APRO RWA Oracle: The World's First AI-Enhanced Oracle for the Trillion-Dollar Unstructured RWA Market

APRO Research

apro.com

September 24, 2025

Abstract

APRO RWA Oracle introduces a dual-layer, AI-native oracle network purpose-built for unstructured Real-World Assets (RWAs). Unlike price-only oracles optimized for numeric feeds, APRO converts documents, images, audio, video, and web artifacts into verifiable, on-chain facts. By separating AI ingestion (Layer 1) from consensus and enforcement (Layer 2), APRO delivers evidence-backed data feeds for high-value, non-standard verticals—pre-IPO equity, collectibles, legal contracts, logistics records, real-estate titles, insurance claims, and more. This paper details APRO's architecture, proof-of-record model, security and incentive mechanisms, and use-case flows, enabling programmable trust across the trillion-dollar unstructured RWA market.

Key Words: Oracle, Real-World Assets, AI, Proof-of-Record, Consensus, Unstructured Data

1 Introduction

APRO Oracle introduces a two-layer, AI-native oracle network purpose-built for **unstructured, non-standard Real-World Assets (RWAs)**. Unlike price-only or structured-feed oracles, APRO ingests **documents, web pages, images, audio/video** and turns them into **verifiable, on-chain facts**. The network separates concerns into:

- **Layer 1 – AI Ingestion & Analysis (L1):** Decentralized nodes perform evidence capture, authenticity checks, multi-modal AI extraction (LLMs/OCR/CV/ASR), confidence scoring, and sign **PoR (Proof of Record/Reserve)** reports.
- **Layer 2 – Audit, Consensus & Enforcement (L2):** Watchdog nodes

recompute, cross‑check, and challenge; on‑chain logic aggregates, finalizes, and **slashes** faulty reports while rewarding correct reporters.

APRO targets **trillion‑dollar, non‑standard RWA verticals**, starting with **pre‑IPO shares** and **collectible cards**, and extending to legal corpus (agreements, court filings), logistics & trade documents, real‑estate registries, insurance claims, and more. This white paper details **capability coverage** and **end‑to‑end processing flows** for each scenario.

---

2. Problem Statement & Design Goals

Non‑standard RWA pain points

The fastest‑growing RWA categories depend on documents and media rather than ready‑made APIs: a cap table lives in PDFs and registrar pages; a rare card's value depends on photos, grading certificates, and auction data; a loan relies on scanned invoices and shipping records. Today's processes are manual and siloed: analysts retype values, reviewers check signatures by eye, and different venues arrive at inconsistent valuations. Existing oracles are optimized for numeric feeds; they do not natively express how a fact was extracted, where it came from in a source file, or how confident the system is.

Design goals

APRO is designed to be evidence‑first and provable. Each reported fact is accompanied by anchors (page/xpath/bbox/frame) pointing to the exact location in the source, hashes of all artifacts, and a reproducible processing receipt (model versions, prompts, parameters). Dual‑layer validation and stochastic recomputation provide defense‑in‑depth, backed by a slashing economy that penalizes low‑quality or dishonest work. Interfaces are intentionally uniform so DeFi and institutional consumers can program against a small set of schemas. Finally, the system practices least‑reveal privacy: chains store minimal digests while full content remains in content‑addressed storage with optional encryption.

The features of this oracle are:

1. **Evidence‑first**: Turn raw, unstructured evidence into structured facts with cryptographic provenance.
2. **Provable processing**: Record model versions, prompts, parameters, and

anchors for **deterministic re‑ runs**.

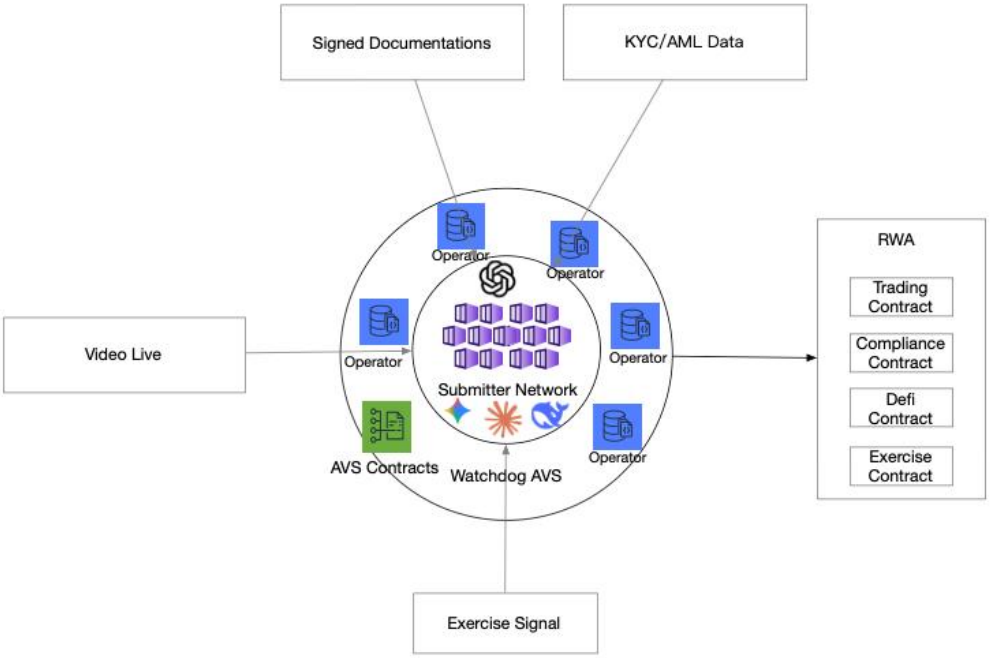
3. **Defense‑ in‑ depth**: Dual‑ layer validation, stochastic recomputation, and slashing‑ backed incentives.

4. **Composable**: Uniform interfaces for DeFi & institutional consumers (price, state, attestations).

5. **Privacy‑ aware**: On‑ chain minimal disclosure; off‑ chain content‑ addressed evidence (IPFS/Arweave/DA).

---

3. System Architecture (Overview)

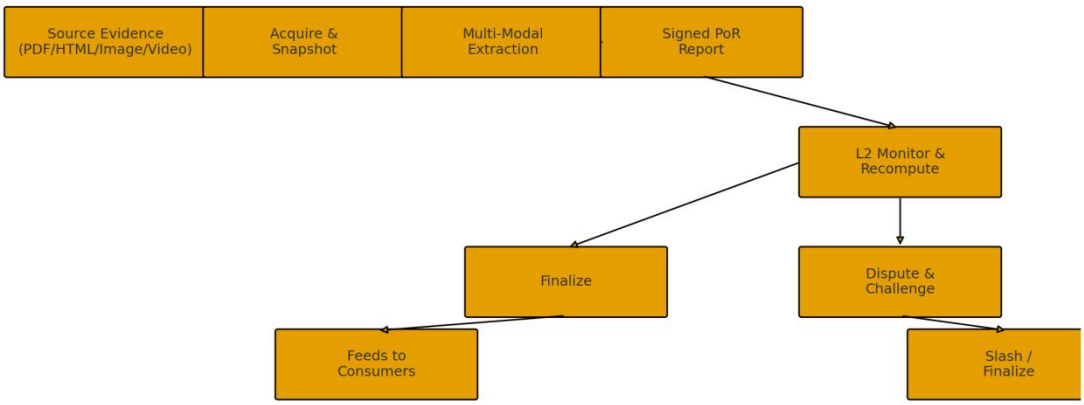


Layer 1 – AI Ingestion. L1 nodes acquire artifacts via secure crawlers, uploads, or delegated retrieval. Every artifact is snapshotted (with content hash, timestamps, TLS fingerprints where applicable) and stored in content‑ addressed backends (e.g., IPFS/Arweave/DA). Nodes run a multi‑ modal pipeline: OCR/ASR converts pixels and audio to text; NLP/LLMs structure the text into schema‑ compliant fields; computer vision models detect object‑ level attributes and forensic signals; rule‑ based validators reconcile totals and cross‑ document invariants. The node then compiles a PoR‑ Report containing evidence URIs and hashes, structured payloads, anchors into the source, model metadata, and per‑ field confidence. The

report is signed and submitted on‑chain or to an L2 inbox.

Layer 2 – Audit & Consensus. L2 watchdogs continuously sample submitted reports and independently recompute them using different model stacks or parameters. Cross‑report consistency rules (e.g., medianization for prices, quorum for categorical facts) drive a deterministic aggregation. A configurable challenge window allows any staked participant to dispute a field by submitting counter‑evidence or a recomputation receipt. If a dispute succeeds, the offending reporter is slashed proportionally to impact; if it fails, frivolous challengers are penalized. Finalized outputs are emitted as on‑chain feeds and can be mirrored across chains through lightweight agents.

High‑level data flow




```
Plain Text
[Source Evidence] → [L1 Acquire + Snapshot] → [L1 AI Pipeline] →
[Signed PoR‑Report]
    → [L2 Monitor] → {No Dispute → Finalize → Feeds}
                    {Dispute → Recompute → Adjudicate →
Slash/Finalize}
```

---

4. Oracle Capability Matrix (Non‑Standard Scenarios)

APRO focuses on high‑value, non‑standard verticals and defines what the

oracle can and will do in each.

Pre‑IPO Shares. Evidence typically includes term sheets, share certificates, board minutes, registrar or transfer‑agent pages, and bank letters. L1 extracts issuer identity, jurisdiction, class structure, authorized/issued/fully‑diluted counts, round terms and dates, and holder‑level positions. Authenticity checks verify digital signatures (PAdES), issuer letterheads, and TLS provenance of registrar pages. Reconciliation rules ensure that holder sums match class totals and the overall cap table; any divergence is flagged with anchored references into the source PDFs. L2 recomputes OCR/LLM on a sample of documents and compares registrar snapshots, reaching field‑level quorum (e.g., ≥2/3 agreement) before finalization. Outputs include a cap‑table digest, last‑round valuation, and a provenance index for downstream contracts.

Collectible Cards. Inputs are high‑resolution front/back images, grading certificates, marketplace listings and historical sales, shipment documents, and vault attestations. L1 performs computer vision identification of set/edition/card number, OCR of grading labels (PSA/BGS/CGC), and image forensics (edge halos, texture anomalies, glare patterns). Certificates are cross‑checked against issuer sites; serials are matched to visual features. Pricing is derived from normalized exchange/auction data, with time‑weighted statistics (floor/median). L2 re‑derives image descriptors, independently verifies certificates, and filters price outliers. Outputs include authenticity and grade, rarity score, floor/median price, vault proofs, and confidence.

Legal/Agreements. Contracts, SPAs, SAFTs, side letters, and court filings are parsed to extract parties, obligations, amounts, dates, jurisdiction, and enforceability signals. Signature validation and certificate‑chain checks support digital execution. L2 conducts clause‑level recomputation and cross‑references public dockets. The feed expresses obligations and key dates with anchors back to clauses.

Logistics/Trade. Bills of lading, invoices, customs docs, tracking pages, and inspection photos converge to shipment state. L1 structures shipper/consignee, HS codes, quantities/weights, and correlates GPS/IoT or carrier events. L2 validates route/time plausibility and performs carrier API checks. Outputs are milestone events, quantity/weight confirmations, and discrepancy flags usable by trade‑finance.

Real Estate. Land‑registry PDFs, deeds, title searches, appraisal reports, and MLS pages are normalized. L1 extracts parcel IDs, ownership, encumbrances, and appraisal ranges; notarizations and registry snapshots are verified. L2 mirrors registry

snapshots and filters appraisal outliers. The feed provides title/encumbrance facts and valuation bands with confidence.

Insurance Claims. Mixed media (photos, videos), adjuster reports, invoices, and police reports inform claim scoring. L1 segments damage via CV, normalizes parts/labor, parses policy terms, and verifies invoices; L2 applies fraud heuristics and peer recomputation. Outputs are claim severity, estimated payout, fraud risk, and decision anchors.

Scenario	Evidence Types	Core Capabilities (L1)	Audit/Consensus (L2)	Outputs/Feeds
Pre- IPO Shares	Term sheets, share certs, cap tables, bank letters, board minutes (PDF/scan), registrar web pages	OCR+LLM extract issuer, class, share count, holder list; detect signatures/stamps; cap- table reconciliation; round/valuation parsing; TLS & registrar checks	Majority agreement on structured cap- table; re- OCR diff; registrar snapshot match; anomaly scoring	Entity registry, cap- table facts, last round valuation, doc anchors, confidence
Collectible Cards	High- res images (front/back), grading certificates, marketplace listings, shipment docs, vault attestations	CV identify set/edition/serial; OCR grading label; glare/edge/watermark forensics; marketplace scrape; comp pricing; Gacha VRF	Cross- node image feature similarity; price outlier filter; chain- of- custody recompute	Auth status, grade, rarity score, floor/median price, vault PoA
Legal/Agreements	Contracts, SPAs, SAFTs, side letters, court filings (PDF/HTML)	Clause extraction; signatory & jurisdiction; date/obligation parsing; e- signature/PAdES validation	Clause- level consensus; signature cert chain checks; docket cross- ref	Clause facts; obligations & dates; enforceable status
Logistics/Trade	B/L, invoices, customs, tracking pages, inspection images	OCR structured docs; HS- code, qty/weight; GPS/IoT correlation; inspection CV	Route & timestamp recompute; carrier API cross- check	Shipment state, milestone events, discrepancy flags
Real Estate	Land registry PDFs, deeds, title searches, appraisal reports, MLS pages	Jurisdiction- aware fielding; parcel ID; lien/mortgage detection; appraisal normalization	Registry mirror snapshot; notarization checks	Title/encumbrance facts; appraisal range; listing comps
Insurance Claims	Photos/videos, adjuster reports, invoices, police reports	Damage segmentation (CV); fraud heuristics; cost normalization; policy term parsing	Peer recompute set; anomaly consensus	Claim severity/amount; fraud risk score; payable decision anchors

5. Detailed Processing Flows by Scenario

5.1 Pre- IPO Shares (Equity in Private Companies)

Pre- IPO equity is documented in term sheets, board resolutions, share certificates, transfer- agent or registrar webpages, and bank confirmations—nearly all of it unstructured. The oracle's entry point is to turn these heterogeneous artefacts into a canonical, machine- readable cap table and a time- stamped valuation record. At Layer 1, APRO's nodes ingest the corpus, verify digital signatures and capture TLS provenance for web sources, and use OCR+LLM templates to extract issuer identity, jurisdiction, share- class structure, authorized/issued/fully diluted counts, holder positions, vesting schedules, and last- round terms. Each field is anchored back to the precise page and bounding box in the original document so that any claim is traceable.

Once the cap table and last- round terms exist as verifiable on- chain objects,

private- market opacity recedes. Buyers and lenders can price secondary transfers against a shared "truth set," diligence time collapses from weeks to minutes, and counterparty risk diminishes because every critical number is linked to evidence. Liquidity improves: fractional trading and collateralized lending gain a sound baseline (e.g., position sizes and anti- dilution clauses become programmable constraints). Compliance conversations also get simpler—auditors and custodians can replay the same anchors the oracle used.

APRO's role. APRO standardizes the cap- table schema, produces a signed Proof- of- Record with anchors and model metadata at L1, and subjects high- value fields (totals, last- round price) to L2 recomputation and quorum before finalization. Discrepancies trigger targeted challenges and slashing proportional to impact. Downstream, APRO exposes feeds—"Latest Round & Valuation," "Authorized/Issued/FD Shares," and "Holder Concentration"—that DeFi markets and institutional ledgers can subscribe to.

Goal: Produce verifiable cap- table and valuation facts for tokenization, lending, and secondary trading.

L1 Workflow

1. **Acquisition**: Ingest term sheets, board minutes, share certs; snapshot registrar/transfer- agent pages. Compute content hashes; store in IPFS/Arweave.
2. **Authenticity**: Validate PDF signatures (PAdES), issuer letterheads; verify registrar TLS; detect scans vs digital originals.
3. **Extraction**: OCR + LLM parse entity name, jurisdiction, share classes, fully diluted counts, vesting schedules, round type/price, liquidation prefs. Build normalized **CapTable JSON**.
4. **Reconciliation**: Sum holder lines = class totals = overall total; cross- doc consistency; flag deltas.
5. **Confidence**: Field- level scores from model entropy, cross- model agreement, registrar corroboration.
6. **Report**: Emit PoR- Report with anchors (page#, bbox) for each field; node signature & model metadata.

L2 Audit

- Random re- OCR & LLM re- parse of a subset; registrar snapshot comparison; quorum rule (e.g., ≥⅔ nodes agree per field).
- Challenge window (e.g., 24–72h); failing fields → partial invalidation and

slashing proportional to impact.

Outputs

- issuerId, jurisdiction, class[] {symbol, authorized, issued, fdiluted}, holders[], lastRound{type, price, date}, valuation, anchors[], confidenceBps.
- **Feeds**: Latest cap‑ table digest, last round valuation, doc provenance index for DeFi integrations.

5.2 Collectible Cards (Sports/Gaming/Art)

Physical collectibles are validated by images, grading certificates, marketplace listings, shipping and vault custody records. The oracle plugs in at authenticity, grade normalization, and price discovery. APRO's L1 pipeline requires high‑ resolution front/back images with EXIF, extracts grading labels via OCR, matches certificate serials against issuer websites, and computes image fingerprints to detect tampering or mismatched serial‑ to‑ image pairs. Simultaneously it normalizes sales and auction data across venues to compute a robust floor and median price.

With authenticity status, grade, and a defensible price band expressed on‑ chain, collectibles become programmable collateral. Gacha pools, lending desks, and AMMs can enforce guardrails (e.g., only PSA‑ 10 with verified custody; LTV tied to the lower of floor/median). Wash‑ trading and price gaming are curtailed by cross‑ venue normalization and on‑ chain transparency. Insurers can price risk more accurately when the facts—grade, provenance, custody—are machine‑ verifiable rather than buried in PDFs and screenshots.

The following subsections specify end‑ to‑ end, APRO oracle‑ mediated workflows that make cards mintable, randomly allocable, financeable, and redeemable in a fully auditable way.

A. Mint & Vault — *From physical existence to an on‑ chain, verifiable fact.*

- **Objective.** Turn "this card exists and is in a vault" into an attested, on‑ chain truth.
- **Oracle data.** (i) **Grading & authentication attestation:** Integrations with PSA/BGS/CGC lookups return certification ID, grade, serial number, and perceptual image hash; an LLM‑ assisted *AI grading* signal can be included as a secondary confidence feature. (ii) **Vault & inventory PoR:** The vault (e.g., Brink's or a third‑ party warehouse) periodically publishes an inventory Merkle root that commits

to each card's unique ID, location, policy number, photo hash, and environmental bands (temperature/humidity). APRO's L1 verifies photo/video via CV/LLM, records the root and signatures on‑chain. (iii) **Anti‑counterfeit binding (optional):** NFC/QR public‑key hash is written into on‑chain metadata; APRO verifies chip signatures via secure challenge–response.

- **On‑chain actions.** Before the first mint of an NFT, the minting contract only accepts **APRO‑signed attestations** tied to the inventory root and grader proof. Any `tokenId` without a valid attestation is rejected.

B. Gacha / Claw (Random Assignment) — *Fair, auditable selection from a vaulted pool.*

- **Objective.** Randomly select a card that is already vaulted and graded; the process must be externally replayable.
- **Oracle data.** (i) **VRF randomness:** A decentralized VRF oracle supplies non‑manipulable randomness. (ii) **Pool snapshot:** A current Merkle root for the eligible set (indexes of allocable cards), with incremental updates as items are drawn. (iii) **Fairness audit:** APRO emits traceable events binding the *request hash, block height, VRF proof,* and the *hit index* so anyone can replay the selection.
- **On‑chain actions.** The `draw()` function finalizes only when a valid VRF receipt is verified **and** the index proves membership against the latest pool root; the contract emits a `Reveal` event including the VRF proof and index.

C. FMV Pricing & Instant Buyback — *Auditable quotes with risk controls.*

- **Objective.** Operationalize "instant buyback at 85%–90% of FMV" in a way that is auditable and risk‑aware.
- **Oracle data.** (i) **Multi‑venue price aggregation:** Crawl eBay/Goldin/TCGplayer and in‑platform fills; compute TWAP/VWAP with robust outlier filters (z‑score/box plot). (ii) **Liquidity thresholds:** For each set/series/grade, compute *reliable sample size* and backtested variance; auto‑down‑weight or flag for manual review when liquidity is thin. (iii) **Buyback quotes:** APRO publishes `fmv`, `buyback_ratio` (dynamically adjustable during campaigns), validity window, and risk flags (circuit‑breaker/limit). (iv) **Pool solvency:** PoR for the buyback pool—reserve Merkle, bank attestations, and on‑chain wallet balances.
- **On‑chain actions.** When a user calls `sellToVault()`, the contract validates the signed quote and its expiry, then checks pool waterline and per‑asset limits before settling.

D. Secondary Trading & DeFi Collateralization — *Price-aware, scarcity-aware risk controls.*

- **Objective.** Make card NFTs usable in lending/AMM while bounded by verifiable price and rarity.
- **Oracle data.** (i) **Collateral/LTV curves:** LTV is parameterized by scarcity, grade, and volatility; with low sample size the oracle automatically lowers LTV or increases haircut. (ii) **Liquidation triggers:** Oracle-pushed TWAPs drive margin calls when thresholds are breached; rarity downgrades (e.g., certificate revoked) propagate as risk events.
- **On-chain actions.** Lending pools read APRO's collateral feed; AMMs can throttle exposure or adjust fees based on volatility and liquidity flags.

E. Burn-to-Redeem & Shipment — *Closed-loop redemption with auditable logistics.*

- **Objective.** Automate and audit the chain from burn to out-of-vault to delivered—or to insured exception.
- **Oracle data.** (i) **Redemption workflow:** APRO listens for Burn events, instructs the vault to release, writes carrier and tracking IDs, and pushes milestones (picked up / customs cleared / out for delivery / delivered / exception). (ii) **Conditional release:** If the shipment is lost or damaged, APRO triggers insurance workflows and marks the NFT as *settled/terminated* accordingly.
- **On-chain actions.** Redemption contracts escrow state transitions and only mark completion when APRO's logistics milestones confirm delivery; exceptions emit events consumable by insurers and refund logic.

5.3 Logistics/Trade (Bills of Lading, Invoices, Tracking)

Bills of lading, invoices, customs declarations, carrier tracking pages, inspection photos, and IoT traces together define shipping truth. The oracle integrates at two layers: document consistency (do quantities/weights/HS codes and consignee data line up?) and journey reconstruction (do timestamps and geos make sense?). APRO's L1 turns documents into structured records and fuses carrier APIs and sensor data to establish milestones; images are analyzed for damage.

Real-time shipment state enables dynamic credit lines, safer invoice factoring, and faster claims decisions. Fraud drops when document sets must reconcile and route

anomalies are flagged immediately. Because state transitions are emitted as events, financiers can subscribe to "goods left origin," "customs cleared," or "arrived/inspected," and reprice risk continuously.

APRO finalizes a "Shipment State" feed only after L2 validates route plausibility and confirms carrier data. Anomaly flags and discrepancy proofs are retained with anchors so insurers and trade‑finance desks can investigate without re‑collecting evidence. The same framework supports supply‑chain ESG attestations where provenance needs to be auditable.

Goal: Provide shipment state and discrepancy flags for trade‑finance, invoice factoring, and insurance.

L1 Workflow: Ingest B/L + invoices; OCR structured fields (shipper/consignee/HS code/qty/weight); correlate carrier tracking + GPS/IoT; image inspection for damages.

L2 Audit: Route/time sanity checks; carrier API recompute; anomaly consensus.

Outputs: `shipmentId`, `milestones[]`, `qty/weight`, `route`, `eta`, `damageFlags`, `confidenceBps`.

5.4 Real Estate (Title, Deeds, Appraisals)

Title status is buried in land‑registry PDFs, deeds, and title searches; valuation lives in appraisals and comparable sales. The oracle extracts parcel identifiers, current ownership, liens/encumbrances, and normalizes appraisal inputs into value bands. APRO's L1 verifies notarizations where present and mirrors registry snapshots for later comparison.

Clear, machine‑verifiable title and valuation bands unlock more granular collateralization and fractional trading. Lenders can program LTV caps against the lower bound of the band and require lien‑free status as a precondition. Secondary markets can respond to registry changes in near‑real time because the oracle publishes change events instead of opaque PDFs.

APRO delivers a dual feed—"Title/Encumbrance" and "Valuation Range"—with field‑level anchors and confidence. L2 filters outlier comps and checks registry mirrors before finalization. Consumers can build liquidation logic or investor disclosures directly on top of these feeds.

5.5 Insurance Claims (Property/Auto)

Claims depend on mixed media (photos, videos), adjuster narratives, invoices, and policy texts. The oracle converts images into damage segments and severity classes, reconciles parts and labor costs to regional tables, validates invoices, and links the result to policy clauses. APRO's L1 composes a coherent, reproducible claim fact‑set from disparate inputs.

Automated, evidence‑backed decisions reduce cycle time and fraud. Carriers can settle straightforward claims immediately while escalating only ambiguous cases. Capital markets and reinsurers gain a transparent stream of claim severity and payable decisions, enabling new forms of parametric protection and risk transfer.

APRO publishes a "Claim Decision" PoR with severity, estimate, fraud score, and anchors, then subjects it to L2 peer recomputation and third‑party checks. Confirmed inaccuracies are penalized swiftly, keeping incentives aligned. The finalized feed becomes a trigger for automated payouts or reinsurance settlements.

---

6. PoR‑Report (Proof of Record/Reserve) Essentials

The PoR‑Report is the core artifact produced by APRO's Layer‑1 nodes and finalized by Layer‑2. It is the **verifiable receipt** that explains *what* fact was published, *from which evidence*, *how* it was computed, and *who* attested to it. This section details the design goals, data model, anchoring semantics, determinism and reproducibility, signature and aggregation rules, audit trail, privacy controls, and validation requirements so developers can implement compatible producers and consumers.

6.1 Design Goals

1. **Traceability.** Every reported field must be traceable to specific bytes/pixels in the original evidence.
2. **Reproducibility.** A third party should be able to re‑run the pipeline (given the evidence and model metadata) and obtain the same result within defined tolerances.
3. **Minimal on‑chain footprint.** Chains store hashes, indices, and compact payloads; heavy artefacts are content‑addressed off‑chain.

4. **Interoperability.** A uniform schema across verticals, extensible via a versioned registry, enables coherent consumption by DeFi and enterprise systems.

5. **Privacy by design.** Sensitive content is redacted or encrypted; the PoR encodes what was hidden and why.

6. **Auditability.** Layer‑2 can audit any PoR with deterministic rules and append results without mutating history.

6.2 Top‑Level Structure

A PoR‑Report consists of seven logical sections:

- **Evidence.** What inputs were processed. Includes URIs (e.g., `ipfs://...`, `ar://...`, `https://...`), `sha256` of each blob, MIME type, size (bytes), capture timestamp, and for web sources the TLS certificate fingerprint and HTTP response digest.
- **Extraction.** The structured result: a `schemaId` (from the registry) and a typed `payload`. Every atomic field in `payload` carries **anchors** pointing into the evidence (page/frame/XPath/bbox) and a per‑field confidence score.
- **Processing.** How the result was produced: `modelId`, `modelCommit` (git SHA or model artifact hash), `containerDigest` (OCI), `promptHash`, decoding parameters (`seed`, `temperature`, `topK`, etc.), and pipeline graph (module ordering).
- **Attestation.** Who stands behind the result: signer addresses, the aggregated signature (e.g., BLS aggregate), quorum parameters (M‑of‑N), submission timestamps, and the reporter's public metadata.
- **Digest & IDs.** Canonical content digests for deduplication and indexing, including a stable `reportId = keccak256(evidenceHashes || schemaId || processingHash || nonce)` and optional `subjectId` for entity‑centric lookup.
- **Audit.** Layer‑2 checks, recomputation receipts, disputes, and outcomes, appended as immutable entries.
- **Versioning.** The PoR standard version (`porUdVersion`) and schema version ensure forward‑compatible parsing.

6.3 Anchoring Semantics

Anchors connect each fact to its origin in the evidence set. APRO mandates **field‑level anchoring** so disputes can target *precise* regions rather than whole documents.

- **PDF.** `{"type":"pdf","doc":0,"page":3,"bbox":[x1,y1,x2,y2]}`; coordinates

use PDF user units after normalization to a canonical DPI (e.g., 72).

- **HTML.** `{"type":"html","doc":1,"xpath":"/html/body/div[2]/table/tr[4]/td[2]"}` or a stable CSS selector; the HTML itself is snapshotted and hashed to avoid DOM drift.
- **Image.** `{"type":"image","doc":2,"bbox":[…],"pHash":"…"}` plus optional keypoints/feature descriptors to resist small transforms.
- **Video/Audio.** `{"type":"video","doc":3,"frames":[210, …, 260],"asrSpan":[tStart,tEnd]}` with a mapping to the transcript.
- **Multi-source.** A single field may cite multiple anchors when cross-validation is used (e.g., price from two auction screenshots). Anchors must include the *doc index* into the Evidence array and remain valid under re-hydration from URIs.

6.4 Reproducibility & Determinism

To make outcomes re-computable, APRO requires:

- **Canonicalization.** A fixed pre-processing order (decode → normalize → OCR/ASR → parse → post-process) and stable library versions; any variability (e.g., OCR beam search) must be reported.
- **Randomness control.** Report `seed` and decoding parameters for LLMs or stochastic vision modules; a **deterministic mode** (greedy or fixed seed) should be available for audit.
- **Tolerance windows.** For numeric fields, Layer-2 defines acceptable deltas (e.g., string-to-number parsing quirks). For categorical fields, string similarity thresholds (e.g., Jaro-Winkler ≥ 0.98) guide equality.
- **Processing hash.** `processingHash = keccak256(modelId || modelCommit || containerDigest || promptHash || params)`; any change yields a different hash and naturally separates results.

6.5 Signatures, Aggregation, and Quorum

- **Signer set.** Each L1 node signs the canonicalized report bytes (CBOR/JSON-Canon). The signature spec (e.g., ECDSA or BLS) is fixed per deployment.
- **Aggregate signature.** A coordinator (or on-chain precompile) aggregates signatures for the same report root to reduce calldata.

- **Quorum.** The PoR includes (M,N) and the contract verifies M distinct authorized signers. Field-level disagreements are supported by packing multiple candidate values with signer bitmaps; Layer-2 resolves them via majority or weighted quorum rules.
- **Replay protection.** submitTs, chain ID, and a domain separator (EIP-712) bind signatures to a context.

6.6 Audit Trail & Dispute Records

The audit section is **append-only**:

- **Checks.** Which fields were re-computed, by whom, with what processingHash and evidence snapshot; pass/fail flags and deltas.
- **Disputes.** Challenge IDs, challenger stakes, evidence of fault (counter-anchors or recomputation digests), and verdicts.
- **Finalization.** A finalization block/tx reference and status (final, disputed, superseded). Consumers can choose to only trust finalized reports or to display disputed status in UI.

6.7 Privacy Controls

- **Redaction.** Fields can be marked redactedIrreversible=true with a salted hash of the removed value for proof of omission without recovery.
- **Encryption.** Sensitive fields include a ciphertext with a KMS key identifier and access policy. APRO never stores plaintext on-chain; decryption happens off-chain under policy.
- **PII labeling.** Each field may carry a sensitivity tag (PII, financial, public) so consumers can decide how to process or display it.

---

7. Privacy & Compliance

- **Least-reveal**: On-chain stores digests + indices; content off-chain with content addressing.
- **PII controls**: Field-level redaction (irreversible) or encrypted fields (reversible, key-gated).

- **Right-to-erasure**: Off-chain encrypted deletion; on-chain immutable hash remains.
- **IP & licensing**: Evidence usage basis recorded (public domain/license/fair use).

---

8. Roadmap

- **Phase 1**: Pre-IPO & Cards schemas; L1 pipeline MVP; L2 dispute contract; basic staking.
- **Phase 2**: Legal & Logistics schemas; advanced forensics; dynamic watchdog sampling.
- **Phase 3**: Real estate & Insurance; TEE/ZK optional proofs; cross-chain feeds.

Appendix A — Example Schemas (abridged)

A.1 Pre-IPO (Cap Table v1)

```JSON
{
  "$id": "apro.schema.cap_table.v1",
  "required": ["issuer", "jurisdiction", "classes", "holders"],
  "properties": {
    "issuer": {"type": "string"},
    "jurisdiction": {"type": "string"},
    "classes": {"type": "array", "items": {"properties": {
      "symbol": {"type": "string"},
      "authorized": {"type": "integer"},
      "issued": {"type": "integer"},
      "fdiluted": {"type": "integer"}
    }}},
    "holders": {"type": "array", "items": {"properties": {
      "name": {"type": "string"},
      "class": {"type": "string"},
```

```
      "qty": {"type": "integer"}
    }}}
  }
}
```

A.2 Cards (Collectible v1)

```JSON
{
  "$id": "apro.schema.cards.v1",
  "required": ["set", "edition", "cardNo", "grade"],
  "properties": {
    "set": {"type": "string"},
    "edition": {"type": "string"},
    "cardNo": {"type": "string"},
    "grade": {"type": "string"},
    "auth": {"type": "string"},
    "floor": {"type": "number"},
    "median": {"type": "number"}
  }
}
```

Appendix B — Example L2 Policy Snippets (informal)

- **Quorum**: field- level acceptance if ≥ 2/3 L1 nodes agree within tolerance.
- **Tolerance**: price fields win by Hampel filter + median absolute deviation.
- **Slashing weights**: proportionate to field criticality (e.g., identity > price > metadata).
- **Escalation**: repeat offenders flagged; increased bond; longer challenge windows.
