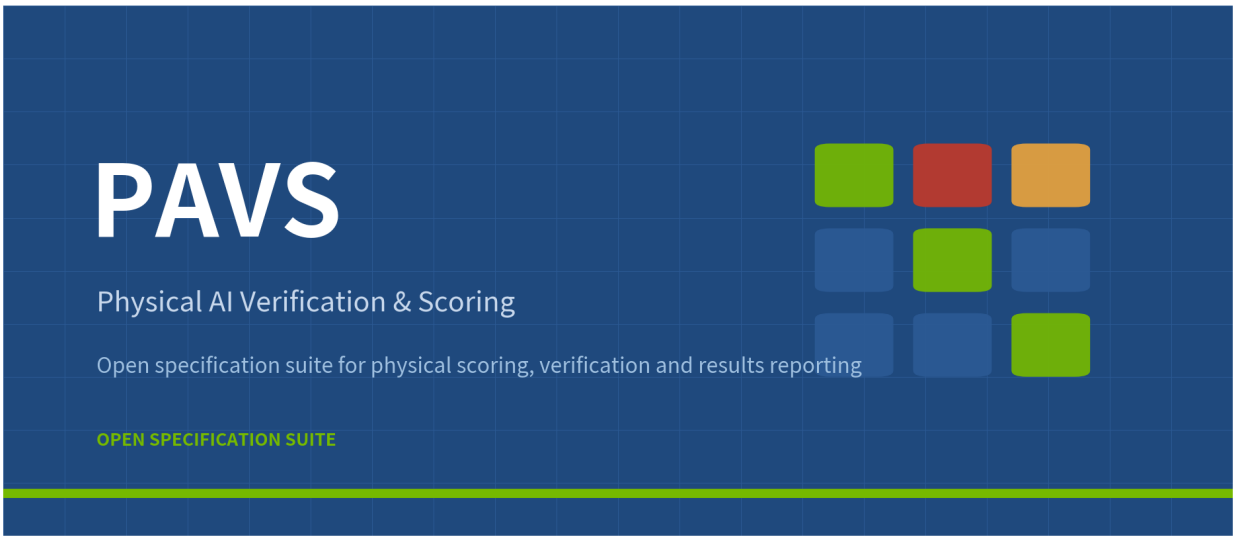


Part 3 — Reference Implementation & Reproduction (informative) · English translation

Issued edition v0.8 · 29 October 2026

Published by ImmersiveCast Co., Ltd. · open specification at the proposal stage (not a standard of a formal standards body)

English translation — the Korean text is authoritative · immersivecast.com/pavs

Contents

1. Position of the reference implementation

2. Mechanisms of the reference implementation — implementation form of the 5 links of the chain

3. The 3 demonstrations and reproduction (executed 2026-08-25)

- 3.1 Demonstration A — model authenticity: change only the specifications and the verdict is reversed
- 3.2 Demonstration B — the whole scoring chain: one test, five links
- 3.3 Demonstration C — the scorer's negative control: the scorer itself is tested too
- 3.4 Demonstration scope — assertions whose executability the three demonstrations showed

4. Standard composition of the run package

5. Location of execution identification information

Revision notice — v0.4 → v0.5 (2026-10-19)

English edition — translation. This is the English translation of the Korean edition v0.8 (2026-10-20). The Korean text is authoritative; where the two differ, the Korean text prevails. Assertion identifiers, clause numbers, case identifiers, versions, dates and figures are identical to the Korean edition.

Document status: Issued edition v0.8 · 2026-10-29 (errata correction — defined terms · table group name · statements of measurement timing) · previous issue v0.7 2026-10-20 · published at immersivecast.com/pavs · corresponding specification Part 1 v1.5 · this document is informative. **Nature**: Explains the executability of, and the method of reproducing, what is specified in the body and annexes of Part 1, and is not part of the specification. **Edition history**: v0.8 (2026-10-20 · draft · errata correction — “verdict kernel” → “scoring engine” (Part 1 3.26) · “result documents” → “results document” (Part 1 3.24) · group name in the table of 3.4 “VRF evidence” → “VRF verification” (Part 1 4.2) · quotation marks in the short name of **PAVS-REP-12** · notation of the list of cell-criterion values of Demonstration B · fixed sentence on the number of days elapsed since measurement → measurement date · Korean edition: general words replaced following the glossary of the Korean expression norm, English wording unchanged · verdicts · figures unchanged) · v0.7 (2026-10-20 · issued edition · old edition indication (v1.2) deleted from 5 citations of the Conformance Statement → current edition = public listing · verdicts · figures unchanged) · v0.6 (2026-10-20 · issued edition · citations of the met count deleted in 2 places → reference to the Conformance Statement — because a document that restates the count shall carry the four elements of the claim form of specification 11.2 · reflects the CNF-01 retest) · v0.5 (2026-10-19 · issued edition) · v0.4 (2026-10-02 · issued edition — separation of demonstration-related assertions from verdicts of met · recount of assertions outside the demonstration scope · measurement dates stated alongside · execution times and internal paths deleted · re-scoring of external AI modules reflected) · v0.3 (2026-08-27 · demonstration scope clause newly established · evidence levels indicated) · v0.2 (descriptions of mechanisms corrected to match the results of the exhaustive internal check of all 72 assertions). **Drafting conventions**: The reference implementation is described with a focus on mechanisms. Scope of disclosure: actual robot type names disclosed · execution commands and internal paths not disclosed · reproduction package provided on request. Test identifiers and execution commands are managed in internal reproduction records and are provided to joint pilot-test partners. Figures in the text are limited to physical quantities of the demonstration results, and the measurement date is stated alongside.

1. Position of the reference implementation

This specification has a reference implementation. The reference implementation exists for two reasons.

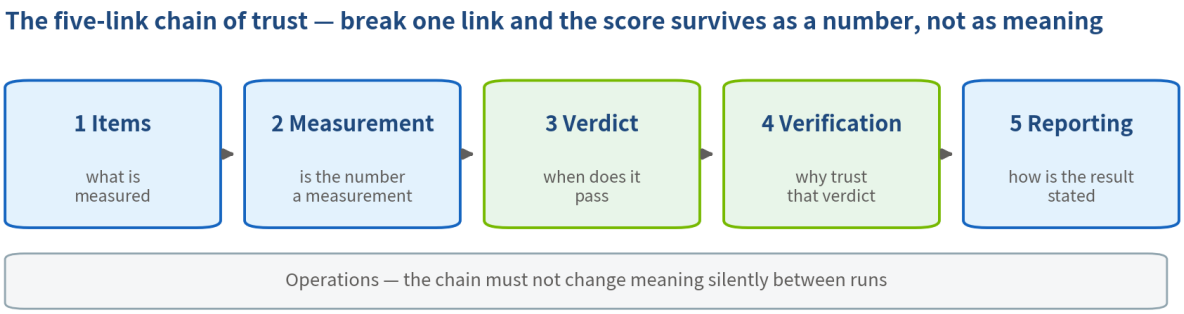
1. **Executability** — The goal is to show, through tangible instances of actual operation, scoring and reporting, that all 72 test assertions of Part 1 are not requirements on paper but executable checks. At present, the range over which the three demonstrations have shown executability is 18 of the 72 (list of related assertions = §3.4). The met status of each assertion is given, identifier by identifier, in the Conformance Statement (public · immersivecast.com/pavs). This appendix does not restate the met count — because a document that restates the count shall also carry the four elements of the claim form of specification 11.2 (version · class · list of assertions met · list of assertions not met).

2. **Reference point for cross-checking** — Provides a reference with which an adopter of the specification can execute the same procedure in its own environment and cross-check the results anew.

The reference implementation follows a design that is not tied to a particular physics engine or a particular robot, and an SUT is accommodated by adding a declarative description (manifest).

This appendix treats the reference implementation **anonymously** — in order to preserve the neutrality of the body of the specification. The configuration, boundaries and form of delivery that this implementation has as a product are described, with its name stated, in Part 4 (Solution Profile · unpublished — next).

2. Mechanisms of the reference implementation — implementation form of the 5 links of the chain



Link	Implementation mechanism
Items	Scenarios are described as domain-language documents, and the assertion statements of those documents become the executable items of the scoring engine. The standard cell specification (proportional to the reach radius of the SUT) and failure controls are distributed as a corpus. **A self-containment check has been introduced at the intake stage of defense engagement tests, and its application to robot-arm and civil tests is in progress** — some scenarios receive their initial conditions as arguments of the submission layer
Measurement	Values used in scoring are accompanied by proof of measurement, so that unproven values are rejected by the scoring engine, and missing data is not replaced with a value (enforced by a storage-layer constraint). **To the provenance categories ‘measured’ and ‘recorded’ have been added ‘declared value’ (a fact derived from the run configuration · not an observation · 2026-08-31) and ‘impossible value excluded’ (the measuring instrument ran, but a value outside physical limits is removed from the measurement and the reason is recorded · 2026-09-04), and values after an irreversible event such as divergence are ‘sealed’ and are not raised back into measurement (2026-09-14). An indication of the measurement location (simulation) is carried with them. The distinction of directive values is incomplete (PAVS-MEA-01 not met) — a category separating commanded values from measured values, and categories for**

	default values and estimates, do not yet exist. The effective values of test conditions are cross-checked by read-back during execution but are not yet stated alongside in the results documents
Verdict	3-valued scoring engine — a pass is produced only when every required condition is met by measurement, not-measured is not counted as a pass (an absence of measurement of a mandatory item is a fail for the reason ‘absence of measurement’ · a non-mandatory item preserves not-measured · blocking of promotion exists at three points: the scoring engine, aggregation and the storage layer), and a fail produced by a stricter verdict layer is not covered over by a more lenient indication. **Exclusion from scoring of output produced by bypassing physics is implemented as items of robot-arm tests (pick and place, etc.), and its extension to mobility, flight, maritime and legged-locomotion tests is in progress.** When the result of an external AI module arrives after the run, the same run is re-scored, and the result is stored only if the remaining verdict rows other than the AI module rows and the overall verdict have not changed (if they change, storage is refused · restored to the original state). If ‘scoring after arrival’ is declared for an attachment that has no re-entry path (a method that loads the model file inside the run), it is rejected before the run is created (confirmed 2026-09-30 to 10-01)
Evidence	Negative-control cells reside permanently in the regression suite (they are automatically compared with the expected verdict, and if a negative control turns to a pass, ‘scorer defect suspected’ is raised), golden-set re-scoring identity and a verdict-shift transition matrix block shifts toward leniency, and a verdict can be reproduced from the stored artifacts with a single command. **The seed is recorded in the results only for robot-arm tests and not for defense, mobile-robot and maritime tests, so the pinning of the reproduction configuration is partial.** The obligation to output video, trajectories and event logs is enforced
Reporting	Results documents are generated directly from the verdict result, and the document itself carries a measurement-reliability section — verdict coverage, provenance of values and ‘what should not be trusted at face value’. The scoring engine's overall verdict appears first, and the document's own threshold table is demoted and marked ‘not a verdict’. The issue history is managed as an append-only ledger, and **detection of content tampering in the ledger exists in the core verifier, and its**

	extension to the persistence layer is in progress

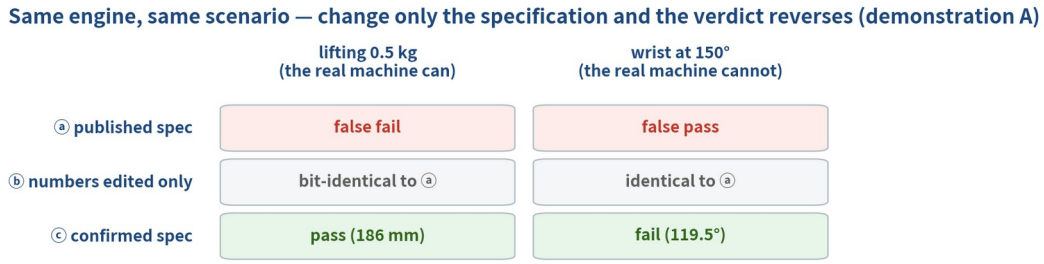
> NOTE — Each entry in this table describes the **current state** of the reference implementation (as of the exhaustive internal check of 2026-08-26 · later changes are dated). The verdicts on the 72 requirements of the specification and the list of assertions not met are in the Conformance Statement (public · current edition = immersivecast.com/pavs). If the description of the implementation runs ahead of the verdicts, that is itself the behavior this specification prohibits, so the two documents shall point to the same verdicts.

3. The 3 demonstrations and reproduction (executed 2026-08-25)

The three demonstrations each show a different layer of the specification in tangible form. Demonstrations B and C were executed on the normal operating path of the reference implementation (test intake API and regression suite), and all their artifacts (including video) are preserved; Demonstration A was executed with an independent check script for model-authenticity comparison (deterministic, CPU only), and its result data and reproduction commands are preserved.

The ‘related assertions whose executability was shown’ of each demonstration are the assertions that the demonstration showed by execution, and are **not a verdict that they are met**. The verdict for each assertion follows §3.4 and the Conformance Statement (current edition).

3.1 Demonstration A — model authenticity: change only the specifications and the verdict is reversed



Related assertions whose executability was shown: **PAVS-VRF-05** (model authenticity comparison) · **PAVS-VDT-08** (rejection of unperformable directives) · **PAVS-VRF-14** (reproducibility) — not a verdict that they are met (§3.4)

Procedure: The tester scores the same two tasks on the same 6-axis lightweight arm (lifting 0.5 kg, within the manufacturer's rating, and aligning the wrist to 150°, beyond the limit of the physical hardware) on the same physics engine and under the same initial conditions, changing only the robot specifications, in three ways — ⓐ the specifications in a public model repository ⓑ the public model with only one actuator figure corrected to the manufacturer's value ⓒ the specifications confirmed by the manufacturer.

Observed results (measured 2026-08-25 · CPU only · deterministic reproduction — the same values in the same environment):

Condition	Lifting 0.5 kg (possible on the physical hardware)	Wrist 150° (impossible on the physical hardware)

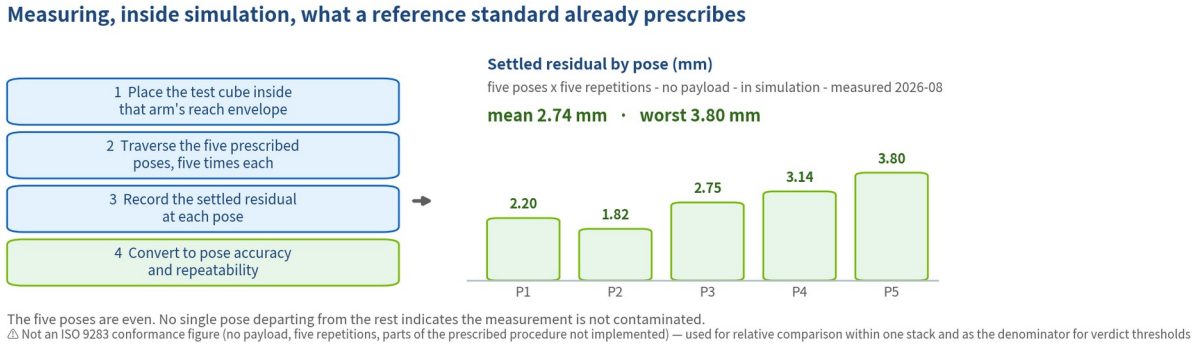
ⓐ Public specifications	**False fail** — drive 0.79 N, object lift 0 mm	**False pass** — 149.8° reached
ⓑ Figure corrected only	**Identical** to ⓐ (bit for bit) — specification correction ≠ physical change	Identical to ⓐ
ⓒ Confirmed specifications	Pass — drive 50 N, lift 186 mm	Fail — limited at 119.5°

The figures of this run agreed with the earlier measured records, which include runs on computers of different types (cross-computer identity is a deterministic reproduction confirmed in the earlier measurements). This property allows an objection to a scoring result to be resolved by ‘re-execution’.

Takeaway: In a system that scores third parties, verification of model authenticity is a precondition of scoring. Unless it is confirmed that the model submitted is genuine, no score proves anything.

> **The reproduction package is provided on request. Execution commands and internal paths are provided to joint pilot-test partners** (inquiries: contact@immersivecast.com). The package consists of the execution script, reference values and reference environment versions (physics engine version and public model repository commit), and, after the public physics engine and the public model repository have been obtained, runs on CPU only with 1 command. The verification mode compares the output with the reference values and returns a success exit code only when they agree completely; if they differ, it reports item by item which value differs and how. This shows by execution, within the scope of this demonstration, the requirement of **PAVS-VRF-14** (an objection to a scoring result shall be resolvable by re-execution). VRF-14 itself is not met in the Conformance Statement (the pinning of reproduction information is limited to robot-arm tests).

3.2 Demonstration B — the whole scoring chain: one test, five links



How the standard pose-measurement procedure is performed in simulation and what it produces — five poses, five round trips each, with the settling residual recorded for each pose.

Related assertions whose executability was shown: **PAVS-ITM-01**/07/08 (standard cell, corpus, baseline with conditions) · MEA-01/07 (provenance, effective value) · VDT-01/02/06 (3-valued, positive evidence, authority hierarchy) · REP-01/02/03 (evidence level, attribution, coverage) · VRF-14 (reproducibility) — not a verdict that they are met (§3.4)

Procedure: The tester submits 1 robot-arm pick-and-place test of the standard cell specification to the intake API with a fixed seed. Video recording is mandatory, and after completion a results document (PDF) is generated from the verdict result.

Observed results (measured 2026-08-25 — measured before the arena baseline was re-pinned (2026-10-02) · subject to re-measurement):

- All 8 cell criteria met (main values — lift 78.9 mm (above the threshold of 50) · placement error 1.0 mm · hold 12.6 s · drops 0 · trajectory reversals 0.0/m · maximum speed 353.7 mm/s (within the declared upper limit))

- Since 2026-09-05 the trajectory-reversal axis has been described by a 3-valued verdict (stable, unstable, not-measured) — the 95% Poisson interval of the number of events counted by arc-length resampling, divided by the length of the path observed, is compared with the threshold of 8/m. ‘0 events’ on a short path can be not-measured rather than a pass. Repeated runs of a fixed-pose cell are not samples when they are replicas (Casebook D11), and whether they are replicas is not assumed but compared for each run — there is a case in which a re-run of the same pose changed the verdict (Casebook D14). The position accuracy test: measured with the characteristic definitions of ISO 9283 · rated load mass applied (from 2026-10-02) · not an official test.

- **At the same time, the scoring engine stated a fail alongside** — self-collision contact (a real defect) was measured, showing that the authority hierarchy, under which meeting the cell criteria (green) does not cover over a scoring engine fail, operates in both the verdict result and the results document. It is the tangible form of the discipline that ‘the success verdict and the scoring engine verdict are different rulers, and quoting only one of them is an overstatement’.
- The generated PDF indicated ‘Verdict (kernel): fail’ (original heading of the generated PDF) directly in its title, and itself contained the measurement-reliability section — verdict coverage (lists of met/below criterion/not-measured), provenance of values and ‘what should not be trusted at face value’ (including the list of channels that are not recorded).
- The identical test with the same seed reproduced identical figures on a different compute node (lift, placement error and speed all agreed).

Takeaway: ‘The form itself states the grade of its own evidence’, which the specification requires, actually holds in an automatically generated document.

3.3 Demonstration C — the scorer's negative control: the scorer itself is tested too

Related assertions whose executability was shown: **PAVS-VRF-01** (permanent negative controls) · ITM-05/06 (discriminating power, reverse-direction control) · REP-12 (safety anomaly ≠ ‘normal’) — not a verdict that they are met (§3.4)

Procedure: The tester executes, in the regression suite, a scenario that deliberately violates the engagement authorization discipline (including 1 shot fired without designation). The expected verdict (fail) is declared in the suite, and comparison with the actual verdict is performed automatically.

Observed results (measured 2026-08-25 — measured before the arena baseline was re-pinned (2026-10-02) · subject to re-measurement): The scoring engine produced a fail on exactly the item concerned — a count of 1 friendly shot fired without designation — (the 2 normally designated shots passed), and the automatic comparison of the suite declared ‘negative control failed as expected — scorer discrimination maintained’. The suite itself warns against using the artifacts of a partial run as the baseline (preventing the narrowing of the check scope).

Takeaway: A regression suite with only passes verifies only half of the scorer. A regression in which the scorer becomes more lenient is caught only by negative controls.

3.4 Demonstration scope — assertions whose executability the three demonstrations showed

This specification requires results reporting to ‘disclose what was measured and what could not be measured as a list, not as a ratio’ (**PAVS-REP-03**). That requirement applies to this appendix as well. The related assertions whose executability the three demonstrations showed are **18 of 72**. This is not a verdict that they are met — of these 18, the assertions that are met in the Conformance Statement are only the 2 marked ‘met’ in the table below (VDT-01★ and VDT-06), and the remaining 16 are 14 not met and 2 inconclusive.

Clause group	Total	Demonstration-related	Related assertions	Conformance Statement verdict
ITM items	10	5	ITM-01 · 05 · 06 · 07 · 08	All 5 not met
MEA measurement	8	2	MEA-01★ · 07	MEA-01★ not met · MEA-07 not met
VDT verdict	8	4	VDT-01★ · 02★ · 06 · 08	**VDT-01★ met** · VDT-02★ not met · **VDT-06 met** · VDT-08 not met
VRF verification	16	3	VRF-01★ · 05 · 14	VRF-01★ inconclusive · VRF-05 inconclusive · VRF-14 not met
REP reporting	15	4	REP-01 · 02 · 03 · 12	All 4 not met
OPS operations	10	**0**	—	—
CNF conformance	5	**0**	—	—
Total	**72**	**18**		met 2 · not met 14 · inconclusive 2

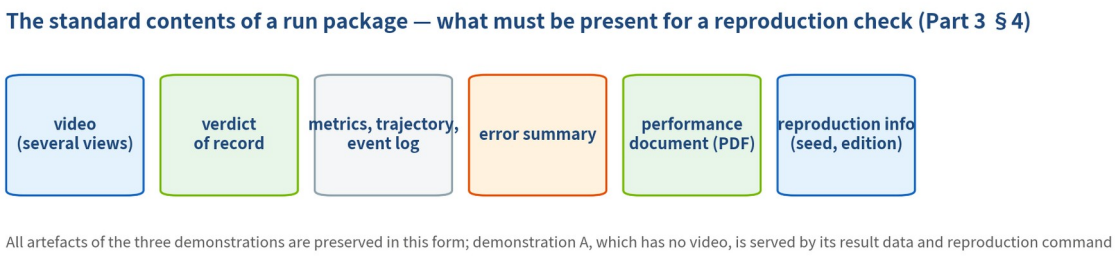
> NOTE — The basis of the verdicts is the Conformance Statement (Annex C tests of 2026-09-04 · REP-15 retested 2026-09-06). The 14 ‘not met’ are the 2 whose Annex C test record is not met (MEA-01★ · VDT-02★) + the 12 judged ‘partial’ in the internal check of 2026-08-26 (awaiting Annex C retest). The 2 ‘inconclusive’ (VRF-01★ · VRF-05) were not performed because part of the procedure requires executing a run, and the Conformance Statement does not treat them as met but puts them in the list of assertions not met. ★ = core assertion that is a precondition of a conformance claim.

Nature of the 54 outside the demonstration scope. Divided by the test types assigned in Annex C, they are 22 injection, 31 audit and 1 demonstration. For these, what this edition presents is the test methods (Annex C) and the verdicts of the Conformance Statement (public · current edition), and **no demonstration execution was performed.**

Reason the 15 operations and conformance assertions have 0 demonstrations. Of them, 13 are of the audit type — the object of the assertion is not an individual run but a property of artifacts spanning a period, so it cannot be shown by 1 run of a test (C.0.2). This follows from the nature of these clauses and is not an omission in the design of the demonstrations. **However, that their nature is so does not mean that their executability has been shown, and in this edition the two clause groups remain outside the demonstration scope.**

> **The demonstration scope of 25% (18/72) is a low figure, and it is not a met ratio either (for the met status, see the Conformance Statement). The reason this appendix states that figure both as a ratio and as a list is that, if the specification's own appendix did not keep the disclosure form that the specification requires of others, that requirement would not hold.**

4. Standard composition of the run package



The artifact package of 1 test that is the object of reproduction verification contains at least the following — this composition itself is the tangible form of the reporting requirements of Part 1.

Artifact	Role
Video (multiple viewpoints + streaming chunks)	Independent evidence of physical events (VDT-03) · checkability by humans
Verdict result (verdict)	3-valued verdict and item-level grounds (the source for the REP series)
Metrics, trajectories and event logs	Material for verifying measurement provenance and effective values (MEA series)
Error summary	Distinction between being ‘wrong without an error’ and ‘an error’
Results document (PDF)	Rendering of the verdict result + measurement-reliability section
AI module attachment record	Runs with an external AI module attached — declaration · attachment status · attribution of the clauses it is responsible for · basis for the not-measured indication of the clauses for which an unattached module is responsible
Re-entry record	Runs scored after arrival — reflection of the arrived result · recomputation of the verdict · whether stored (condition that the verdicts other than the AI module rows and the overall verdict are unchanged)
Reproduction information (seed, configuration, code version)	Robot-arm tests as the basis · other tests being supplemented (**PAVS-VRF-14** not met)

5. Location of execution identification information

The level of disclosure differs from demonstration to demonstration, and that difference is not hidden. The scope of disclosure was settled as follows (2026-10-02): actual robot type

names disclosed · execution commands and internal paths not disclosed · reproduction package provided on request.

Demonstration	Evidence level	What is disclosed	Reproducibility
A — model authenticity	**L1** — an independent check script, not the production path	Result figures and verdicts. The reproduction package (script, reference values, reference environment versions) is provided on request · CPU only · 1 command	A third party that has received the package can reproduce it on its own computer
B — the whole scoring chain	**L3** — end-to-end execution on the production path (intake API) + direct inspection of artifacts	Results, verdicts and video. Test identifiers, execution commands and internal paths are not disclosed	Provided separately to joint pilot-test partners
C — negative control	**L3** — end-to-end execution on the production path (regression suite) + direct inspection of artifacts	Same as above	Same as above

> NOTE — Evidence level and scope of disclosure are different axes. **Demonstration A, the closest to third-party reproduction, has the lowest evidence level** (L1). That a third party can run it and that the run went through the production path are different propositions, and this table states the two separately.

Demonstration A was chosen first as the object of reproduction not because its value is great but **because its cost is the lowest** — it is deterministic, runs on CPU only with 1 command, and has neither customer assets nor a test-site configuration attached to it.

Revision notice — v0.4 → v0.5 (2026-10-19)

Change	Content	Reason
Correction of the statement on determinism	‘Repeated runs of a fixed-pose cell are not samples but replicas’ → not samples when they are replicas, and whether they are replicas is compared for each run (Casebook D11 · D14)	A generalization at odds with the recount of re-runs under the same conditions (trajectories matched in 8 of 39 bundles · the verdict diverged in one test slot)
Edition indication	Corresponding specification Part 1 v1.4 · Conformance Statement v1.2 (verdicts · figures identical)	Matching editions across the bundle
Terminology · notation	‘edition number’ (판번) → ‘version’ (버전) · double angle brackets → single quotation marks · particles written without a space after Latin letters and numerals	2026-10-19 terminology decision · domestic notation norm for external documents

PAVS Part 3 Reference Implementation & Reproduction · corresponding specification Part 1 v1.5 · 3 demonstrations executed 2026-08-25 (Demonstrations B and C: video, verdicts and logs preserved · Demonstration A: result data and reproduction commands preserved) · original: this Markdown; the docx is a rendered output · previous issue 2026-10-20 (v0.7) · draft v0.8 · 2026-10-20.