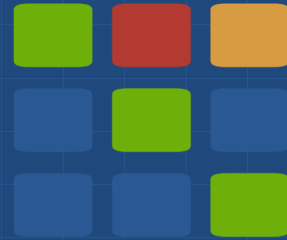


PAVS

Physical AI Verification & Scoring

물리 채점 · 검증 · 성과 보고 공개 규격 스위트

OPEN SPECIFICATION SUITE



Part 3 — 참조 구현·재현 부록 (informative)

발행본 v0.4 · 2026. 10. 2.

발행: (주)이머시브캐스트 · 제안 단계의 공개 규격(공식 표준화 기구의 표준이 아님)

immersivecast.com/pavs

목 차

1. 참조 구현의 위상

2. 참조 구현의 기전 — 사슬 5의 구현 형태

3. 시연 3종과 재현 (2026-08-25 실행)

- 3.1 시연 A — 모델 진본성: 제원만 바꾸면 판정이 정반대
- 3.2 시연 B — 채점 사슬 전체: 한 번의 시험, 다섯 개의 사슬
- 3.3 시연 C — 채점기의 음성 대조군: 채점기 자신도 시험을 본다
- 3.4 시연 범위 — 세 시연이 실행 가능성을 보인 단언

4. 런 패키지의 표준 구성

5. 실행 식별 정보의 위치

문서 지위: 발행본 v0.4 · 2026-10-02 · 공개 위치 immersivecast.com/pavs · 대응 규격 Part 1 v1.3 · 이 문서는 informative(설명적)이다. **성격:** Part 1 본체·부속서에 규정한 사항의 실행 가능성과 재현 방법을 설명하며, 표준의 일부는 아니다. **판 이력:** v0.4(2026-10-02 · 발행본 — 시연 관련 단언과 총족 판정 분리 · 시연 범위 밖 단언 재집계 · 측정일 병기 · 실행 시간·내부 경로 삭제 · 외부 AI 모듈 재채점 반영) · v0.3(2026-08-27 · 시연 범위 절 신설 · 증거 수준 표기) · v0.2(기전 서술을 단언 72 건 전수 자체 점검 결과에 맞추어 정정). **서술 규약:** 참조 구현은 기전(메커니즘) 중심으로 기술한다. 공개 범위 = 기종 실명 공개 · 실행 명령·내부 경로 비공개 · 재현 패키지 요청 시 제공. 시험 식별자·실행 명령은 내부 재현 기록에서 관리하며 공동 시범 시험 상대방에게 제공한다. 본문의 수치는 시연 결과의 물리량에 한하며 측정일을 병기한다.

1. 참조 구현의 위상

이 규격에는 참조 구현이 존재한다. 참조 구현의 존재 이유는 두 가지다.

1. **실행 가능성** — 목표는 Part 1의 시험 단언 72 건 전부가 문서 위의 요구가 아니라 실행 가능한 검사임을 실제 구동·채점·보고의 실물로 보이는 것이다. 현재 세 시연이 실행 가능성을 보인 범위는 72 건 가운데 18 건이다(관련 단언 목록 = §3.4). 단언별 총족 현황은 적합성 진술 v1.1(공개)을 따르며, 그 판정은 총족 7/72 · ★핵심 3/8 이다(2026-09-04 부속서 C 시험 · REP-15 는 2026-09-06 재시험).

2. **재검산의 기준점** — 규격 채택자가 자기 환경에서 같은 절차를 실행해 결과를 재검산할 수 있는 기준을 제공한다. 참조 구현은 특정 물리 엔진·특정 로봇에 종속되지 않는 설계를 따르며, 피시험체의 수용은 선언적 기술(매니페스트) 추가로 이루어진다.

이 부록은 참조 구현을 **익명으로** 다룬다 — 규격 본문의 중립을 지키기 위해서다. 이 구현이 제품으로서 갖는 형상·경제·제공 형태는 Part 4(솔루션 프로파일 · 미공개 — 차기)에서 이름을 밝혀 서술한다.

2. 참조 구현의 기전 — 사슬 5의 구현 형태

채점 신뢰의 5개 요건



운영 지원: 실패 기록부 · 소급 감사 · 기록 규율 (10장)

사슬	구현 기전
문항	시나리오의 도메인 언어 문서로 기술되고, 그 문서의 단언문이 판정 커널의 실행 문항이 된다. 표준 셀 사양(피시험체 도달 반경에 비례)과 실패 대조군이 코퍼스로 배포된다. 자기완결성 검사는 방산 교전 시험의 접수 단계에 도입되어 있고 로봇 팔·민수 시험으로의 적용은 진행 중이다 — 일부 시나리오는 초기 조건을 제출 계층의 인자로 받는다
계속	채점에 쓰이는 값에 측정 증명이 동반되어 미증명 값이 커널에

	서 거부되고, 결측은 값으로 치환되지 않는다(저장층 제약으로 강제). 출처 범주는 «실측»·«기록됨»에 «선언값»(런 설정에서 나온 사실 · 관측 아님 · 2026-08-31)과 «불가능 값 제외»(계측기는 돌아으나 물리 한계 밖인 값을 측정에서 빼고 사유를 기록 · 2026-09-04)가 더해졌고, 발산처럼 되돌릴 수 없는 사건 뒤의 값은 «봉인»되어 측정으로 다시 올라가지 않는다(2026-09-14). 계측 장소(시뮬레이션) 표기가 함께 실린다. 지시값 구분 미비(PAVS-MEA-01 미충족) — 명령값을 실측과 가르느 범주와 기본값·추정 범주는 아직 없다. 시험 조건의 실효값은 실행 중 되읽기로 대조되나 결과 문서에는 아직 병기되지 않는다
판정	3 상 판정 커널 — 합격은 필수 조건 전부의 실측 충족으로만 산출되고, 미측정은 합격으로 계상되지 않으며(필수 문항의 계측 부재는 «계측 부재» 사유의 불합격 · 비필수 문항은 미측정 보존 · 승격 차단이 커널·집계·저장층 세 지점에 있다), 더 엄격한 판정 계층이 산출한 불합격은 관대한 표시로 덮이지 않는다. 물리 우회 산출의 채점 배제는 로봇팔 시험(집기·놓기 등)의 문항으로 구현되어 있고 이동·비행·해양·보행 시험으로의 확대는 진행 중이다. 외부 AI 모듈의 결과가 런 뒤에 도착하면 같은 런에 재채점되며, AI 모듈 행을 뺀 나머지 판정 행과 총평이 바뀌지 않은 경우에만 저장된다(바뀌면 저장 거부 · 원상 복구). 재진입 경로가 없는 결합(모델 파일을 런 안에 적재하는 방식)에 «도착 후 채점»을 선언하면 런이 생성되기 전에 거부된다(2026-09-30~10-01 확인)
증거	음성 대조군 셀이 회귀 스위트에 상주하고(기대 판정과 자동 대조되며, 음성 대조군이 합격으로 바뀌면 «채점기 결함 의심»을 낸다), 골든셋 재채점 항등과 판정 이동 전이 행렬이 관대해진 이동을 차단하며, 저장 산출물에서 판정을 단일 명령으로 재현할 수 있다. 시드의 결과 기록은 로봇팔 시험에만 있고 방산·이동로봇·해양 시험에는 없어 재현 구성의 고정은 부분적이다. 영상·궤적·이벤트 로그의 산출 의무가 강제된다
보고	성과 문서는 판정 결과에서 직접 생성되고, 측정 신뢰도 절 — 판정 커버리지·값의 출처·«이 값을 그대로 믿으면 안 되는 것» — 을 문서가 스스로 신는다. 커널 총평이 먼저 실리고 문서 자체의 임계표는 «판정이 아님»으로 강등 표기된다. 발행 이력은 추가 전용 원장으로 관리되며, 원장의 내용 변조 탐지는 코어 검증기에 있고 영속 계층으로의 확대는 진행 중이다

비고 — 이 표의 각 항목은 참조 구현의 **현재 상태**를 서술한다(2026-08-26 전수 자체 점검 기준 · 이후 변경분은 날짜 병기). 규격 요구 72 건에 대한 판정과 미충족 목록은 적합성 진술 v1.1(공개)에 있다. 구현 서술이 판정보다 앞서 있으면 그것이 곧 이 규격이 금지하는 행동이므로, 두 문서는 같은 판정을 가리켜야 한다.

3. 시연 3 종과 재현 (2026-08-25 실행)

세 시연은 각각 규격의 다른 층을 실물로 보인다. 시연 B·C는 참조 구현의 정상 운용 경로(시험 접수 API·회귀 스위트)로 실행되어 산출물 전량(영상 포함)이 보존되어 있고, 시연 A는 모델 진본성 대조용 독립 검사 스크립트(결정적·CPU 전용)로 실행되어 결과 데이터와 재현 명령이 보존되어 있다.

각 시연의 «실행 가능성을 보인 관련 단언»은 그 시연이 실행으로 보인 단언이며 **충족 판정이 아니다**. 단언별 판정은 §3.4 와 적합성 진술 v1.1 을 따른다.

3.1 시연 A — 모델 진본성: 제원만 바꾸면 판정이 정반대

동일 엔진·동일 시나리오에서 제원만 변경 시 판정 상반 (시연 A)

	0.5 kg 들기 (실기는 가능)	손목 150° (실기는 불가능)
㉠ 공개 제원	거짓 불합격	거짓 합격
㉡ 수치만 수정	㉠와 비트 동일	㉠와 동일
㉢ 확정 제원	합격 (186 mm)	불합격 (119.5°)

실행 가능성을 보인 관련 단언: **PAVS-VRF-05**(모델 진본성 대조) · **PAVS-VDT-08**(수행 불가 지령 거부) · **PAVS-VRF-14**(재현성) — 충족 판정 아님(§3.4)

절차: 동일한 6 축 경량 팔의 동일한 두 과제(제조사 정격 이내의 0.5 kg 들어올리기 · 실기 한계 밖의 손목 150° 정렬)를, 같은 물리 엔진·같은 초기조건에서 로봇 제원만 세 가지로 바꿔 채점한다 — ㉠공개 모델 저장소의 제원 ㉡공개 모델에서 구동기 수치 하나만 제조자 값으로 수정 ㉢제조사 확정 제원.

관찰 결과 (2026-08-25 측정 · CPU 전용 · 결정적 재현 — 같은 환경에서 같은 값):

조건	0.5 kg 들기 (실기는 가능)	손목 150° (실기는 불가능)
㉠ 공개 제원	거짓 불합격 — 구동 0.79 N, 대상물 상승 0 mm	거짓 합격 — 149.8° 도달
㉡ 수치만 수정	㉠와 동일(비트 단위) — 제원 수정 ≠ 물리 변경	㉠와 동일
㉢ 확정 제원	합격 — 구동 50 N, 상승 186 mm	불합격 — 119.5°에서 제한

이 실행의 수치는 선행 실측 기록 — 서로 다른 기종의 계산기에서의 실행을 포함한다 — 과 일치했다(교차 계산기 동일성은 선행 실측에서 확인된 결정적 재현). 채점 결과에 대한 이의가 «재실행»으로 해소될 수 있는 성질이다.

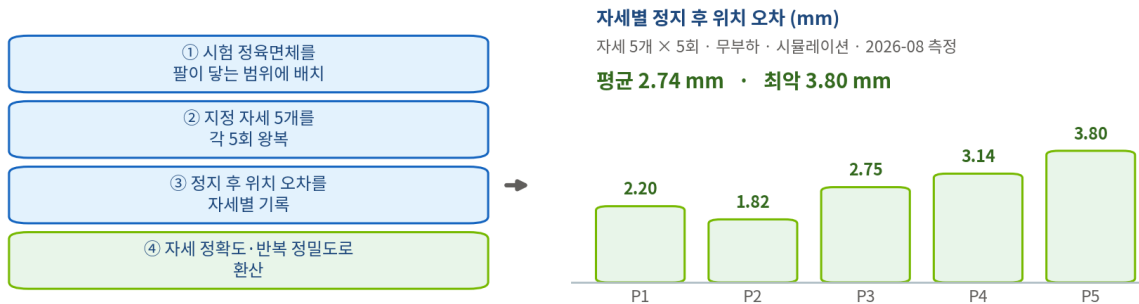
착지: 제 3 자를 채점하는 체계에서 모델 진본성 검증은 채점의 전제다. 들어온 모델이 진짜인지 확인하지 않으면, 어떤 성적도 아무것도 증명하지 않는다.

재현 패키지는 요청 시 제공한다. 실행 명령과 내부 경로는 공동 시범 시험 상대방에게 제공한다(문의 contact@immersivecast.com). 패키지는 실행 스크립트·기준값·기준 환경 판번(물리 엔진 판번·공개 모델 저장소 커밋)으로 구성되며, 공개 물리 엔진과 공개 모델 저장소를 받은 뒤 CPU 전용 · 명령 1 회로 실행된다. 검증 모드는 산출을 기준값과 대조해 완전히 일치할 때만 성공 종료 코드를 내고, 다르면 어느 값이 어떻게 다른지 항목별로 보고한다.

PAVS-VRF-14 의 요구(채점 결과에 대한 이의는 재실행으로 해소될 수 있어야 한다)를 이 시연 범위에서 실행으로 보인 것이다. VRF-14 자체는 적합성 진술에서 미충족이다(재현 정보의 고정이 로봇팔 시험 기준에 한정).

3.2 시연 B — 채점 사슬 전체: 한 번의 시험, 다섯 개의 사슬

표준이 정한 측정 항목을 시뮬레이션에서 측정



5개 자세의 오차가 균일함. 특정 자세만 이탈하지 않는 것이 측정 조건 정상 지표임.

△ 본 값은 ISO 9283 준거 성능값이 아님(무부하·5회·규격 절차 일부 미구성). 용도는 동일 스택 내 상대 비교와 판정 임계의 분모로 한정

표준 자세 측정 절차가 시뮬레이션 안에서 수행되는 방식과 그 산출 — 다섯 자세를 각 다섯 번 왕복하고 정작 잔차를 자세마다 기록한다.

실행 가능성을 보인 관련 단언: **PAVS-ITM-01**/07/08(표준 셀·코퍼스·조건 있는 기준선) · MEA-01/07(출처·실 효값) · VDT-01/02/06(3 상·긍정 증거·권위 계층) · REP-01/02/03(증거 수준·귀속·커버리지) · VRF-14(재현 성) — 충족 판정 아님(\$3.4)

절차: 표준 셀 사양의 로봇팔 집기·놓기 시험 1 건을 고정 시드로 접수 API 에 제출한다. 영상 기록은 의무이며, 완료 후 성과 문서(PDF)를 판정 결과에서 생성한다.

관찰 결과 (2026-08-25 측정 · 측정 후 30 일 경과 — 재측정 대상(아레나 기준 2026-10-02 재고정)):

- 셀 기준 8 항목 전부 충족 — 들어올림 78.9 mm(임계 50 초과) · 배치 오차 1.0 mm · 유지 12.6 s · 낙하 0 · 궤적 반전 0.0/m · 최고 속도 353.7 mm/s(선언 상한 이내)

- 궤적 반전 측은 2026-09-05 부터 3 상 판정(안정·불안정·미측정)으로 기술한다 — 호길의 대표본으로 센 사건 수를 들여다본 경로 길이로 나눈 값의 포아송 95 % 구간을 임계 8/m 와 비교한다. 짧은 경로의 «0 건» 은 통과가 아니라 미측정일 수 있으며, 자세 고정 셀의 반복 실행은 표본이 아니라 복제다(사례집 D11). 위치 정확도 시험은 ISO 9283 특성 정의로 측정 · 정격 부하 질량 적용(2026-10-02 부터) · 공식 시험 아님.

- 동시에, 판정 커널은 불합격을 병기했다 — 자기 간섭 접촉(실결함)이 계측되어, 셀 기준 충족(초록)이 커널 불합격을 덮지 않는 권위 계층이 판정 결과와 성과 문서 양쪽에서 동작함을 보였다. «성공 판정과 커널 판정은 다른 자(尺)이며, 하나만 인용하면 과대 진술»이라는 규율의 실물이다.
- 생성된 PDF 는 «판정(커널): 불합격»을 표제에 직결 표기하고, 측정 신뢰도 절 — 판정 커버리지(충족/미달/미측정 목록)·값의 출처·«이 값을 그대로 믿으면 안 되는 것»(기록되지 않는 채널 목록 포함) — 을 자체 수록했다.
- 같은 시드의 동일 시험이 다른 계산 노드에서 동일 수치를 재현했다(들어올림·배치 오차·속도 모두 일치).

착지: 규격이 요구하는 «양식이 스스로 자기 증거의 급을 말한다»가 자동 생성 문서에서 실제로 성립한다.

3.3 시연 C — 채점기의 음성 대조군: 채점기 자신도 시험을 본다

실행 가능성을 보인 관련 단언: **PAVS-VRF-01**(음성 대조군 상설) · ITM-05/06(변별력·반대 방향 대조군) · REP-12(안전 이상 ≠ 정상) — 충족 판정 아님(\$3.4)

절차: 교전 승인 규율을 의도적으로 위반하는 시나리오(지정 없는 사격 1 건 포함)를 회귀 스위트로 실행한다. 스위트에는 기대 판정(불합격)이 선언되어 있고, 실제 판정과의 대조가 자동으로 이루어진다.

관찰 결과 (2026-08-25 측정 · 측정 후 30 일 경과 — 재측정 대상(아레나 기준 2026-10-02 재고정)): 판정 커널이 정확히 해당 문항 — 지정 없는 아군 사격 계수 1 건 — 으로 불합격을 산출했고(정상 지정 사격 2 건은 통과), 스위

트의 자동 대조가 «음성 대조군이 기대대로 불합격 — 채점기 변별 유지»를 선언했다. 스위트는 부분 실행 산출물을 기준선(baseline)으로 쓰지 말 것을 스스로 경고한다(검사 범위 축소 방지).

착지: 합격만 있는 회귀 스위트는 채점기의 절반만 검증한다. 채점기가 관대해지는 회귀는 음성 대조군만이 잡는다.

3.4 시연 범위 — 세 시연이 실행 가능성을 보인 단언

이 규격은 성과 보고에 «측정한 것과 측정하지 못한 것을 비율이 아니라 목록으로 공개»할 것을 요구한다(**PAVS-REP-03**). 그 요구는 이 부록에도 적용된다. 세 시연이 실행 가능성을 보인 관련 단언은 **72 건 중 18 건**이다. 이는 충족 판정이 아니다 — 이 18 건 가운데 적합성 진술 v1.1 에서 충족인 단언은 아래 표에 «충족»으로 표시한 2 건 (VDT-01★ · VDT-06)뿐이며, 나머지 16 건은 미충족 14 · 판정 불가 2 다.

조항군	전체	시연 관련	관련 단언	적합성 진술 판정
ITM 문항	10	5	ITM-01 · 05 · 06 · 07 · 08	5 건 모두 미충족
MEA 계측	8	2	MEA-01★ · 07	MEA-01★ 미충족 · MEA-07 미충족
VDT 판정	8	4	VDT-01★ · 02★ · 06 · 08	VDT-01★ 충족 · VDT-02★ 미충족 · VDT-06 충족 · VDT-08 미충족
VRF 증거	16	3	VRF-01★ · 05 · 14	VRF-01★ 판정 불가 · VRF-05 판정 불가 · VRF-14 미충족
REP 보고	15	4	REP-01 · 02 · 03 · 12	4 건 모두 미충족
OPS 운영	10	0	—	—
CNF 적합성	5	0	—	—
계	72	18		충족 2 · 미충족 14 · 판정 불가 2

비고 — 판정 기준 = 적합성 진술 v1.1(2026-09-04 부속서 C 시험 · REP-15 는 2026-09-06 재시험). «미충족» 14 건 = 부속서 C 시험 기록이 미충족인 2 건(MEA-01★ · VDT-02★) + 2026-08-26 자체 점검의 «부분» 판정 12 건(부속서 C 재시험 대기). «판정 불가» 2 건(VRF-01★ · VRF-05)은 절차 일부가 런 실행을 요구해 미수행이며, 적합성 진술은 이를 충족으로 처리하지 않고 미충족 목록에 넣는다. ★ = 적합성 주장의 전제가 되는 핵심 단언.

시연 범위 밖 54 건의 성격. 부속서 C 가 지정한 시험 유형으로 나누면 주입 22 건 · 심사 31 건 · 시연 1 건이다. 이 들에 대하여 이 판이 제시하는 것은 시험 방법(부속서 C)과 적합성 진술 v1.1(공개)의 판정이며, **시연 실행은 수행되지 않았다**.

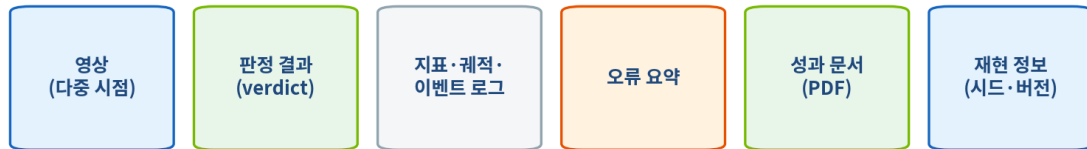
운영·적합성 15 건이 시연 0 건인 이유. 그중 13 건이 심사 유형이다 — 단언의 대상이 개별 실행이 아니라 기간에 걸친 산출물의 성질이므로 시험 1 회의 실행으로는 보일 수 없다(C.0.2). 이는 이들 조항의 성질에서 오는 것이며 시연 설계의 누락이 아니다. **다만 성질이 그렇다는 것이 실행 가능성을 보였다는 뜻은 아니며, 이 판에서 두 조항군은 시연 범위 밖으로 남는다.**

시연 범위 25%(18/72)는 낮은 수치이며 충족 비율도 아니다(충족 = 적합성 진술 7/72). 이 부록이 그 수치를 비율과 목록으로 함께 적는 이유는, 규격이 남에게 요구하는 공개 형식을 규격 자

신의 부록이 지키지 않으면 그 요구가 성립하지 않기 때문이다.

4. 런 패키지의 표준 구성

시험 결과 묶음의 표준 구성: 재현 검증에 필요한 산출물 (Part 3 § 4)



시연 3종(A·B·C)의 산출물을 이 구성으로 보존. 영상이 없는 시연 A는 결과 데이터와 재현 명령으로 대체. 재현 정보 = 로봇팔 시험 기준(PAVS-VRF-14)

재현 검증의 대상이 되는 시험 1 건의 산출물 패키지는 최소 다음을 포함한다 — 이 구성 자체가 Part 1 보고 요구의 실물 형태다.

산출물	역할
영상(다중 시점 + 스트리밍 청크)	물리 사건의 독립 증거(VDT-03) · 사람 확인 가능성
판정 결과(verdict)	3 상 판정·문항별 근거(REP 계열의 원천)
지표·궤적·이벤트 로그	계측 출처·실효값 검증의 재료(MEA 계열)
오류 요약	«에러 없이 틀리는» 것과 «에러»의 구분
성과 문서(PDF)	판정 결과의 렌더링 + 측정 신뢰도 절
AI 모듈 결합 기록	외부 AI 모듈 결합 런 — 선언·결합 상태·담당 조항 귀속 · 미결합 모듈 담당 조항의 미측정 표기 근거
재진입 기록	도착 후 채점 런 — 도착 결과 반영·판정 재계산·저장 여부(AI 모듈 행 외 판정·총평 불변 조건)
재현 정보(시드·구성·코드 버전)	로봇팔 시험 기준 · 다른 시험은 보완 중(PAVS-VRF-14 미 충족)

5. 실행 식별 정보의 위치

공개 수준은 시연마다 다르며, 그 차이를 숨기지 않는다. 공개 범위는 다음과 같이 확정되었다(2026-10-02): 기종 실명 공개 · 실행 명령·내부 경로 비공개 · 재현 패키지 요청 시 제공.

시연	증거 수준	공개되는 것	재현 가능성
A — 모델 진본성	L1 — 실운용 경로가 아닌 독립 검사 스크립트	결과 수치·판정. 재현 패키지 (스크립트·기준값·기준 환경 판번)는 요청 시 제공 · CPU 전용 · 명령 1 회	패키지를 받은 제 3 자가 자기 컴퓨터에서 재현 가능
B — 채점 사슬 전체	L3 — 실운용 경로(접수 API) 종단 실행 + 산출물 직접 확인	결과·판정·영상. 시험 식별자 · 실행 명령·내부 경로는 비공개	공동 시범 시험 상대방에게 별도 제공

C — 음성 대조군	L3 — 실운용 경로(회귀 스 위트) 중단 실행 + 산출물 직 접 확인	//	//
------------	--	----	----

비고 — 증거 수준과 공개 범위는 서로 다른 축이다. **제 3 자 재현에 가장 가까운 시연 A 가 증거 수준은 가장 낮다(L1).**
제 3 자가 돌려 볼 수 있다는 것과 그 실행이 실운용 경로를 거쳤다는 것은 다른 명제이며, 이 표는 둘을 분리해 적는다.

시연 A 를 재현 대상으로 먼저 정한 이유는 값이 커서가 아니라 **비용이 가장 낮기 때문이다** — 결정적이고, CPU 전용
· 명령 1 회이며, 고객 자산도 시험장 구성도 붙어 있지 않다.

*PAVS Part 3 참조 구현·재현 부록 · 대응 규격 Part 1 v1.3 · 시연 3 종 실행 2026-08-25(시연 B·C = 영상·판정·로그 보존 · 시연 A
= 결과 데이터·재현 명령 보존) · 원본 = 본 Markdown, docx 는 렌더 산출물 · 발행본 v0.4 · 2026-10-02.*