R-EVAL · MULTI-AGENT EVALUATION PLATFORM

R-Eval Evaluation Platform

From telemetry to auditable evidence

Functional Description v1.0 · 2026-11 · Delivery: SaaS / on-premise

R-Eval ingests R-Box data and performs temporal alignment, event slicing, metric computation, review and reporting. Its value is not to replace judges, but to turn raw telemetry into evidence that is computable, explainable, reproducible and auditable.

> Current version: r-eval.engine.v1 / rbox.telemetry.v1. Standard file import (CSV / JSON) and R-Box protocol adaptation are supported; live in-vehicle streaming and vendor-private protocols are delivered per project.

01 Core functions

Module	Capability
Data import	CSV / JSON; rbox.telemetry.v1; ≤100,000 rows and ≤20 MB per batch; per-row errors and content-hash dedup
Evaluation engine	r-eval.engine.v1; RRI per vehicle, CRI for the pair; weights configurable and confirmed by the rules committee
Temporal alignment	Clock offset estimated from shared seq; 150 ms tolerance nearest-neighbor pairing; stores clockOffsetMs and pairedSampleCount
Event slicing	Faults, manual takeover, cooperative tasks, invalid data and checkpoints become time-stamped events
Dual-vehicle replay	Time cursor, event jump and evidence JSON inspection; frame-level localization
Review & lock	Automatic score / manual adjustment / final decision kept separate; adjustments require a reason and retain operator and time
Reports	HTML (reading/printing), JSON (system exchange), CSV (analysis); includes protocol and engine versions and limitations
Data governance	Anonymized datasets (random salt + SHA-256, relative time); access levels private/authorized/public; read/export/reuse grants
Audit	Who did what to which entity and when; request id, status and latency; separate from request logs

02 Metrics

Metric	Meaning	Basis
RRI	Robot Reliability Index	Fault-injection results and task process data; computed per vehicle, then averaged
CRI	Coordination Reliability Index	Formation keeping, task-allocation success, link stability, fault compensation
Data quality	Evidence confidence	100 × (0.6 × valid-record ratio + 0.4 × mean time confidence)
Physics residual	Physical-consistency hint	Speed jumps, position–velocity deviation, negative formation distance; flags for review only

03 Architecture

Single-repository full stack: the browser calls typed tRPC through TanStack Query; the Hono server handles auth, permissions and validation; Drizzle ORM accesses MySQL; files go to object storage. Shared contracts keep import protocols, scoring evidence, reports and dataset structures consistent across front and back end.

Layer	Implementation
Frontend	React workspace + public event dashboard; 9 core routes; page-level lazy loading; local scrolling on mobile
API	/api/trpc; publicQuery / authedQuery / adminQuery
Data	MySQL; object storage for raw files, reports and datasets; database stores object keys and hashes only
Observability	Request id; structured logs; stack traces never returned to callers

04 Data protocol and compatibility

- Protocol: rbox.telemetry.v1. CSV supports quotes, commas and ISO time; JSON supports standard record arrays.
- Batch limits: 100,000 rows and 20 MB. Unknown R-Box, wrong protocol, invalid time or values are rejected or reported per row.
- Batches are registered before processing; on failure, written telemetry and objects are cleaned up without masking the original error.
- Reports and datasets record protocol version, engine version and generation time for cross-version traceability.

05 Workflow

- Create an event; register teams, vehicles and unique R-Box IDs.
- Import CSV / JSON; inspect valid rows, invalid rows and per-row errors.
- Run evaluation; review ranking, RRI / CRI breakdown, clock offset, paired samples and data quality.
- Dual-vehicle replay: drag the time cursor, jump to events, inspect evidence JSON.
- For manual decisions, enter the adjusted value and reason, then complete review and lock.
- Generate reports; create anonymized datasets, set access level and grant permissions.
- Publish the public dashboard; manage roles and audit key actions.

06 Deployment

Item	Requirement
Runtime	Node.js 20.12+; MySQL; npm
Deployment	Two-stage Docker build; npm start; default port 3000
Storage	Object storage for raw files, reports and datasets; downloads via short-lived signed URLs
Before go-live	Schema sync, HTTPS, backup/restore drill, monitoring, key rotation, peak-load test

07 Known boundaries

- This delivery is a complete runnable software baseline; live in-vehicle streaming, wireless links and vendor protocols are integrated per project.
- Checkpoints, speed, formation distance, recovery window and weights are engine-v1 baselines; confirm and version them with the rules committee.

- Automatic scoring and physics residuals are supporting evidence, not a substitute for judges; manual adjustment and final lock preserve the chain of responsibility.
- Anonymization reduces direct exposure but does not guarantee non-re-identification under all external data combinations; review before publication.

Contact: info@r-power.hk · r-power.hk · R-Power Technologies Limited (Reg. No. 79936367)