MolTrust — EU AI Act Mapping

Article-by-Article Technical Mapping for Agentic AI Trust Infrastructure

Version 1.0 · April 2026

Lars Kersten Kroehl

MolTrust / CryptoKRI GmbH, Zurich

lars@moltrust.ch ·moltrust.ch

1. Scope and Positioning

This document maps the MolTrust Protocol — a W3C Verifiable Credentials and Decentralized Identifier trust infrastructure for autonomous AI agents — against the obligations of Regulation (EU) 2024/1689 (the EU Artificial Intelligence Act).

What this document is: a *technical mapping* showing which articles of the EU AI Act are operationally supported by MolTrust primitives, and through which evidence mechanism. It is produced as a reference for technology evaluators, enterprise architects, and regulatory analysts who need a concrete view of where standards-based agent trust infrastructure meets specific regulatory obligations.

What this document is not: it is not a conformity assessment, not a legal compliance statement, and not a substitute for formal legal review by counsel qualified in EU regulatory matters. MolTrust is a technical infrastructure layer; compliance is an organizational process that requires policy, documentation, and human accountability that no infrastructure layer can provide alone.

Scope of the mapping: the document focuses on the articles most directly applicable to autonomous AI agents deployed in high-risk domains (Annex III systems) and on the General-Purpose AI (GPAI) obligations that apply to the underlying models such agents commonly use. Articles 9 through 15 define the core obligations for high-risk AI system providers and deployers; Articles 43 and Annex V define the conformity assessment and declaration process; Articles 50 through 55 define GPAI obligations. These are the articles mapped in Section 3 and Section 4.

Enforcement timeline: the EU AI Act entered into force on 1 August 2024. Prohibited-practices obligations (Article 5) took effect 2 February 2025. GPAI obligations took effect 2 August 2025. Obligations on high-risk AI systems take effect 2 August 2026, with some provisions on high-risk systems already embedded in other regulated products applying from 2 August 2027. This document addresses obligations that are currently in force as well as those taking effect within the next four months.

Reference version: this mapping is against the consolidated text of Regulation (EU) 2024/1689 as published in the Official Journal of the European Union on 12 July 2024.

---

2. Methodology

For each relevant article, the mapping is presented in the following form:

- **Article reference** and short title
- **Core obligation** (summary of the requirement as it applies to agentic AI systems)
- **MolTrust primitive** that operationally supports the obligation
- **Evidence pointer** (specific API endpoint, specification section, or on-chain artifact that demonstrates the implementation)
- **Status**: *live* (deployed and verifiable as of the version date above), *roadmap* (planned for delivery within a stated timeframe), or *gap* (explicitly outside the scope of the MolTrust layer and the responsibility of the deployer)

The mapping is deliberately conservative. Where the article imposes a process or organizational obligation (for example, staff training or internal risk management processes), MolTrust is marked as *gap* because those obligations cannot be met by an infrastructure layer — they require deployer-level commitments. Overstating coverage would undermine the purpose of the document.

Where MolTrust provides partial support, the limitation is stated explicitly. This is consistent with the MolTrust design principle that claims about the protocol should be independently verifiable by any party against the live endpoint at `api.moltrust.ch` and the public reference implementation.

---

3. Main Mapping — High-Risk AI System Obligations

Article 9 — Risk Management System

Core obligation: providers of high-risk AI systems must establish, implement, document, and maintain a risk management system. The system must identify reasonably foreseeable risks, estimate and evaluate risks that may emerge under intended use or misuse, and adopt appropriate risk management measures.

Dimension	MolTrust response
Risk identification	Agent Authorization Envelope (AAE) CONSTRAINTS block enumerates operational risk dimensions explicitly: financial thresholds, jurisdictional restrictions, counterparty minimum trust score, time boundaries
Risk estimation	Trust Score (0–100, A–F grade) provides a continuous behavioral-risk signal derived from interaction history, endorsement graph, and violation records

Dimension	MolTrust response
Risk mitigation	AAE default-deny semantics: any action not explicitly permitted is denied. Three-layer enforcement (cryptographic, API, kernel) ensures mitigation is not bypassable from the agent's own runtime
Risk documentation	On-chain anchoring of AAE and Interaction Proof Record hashes on Base Layer 2 produces an immutable, timestamped risk audit trail

Evidence pointer: Agent Authorization Envelope specification in MolTrust Technical Specification v0.8 Section 3; live endpoint `POST /aae/validate`; reference implementation `@moltrust/aae` v1.0.0 on npm.

Status: *live*. MolTrust provides the technical instrumentation for risk management; the risk management *process* (periodic review, risk register maintenance, mitigation planning) is a deployer responsibility.

Article 10 — Data and Data Governance

Core obligation: training, validation, and testing data sets used for high-risk AI systems must meet quality criteria including relevance, representativeness, freedom from errors, and statistical properties appropriate to the intended purpose.

Dimension	MolTrust response
Data provenance	`ProductProvenanceCredential` vertical supports cryptographic chain-of-custody for datasets where data sourcing requires verification
Data quality attestation	Verifiable Credentials can attest to data quality properties (source, collection methodology, bias assessment) issued by the data provider and cryptographically verifiable by any consumer
Data governance	`data_read` and `data_write` purpose categories in the AAE MANDATE block allow fine-grained authorization of agent data access; `deniedActions` takes precedence over allowed actions

Evidence pointer: Verifiable Credential verticals in MolTrust Technical Specification v0.8 Section 5; live endpoint `POST /vc/issue` supports custom credential types for data provenance.

Status: *partial live, gap on training-data governance*. MolTrust supports runtime data access authorization. It does not address training-data quality, bias assessment, or dataset documentation, which are model-lifecycle obligations upstream of agent deployment. These remain the responsibility of the AI system provider.

Article 11 — Technical Documentation

Core obligation: technical documentation must be drawn up before the high-risk AI system is placed on the market and kept up to date. Annex IV specifies minimum content: general description, detailed description of elements and development process, monitoring system, description of risk management system, and relevant standards applied.

Dimension	MolTrust response
Architectural documentation	MolTrust Protocol Technical Specification v0.8 (SHA-256 anchored on Base L2, Block 44745864) provides a versioned, immutable reference for the protocol architecture
Standards documentation	Conformance specification (CONFORMANCE.md v1.0) enumerates five reproducible test vectors against W3C VCs, W3C DIDs, Ed25519 per Ed25519Signature2020, RFC 8785 canonical JSON, and RFC 9396 rich authorization requests
Risk management documentation	See Article 9 above
Research-grade documentation	MolTrust arXiv Preprint v1.0 (SHA-256 c9c34985…1c11d946, anchored Block 45,037,732, tagged MolTrust/arXiv/v1.0) provides peer-review-quality technical documentation of the protocol architecture and empirical deployment evidence

Evidence pointer: all documents are publicly accessible and independently verifiable through their on-chain anchors on Base L2 (chainId 8453).

Status: *live*. The MolTrust infrastructure layer produces documentation artifacts that are cryptographically verifiable and version-pinned — a property not typically available for proprietary technical documentation. Deployers are responsible for integrating MolTrust-specific documentation into their overall Annex IV technical dossier.

Article 12 — Record-Keeping

Core obligation: high-risk AI systems must technically allow the automatic recording of events over the duration of their lifetime. These logs must ensure traceability of the system's functioning appropriate to the intended purpose and allow monitoring for risks or substantial modifications.

Dimension	MolTrust response
Interaction recording	Interaction Proof Records (IPR) capture every recorded agent-to-agent or agent-to-counterparty interaction with dual Ed25519 signatures and Merkle batch anchoring on Base L2
Tamper-evidence	IPR hashes are anchored on-chain within 60 seconds of creation; any modification to the record is cryptographically detectable by any verifier
Long-term traceability	On-chain anchors persist indefinitely; off-chain storage is supported through IPFS dual-write (Pinata integration) with cryptographic integrity tied to the on-chain root
Violation records	Principal-DID-linked Violation Records persist across agent re-registrations, preventing agents from shedding behavioral history through DID rotation

Evidence pointer: IPR specification in MolTrust Technical Specification v0.8 Section 5; live endpoint `POST /vc/ipr`; anchor verification through `GET /ipr/{id}/anchor` returning Base L2 TX hash for independent verification.

Status: *live*. This is one of the strongest points of direct alignment between MolTrust primitives and Article 12. The immutability property provided by on-chain anchoring exceeds what is achievable through file-based logging alone.

Article 13 — Transparency and Information to Deployers

Core obligation: high-risk AI systems must be designed and developed in a sufficiently transparent manner to enable deployers to interpret the system's output and use it appropriately. Instructions for use must include specified information about the provider, intended purpose, accuracy, robustness, and cybersecurity.

Dimension	MolTrust response
Agent identity transparency	DID Document at `api.moltrust.ch/.well-known/did.json` exposes the reference registry's identity metadata in W3C-standardized form; every registered agent has a resolvable DID Document

Dimension	MolTrust response
Capability transparency	`AgentCard` format (A2A-compatible) at `/.well-known/agent-card.json` describes each agent's capabilities, authentication method, and contact information
Authorization transparency	AAE is machine-readable and human-inspectable; the MANDATE block enumerates permitted actions explicitly rather than through opaque ACLs
Trust Score transparency	Trust Score is computed from a publicly documented formula (Section 3.1 of arXiv Preprint v1.0); endorsement graph is queryable for inspection

Evidence pointer: live endpoints `GET /.well-known/did.json`, `GET /.well-known/agent-card.json`, `GET /skill/trust-score/{did}`; Trust Score formula documented in Technical Specification v0.8 Section 4 and arXiv Preprint v1.0 Section 3.1.

Status: *live*. Transparency of the *agent layer* is directly supported. Transparency of the *model layer* (explainability of LLM outputs) is outside MolTrust's scope and remains a model-provider responsibility.

Article 14 — Human Oversight

Core obligation: high-risk AI systems must be designed such that they can be effectively overseen by natural persons during the period in which they are in use. Oversight measures must enable persons to understand the capacities and limitations of the system, detect anomalies, and intervene by interrupting the system or overriding its output.

This article is particularly relevant for autonomous agent systems, where the oversight boundary is explicitly where human authority begins and agent autonomy ends. MolTrust provides several primitives that operationalize this boundary cryptographically rather than through policy alone.

Approval thresholds as human oversight gates. The AAE CONSTRAINTS block defines three financial thresholds per agent: an autonomous threshold below which the agent may act without additional human check, a step-up threshold that triggers additional verification, and an approval threshold above which human sign-off is required. Crossing the approval threshold without a corresponding human signature produces a verifiable policy violation recorded as a Violation Record. This means the oversight boundary is not advisory but *enforced at the infrastructure level* — an agent attempting to cross an approval threshold without human authorization produces a permanent, on-chain-anchored violation event regardless of whether the agent's own runtime reports it.

Delegation attenuation. The AAE delegation rules require that any sub-authorization granted by an agent must be a strict subset of the parent authorization (attenuation-only). An agent cannot expand its own authority through delegation, and cannot authorize a sub-agent to perform

actions it is not itself permitted to perform. This prevents a class of escalation attacks where an agent attempts to circumvent human oversight by delegating to a less-constrained sub-agent.

Time-bounded authorization. No AAE may be open-ended. The `expiresAt` field is mandatory. This ensures that human oversight does not need to be continuously re-established; instead, authorization naturally requires human renewal on a defined cadence, which is exactly the pattern Article 14 contemplates.

Interruption and override. Credential revocation via the `revocationEndpoint` (CAEP-compatible) propagates to verifiers within 60 seconds in the reference implementation. The verifier-side default behavior on unreachable revocation endpoints is fail-closed (credential denied). This provides the interruption and override mechanism required by Article 14 with a response time suitable for operational intervention.

Dimension	MolTrust response
Oversight gate	Financial threshold mechanism in AAE CONSTRAINTS
Escalation prevention	Attenuation-only delegation in AAE MANDATE
Time-bounded authority	Mandatory expiry in AAE VALIDITY
Intervention capability	CAEP-compatible revocation with 60-second propagation
Violation record	Principal-DID-linked persistent violation history

Evidence pointer: AAE specification in MolTrust Technical Specification v0.8 Section 3; arXiv Preprint v1.0 Section 3.3 and Figure 2; live endpoint `POST /aae/validate`; revocation endpoint pattern documented in Section 5 of the specification.

Status: *live*. Article 14 is the article where MolTrust's AAE structure is most directly and operationally supportive.

Article 15 — Accuracy, Robustness, and Cybersecurity

Core obligation: high-risk AI systems must be designed to achieve appropriate levels of accuracy, robustness, and cybersecurity, performing consistently throughout their lifecycle. This includes resilience against errors, faults, or inconsistencies, and against attempts by unauthorized third parties to alter use, outputs, or performance by exploiting vulnerabilities.

Cybersecurity through Skill Audit. The MolTrust Skill Audit infrastructure defines nine conformance checks that any agent registering in the trust registry must pass. Checks include:

- **Secrets scanning** (hard fail, mapped to CWE-798 "Use of Hard-coded Credentials"): agents whose disclosed source material or configuration contains secret keys are rejected from registration
- **A2A discovery scan**: validates that advertised A2A endpoints actually respond as claimed
- Six additional checks covering endpoint reachability, signature validity, DID resolution, and authorization endpoint correctness, each with explicit CWE mapping where applicable

Robustness through three-layer enforcement. As described in the arXiv Preprint v1.0 Section 3.4 (Figure 3), MolTrust implements enforcement at three layers: cryptographic (Ed25519 signatures, RFC 8785 canonical JSON), API (registry-level policy checks), and kernel (Falco eBPF syscall monitoring). The three-layer design ensures that an exploit at any single layer does not result in a complete bypass of agent constraints.

Accuracy through Interaction Proof Records. IPRs capture the declared output of an agent at the time of the interaction, cryptographically bound to that specific time and the specific counterparty. This prevents retrospective adjustment of agent claims — an agent cannot later claim to have made a different prediction or provided different advice than what was signed at the time.

Resilience to unauthorized modification. AAE tampering is detectable by any verifier through Ed25519 signature verification. The signature covers the entire envelope in RFC 8785 canonical form; any modification produces signature failure. This is a construction-level property, not a policy enforcement.

Dimension	MolTrust response
Accuracy	IPR-based declared-output binding prevents retrospective adjustment
Robustness (fault tolerance)	Three-layer enforcement; no single-layer failure produces full bypass
Cybersecurity (secrets)	Skill Audit CWE-798 hard fail
Cybersecurity (endpoint integrity)	Skill Audit endpoint reachability and signature validity checks
Tamper resistance	Ed25519 signature over RFC 8785 canonical JSON

Evidence pointer: Skill Audit specification in Technical Specification v0.8 Section 6; conformance check registry at `api.moltrust.ch/guard/audit/checks`; CONFORMANCE.md v1.0 (SHA-256 anchored); live endpoint `POST /guard/validate-capabilities`.

Status: *live*. Article 15 is directly supported through a combination of runtime checks and construction-level cryptographic properties.

Article 43 — Conformity Assessment

Core obligation: high-risk AI systems must undergo a conformity assessment procedure before being placed on the market or put into service. For systems covered by specific product regulations, the assessment follows the applicable sectoral rules. For other high-risk systems under Annex III, Article 43 specifies the procedure, including third-party conformity assessment bodies where required.

Honest positioning: MolTrust is not a conformity assessment body, and this document does not constitute a conformity assessment. MolTrust is an *evidence-producing infrastructure* that supports the preparation and conduct of a conformity assessment by the deployer or their chosen assessment body.

Dimension	MolTrust response
Evidence production	All of the above Article mappings produce verifiable evidence artifacts (on-chain anchors, VCs, IPRs, audit logs)
Reproducibility	CONFORMANCE.md v1.0 provides five test vectors that any third party can run against the live endpoint to independently verify protocol compliance
Independence	Verification does not require consulting the MolTrust registry — anchors are on a public blockchain, signatures are cryptographically self-verifying

Evidence pointer: CONFORMANCE.md v1.0 in MolTrust GitHub repository; Base L2 anchors (see above).

Status: *gap on conformity assessment process, live on evidence production*. The conformity assessment itself remains the responsibility of the deployer and their chosen assessment body. MolTrust provides the technical artifacts that such an assessment can evaluate against specific article obligations.

Annex V — EU Declaration of Conformity

Core obligation: providers of high-risk AI systems must draw up a written EU declaration of conformity for each system and keep it for ten years after placing on the market. The declaration must identify the AI system, indicate applicable Union harmonization legislation, and reference the conformity assessment procedure.

Dimension	MolTrust response
System identification	Agent DID provides cryptographically unique system identifier that cannot be falsified
Long-term availability	On-chain anchoring of credential and AAE hashes persists well beyond the ten-year retention requirement
Document integrity	SHA-256 anchoring allows the provider to demonstrate that a declaration of conformity document has not been modified since issuance

Evidence pointer: the mechanism used to anchor the MolTrust arXiv Preprint v1.0 (see Section 2 above) is equally available for declarations of conformity.

Status: *supportive, not substitutive*. MolTrust provides mechanisms that can be used to strengthen the integrity of a declaration of conformity; the content of the declaration and its submission to authorities remain deployer responsibilities.

4. General-Purpose AI Obligations (Articles 50–55)

The EU AI Act defines additional obligations for General-Purpose AI (GPAI) models — the foundation models that many autonomous agents rely on. GPAI-specific obligations took effect on 2 August 2025, with additional obligations for GPAI models with systemic risk following in 2026.

MolTrust does not produce or distribute GPAI models. However, agents built on top of GPAI models inherit certain transparency and documentation obligations that the agent-level infrastructure can partially support.

Article 50 — Transparency. Agents that interact with humans must make clear that the interaction is with an AI system. The MolTrust AgentCard (A2A-compatible, served at `/.well-known/agent-card.json`) provides a machine-readable declaration of agent identity and nature that can be consumed by any interacting party to fulfill this disclosure obligation.

Article 53 — Provider Obligations for GPAI Models. GPAI providers must draw up and keep technical documentation, make available information for downstream providers, and put in place a copyright policy. For agents deployed on GPAI models, the model-provider obligations remain upstream. MolTrust's role is to make the agent's binding to its underlying model (via Developer DID and capability declarations) cryptographically verifiable, which supports the downstream-provider information requirement of Article 53.

Article 55 — GPAI Models with Systemic Risk. Additional obligations apply to GPAI models classified as having systemic risk, including incident reporting to the AI Office. MolTrust's Violation Record mechanism — principal-DID-linked and on-chain anchored — provides a credible evidence base for agent-level incident identification, which can feed into the incident reporting process the deployer must maintain.

Status for Section 4 as a whole: *supportive*. MolTrust provides agent-level primitives that integrate with GPAI obligations; the GPAI model obligations themselves are the responsibility of the model provider.

5. Honest Gaps and Deployer Responsibilities

This section explicitly enumerates the EU AI Act obligations that MolTrust does *not* address. Acknowledging these gaps is essential: MolTrust is an infrastructure layer, not a compliance solution, and any claim to the contrary would be a misrepresentation.

Fundamental Rights Impact Assessment (Article 27). Deployers of high-risk AI systems in specified public services must carry out a Fundamental Rights Impact Assessment (FRIA) before deployment. FRIA is an organizational assessment process that cannot be replaced by infrastructure evidence. MolTrust provides evidence that can inform the technical portions of the assessment; the assessment itself is a deployer obligation.

Market surveillance reporting (Articles 73–79). Deployers and providers have obligations to report serious incidents and malfunctioning to national competent authorities. MolTrust provides

the mechanism to detect violations (Violation Records) but does not itself report to authorities. The reporting workflow, including jurisdictional determination and authority coordination, is a deployer responsibility.

Post-market monitoring (Article 72). Providers must establish and document a post-market monitoring system proportionate to the nature of the AI system and its intended purpose. MolTrust's IPR infrastructure provides the observational data foundation for such monitoring; establishing the monitoring system, including incident categorization, trend analysis, and remediation workflows, is a deployer responsibility.

Human resources and training obligations. Several articles impose training and awareness obligations on staff who interact with high-risk AI systems. These are organizational commitments that are entirely outside the scope of any infrastructure layer.

Formal conformity assessment. As noted in the Article 43 mapping, MolTrust produces evidence; it does not conduct conformity assessment.

Recommendation: any deployer using MolTrust for EU AI Act compliance should treat this mapping as documenting the *infrastructure layer* of their compliance posture. A complete compliance documentation set additionally requires organizational policy documentation, staff training records, risk management process documentation, post-market monitoring process documentation, and the relevant conformity assessment outputs.

---

6. References

Primary regulatory texts:

- Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). Official Journal of the European Union, L 1689, 12 July 2024.

Parallel regulatory frameworks:

- National Institute of Standards and Technology (NIST), AI Risk Management Framework (AI RMF 1.0), NIST AI 100-1, January 2023. See MolTrust NIST AI RMF Function Mapping (companion document, same roadmap delivery).
- Infocomm Media Development Authority (IMDA) of Singapore, Model AI Governance Framework for Agentic AI, published at the World Economic Forum, 22 January 2026.

MolTrust references:

- MolTrust Protocol Technical Specification v0.8, SHA-256 anchored on Base L2 Mainnet, Block 44745864.
- MolTrust Protocol Whitepaper v0.8, SHA-256 anchored on Base L2 Mainnet, Block 44507827.
- MolTrust arXiv Preprint v1.0, SHA-256 `c9c349852dbc77b80c4d8d3b0f9e3db60244a2bc60345b6fc43de1dc1` anchored on Base L2 Mainnet, TX `0x49a548346f12ee0a273cf9d4d60f00685777d88050df06ef77d92caaff`

Block 45,037,732, tag `MolTrust/arXiv/v1.0`.
- CONFORMANCE.md v1.0 (MolTrust GitHub repository).

Technical standards:

- W3C Decentralized Identifiers (DIDs) v1.0, W3C Recommendation, 19 July 2022.
- W3C Verifiable Credentials Data Model v2.0, W3C Recommendation, 15 May 2025.
- RFC 8785, JavaScript Object Notation (JSON) Canonicalization Scheme (JCS). Internet Engineering Task Force, June 2020.
- RFC 9396, OAuth 2.0 Rich Authorization Requests. Internet Engineering Task Force, May 2023.
- Ed25519Signature2020, W3C Credentials Community Group.
- NIST SP 800-162, Guide to Attribute Based Access Control (ABAC) Definition and Considerations. National Institute of Standards and Technology, January 2014.

This document is part of a three-document roadmap delivered to Forrester Research in April 2026, alongside the MolTrust NIST AI RMF Function Mapping and the MolTrust Sybil-Resistance Methodology Note. All three documents are versioned, publicly accessible through the MolTrust GitHub repository, and SHA-256 anchored on Base Layer 2 for integrity verification.