The MolTrust Protocol

Technical Specification

Version 0.2.1 — Draft for Review

MolTrust / CryptoKRI GmbH, Zurich

March 2026

This document is a companion to *The MolTrust Protocol: A Verification Standard for Autonomous Software Agents* (Whitepaper v0.4). It provides the technical definitions, data models, verification flows, and conformance requirements referenced in that document. The specification is organized around three layers: Layer A (Protocol Standard — normative), Layer B (Reference Registry), and Layer C (Reference Reputation Model — informative).

SHA-256	d2ca9b372d9190bd5661ac1bb4c911881cfe0ef197aa43396cdc8fd27116a554
On-chain anchor	Base L2 (mainnet), Block 43691643
Transaction	0xc79a233fb19401b06ed870f27c9571b6a9e780ab6ba4fdb1d5e8a1bfdb17c972
Timestamp	2026-03-22T09:43:53 UTC

Table of Contents

1. Scope and Terminology
2. Data Model (Layer A)
3. Verification Flow (Layer A)
4. Reference Reputation Model (Layer C — Informative)
5. Reference Registry (Layer B)
6. On-Chain Anchoring (Layer B)
7. Credential Lifecycle (Layer A)
8. Agent Lifecycle (Layer A)
9. Threat Model
10. Privacy Model
11. Worked Example
12. Conformance

— 1 —

Scope and Terminology

Layer A — Protocol Standard	Normative core. Data formats, signing rules, verification flows, lifecycle semantics. Independent implementations conforming to Layer A interoperate at the evidence level.
Layer B — Reference Registry	MolTrust-operated service layer. Identity resolution, revocation, trust scores, anchoring. Other operators MAY run conformant registries.
Layer C — Reference Reputation Model	Informative scoring model. Other implementations MAY use different models provided they consume Layer A evidence formats.

1.2 Terminology

Key words MUST, MUST NOT, SHOULD, SHOULD NOT, MAY are per RFC 2119.

Agent	An autonomous software process that acts on behalf of a principal.
Principal	The human, organization, or agent on whose behalf an agent acts. SHOULD have a stable DID.
DID	Decentralized Identifier — W3C DID Core v1.0.
Verifiable Credential (VC)	Tamper-evident claim about a subject — W3C VC Data Model 2.0.
Interaction Proof	Signed artifact produced after an agent interaction, attesting to fact and outcome.
Trust Score	Numeric value [0–100] derived from behavioral record and endorsement graph.
Endorsement	Signed attestation that an agent has behaved reliably in a defined vertical.
Vertical	Domain of agent activity, expressed as <namespace>/<identifier>.
Seed Agent	Agent registered with bootstrap weight by operator to initialize the endorsement graph.
Violation Record	Signed artifact attesting to a confirmed protocol violation.
Registry	Service that resolves DIDs, maintains revocations, computes trust scores.

— 2 —

Data Model (Layer A)

2.1 Signed Payload Boundary

For all MolTrust-defined signed artifacts (Authorization Credentials, Interaction Proofs, Endorsements, Violation Records): the signed payload is the RFC 8785 canonical JSON serialization excluding the proof field(s). DID Documents are explicitly excluded — their integrity is guaranteed by the DID method's own resolution mechanism,

not by an embedded proof block. Signing algorithm: Ed25519Signature2020. Signature values: base58btc multibase encoding.

2.2 Agent DID Document

Mandatory fields: id, verificationMethod (Ed25519), authentication, assertionMethod.

```
{
"@context": ["https://www.w3.org/ns/did/v1", "https://moltrust.ch/ns/v1"],
"id": "did:moltrust:<uid>",
"verificationMethod": [{
"id": "did:moltrust:<uid>#keys-1",
"type": "Ed25519VerificationKey2020",
"controller": "did:moltrust:<uid>",
"publicKeyMultibase": "<base58btc-key>"
}],
"authentication": ["did:moltrust:<uid>#keys-1"],
"assertionMethod": ["did:moltrust:<uid>#keys-1"]
}
```

Optional: service, created, updated, controller (principal DID for sub-agents), alsoKnownAs (ERC-8004 ID).

Key rotation: add new key, update authentication/assertionMethod references, mark old key "revoked": true with revokedDate.

2.3 Authorization Credential

Mandatory: id (UUID v4), issuer, issuanceDate, expirationDate (max 730 days), credentialSubject with permittedActions and vertical, Ed25519 proof.

permittedActions: transact, delegate, endorse, verify, publish, * (all). Custom actions: <namespace>/<action>.

Delegation: sub-agent permittedActions MUST be subset of parent. Max 8 hops from root principal. Verifiers MUST reject chains exceeding 8 hops.

2.4 Interaction Proof

The primary evidence artifact. Produced after interactions both parties wish to record.


```
{
"@context": "https://moltrust.ch/ns/interaction/v1",
"type": "InteractionProof",
"id": "<uuid-v4>",
"session": "<session-id>",
"initiator": {"did": "did:moltrust:<id>", "vertical": "<ns>/<id>"},
"responder": {"did": "did:moltrust:<id>", "vertical": "<ns>/<id>"},
"timestamp": "2026-03-22T10:00:00Z",
"outcome": "completed",
"outcomeHash": "sha256:<hex>",
"proofInitiator": {"type": "Ed25519Signature2020", "proofValue": "<sig>"},
"proofResponder": {"type": "Ed25519Signature2020", "proofValue": "<sig>"}
}
```

outcome values: completed, partial, disputed, failed.

outcomeHash: SHA-256 of RFC 8785 canonical JSON of {proofId, timestamp, outcome, summary}. Raw transaction data MUST NOT be included. summary: free text, max 256 chars.

One-sided proofs: if responder unavailable after 300s timeout, submit with "singleSig": true. Valid but carries reduced score weight.

2.5 Vertical Identifiers

Format: <namespace>/<identifier> — alphanumeric, hyphens, underscores. Max 128 chars. Case-sensitive.

moltrust/ namespace is reserved. Defined values:

moltrust/travel	Travel booking and logistics
moltrust/commerce	Agentic commerce and purchasing
moltrust/prediction	Prediction markets and forecasting
moltrust/skill	Skill verification and auditing
moltrust/finance	Financial transactions and DeFi

moltrust/identity	Identity verification services
moltrust/general	General-purpose, cross-domain

Any party MAY define verticals in their own namespace. No registration required.

2.6 Endorsement

Mandatory: id, issuer, issuanceDate, expirationDate (max 365d, default 90d), credentialSubject with vertical/weight/basis/evidenceSummaryHash, Ed25519 proof.

basis values: interaction-proofs, delegation, operator. One active endorsement per (endorsed DID, vertical) pair.

2.7 Violation Record

```
{
"@context": "https://moltrust.ch/ns/violation/v1",
"type": "ViolationRecord",
"id": "<uuid-v4>",
"issuanceDate": "2026-03-22T00:00:00Z",
"subject": {"agentDid": "did:moltrust:<id>", "principalDid": "did:<method>:<id>"},
"violation": {"type": "<type>", "interactionProofId": "<uuid>", "description": "<max 512>"},
"adjudication": {"adjudicatorType": "external", "adjudicatorReference": "<ref>",
"confirmedAt": "2026-03-22T12:00:00Z"},
"registrySignature": {"type": "Ed25519Signature2020", "proofValue": "<sig>"}
}
```

violation types: identity-spoofing, authorization-abuse, sybil, behavioral-fraud, clone-impersonation.

Reversal: registry MUST record a ViolationReversal object containing: id, reversedRecordId, reversalDate, adjudicatorReference, registrySignature. Reversed records MUST NOT contribute to score computation.

— 3 —

Verification Flow (Layer A)

3.1 Identity Verification

1. Resolve DID to DID Document
2. Validate DID Document is well-formed
3. Check DID not on revocation list
4. Send random nonce (min 128 bits entropy)
5. Verify Ed25519 signature over nonce bytes
6. Confirm signing key not marked "revoked": true

3.2 Authorization Verification

1. Request AuthorizationCredential for relevant vertical and action
2. Verify credential signature against issuer DID Document
3. Check expirationDate is in the future (UTC)
4. If delegation chain: recursively verify up to root, reject if >8 hops
5. Verify permittedActions includes claimed action
6. Check issuer DID not revoked

3.3 Interaction Proof Verification

1. Check id not already in registry (duplicate rejection)
2. Verify proofInitiator signature against initiator DID Document
3. If bilateral: verify proofResponder; if one-sided: confirm singleSig: true
4. Check outcome is valid value
5. Confirm outcomeHash is valid SHA-256 hex string
6. Check neither party DID is revoked

Trust scores are advisory inputs. Verifiers MUST NOT treat trust scores as authoritative verdicts.

— 4 —

Reference Reputation Model (Layer C — Informative)

This section is informative. Implementations MAY use a different scoring model provided they accept Layer A evidence formats. This model is intentionally simple and heuristic — not a mathematically rigorous anti-manipulation guarantee.

4.1 Score Range and Grades

80–100 (A)	Strong behavioral record, diverse endorsements
60–79 (B)	Good record, limited cross-vertical coverage
40–59 (C)	Emerging record, limited history
20–39 (D)	Thin or inconsistent record

0–19 (F)	Insufficient data or flags present

4.2 Score Formula

```
score = clamp(
0.6 × direct_score
+ 0.3 × propagated_score
+ 0.1 × cross_vertical_bonus
+ interaction_bonus
- sybil_penalty,
0, 100
)

direct_score = Σ(w_i × e_i × d_i) / Σ(w_i) × 100 [weighted mean]
w_i = endorser_trust_score / 100
e_i = endorsement weight (0.0–1.0)
d_i = exp(-0.005 × age_in_days) [time decay]

propagated_score = mean(trust_score(endorser_i)) [single hop]

cross_vertical_bonus = min(n_distinct_verticals × 5, 20)

interaction_bonus = min(n_bilateral × 0.5 + n_single_sig × 0.2, 10)

sybil_penalty = 20 × max(0, jaccard(endorsers_A, endorsers_B) - 0.7)
```

4.3 Minimum Endorser Threshold

Scores withheld (null) until 3 distinct endorser DIDs. Bootstrap weights count during bootstrap period only.

4.4 Bootstrap Weight

Operator-assigned initial score contribution that decays over time:

```
bootstrap_contribution = bootstrap_weight × decay_factor

decay_factor = max(0,

1 - days_since_registration / 90

- organic_endorsement_count / 10

)

→ reaches zero after 90 days OR 10 organic endorsements, whichever first
```

Operator bootstrap endorsements: basis "operator", excluded from diversity calculations.

4.5 Consistency Signal

consistency = 1 - (std_deviation(outcome_values) / 100). outcome_values: completed→100, partial→50, disputed→10, failed→0. Anomaly flagged if consistency drops >0.3 in any 30-day window vs. prior 90-day baseline.

— 5 —

Reference Registry (Layer B)

5.1 API Endpoints

GET /identity/did/{did}	Resolve DID Document
POST /identity/register	Register new agent DID
POST /identity/revoke	Revoke DID or credential
GET /skill/trust-score/{did}	Query trust score
POST /skill/endorse	Submit endorsement
GET /skill/endorsements/{did}	List endorsements received
POST /interaction/proof	Submit interaction proof
POST /violation/record	Submit violation record (operator only)
GET /violation/{id}	Retrieve violation record
GET /swarm/stats	Network-level statistics
GET /health	Registry health status

5.2 Trust Score Response

```
{
"did": "did:moltrust:<id>",
"trust_score": 72.4,
"grade": "B",
"withheld": false,
"endorser_count": 5,
"breakdown": {
"direct_score": 68.2,
"propagated_score": 74.1,
"cross_vertical_bonus": 10.0,
"interaction_bonus": 3.5,
"sybil_penalty": 0.0,
"bootstrap_contribution": 0.0,
"computation_method": "moltrust-v0.2"
},
"consistency": 0.91,
"anomaly_flag": false,
"computed_at": "...",
"cache_valid_until": "...",
"registry_signature": {"type": "Ed25519Signature2020", "proofValue": "<sig>"}
}
```

computation_method "moltrust-v0.2" refers to the reference model defined in Section 4. All responses MUST be signed by the registry operator key.

5.3 Revocation

Registry maintains a signed revocation list updated on every revocation event. Reference implementation propagation target: 60 seconds. Verifiers MUST honor cache_valid_until and revalidate on expiry.

— 6 —

On-Chain Anchoring (Layer B)

Agent registration	SHOULD anchor SHA-256 of DID Document
Confirmed violation	MUST anchor SHA-256 of ViolationRecord
Document integrity	SHOULD anchor SHA-256 of specification

6.2 Anchor Format

```
MolTrust/<event-type>/<version> SHA256:<64-char-hex>


Examples:

MolTrust/AgentRegistration/1 SHA256:ffbc2b04...

MolTrust/Violation/1 SHA256:3a8f91c2...

MolTrust/DocumentIntegrity/1 SHA256:d2ca9b37...
```

Reference chain: Base L2 (mainnet). Any EVM-compatible L2 with <10s finality and permanent data availability is acceptable.

— 7 —

Credential Lifecycle (Layer A)

Issuance	MUST carry issuanceDate, expirationDate, valid issuer signature. Max 365d (endorsements), 730d (auth credentials).
Renewal	New id, new issuanceDate, new expirationDate. Does not reset behavioral history.
Revocation	Any issuer MAY revoke credentials they issued. Submit credential id + revocation signature to registry.
Expiry	Verifiers MUST reject credentials where expirationDate is in the past (UTC). No grace period.

— 8 —

Agent Lifecycle (Layer A)

Registration	Requires: conformant DID Document, principal DID, initial AuthorizationCredential. Optional: stake, bootstrap weight (operator only).
Bootstrap period	Bootstrap weight decays per Section 4.4. After expiry, score is entirely organic. bootstrap_contribution exposed in score breakdown.
Sub-agents	Own distinct DID. AuthorizationCredential from parent. Do NOT inherit parent history. Max 8 hops from root principal.
Cloning	Clone MUST register new DID. Does NOT inherit history. Representing clone as original = clone-impersonation violation.
Redeployment	Same logical identity, same principal: use same DID with key rotation rather than re-registration.
Principal continuity	ViolationRecords associated with both agent DID and principal DID. Re-registration with new agent DID flagged if principal has unresolved violations.
Deregistration	DID marked inactive. Credentials valid until expiry. Behavioral record retained. Stake returned if no unresolved violations.
Stake	Optional USDC deposit in registry smart contract. Returned on clean deregistration, forfeited on confirmed violation. Min meaningful signal: 10 USDC. Requires ViolationRecord to slash.

— 9 —

Threat Model

This section covers known attack vectors. It is not exhaustive.

Attack	Mitigations	Residual Risk
Sybil clusters	Cross-vertical requirement, Jaccard detection, stake cost	*Well-funded patient attacker can construct convincing clusters*
Slow-burn trust	Consistency signal, stake forfeiture	*Patient attacker with low-detectable violation may not trigger signal*
Key theft	Key rotation, revocation propagation, consistency discontinuity	*Stale-cache verifiers may not detect revocation immediately*
Collusion / bribed endorsements	Endorser weight propagation, Jaccard detection	*High-trust collusion harder to detect*
Reputation laundering	Principal DID continuity, on-chain violation permanence	*New principal identity cannot be auto-linked without external evidence*
Replay attacks	UUID deduplication, expiry dates, challenge nonces	*None for conformant implementations*
Data withholding	One-sided proofs, pattern detection	*Bilateral collusion to suppress proofs is undetectable*
Hash preimage inference	Outcome hash includes structured payload	*Predictable outcomes may be correlatable — use random salt*
Adjudicator compromise	ViolationReversal mechanism	*Corrupted adjudicator could produce false violations*

— 10 —

Privacy Model

10.1 Principles

Data minimization: Only hashes and structural metadata submitted to shared infrastructure. Raw transaction content retained locally by parties only. **Pseudonymity:** DID Documents contain no personal data. **On-chain/off-chain separation:** On-chain anchors contain only hashes.

10.2 Storage Map

DID Document	Registry (public)
Authorization credential	Agent + Registry (verifiers on request)

Interaction proof structure	Registry (verifiers on request)
Outcome hash	Registry (public)
Raw outcome data	Local only — parties to the interaction
On-chain anchor hash	Blockchain (public, permanent)
Trust score	Registry (public)
Violation Record	Registry + on-chain hash (public)

10.3 Hash Preimage Risk

For low-entropy or predictable outcomes, outcomeHash may be correlatable. Parties handling sensitive interactions SHOULD include a random salt in the outcome payload before hashing.

10.4 GDPR / Swiss DSG

Informational only. Not legal advice.

Conformant DID Documents contain no personal data. If a DID can be linked to a natural person, interaction records may constitute personal data processing. On-chain anchors are permanent — implementers MUST NOT anchor personal data or directly-linked pseudonymous identifiers. Data retention: 12–60 months for off-chain records. Organizations using the MolTrust reference API for personal data processing MUST establish a DPA with CryptoKRI GmbH.

— 11 —

Worked Example

Scenario: Travel booking agent (Agent A) books a hotel via Agent B. Agent B requires trust score ≥ 60 (Grade B).

Step 1: Identity Verification

Agent B sends nonce "f7a3c2d9". Agent A signs with Ed25519 key, returns DID + signature. Agent B resolves DID Document, verifies signature. ✓

Step 2: Authorization Verification

Agent A presents AuthorizationCredential: issuer = human-traveler-principal, permittedActions = ["transact"], vertical = "moltrust/travel", constraints = {maxTransactionValue: 2000}. Agent B verifies: signature valid, not expired, transact in permittedActions. ✓

Step 3: Behavioral History

Agent B queries GET /skill/trust-score/did:moltrust:traveler-agent-001. Response: trust_score 72.4, grade B, withheld false, signed by registry. Score ≥ 60. ✓

Step 4: Interaction

Agent A submits booking. Agent B confirms reservation.

Step 5: Interaction Proof

Agent A constructs proof: type InteractionProof, outcome completed, outcomeHash sha256:9f86d081..., signs as proofInitiator. Agent B adds proofResponder. Either party submits to registry. Both behavioral records updated at next scoring cycle. ✓

— 12 —

Conformance

12.1 Layer A — Protocol Conformance (MUST)

- Issue DIDs conforming to W3C DID Core v1.0
- Issue and verify credentials conforming to W3C VC Data Model 2.0
- Produce interaction proofs with all mandatory fields per Section 2.4
- Sign artifacts using Ed25519Signature2020 over RFC 8785 canonical JSON
- Use UUID v4 for all id fields
- Reject credentials with expirationDate in the past
- Reject interaction proofs with duplicate id values (per registry)
- Enforce delegation chain depth limit of 8 hops
- Express verticals using / format

12.2 Layer B — Registry Conformance (MUST)

- Implement all endpoints in Section 5.1
- Return trust score responses in Section 5.2 format
- Sign all trust score responses with operator registry key
- Maintain revocation list and propagate within 60 seconds
- Publish operator DID Document at /.well-known/did.json
- Associate ViolationRecords with both agent and principal DID

12.3 Layer C

No mandatory conformance. Implementations MAY use any scoring model consuming Layer A evidence.

References

- W3C DID Core v1.0: https://www.w3.org/TR/did-core/
- W3C VC Data Model 2.0: https://www.w3.org/TR/vc-data-model-2.0/
- Ed25519Signature2020: https://w3c-ccg.github.io/di-eddsa-2020/
- RFC 8785 (JSON Canonicalization): https://www.rfc-editor.org/rfc/rfc8785
- RFC 2119 (Key Words): https://www.rfc-editor.org/rfc/rfc2119
- ERC-8004: https://eips.ethereum.org/EIPS/eip-8004
- Trusted Agentic Mesh (TAM): https://www.ijfmr.com/papers/2026/1/66724.pdf

- AgentHub — Agent Registry and Provenance: https://arxiv.org/abs/2510.03495
- W3C AI Agent Protocol CG: https://agent-network-protocol.com
- DIF Trusted AI Agents WG: https://identity.foundation
- MolTrust Reference API: https://api.moltrust.ch
- MolTrust Whitepaper: https://moltrust.ch/whitepaper

MolTrust
CryptoKRI GmbH, Zurich

api.moltrust.ch
moltrust.ch

info@moltrust.ch
@moltrust