover missing outcomes, duplicate applications, population coverage and differences in relevant selection or error rates. Record their tolerances, rationale and treatment of uncertainty. An enforceable percentage is not automatically a legally sufficient one.

Article 10(2) requires assessment of data availability, quantity and suitability, and identification of relevant gaps or shortcomings and how to address them.¹¹ Name missing populations or unreliable labels, explain their effect on intended use and assign an owner to the response. More rows from the same unsuitable source may not close the gap. Carry unresolved limitations into the release decision.

Group-level results may reveal a disparity without showing how to correct it. Reweighting, relabelling or targeted collection can require record-level links between protected attributes, model inputs and outcomes. Permission for a statistical comparison does not automatically cover those operations. The model owner should explain what the measurements establish, what remains uncertain and whether correction needs a revised processing decision. Record the measures used to detect, prevent and mitigate identified bias, their owners and the evidence of their effects.¹¹

5.2 Make assessment govern release

The Business Vault can hold integrated data and queries that measure coverage, quality and bias. An AI-Mart gate can withhold extracts that fail the agreed checks. Practitioner work describes the AI-Mart as a location for curated, validated training data; the Act requires no such named layer.³⁶ An export service or training pipeline can enforce the same decision. The assessment must govern the route that actually supplies data to the model.

The release record should identify dataset and model versions, assessment results, unresolved limitations and the decision owner. Engineering supplies execution evidence; the model owner explains the consequences for intended use. A completed job cannot substitute for a missing or failed assessment. Where an exception is allowed, document it and keep it within applicable requirements and the approved scope. An override cannot make unlawful processing lawful.

5.3 Check deployer-controlled inputs separately

An employer may control the applicant records supplied to the recruitment system while the provider controls its training dataset. Article 26(4) requires relevant and sufficiently representative inputs insofar as the deployer controls them.¹² The employer's data owner

should identify that boundary, apply the provider's input specifications and assess whether the records fit the intended recruitment setting. This responsibility is distinct from the provider's training-data programme.

For each input batch, record its source, expected fields, quality and coverage checks, known limitations and accountable reviewer. Test an incomplete or out-of-scope batch before it reaches the live input route. The process should hold or reject it, record the reason and escalate to the system owner. Schema checks alone cannot establish representativeness; deployers should not claim control or evidence over data they cannot inspect.

5.4 Assess what may be retained

Small groups and overlapping queries can expose membership, hidden attributes or individual records. Aggregate data are not inherently anonymous.³⁷ Cell suppression, minimum group sizes and differential-privacy noise may reduce disclosure risk while weakening the audit's signal. Configure them for the use case and the GDPR test of whether identification remains reasonably likely; there is no universal "safe" cell count.²⁷ The responsible owner should record the disclosure assessment and its effect on the conclusions.

Computed satellites in the Business Vault can retain lawful audit results linked to an audit run or model version.³⁸ The results still need sufficient anonymity or another lawful basis. They cannot support every future metric once protected records have been deleted. Record that limit, the method and the supporting evidence that may remain.

Test. Add a known quality defect to a non-personal test extract. Separately, omit its assessment result. Verify that both attempts are refused at the model-input route and produce usable failure records. Test the report's small-group and overlapping-query controls, and check that its interpretation accounts for any suppression or noise.

6. Connect datasets, model use and evidence

6.1 Make each delivery traceable

For each released recruitment model, retain a trace to the dataset version, the changes made to it and the decisions permitting its use.

Hubs hold business identifiers, links represent relationships, and satellites preserve descriptive history. Normal satellite loads append rows, supporting a historical system of record and parallel loading; controlled deletion remains technically possible.^{39,40,41} Give each dataset release, model version and bias-review campaign a stable identifier. Connect them through links and attach the relevant decisions, transformations and results through satellites.

The data engineering owner should be able to trace a delivery from its source and collection context through preparation, Business Vault measurements and the AI-Mart release decision to the receiving model version. Attach the approved purpose and retention rule to the campaign; `record_source` and `load_date` do not convey them. Record the transformation version and checks that accepted the extract. Keep enough version information to explain a historical result without retaining protected values beyond their lawful deletion point.

Test. Ask a reviewer who did not build the pipeline to identify a model version's dataset, transformation rules, check results and release decision. Change the dataset version or fail a release check: the trace should distinguish the new attempt, and the failed extract should not reach the model. These tests do not establish that the metrics or thresholds satisfy the law.

6.2 Control model logs

Where Article 12 applies, high-risk systems must support automatic event logging throughout their lifetime.⁴² Paragraphs 1 and 2 contain the general requirements; paragraph 3's detailed list applies to specified biometric systems.⁴³ Choose fields for the applicable requirements. Lifetime logging capability does not require lifetime retention of every event: assess retention separately.

Training and inference logs can return to the warehouse as satellites linked to the run or model version, following published Data Vault practice.⁴⁴ Select useful evidence—potentially inputs, transformations, parameters, outputs and confidence scores—without copying every input into a permanent log. Model and platform owners should agree on event meanings, access restrictions and retention. Omit sensitive values where the logging purpose can be met without them; otherwise include those log satellites in the sensitive perimeter and erasure plan.

A source-system audit trail does not establish who queried the warehouse. Status tracking satellites can preserve source-supplied create, read, update and delete events; warehouse query and access logs describe a different boundary. Verify that the chosen evidence identifies the relevant user, operation and data scope without unnecessarily copying protected values.⁴⁵

A privileged operator may change history despite append-only loading. Protect change and access logs, restrict mutation rights and provide appropriate integrity checks. Record authorised erasure without preserving the deleted payload. Test both a forbidden mutation and an authorised erasure: the former should be blocked or detected, while the latter should leave the remaining evidence usable.

Article 19(1) assigns retention of provider-controlled logs to the provider; Article 26(6) addresses the deployer. Each requires a period appropriate to the purpose of at least six

months, unless applicable law, particularly data-protection law, provides otherwise.⁴⁶ Identify which model-operation logs the recruitment-model provider controls and which use logs the employer controls. Distinguish ordinary operational evidence from protected payloads subject to the assessed erasure scope.

6.3 Retain evidence lawfully

A computed satellite may retain a lawful result, its method, campaign identifier and decision record after deletion.³⁸ Small groups or overlapping results may still expose individuals: aggregation alone does not establish anonymity.^{37,27} Record the anonymity assessment or other lawful basis for retention, and the resulting limits on re-audit. A saved recruitment-disparity result cannot recreate deleted records for every later metric or model version.

7. Make retention and erasure executable

7.1 Connect both deletion triggers

Article 4a(1)(e) requires deletion at the earlier of bias correction and retention expiry.⁴⁷ A scheduler can detect a deadline, but cannot decide that bias has been corrected. The model owner and legal or governance function must define, validate and record that state. Engineering then connects the lifecycle event and retention deadline to one erasure workflow. Van Bakkum notes the condition's close relationship to GDPR storage limitation: the retention decision needs justification as well as an executable deadline.⁴⁸

Before loading, define whether deletion covers a campaign, model version, subject or narrower set with a common purpose and trigger. Map that scope to satellite rows, features, intermediate tables, marts, logs, exports and recoverable copies. Assign an operational owner to complete deletion across that inventory and escalate failures. Record the scope, trigger, affected components and outcome without copying the protected value.

7.2 Match the pattern to the deletion unit

A table, a row set and an encryption key define different deletion boundaries. Match the pattern to the approved scope and recovery behaviour.

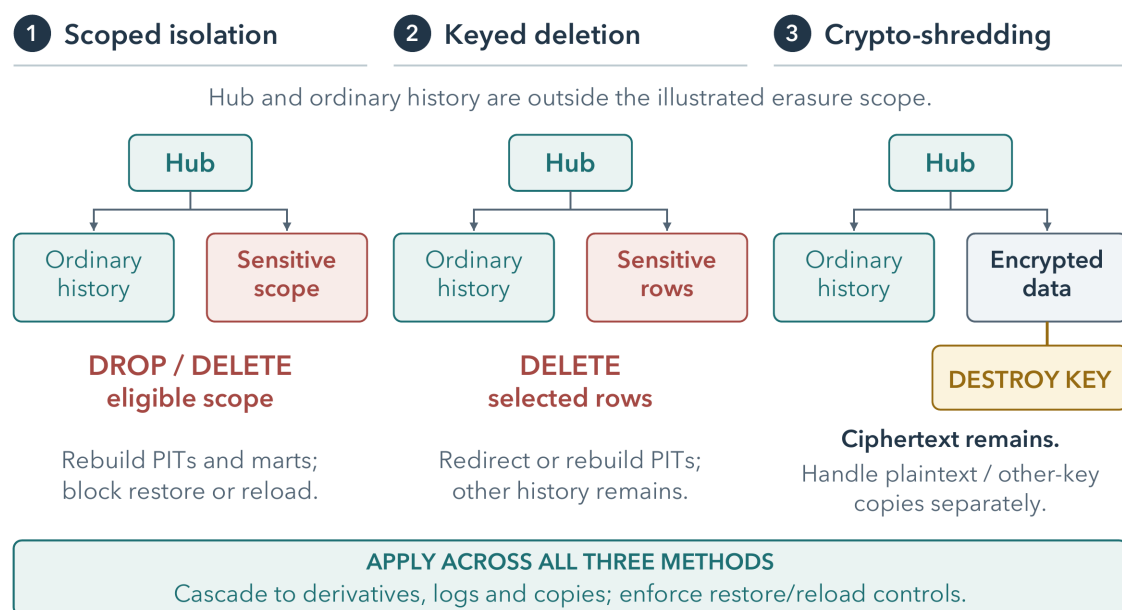


Figure 3. Three controlled erasure patterns. Define the sensitive scope before loading; remove an eligible partition or row set, delete selected rows, or destroy the relevant encryption key. Each pattern also needs dependency, recovery and reload controls.

Scoped isolation places the protected attribute in a dedicated satellite, partition or row set, following a published practitioner approach to personal-data separation.⁴⁹ Drop a whole table only when every row shares the relevant purpose, retention rule and trigger. A table shared by recruitment campaigns may need partitions or selected-row deletion instead. Erasure must also remove affected derivatives; surviving identifiers do not thereby become non-personal.

Keyed row deletion combines the approved campaign or other deletion scope with subject identifiers and historical versions. A subject's deterministic business-key hash may recur across campaigns; it cannot identify the deletion unit by itself. Its ordinary role is deterministic identification and parallel loading, not erasure.⁵⁰ The schema and dependency rules make selective deletion possible. Test that all intended historical versions disappear and that another campaign's history remains intact where legally retainable. Deleting only the latest row may leave the attribute in earlier records.

Crypto-shredding destroys separately managed key material, making covered ciphertext unreadable while leaving the rows in place. Scope keys to the deletion unit and account for every recoverable key copy. Only data encrypted under that key or key hierarchy are affected: plaintext exports, decrypted caches and independently encrypted copies need separate treatment. Its legal acceptability remains uncertain. The cited literature records contrary supervisory views, so assess the method for the jurisdiction and system.⁵¹

7.3 Repair references and prevent reappearance

A point-in-time (PIT) table can retain a deleted satellite row's hash key and load timestamp. A ghost record is a placeholder row; it does not update those references. Rebuild affected entries or redirect them to the ghost, then check the joins.⁵² Old PIT

snapshots can ordinarily be reconstructed from the Raw Vault.⁵³ Erasure must prevent reconstruction of the protected value, not just remove a query structure.

A restore, retry or source refresh must not make an erased value available again. Apply the assessed scope to backups and other recoverable copies. A suppression record can coordinate this only if its content and use are appropriate; do not retain the deleted attribute inside the deletion log. Have the recovery owner demonstrate the restoration procedure before the campaign enters production.

A completed campaign deletion may leave an Article 17 GDPR request unresolved. That right has its own grounds and exceptions and may reach hub identifiers, business keys, links or other satellites left intact by Article 4a's attribute-focused deletion.⁵⁴ The literature on resistant storage likewise treats erasure as a wider architectural problem.⁵⁵ Assess the broader request against an inventory of personal data; an empty bias satellite is not a completion criterion.

7.4 Resolve log retention and model effects

Article 26(6) requires deployers to retain automatically generated logs under their control for a purpose-appropriate period of at least six months, unless applicable law, particularly data-protection law, provides otherwise.⁵⁶ Campaign closure does not authorise discarding all operational logs. Assess which special-category values fall within Article 4a's earlier trigger and which other records can or must remain. Encode those scopes separately.

Deleting training data does not automatically remove its effect on the model. Reweighting, relabelling or sampling may use a protected attribute absent from the final inputs. Full retraining is the reference point for comparison; exact and approximate unlearning differ in cost, access requirements and limits.⁵⁷ Record the attribute's influence and assess the need for retraining, unlearning, replacement or another documented response. Whether the model retains or exposes personal data requires a case-specific assessment of extraction and query risks.⁵⁸ A warehouse deletion test cannot answer that question.

8. Test, operate and hand over the controls

8.1 Test the whole path, including failure

With suitable test data, run a campaign through permitted collection, profiling, release, logging and deletion. Trigger bias correction before the retention deadline, then reverse the order in a separate run. Verify the primary and derived stores, PIT repair, logging treatment and lawful retained results. Inspect the data and query behaviour, not just the job status.

Interrupt a deletion and retry it. Restore an older copy under the documented procedure and replay a source load; neither should reintroduce the protected value into use. Attempt an unauthorised read or export with an ordinary account: both must be denied. Test privileged modification under agreed controls, specifying whether prevention or detection is achievable and how to escalate and remedy tampering. Detecting a completed unauthorised disclosure does not pass an access-control test. Record failures and responsible owners.

8.2 Hand over decisions, evidence and responsibility

The handover should cover applicability and lawful processing; dataset and model relationships; sensitive scope; release checks; access and egress rules; log fields and retention; deletion triggers and method; recovery and reload; test results; and unresolved issues. Name an accountable function for each operational responsibility.

Operations need a runbook for recognising triggers, inspecting progress, recovering from partial failure, checking remaining data and escalating unresolved components. The model owner needs the separate decision on model effects and re-audit limits. Give governance the test results, failures and judgments still open. Review the records when sources, model use, transformations, retention rules or recovery arrangements change.

8.3 Support wider system obligations

Send the provider's documentation owner the dataset versions, collection context, transformation history, validation results and known limitations needed for the technical file. Connect runbooks and review records to the quality-management system. Articles 11, 16 and 17 require more documentation and procedures than a catalog can supply; Article 18's ten-year rule covers the listed documentation, not blanket retention of personal training data.⁴⁶

Supply input specifications and performance limitations to the team writing deployer instructions. Give human-oversight owners intelligible evidence and intervention paths; an automated release gate is not human oversight. Work with model, platform and security teams on accuracy, robustness, feedback and poisoning risks. These contributions support Articles 13–15; a warehouse quality score cannot establish compliance with them.⁵⁹

After release, link monitoring results and reported problems to the affected model, dataset and processing versions. Send significant signals to the monitoring or incident owner with timestamps and usable evidence. A threshold breach is not automatically a reportable serious incident; an unreviewed pipeline should not make that determination. For deployers within Article 27's defined cohort, subject to its exceptions, supply the data-flow, population and safeguard facts for the fundamental-rights impact assessment.

Assess the applicability and timing of Articles 72-73 separately, including their transitions.⁶⁰

Test. Select a monitoring alert and check that an accountable person can identify the system and dataset version, retrieve the evidence and follow the review and escalation route. Confirm that the instructions and technical file describe the same released configuration. Passing this test does not determine whether an incident is legally reportable.

8.4 Keep the limits visible

Review the affected controls when permissions, pipelines, models or recovery procedures change. Verify that the safeguards still work where the data are processed.

Successful tests establish neither strict necessity nor an Article 6 basis.¹⁸ Article 4a(2) does not create a general duty to conduct bias testing outside the applicable obligations.⁶¹ Record the assessed high-risk scope, application dates, exclusions and distinct general-purpose-model regime in the handover.¹⁰ Quality management, conformity assessment, technical documentation and organisational accountability extend beyond the warehouse and require continuing operation.

9. Conclusion

The warehouse's capacity to remember is valuable precisely where deletion hurts: reconstructing a decision, checking a new source of bias, or explaining an earlier result. GDPR's restrictions and erasure duties still apply, alongside the AI Act's applicable demands. Controlled deletion is possible, but its scope determines which evidence survives and which questions can no longer be answered. Those choices belong in the design, with an owner and a lawful purpose for what remains.

Choose the deletion unit before loading. Connect the earlier of the two Article 4a triggers to a workflow that reaches derivatives, logs and recoverable copies. Test whether a restore or fresh source load brings the value back. Preserve the method, decisions and results that may lawfully remain, and say what they can no longer prove. A useful Vault must be able to explain its history—and carry out a justified decision to erase part of it.

Appendix. Legal background and interpretive limits

A.1 From Article 10(5) to Article 4a

The original exception arrived without a clear account of its origin. Arnoud Engelfriet traces Article 10(5) to a Commission proposal whose accompanying materials left that question unanswered. The EDPB and EDPS challenged its interaction with the GDPR in

2021. Parliament subsequently added safeguards and the exceptional character of the permission; the Council added the Article 9(2)(g) reference reflected in recital 70 of the 2024 AI Act.^{62,63}

Regulation 2026/1744 deleted Article 10(5) and inserted Article 4a, retaining the six safeguards while widening the categories of actors able to invoke the permission. The amended Article 10 cross-references point to Article 4a(1). Article 2(7) continues to preserve the GDPR subject to the specified AI Act adjustments, and recital 9 links bias detection and correction to substantial public interest and GDPR Article 9(2)(g). The architecture must operate within that legal framework.^{64,65,66}

A.2 Article 6 requires a separate basis

Regulation 2026/1744 supplies no general Article 6 ground for bias processing. Its GDPR references include Article 9(2)(g) in recital 9 and Article 35 in the amended impact-assessment provision; Article 4a operates in addition to the GDPR. The earlier regulatory-sandbox recital, by comparison, expressly mentions both Article 6(4) and Article 9(2)(g). These differences support a separate assessment of the Article 6 ground and Article 9 route, rather than deriving both from Article 4a.^{16,17,67}

Engelfriet reads the original exception as a tightly bounded, dependent derogation whose exceptional character points to serious systemic bias. That is his interpretation, not an express statutory threshold. Article 4a(2)'s broader wording also matters. Where the grounds or safeguards cannot be established, the available choices are another lawful technique, a narrower activity, or no processing.^{68,18}

A.3 Distinguish enacted law from the GDPR proposal

The separate GDPR Digital Omnibus proposed a route for special-category data encountered incidentally during AI development and operation, together with an AI-specific legitimate-interest provision. Unlike the enacted AI Omnibus, it remains a proposal. Its provisions cannot authorise present processing.⁶⁹

The EDPB and EDPS considered a dedicated AI legitimate-interest provision unnecessary: the existing framework already operates case by case. They also requested clearer interaction among the proposed sensitive-data routes. Council document ST 10729/26, dated 22 June 2026, records the compromise then under discussion. Keep permission records versioned as the law changes; a possible future permission has no place in the current processing configuration.⁷⁰

A.4 Put safeguards where processing occurs

A safeguard written in a policy may fail where the data are actually processed. Engelfriet connects structural protection to the Court of Justice's reasoning in *La Quadrature du Net*, a judgment arising in a different retention context. The connection supports an

attributed architectural argument; the Court did not prescribe Data Vault, a partition or an AI-Mart.^{71,72,73}

A duty to mitigate bias can coexist with restrictions on the data needed to investigate it. Article 4a(2)'s optional testing remains optional, and duties on providers of high-risk systems cannot be extended to general-purpose models by analogy. Operating controls therefore depend on the assessed actor, system and purpose throughout their use.^{74,61,10}

Notes

1. Regulation (EU) 2026/1744 of the European Parliament and of the Council of 8 July 2026 amending Regulations (EU) 2024/1689, (EU) 2018/1139 and (EU) 2023/1230 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI), OJ L, 24 July 2026 (ELI: [data.europa.eu/eli/reg/2026/1744/oj](https://eur-lex.europa.eu/eli/reg/2026/1744/oj)); in force on the third day following publication (Art. 4).
2. Reg. 2026/1744, Art. 1(9)(b) ("paragraph 5 is deleted").
3. Reg. 2026/1744, Art. 1(6), inserting Art. 4a into Regulation (EU) 2024/1689. Article citations to the AI Act herein are to the consolidated text 02024R1689 – EN – 27.07.2026 – 001.001, except recitals, which are cited to the respective OJ texts.
4. AI Act, Art. 4a(2).
5. AI Act, Art. 4a(2), closing sentence.
6. AI Act, Art. 10(2)(f) and (g).
7. AI Act, Art. 10(3).
8. Reg. 2026/1744, recital 9 (final sentence: the legal basis "should apply from the date of entry into application of that Regulation"); AI Act, Art. 113 (Chapters I and II apply from 2 February 2025; the amended point (a) carves out the new Art. 5 prohibitions, which apply from 2 December 2026 – immaterial to Art. 4a).
9. AI Act, Art. 113(c), as amended by Reg. 2026/1744, Art. 1(40) (2 December 2027 for systems under Art. 6(2); 2 August 2028 for systems under Art. 6(1)); AI Act, Art. 2(2) (for systems related to products listed in Annex I Section B, only Art. 6(1), Art. 60a, and Arts. 102-112 apply, so Art. 10 does not).
10. AI Act, Art. 113(c), Arts. 51-56, and Chapter V; Art. 2(2) (limiting the application of Chapter III duties to systems related to products listed in Annex I Section B).
11. AI Act, Art. 10(1), (2)(a)-(h) and (6), as amended by Regulation (EU) 2026/1744, Art. 1(9). Paragraph 6 limits the specified requirements to testing datasets for the development of high-risk systems not using model-training techniques. Consolidated text: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02024R1689-20260727>.
12. AI Act, Art. 26(1), (4): use according to instructions; relevant and sufficiently representative inputs, in view of intended purpose, insofar as the deployer controls them.
13. Regulation (EU) 2016/679 (GDPR), OJ L 119, 4 May 2016, Art. 9(1).
14. GDPR, Arts. 6(1), 6(3), 9(1)-(2), 17(1)-(3). Article 6(1)(f) excludes public-authority processing in performance of public tasks; points (c)/(e) require the Article 6(3) legal basis. Case C-667/21 *Krankenversicherung Nordrhein* EU:C:2023:1022, paras 73-79: concurrent Articles 5, 6 and 9 requirements. <https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng>.
15. Engelfriet (n 62) 3-5; Case C-667/21 *Krankenversicherung Nordrhein* EU:C:2023:1022, paras 73-78, esp. 78, and operative part, point 3 (processing under Art. 9(2)(h) "must, in order to be lawful, not only comply with the requirements arising from that provision, but must also satisfy at least one of the conditions of lawfulness set out in Article 6(1)"). The article expressly "does not attempt to resolve which legal basis under Article 6 GDPR might apply to bias mitigation activities": *ibid* 2.
16. Reg. 2026/1744, *passim*; the only provision-level GDPR citations are recital 9 (Art. 9(2)(g)) and Art. 1(13) (amending AI Act Art. 27(4), citing GDPR Art. 35).

17. AI Act, Art. 4a(1), chapeau.
18. Engelfriet (n 62) 2, 4 (an Article 6 ground assumed to exist elsewhere; the question expressly left unresolved) and 10 (the fallback where the conditions are unmet: less effective techniques or the risk of non-compliance).
19. Van Bekkum (n 24), §4.6.2 (Article 6(1)(c) GDPR juncto Article 10(2) AI Act as "a concrete legal obligation").
20. Case C-621/22 *Koninklijke Nederlandse Lawn Tennisbond* EU:C:2024:858, paras 36-57 (the three cumulative conditions at para 37; commercial interest at paras 47-49) and operative part (a purely commercial interest may qualify as a legitimate interest provided it is lawful, the processing is strictly necessary for it, and the data subject's interests do not override it).
21. AI Act, Art. 4a(1)(a).
22. AI Act, Art. 4a(1)(f).
23. Žliobaitė & Custers, 'Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models' (2016) 24 *Artificial Intelligence and Law* 183; Veale & Binns, 'Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data' (2017) 4(2) *Big Data & Society*, art 2053951717743530.
24. On the necessity of the protected attribute for proxy validation, see M. van Bekkum, 'Using sensitive data to de-bias AI systems: Article 10(5) of the EU AI Act' (2025) 56 *Computer Law & Security Review* 106115, §2 (a provider "must collect ethnicity data" to establish that an attribute is a proxy).
25. Veale & Binns (n 23), proposing trusted third parties that hold sensitive attributes and return fairness statistics.
26. AI Act, Art. 4a(1)(d).
27. GDPR, recital 26 (identifiability assessed by "all the means reasonably likely to be used"; pseudonymised data remains personal data). On the personal/non-personal boundary see M. Finck & F. Pallas, 'They who must not be identified – distinguishing personal from non-personal data under the GDPR' (2020) 10(1) *International Data Privacy Law* 11, 12-13 ("perfect anonymization is impossible" and the legal definition "needs to embrace the remaining risk"; the recital 26 reasonable-likelihood test as the operative boundary).
28. AI Act, Art. 4a(1)(b).
29. AI Act, Art. 4a(1)(c).
30. AI Act, Art. 4a(1)(e).
31. AI Act, Art. 10(2)(b).
32. AI Act, Art. 10(2)(d).
33. AI Act, Art. 10(2)(c).
34. AI Act, Art. 10(3).
35. AI Act, Art. 10(4).
36. Scalefree International GmbH, 'AI Act Insight: Ensuring Responsible AI for Your Business' (scalefree.com/blog/artificial-intelligence/ai-act-insight-ensuring-responsible-ai-for-your-business/, published 29 October 2024, last updated 21 May 2026; accessed 2 September 2026): the AI-Mart as "a specialized data mart" for curated, validated AI training data, with model logs loaded back into the warehouse.
37. A. Gadotti, L. Rocher, F. Houssiau, A.-M. Crețu & Y.-A. de Montjoye, 'Anonymization: The imperfect science of using data while preserving privacy' (2024) 10(29) *Science Advances* eadn7053 (taxonomy of attacks on aggregate data: membership inference, attribute inference, and reconstruction, with the differencing attack described as "a very simple kind of inference attack"; aggregate data "do not inherently protect against privacy attacks").
38. Linstedt & Olschimke (n 41) ch 5.3.6 (computed satellites in the Business Vault).
39. D. Krneta, V. Jovanović & Z. Marjanović, 'A direct approach to physical Data Vault design' (2014) 11(2) *Computer Science and Information Systems* 569, 569-570.
40. A. Vines & R.-E. Samoilă, 'An Overview of Data Vault Methodology and Its Benefits' (2023) 27(2) *Informatica Economica* 15, 17-21 (hash keys enabling parallel processing of hubs and satellites; record source and load date mechanics); Linstedt & Olschimke (n 41) chs 4.3-4.5.

41. D. Linstedt & M. Olschimke, *Building a Scalable Data Warehouse with Data Vault 2.0* (Morgan Kaufmann 2016) ch 4.5.1 ("Because the history of the data needs to be preserved, you are not allowed to update or modify the data in the satellite. The only exception to this rule is the Load End Date attribute") and ch 6 (PIT and bridge tables as Business Vault query-assistant entities).
42. AI Act, Art. 12(1).
43. AI Act, Art. 12(2)–(3); the detailed list in Art. 12(3) applies to systems referred to in Annex III, point 1(a).
44. Scalefree (n 36).
45. Linstedt & Olschimke (n 41) ch 5.3.3 (status tracking satellites loading CRUD audit trails; logging of reads "for security reasons to provide information about who accessed which data").
46. AI Act, Art. 11 and Annex IV; Arts. 16–17; Art. 18(1) (listed documentation until ten years after placement on the market or putting into service); Arts. 19(1), 26(6) (automatic logs under the actor's control, retained for a purpose-appropriate period of at least six months unless applicable law, particularly data-protection law, provides otherwise). Documentation retention does not itself authorise retention of personal training data.
47. AI Act, Art. 4a(1)(e) (formerly Art. 10(5)(e)).
48. Van Bakkum (n 24), §4.6.1.
49. Scalefree (n 54).
50. Linstedt & Olschimke (n 41) ch 4.3.2.1: the hash key "replaces the sequence number from the Data Vault 1.0 standard", is computed from the business key, "can be regenerated", and is cross-platform; that is, no central sequencer is needed and loads parallelize deterministically.
51. Belen-Saglam et al. (n 55) §4.6.1 ('Hashing out') and accompanying text (encryption-based key deletion as a proposed right-to-be-forgotten implementation, with contrary views recorded). The technical effect extends only to data that remain encrypted under the destroyed key material; plaintext or independently encrypted copies require separate treatment. Politou, Alepis & Patsakis (n 55) address erasure in resistant infrastructures.
52. Linstedt & Olschimke (n 41) ch 6.1.1 (PIT entries use a satellite hash key and load timestamp; ghost records replace missing references so equi-joins remain possible). After a referenced real row is erased, affected PIT entries must be rebuilt or redirected to the ghost record; the ghost's mere existence does not change a stale PIT reference.
53. Linstedt & Olschimke (n 41) ch 6.1.2 (managed and logarithmic PIT windows; deleted snapshots "could be rebuilt using the same algorithm that has built them in the past").
54. See the published material of Scalefree International GmbH, cited as one example of practitioner documentation: 'Use Data Vault 2.0 to Tackle GDPR' (published 20 December 2024, updated 11 May 2026), which describes separating privacy-relevant attributes and deleting records from the personal-data satellite while retaining other warehouse history; and 'Implementing GDPR in Data Warehousing' (published 11 July 2022, updated 16 April 2026). Both accessed 2 September 2026. The same material notes that a business key may itself contain personal data; retained structures must therefore be assessed separately.
55. E. Politou, E. Alepis & C. Patsakis, 'Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions' (2018) 4(1) *Journal of Cybersecurity* ty001; R. Belen-Saglam et al., 'A systematic literature review of the tension between the GDPR and public blockchain systems' (2023) 4(2) *Blockchain: Research and Applications* 100129; A. Zafar, 'Reconciling blockchain technology and data protection laws: regulatory challenges, technical solutions, and practical pathways' (2025) 11(1) *Journal of Cybersecurity* tyaf002 (abstract).
56. AI Act, Art. 26(6) (emphasis added).
57. H. Xu, T. Zhu, L. Zhang, W. Zhou & P.S. Yu, 'Machine Unlearning: A Survey' (2024) 56(1) *ACM Computing Surveys*, art 9, DOI 10.1145/3603620 (distinguishing retraining, exact unlearning, and approximate unlearning, and discussing their cost and limitations).
58. I. Marco-Pérez, B. Pérez, Á.L. Rubio García & M.A. Zapata, 'The Many Faces of Data Deletion: On the Significance and Implications of Deleting Data' (2026) 58(7) *ACM Computing Surveys*, DOI 10.1145/3779299; EDPB Opinion 28/2024 on certain data-protection aspects related to the processing of personal data in the context of AI models, paras 29–49 (whether a model is

anonymous or contains personal data requires a case-specific assessment of extraction and query risks).

59. AI Act, Arts. 13(1)–(3), 14, 15 and 26(2): instructions, human oversight, accuracy, robustness and cybersecurity; competent, trained and authorised deployer oversight. These are system-level duties, not prescribed warehouse designs.
60. AI Act, Arts. 27, 26(9), 72, 73, 26(5), 111 and 113. Article 27 concerns the specified deployer cohort; Article 26(9) applies where a DPIA is required. Article 72 addresses provider monitoring; Article 26(5), deployer monitoring and escalation. These obligations depend on actor, system, statutory exceptions and transitions. Articles 72–73 fall outside Chapter III's specific deferral. Relevant sectoral provisions also require assessment.
61. AI Act, Art. 4a(2), closing sentence.
62. A. Engelfriet, 'Permitted by design? Article 10(5) AIA, sensitive data, and the legal illusion of bias correction' (2026) 16(1) International Data Privacy Law ipaf035, 1–10 (DOI: 10.1093/idpl/ipaf035), 2–3 ("no publicly stated rationale for why such a clause was included"); EDPB-EDPS Joint Opinion 5/2021 (18 June 2021), quoted *ibid* 2 and 4.
63. Engelfriet (n 62) 2–3; AI Act (2024 text), recital 70, OJ L 2024/1689, 12 July 2024.
64. Reg. 2026/1744, Art. 1(9)(b) and Art. 1(6).
65. AI Act, Arts. 10(1), 10(6), 2(7), as amended.
66. Reg. 2026/1744, recital 9.
67. AI Act (2024 text), recital 140 ("in accordance with Article 6(4) and Article 9(2), point (g), of Regulation (EU) 2016/679"); Engelfriet (n 62) 4.
68. Engelfriet (n 62) 1, 9–10 (Article 10(5) "best understood not as a general permission for fairness-oriented processing, but as a clause of last resort, invocable only in narrowly circumscribed cases where systemic bias rises to the level of a serious incident").
69. Proposal for a Regulation amending Regulation (EU) 2016/679 and others as regards simplification of the digital legislative framework (Digital Omnibus), COM(2025) 837 final, 19 November 2025, Art. 3(3) (proposed GDPR Art. 9(2)(k) for special-category data that occurs in AI development, subject to proposed Art. 9(5) safeguards) and Art. 3(15) (proposed Art. 88c on legitimate interests in AI development and operation). The proposal remains pending and is not cited as law.
70. EDPB-EDPS Joint Opinion 2/2026 on the Digital Omnibus (10 February 2026), paras 39–41 (a dedicated AI legitimate-interest provision is unnecessary because the existing Art. 6(1)(f) framework applies case by case) and para 52 (requesting clearer interaction among the proposed Art. 9 routes). Council document ST 10729/26, 22 June 2026, *Mandate for negotiations with the European Parliament*, records the Presidency compromise submitted to Coreper for confirmation; it is not evidence that a negotiating mandate was withdrawn on 30 June 2026. On the Commission's description of the AI-Omnibus and GDPR-Omnibus provisions as independent, see Council working document WK 251/2026 INIT, 9 January 2026.
71. Engelfriet (n 62) 7–8.
72. Joined Cases C-511/18, C-512/18 and C-520/18 *La Quadrature du Net and Others* EU:C:2020:791, para 132 (in the context of paras 130–136).
73. Engelfriet (n 62) 8: safeguards should be 'embedded into the architecture of processing, not appended post hoc' - his extension of the *La Quadrature* line, adopted here with attribution. See also GDPR, Art. 25 (data protection by design and by default).
74. Engelfriet (n 62) 10 ("The compromise, however, comes at a cost ... the obligation to mitigate persists, yet the lawful ability to process the data needed to fulfil it is curtailed").

Note on sources

AI Act articles cite consolidated text 02024R1689–27.07.2026–001.001; recitals cite the Official Journal instrument that introduced them. The consolidation is a documentation tool; the Official Journal instruments remain authoritative. GDPR articles use the consolidated text with the 2018 corrigendum; recitals use OJ L 119. Legislative status and the expanded statutory map were rechecked against the cited official sources on 2 September 2026. Practitioner references retain their recorded publication, update and access dates. The consultancy material supplies public examples of Data Vault practice. The recruitment example, architecture choices, evidence packages and acceptance checks are illustrative proposals, not prescribed statutory designs. Views and errors are the author's.