Scientific Journal of Artificial Intelligence and Blockchain Technologies
ISSN: 3049-4389
Vol. 1, Issue 3, Jul – Sep 2024 || PP. 10-19 https://doi.org/10.63345/sjaibt.v1.i3.102

Data Sovereignty in Global AI-Blockchain Infrastructure

Prof (Dr) Ajay Shriram Kushwaha

Sharda University

Knowledge Park III, Greater Noida, U.P. 201310, India

kushwaha.ajay22@gmail.com

Date of Submission: 22-06-2024 Date of Acceptance: 23-06-2024 Date of Publication: 03-07-2024

ABSTRACT

As artificial intelligence (AI) systems scale across borders and decentralized ledgers interconnect global networks, a central challenge emerges: how to ensure data sovereignty—the ability of jurisdictions, organizations, and individuals to exert legitimate control over data—without stifling innovation or undermining the integrity and utility of distributed architectures. This manuscript proposes a comprehensive, practice-oriented blueprint for embedding data sovereignty into AI-blockchain infrastructure. We first synthesize the legal and policy landscape shaping cross-border processing (e.g., GDPR, Schrems II, SCCs, EU–U.S. Data Privacy Framework, EU AI Act, Data Governance Act, Data Act, CLOUD Act, India's DPDP Act, China's PIPL, OECD/G7 initiatives, and Global CBPR). We then examine technical levers—on-chain/off-chain partitioning, permissioned topologies, sovereign cloud patterns, privacy-enhancing computation, verifiable provenance, and risk management frameworks—to operationalize jurisdictional constraints without losing decentralization benefits. Building on this review, we introduce SOVEREIGN-Stack, a governance-by-design methodology spanning eight layers (identity, consent, data classification, locality & routing, compute & model governance, ledger governance, transfer mechanisms, and assurance/tooling). Two applied vignettes—in health analytics spanning the EU, U.S., and India, and a permissioned supply-chain ledger touching the EU and APAC—demonstrate how the approach balances legal obligations (erasure, purpose limitation, transfer restrictions) with architectural needs (immutability, integrity, transparency). We conclude with a set of implementation checkpoints and maturity indicators that organizations can use to align AI-blockchain roadmaps with evolving global rules while maintaining verifiability, auditability, and performance.

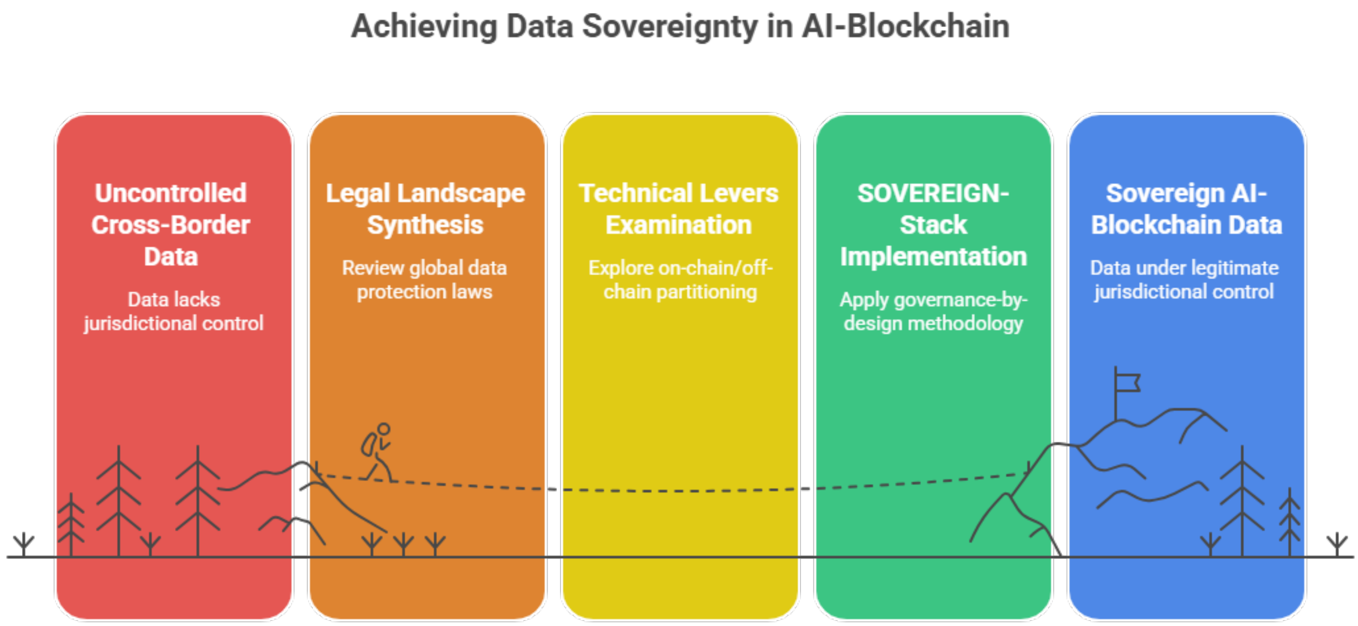
Achieving Data Sovereignty in AI-Blockchain
Uncontrolled Cross-Border Data
Data lacks jurisdictional control
Legal Landscape Synthesis
Review global data protection laws
Technical Levers Examination
Explore on-chain/off-chain partitioning
SOVEREIGN-Stack Implementation
Apply governance-by-design methodology
Sovereign AI-Blockchain Data
Data under legitimate jurisdictional control

Figure-1.Achieving Data Sovereignty in AI-Blockchain

Keywords

Data Sovereignty, Cross-Border Data Flows, AI Governance, Blockchain Compliance, Privacy-Enhancing Technologies, Federated Learning, Sovereign Cloud, GDPR, EU AI Act, Global CBPR

Introduction

AI-blockchain convergence has intensified long-standing tensions between open, verifiable computation and jurisdiction-specific limits on personal and sensitive data. On one side, decentralized ledgers, append-only logs, and tamper-evident states maximize transparency and auditability; on the other, privacy and data protection regimes demand purpose limitation, minimization, erasure/rectification, and transfer controls. The stakes increased in 2024–2025 as the EU Artificial Intelligence Act introduced binding risk-based obligations for AI systems and general-purpose AI (GPAI) models, with enforcement milestones starting in 2025 for transparency and related duties. These provisions, combined with existing data laws, are reshaping global infrastructure decisions for model training, inference, storage, and synchronization across borders.

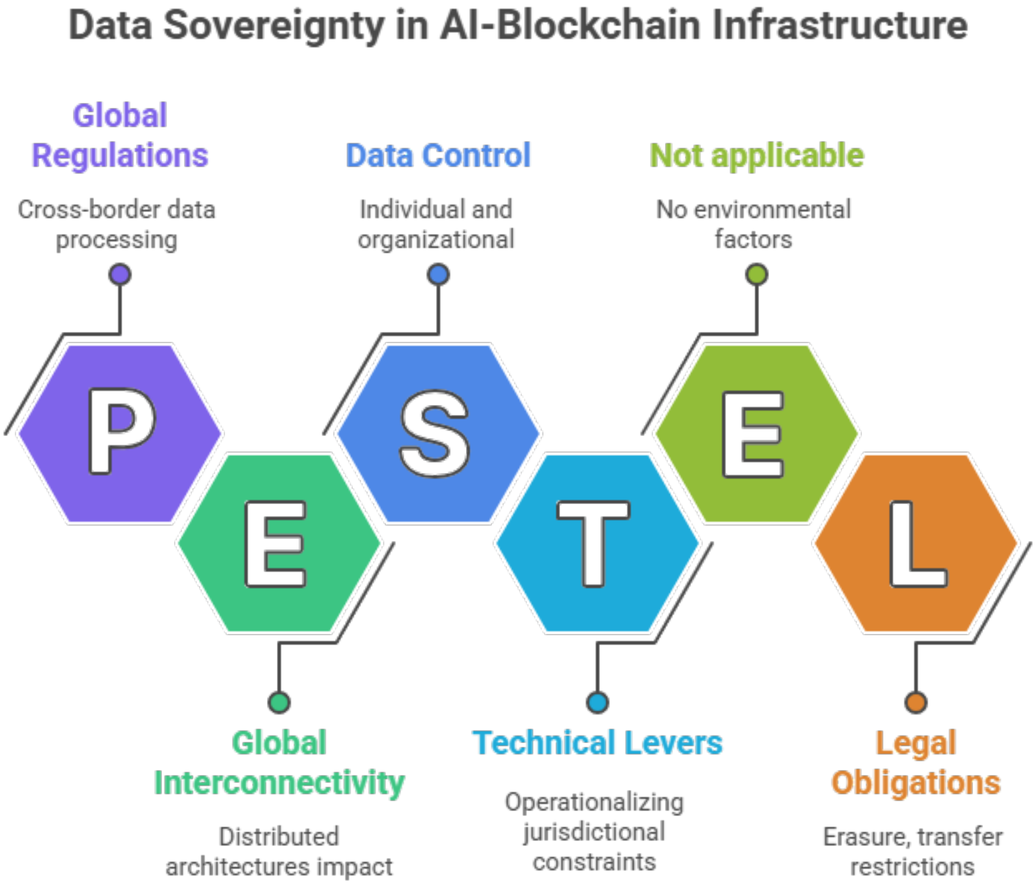
Data Sovereignty in AI-Blockchain Infrastructure
Global Regulations
Cross-border data processing
P
Data Control
Individual and organizational
S
Not applicable
No environmental factors
E
E
Global Interconnectivity
Distributed architectures impact
T
Technical Levers
Operationalizing jurisdictional constraints
L
Legal Obligations
Erasure, transfer restrictions

Figure-2.Data Sovereignty in AI-Blockchain Infrastructure

Data sovereignty in this context has three intersecting layers:

1. **Jurisdictional sovereignty:** Which legal regime governs which data and processing action (collection, training, inference, logging, sharing), and which transfer mechanisms are lawful? (e.g., GDPR, SCCs, EU–U.S. Data Privacy Framework, CLOUD Act, DPDP Act, PIPL).
2. **Organizational sovereignty:** How enterprises assert policy control over where data rests, who can access it, and how derived artifacts (models, embeddings, proofs) flow across cloud regions and validator sets (e.g., sovereign cloud patterns, Gaia-X).
3. **Individual sovereignty:** How data principals retain meaningful agency via consent, redress, portability, and erasure—especially when records are replicated across ledgers and model pipelines. The tension with blockchain immutability is explicit in EU guidance.

The question is not whether to choose privacy or decentralization, but **how to make them cohere**—with governance engineered into the stack, rather than bolted on later.

LITERATURE REVIEW

1) Regulatory pillars for cross-border AI-blockchain

GDPR sets core principles, including the right to erasure (Article 17) and constraints on transfers to third countries absent adequate safeguards. Schrems II (CJEU, 2020) invalidated the Privacy Shield and tightened expectations on supplemental measures when relying on Standard Contractual Clauses (SCCs). In 2023, the European Commission adopted an adequacy decision for the EU–U.S. Data Privacy Framework (DPF) for certified organizations, restoring a transfer pathway, alongside updated 2021 SCCs.

The EU AI Act (Regulation 2024/1689), published in the Official Journal on 12 July 2024, introduces risk-tier obligations for AI, with GPAI transparency and copyright-related duties beginning to apply from August 2, 2025, and additional requirements phasing in later for models with systemic risk. These dates matter for platform roadmaps that combine model services with ledger-backed provenance.

Beyond the GDPR/AI Act, the EU's Data Governance Act (DGA) and Data Act aim to create trusted mechanisms for data sharing, re-use, and access, while addressing cloud switching and interoperability—key issues for multi-region deployments.

Outside the EU, cross-border obligations and state access rules shape architecture: the U.S. CLOUD Act clarifies law-enforcement access and bilateral executive agreements; India's Digital Personal Data Protection (DPDP) Act, 2023 sets national rules with evolving transfer conditions; and China's PIPL asserts extraterritorial reach for handling Chinese personal information outside China.

Complementing binding law are soft-law frameworks and cooperative instruments: the OECD AI Principles (adopted 2019; updated 2024) for trustworthy AI; the G7 Hiroshima Leaders' Communiqué endorsing Data Free Flow with Trust (DFFT); and the Global CBPR Forum (2025 launch of certifications), which seeks interoperability via accountability-based certifications for data flows.

2) Guidance at the AI–blockchain interface

European regulators have explicitly addressed blockchain under the GDPR. In EDPB Guidelines 02/2025, the Board underscores that encryption does not absolve GDPR duties and that persistent on-chain personal data aggravate retention and erasure risks, thus favoring designs that keep personal data off-chain with on-chain references/hashes where possible.

GPAI obligations under the EU AI Act (e.g., transparency, documentation, risk controls) interact with data sovereignty because model developers must evidence lawful sourcing and risk governance, potentially requiring provenance records, dataset governance, and territorial controls tied to model cards and evaluations.

3) Technical and architectural literature

From the ledger side, ISO 22739:2020 standardizes blockchain/DLT terminology; permissioned frameworks such as Hyperledger Fabric implement private data collections to restrict data dissemination to authorized organizations, facilitating jurisdiction-aware channeling of sensitive data.

From the AI governance side, the NIST AI Risk Management Framework (AI RMF 1.0) provides a widely used scaffold for mapping AI risks and controls—privacy, security, explainability, and accountability—across the AI lifecycle, adaptable to jurisdiction-specific duties.

Sovereign cloud and federated data spaces initiatives such as Gaia-X seek to preserve data control through federated identity, policy, and interoperability labels—useful for binding regional policy to infrastructure while enabling cross-provider ecosystems.

METHODOLOGY

We propose **SOVEREIGN-Stack**, a governance-by-design methodology that embeds data sovereignty across eight layers. Rather than treating compliance as a post-hoc gate, SOVEREIGN integrates legal, organizational, and technical controls into the architecture.

Layer 1 — Identity & Role Binding (Who?)

- Establish decentralized, attestable identities for organizations and services; bind **roles** (controller/processor, data fiduciary/processor equivalents) to infrastructure components.
- Associate every node/operator with a **jurisdiction profile** (e.g., EU, U.S., India, China) and attach a **data processing role** to smart-contract functions and AI services.
- Use **short-lived credentials** and policy-bound keys so that revocation and key rotation mirror consent and purpose changes.

Layer 2 — Consent & Legitimate Basis (Why?)

- Capture legal bases (consent, contractual necessity, legitimate uses, etc.) and purpose tags as machine-readable **policy claims**.
- Store **consent artifacts** off-chain with on-chain receipts; ensure consent scope/version is verifiable during training or inference calls.
- Implement **data-use ontologies** (e.g., "diagnosis," "fraud detection," "route optimization") aligned to allowable purposes per jurisdiction.

Layer 3 — Data Classification & Minimization (What?)

- Classify inputs, features, and derived artifacts (embeddings, gradients, model cards, prompts, logs) across **sensitivity tiers** (personal, special category, trade secret, critical infrastructure).

- Generate **data lineage graphs** linking each artifact to source datasets, consents, transfer mechanism, and jurisdictional constraints.

Layer 4 — Locality, Routing & Storage (Where?)

- Adopt **geo-fenced storage** and **region-bound processing** for governed data; replicate only **non-sensitive** proofs/metadata across ledgers.
- Use **on-chain/off-chain partitioning**: keep personal data **off-chain**, commit only salted hashes or tokens; use **private data collections** or permissioned channels for sensitive transactions.
- Bind node placement to **legal risk maps** (e.g., EU nodes for EU-personal data; APAC nodes for APAC-restricted datasets).

Layer 5 — Compute & Model Governance (How?)

- Choose **PETs** (federated learning, secure enclaves, MPC, differential privacy) as fit-for-purpose controls to keep personal data in-region while sharing computed insights/gradients.
- Implement **dataset governance**: lawful sourcing records, data quality checks, representativeness notes, and redress pathways mapped to **AI RMF** functions (govern, map, measure, manage).
- Maintain **model cards** with training data categories, jurisdictions used, transfer mechanisms, and risk mitigations (e.g., synthetic data in non-adequate destinations).

Layer 6 — Ledger Governance & Lifecycle (When/How long?)

- Make **data retention** an explicit ledger policy: define purge/expiry semantics for off-chain stores; ensure **cryptographic tombstones** record de-linking/erasure events.
- Use **selective disclosure** and **zero-knowledge attestation** for compliance proofs (e.g., “this node stores no EU personal data” without revealing raw data).

Layer 7 — Cross-Border Transfer Mechanisms (Under what legal cover?)

- Encode which mechanism applies to each flow (e.g., **SCCs**, **EU–U.S. DPF** certification ID, **Global CBPR** certification, **intra-group agreements**). Attach mechanism IDs to the transaction or model deployment metadata.
- Build **automated flow blockers**: if the receiving node lacks appropriate safeguards or certification, the pipeline halts or routes to an alternative.
- For U.S. providers, track **CLOUD Act** exposure states and conflict-of-law escalation routes.

Layer 8 — Assurance, Audit & Redress (Prove it.)

- Offer **verifiable logs** and **regulatory-grade dashboards**: proofs of region-bound processing, consent checks, data minimization, and transfer controls.
- Support **Data Subject Request** workflows (access, erasure) by ensuring references on-chain can be irreversibly de-linked from off-chain personal data, honoring **Article 17** constraints.

Process Phases:

1. **Jurisdiction & Risk Mapping** (laws, sector rules, supervisory expectations);
2. **Architecture Selection** (permissioned vs. permissionless overlays, off-chain stores, sovereign cloud);
3. **Control Design** (PETs, policy engines, identity/role bindings);
4. **Compliance Simulation** (what-if tests under Schrems II-style scrutiny);
5. **Pilot & Evaluate** (KPIs: transfer compliance rate, erasure latency, audit completeness, model-data traceability);
6. **Scale with Monitoring** (control drift detection, obligations calendar for AI Act/DPA milestones).

RESULTS

Vignette A: Cross-regional health analytics (EU → U.S. → India)

Scenario. A research consortium trains diagnostic models on EU hospital data, operates inference in the U.S. (for specialized GPU clusters), and conducts post-market monitoring with clinics in India. The consortium must respect **GDPR** across training, ensure lawful **EU→U.S.** transfer, and support India-based services under the **DPDP Act**.

SOVEREIGN-Stack application

- **Locality:** All EU personal data remain in EU sovereign cloud stores; EU model training uses **federated learning** so raw data never leave EU regions; only **aggregated updates** and **privacy-screened model weights** transit.
- **Transfer mechanism:** The U.S. inference service is either **DPF-certified** or covered by **SCCs** with supplemental controls; metadata contains the mechanism ID and a pointer to a DPIA summary.
- **Erasure & purpose:** Off-chain patient records are keyed to revocable tokens; on erasure request, tokens are burned and off-chain data deleted; inference caches are purged according to EU retention rules; on-chain artifacts contain only salted digests.
- **India leg:** For services delivered in India, the **DPDP Act** duties (notice, lawful purpose, children's data safeguards) are captured in policy claims; if the flow would move EU personal data to India, automated blockers enforce EU transfer rules, routing instead through EU nodes.

Observed outcomes (qualitative)

- **Compliance coverage** improves: transfer flows are provably tied to lawful mechanisms; erasure is technically enforceable via token de-linking.
- **Risk reduction:** Federated updates minimize data export risk; auditability increases via on-chain receipts and off-chain evidence folders.
- **Performance trade-offs:** Some latency added by geo-fencing and PET orchestration; mitigated by regional edge inference for EU users.

Vignette B: Permissioned supply-chain ledger (EU ↔ APAC)

Scenario

A consortium of manufacturers and logistics providers operates a permissioned ledger to trace parts provenance. Participants span EU and APAC (Japan, Singapore). The system aggregates IoT telemetry and operator attestations and occasionally feeds an AI model that predicts defect risk.

SOVEREIGN-Stack application

- **Private data collections:** Sensitive business data (supplier bids, employee IDs) reside in **Fabric** private collections—shared only among entitled organizations, with purge policies. Public channel stores hash commitments to support dispute resolution.
- **Global CBPR alignment:** APAC participants seeking a common baseline adopt **Global CBPR** certification; the system attaches certification references to nodes, enabling **policy-aware routing** of telemetry and model features across borders.
- **Model governance:** The defect-risk model uses **region-filtered features**: EU personal data never feed APAC training runs; synthetic data or statistics may cross where lawful; model cards document sources and jurisdictions per **AI RMF** practice.

Observed outcomes (qualitative)

- **Interoperability with control:** Partners exchange verifiable proofs without exposing raw confidential or personal data.
- **Regulatory readiness:** DSR handling is feasible (IDs tokenized); retention enforced via purgeable off-chain stores; DGA/Data Act-style sharing mechanisms are easier to implement downstream.

DISCUSSION

A common objection is that **blockchain immutability conflicts with erasure/rectification**. The rapidly maturing guidance suggests a practical reconciliation: keep personal data **off-chain**, record only **non-personal proofs** on-chain, and ensure that off-chain stores implement **cryptographic erasure** and **token de-linking**. When a data subject requests erasure, controllers delete off-chain records and render the on-chain pointer useless (e.g., via key destruction or revocation), thereby complying with Article 17 while keeping the integrity of the ledger's audit trail intact.

Cross-border is similar: encode the **legal basis and transfer mechanism** into the pipeline so that routing decisions are automatic; block or re-route flows to nodes lacking **SCCs**, **adequacy**, **DPF certification**, or **CBPR/PRP** status. This **policy-in-the-loop** design transforms compliance from a manual legal review into a continuous, verifiable control.

CONCLUSION

Global AI-blockchain infrastructure does not have to choose between sovereignty and scale. By shifting from **application-level compliance** to **architecture-level governance**, organizations can pre-wire jurisdictional logic into identity, consent, data classification, locality, compute, and ledger layers—turning laws and standards into machine-enforceable constraints and auditable evidence. The **SOVEREIGN-Stack** methodology offered here provides a structured pathway:

- map jurisdictions and roles;
- partition data and computation with **off-chain first** defaults and **permissioned** overlays;
- attach legal mechanisms (SCCs, DPF, CBPR) to flows;
- operationalize AI governance using **NIST AI RMF**; and
- expose verifiable controls for regulators and partners.

Regulatory momentum (EU AI Act milestones, DGA/Data Act operationalization, evolving transfer frameworks like DPF and Global CBPR) will continue to shape design choices. Teams that treat sovereignty as a **design affordance**—not a constraint—can achieve trustworthy AI, resilient ledgers, and interoperable ecosystems that earn the legal and social license to operate across borders

REFERENCES

- *European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). EUR-Lex.* *https://eur-lex.europa.eu/eli/reg/2024/1689/oj*
- *European Commission. (2025, August 1). EU rules on general-purpose AI models start to apply, bringing more transparency, safety and accountability.* *https://digital-strategy.ec.europa.eu*
- *National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1).* *https://doi.org/10.6028/NIST.AI.100-1*
- *European Union. (2016). Regulation (EU) 2016/679 (General Data Protection Regulation). EUR-Lex.* *https://eur-lex.europa.eu/eli/reg/2016/679/oj*
- *Court of Justice of the European Union. (2020, July 16). Press Release No 91/20: Data Protection Commissioner v Facebook Ireland and Maximillian Schrems (Schrems II).* *https://curia.europa.eu*
- *European Commission. (2021, June 4). Implementing Decision (EU) 2021/914 on standard contractual clauses for the transfer of personal data to third countries. EUR-Lex.* *https://eur-lex.europa.eu/eli/dec_impl/2021/914/oj*
- *European Commission. (2023, July 10). Implementing Decision (EU) 2023/1795 on the adequate level of protection under the EU–U.S. Data Privacy Framework. EUR-Lex.*
- *U.S. Department of Justice. (n.d.). CLOUD Act resources.* *https://www.justice.gov/criminal/cloud-act-resources*
- *Ministry of Electronics and Information Technology (India). (2023). Digital Personal Data Protection Act, 2023.* *https://www.meity.gov.in*
- *Stanford DigiChina. (2021). Translation: Personal Information Protection Law (PIPL) of the People's Republic of China.* *https://digichina.stanford.edu*

- *OECD. (2019/2024). OECD AI Principles. https://oecd.ai/en/ai-principles*
- *G7. (2023, May 20). Hiroshima Leaders' Communiqué (DFFT). Council of the EU. https://www.consilium.europa.eu*
- *Global CBPR Forum. (2025). Global CBPR Forum—Building digital trust through partnerships. https://www.globalcbpr.org*
- *European Union. (2022). Regulation (EU) 2022/868 (Data Governance Act). EUR-Lex. https://eur-lex.europa.eu/eli/reg/2022/868/oj*
- *European Union. (2023). Regulation (EU) 2023/2854 (Data Act). EUR-Lex. https://eur-lex.europa.eu/eli/reg/2023/2854/oj*
- *Hyperledger Fabric. (n.d.). Private data collection definition. https://hyperledger-fabric.readthedocs.io*
- *International Organization for Standardization. (2020). ISO 22739:2020—Blockchain and distributed ledger technologies—Vocabulary. https://www.iso.org/obp/ui*
- *European Data Protection Board. (2025, April 8). Guidelines 02/2025 on processing of personal data through blockchain technologies. https://www.edpb.europa.eu*
- *Gaia-X AISBL. (n.d.). About Gaia-X: A federated secure data infrastructure. https://gaia-x.eu/about/*
- *European Commission. (2025, July 10). The General-Purpose AI Code of Practice (GPAI). https://digital-strategy.ec.europa.eu*