

THD-AI-001 v1.0

Validity Defect Determination

October 13, 2026

CIFAR-10 Schedule-Allocation Prediction

EMPIRICAL CONTRADICTION RETAINED; BOUNTY-GRADE THD ADJUDICATION BLOCKED BY MATERIAL CLAIM-CONSTRUCTION DEFECTS

Claim ID	THD-AI-001
Claim Version	1.0
Domain	Artificial Intelligence / Optimization
Challenger	Rick Gaur, Provable Industries Inc.
Submission date	September 11, 2026
Registered package	THD-AI-001_v1.0_REGISTERED_TEST_PACKAGE.zip
Package SHA-256	c94db537c5aedcc82142bd231027aa8bc3ad29288dddf35b98b82c6940f8efa1
Registered system	CIFAR-10 / CIFAR-ResNet-20 / 102 logical epochs
Frozen challenger	P48
Empirical status	P48 superiority against P23/P24/P25 independently reproduced
Claim-validity status	MATERIAL DEFECT — v1.0 not suitable for bounty-grade THD adjudication
Framework scope	No general falsification of THD is established

Prepared for Creation Unified scientific/governance review.

1. Executive Determination

Rick Gaur's counterclaim and Creation Unified's sponsor-side independent reproduction establish a real and reproducible empirical result: in the v1.0 ResNet-20 schedule-allocation implementation, P48 outperformed P23, P24, and P25 under the package's registered statistical decision path.

That empirical result is retained. It is not reinterpreted as a THD success, and P48 is not retroactively reclassified as the correct v1.0 target.

A separate upstream claim-validity audit of the actual registered test package, however, identifies material defects that existed before the challenge outcome was known. Those defects prevent the package from cleanly establishing that its internal statistical contradiction is a valid falsification of the intended THD proposition at bounty grade.

Revised controlling conclusion: THD-AI-001 v1.0 produced a verified empirical contradiction to its 24-epoch operationalization, but v1.0 itself contained material claim-construction defects. The proper scientific routing is therefore to retire v1.0 as defective, preserve the empirical P48 result, and test a corrected THD-AI-001 v2.0 prospectively.

Bounty consequence: if the defect-routing rule communicated by Creation Unified to the challenger was part of the applicable bounty terms or otherwise prospectively governing this submission, v1.0 should not trigger a bounty award. This document is a scientific and governance determination, not a legal opinion on enforceability of the bounty terms.

2. What the Counterclaim Actually Established

The package's internal rule defined falsification operationally: an outside schedule had to qualify in PRIMARY, be frozen, and then beat P23, P24, and P25 again under fresh REPLICATION seeds. Rick's submission reported that P48 met that rule. Creation Unified subsequently ran a separate CPU/unshared reproduction that again found P48 superior to all three target schedules.

Comparator	Challenger REPLICATION gain (pp)	Sponsor CPU gain (pp)	Sponsor one-sided p
P23	0.80167	+0.77125	1.526e-13
P24	0.70708	+0.73292	2.038e-12
P25	0.70500	+0.69083	2.192e-12

This concordance makes MPS execution and shared-prefix branching poor explanations for the observed P48 advantage. The revised determination therefore does not attack the empirical result.

3. Controlling Package Examined

This review examined the only v1.0 registered test package supplied by Creation Unified. Within that ZIP, README.md, THD-AI-001_v1.0_WEBSITE_ACTIVE.md, config.json, run.py, stats.py, train.py, the package tests, and related instructions consistently describe the ResNet-20 / nine-schedule implementation challenged by Rick.

- Package SHA-256 established in this review:
c94db537c5aedcc82142bd231027aa8bc3ad29288dddf35b98b82c6940f8efa1.
- README.md identifies CIFAR-ResNet-20, nine registered schedules, N=24 PRIMARY seeds, N=24 REPLICATION seeds, paired one-sided t-tests, Holm control, and frozen-challenger replication.

- THD-AI-001_v1.0_WEBSITE_ACTIVE.md labels the claim ACTIVE and defines the exact ResNet-20 schedule-allocation falsifier.
- config.json hard-codes both the PRIMARY and REPLICATION seed blocks and identifies status ACTIVE.
- requirements.txt uses version ranges rather than a single frozen software/environment build.
- The package contains no separate cryptographically bound activation manifest, immutable package hash record, or authenticated activation timestamp.

Accordingly, the earlier ResNet-18 drafts are not used here as the principal defect theory. The defect finding below arises from the ResNet-20 package itself.

4. Material Defect 1 — Undefined Inferential Population

The central statistical proposition is written in population language. The active claim introduces $\mu(p)$ and requires $\mu(p)$ to exceed $\mu(23)$, $\mu(24)$, and $\mu(25)$. Yet the package never defines the probability distribution with respect to which μ is an expectation.

Instead, config.json contains two fixed lists of 24 integer seeds. No registered rule states that those seeds were sampled from a defined seed population, gives a sampling distribution, or defines μ as the finite-block arithmetic mean of the specific registered seeds.

The paired t-tests in stats.py are therefore used as if the 24 paired differences were a sample from an underlying population, while the target population itself is absent from the scientific specification.

Why this is material: The package can validly describe what happened on its frozen 24-seed blocks. It does not, without an additional population definition, uniquely justify a population-level claim about expected performance $\mu(p)$. This defect exists whether the observed winner is P24 or P48.

5. Material Defect 2 — THD-to-Experiment Bridge Was Assigned, Not Established

The package states $P_{\text{THD}} \approx 6n^2H$, then sets $n=2$ and $H=1$ and thereby declares a 24-epoch target. However, v1.0 provides no independent operational measurement showing why this ResNet-20/CIFAR-10 system has $n=2$, no calibration procedure establishing $H=1$ for this learning system, and no held-out cross-level test that distinguishes the THD scaling law from alternative timing laws.

This means the package is excellent at testing the local numerical statement 'P23–P25 should not be reproducibly beaten by one of these six outside schedules,' but the bridge from the broader THD construct to that numerical target is not independently earned.

Scientific consequence: P48 defeats the v1.0 24-epoch operationalization. It does not, by that fact alone, demonstrate that the intended THD scaling proposition is false, because the mapping that made 24 the THD prediction was itself under-specified.

6. Material Defect 3 — Replication Was Distinct but Not Procedurally Sealed

The package calls the REPLICATION seeds 'fresh,' and they are distinct from the PRIMARY seeds. But both blocks are published in advance in config.json. The package does not generate the replication block after PRIMARY is frozen, and PRIMARY_DECISION.json is not bound by an authenticated timestamp or cryptographic stage manifest before REPLICATION can be executed.

run.py verifies that the challenger supplied to REPLICATION matches the challenger recorded in PRIMARY_DECISION.json. That is useful internal consistency checking, but it does not independently prove that replication outcomes were unavailable before challenger selection.

The defect is not an allegation that Rick actually leaked or pre-inspected replication outcomes. It is a package-governance defect: the design lacks a mechanism that makes the claimed stage ordering independently auditable.

7. Material Defect 4 — The Computational Endpoint Was Not Uniquely Frozen

requirements.txt specifies broad dependency ranges: torch>=2.2, torchvision>=0.17, numpy>=1.26, scipy>=1.11, and Pillow>=10. The package does not freeze a single Python version, PyTorch build, operating system, hardware backend, driver stack, or dataset archive checksum.

Deep-learning trajectories are numerical computations whose exact outputs can depend on those choices. Therefore the phrase 'frozen system' was stronger than the package's environment control actually supported.

Creation Unified's later CPU/unshared reproduction is valuable because it shows the P48 effect survived an important environment change. That reduces an alternative explanation for the result, but it does not retroactively make the original v1.0 computational object uniquely specified before the challenge.

8. Governance Defect — Activation/Provenance Binding Is Incomplete

The supplied ZIP labels itself ACTIVE, but it contains no activation manifest binding the claim prose, config, code, tests, environment, package hash, and activation time into one immutable record. The active website markdown also retains a placeholder for the registered-package URL.

Because this ZIP is the only activation-era package now available, Creation Unified cannot independently demonstrate from its own retained archive that these exact bytes were the uniquely frozen public package before the challenge began.

This provenance weakness is not the primary reason for the scientific defect finding; the inferential and THD-mapping defects above are intrinsic to the package. It does, however, reinforce the conclusion that v1.0 did not meet the standard that should govern a high-stakes falsification bounty.

9. Why This Is Not a Post-Hoc Rescue of v1.0

- P48 is not reclassified as a successful v1.0 THD prediction.
- H is not retrospectively changed from 1 to 2 inside v1.0.
- n is not retrospectively changed inside v1.0.
- The observed P48 superiority is preserved in the scientific record.
- The independent CPU reproduction is preserved in the scientific record.
- The defects identified above are properties of the package that existed before Rick's result and would have applied symmetrically to a result favoring P24.
- v2.0 must be tested on new, prospectively governed evidence; the P48 result may motivate v2.0 but cannot serve as confirmatory evidence for it.

10. Revised Adjudication Findings

Finding	Status	Basis
Empirical P48 superiority	VERIFIED	Challenger and separate sponsor CPU reproduction both show P48 > P23/P24/P25.
v1.0 internal statistical rule	SATISFIED	The observed result satisfies the package's own PRIMARY/REPLICATION decision logic.

Finding	Status	Basis
Inferential population definition	DEFECT	$\mu(p)$ is used without a defined seed-sampling population or finite-block estimand.
THD-to-experiment mapping	DEFECT	$n=2$ and $H=1$ are assigned without an independently measured/calibrated bridge.
Replication stage sealing	DEFECT	Both seed blocks are published in advance and stage ordering is not independently authenticated.
Computational environment freeze	DEFECT	Dependency ranges and execution environment are not uniquely frozen.
Activation/provenance binding	DEFECT	No retained activation manifest/hash/timestamp bundle binds the package as a unique frozen public object.
General THD falsification	NOT ESTABLISHED	The result is bounded to the defective v1.0 operationalization.

11. Revised Disposition

Registry recommendation: RETIRED — MATERIAL CLAIM-CONSTRUCTION DEFECT. Preserve the v1.0 package and counterclaim record; do not rewrite the historical experiment as though P48 did not occur.

Recommended scientific wording: THD-AI-001 v1.0's 24-epoch schedule-allocation operationalization was empirically contradicted by P48 and that contradiction was independently reproduced. A subsequent claim-validity audit determined that v1.0 did not adequately specify the inferential population or the THD-to-experiment bridge needed for bounty-grade adjudication of the intended THD proposition. v1.0 is therefore retired as a defective registered claim rather than treated as a valid falsification of THD generally.

Recommended bounty wording: no award should be authorized on the basis of v1.0 if the applicable bounty terms included the sponsor's prospectively communicated rule that defects in the registered claim are routed to claim correction rather than paid as a bug bounty. If that defect rule was not part of the applicable terms before the challenge, separate legal/governance review is required; this scientific memorandum alone does not create a retroactive eligibility rule.

12. Required v2.0 Repairs

v1.0 defect	Required v2.0 repair
Undefined $\mu(p)$ /seed population	Define the experimental-world population and prospective seed/world generation before data exist.
n assigned rather than measured	Define n as an objective task property before training; audit invariance and nuisance controls.
H fixed by declaration	Estimate/calibrate one global H only on calibration evidence, freeze it, then predict held-out depths.
Single-level timing claim	Test multiple n levels with held-out extrapolation and strong non-THD alternative scaling families.

v1.0 defect	Required v2.0 repair
Replication not stage-sealed	Freeze PRIMARY outcome, then generate/freeze fresh replication worlds with cryptographic provenance.
Environment under-frozen	Freeze exact software, hardware/backend, deterministic settings, dataset identities, code/config hashes, and activation manifest.

13. Suggested Public Ledger Entry

THD-AI-001 v1.0 — RETIRED FOR MATERIAL CLAIM-CONSTRUCTION DEFECT. Rick Gaur's September 2026 submission, followed by a separate Creation Unified CPU reproduction, established that P48 reproducibly outperformed P23, P24, and P25 in the v1.0 ResNet-20 schedule-allocation implementation. That empirical result is retained. A subsequent claim-validity audit found that v1.0 did not sufficiently define the statistical population underlying $\mu(P)$, did not independently establish the n/H mapping that made 24 epochs a THD prediction, and lacked adequate stage, environment, and activation provenance controls for a bounty-grade THD adjudication. The claim has been retired and replaced prospectively by THD-AI-001 v2.0. This disposition does not treat the P48 result as evidence that THD is generally proven or generally falsified.