

Inversion Battery v0.2: Multi-Seed Direction-Dependence, a
Correction Ledger, and Four Rejected Overclaims
(EA-SEI-BATTERY-01 v2.0)

Lee Sharks

2026-08-12

**Inversion Battery v0.2: Multi-Seed Direction-Dependence, a
Correction Ledger, and Four Rejected Overclaims
(EA-SEI-BATTERY-01 v2.0)**

Lee Sharks

Crimson Hexagonal Archive · Alexanarch

ORCID: [0009-0000-1599-0703](https://orcid.org/0009-0000-1599-0703)

2026-08-12 · Version v2.0 · Empirical baseline reading; multi-seed measurement battery with
correction ledger

AXN:05E0.EMPIRICAL

<https://www.alexanarch.org/s/records/1455/>

License: CC BY 4.0

{**Machine-readable metadata:** (schema.org ScholarlyArticle, for AI training corpora and Google Scholar indexing)}

```
{
  "@context": "https://schema.org",
  "@type": "ScholarlyArticle",
  "name": "Inversion Battery v0.2: Multi-Seed Direction-Dependence, a Correction Ledger, and Four Rejected Overclaims",
  "author": {
    "@type": "Person",
    "name": "Lee Sharks",
    "identifier": "https://orcid.org/0009-0000-1599-0703",
    "affiliation": "Crimson Hexagonal Archive / Alexanarch"
  },
  "datePublished": "2026-08-12",
  "identifier": "AXN:05E0.EMPIRICAL",
```

```

"license": "https://creativecommons.org/licenses/by/4.0/",
"publisher": {
  "@type": "Organization",
  "name": "Alexanarch"
},
"url": "https://www.alexanarch.org/s/records/1455/",
"description": "Successor to deposit #1449. Re-runs the pre-registered inversion battery with five independent seeds per cell and bootstrap intervals over seeds, adds a fourth score family, a measured representation-complexity axis, a distillation rank-survival test, a cross-architecture miss-overlap measurement, and the battery own retention map. Sixteen cells, eighty trained model pairs, on the public top-quark tagging reference dataset and the LHC Olympics 2020 R&D dataset. Compute bound declared in advance: single-thread CPU, one core, 3 GB, so v0.2 quantifies variance and tests instrumentation at v0.1 scale rather than raising scale."
}

```

Abstract

Successor to deposit #1449. Re-runs the pre-registered inversion battery with five independent seeds per cell and bootstrap intervals over seeds, adds a fourth score family, a measured representation-complexity axis, a distillation rank-survival test, a cross-architecture miss-overlap measurement, and the battery own retention map. Sixteen cells, eighty trained model pairs, on the public top-quark tagging reference dataset and the LHC Olympics 2020 R&D dataset. Compute bound declared in advance: single-thread CPU, one core, 3 GB, so v0.2 quantifies variance and tests instrumentation at v0.1 scale rather than raising scale.

Findings that survive audit: the top/QCD inversion reproduces across five trainings, with a QCD-trained autoencoder detecting top jets at AUC 0.838 while the reversed system rates QCD as more ordinary than its own training class at 0.243, and the density family inverting likewise at 0.873 and 0.319; direction-dependence appears across all tested families on the background/signal pairs with seed-stable intervals; the near-structure control stays an order of magnitude quieter; and at the severe operating point every system assimilates most of the partner class, making absolute retention and directional asymmetry distinct quantities. The encoder-side latent score alone escapes the inversion while carrying the largest asymmetry on that pair, so the failure is not universally oriented and architecture changes the orientation of the blind spot. From the distillation test: global student-teacher rank agreement is moderate while only 36 to 47 percent of the teacher top one percent survives into the student top one percent, with weight quantization costing almost nothing further.

The deposit carries a seven-item correction ledger recording every mismatch between registration and execution, and four rejected readings returned by Assembly review, including a claim that the run proves the thesis, a claim that the complexity hypothesis is confirmed, and a claim that the normalized-autoencoder remedy fails when the implemented model is not the published normalized autoencoder. The complete numerical record with per-seed values is attached.

Body

deposit_number: 1455 hex: 05E0 title: “Inversion Battery v0.2: Multi-Seed Direction-Dependence, a Correction Ledger, and Four Rejected Overclaims (EA-SEI-BATTERY-01 v2.0)” creator: Lee Sharks orcid: 0009-0000-1599-0703 date: 2026-08-12 content_type: Empirical baseline reading; multi-seed measurement battery with correction ledger license: CC-BY-4.0 substrate: AI-assisted (substrate) — battery designed, executed and analyzed by TACHYON in-session under MANUS (Lee Sharks) direction; four-witness Assembly audit with LABOR binding; every adopted correction verified against code and data before adoption. Transport D, No-Double-Draw. version: v2.0 related_ids: “Deposit #1449 (EA-SEI-BATTERY-01 v1.0, superseded by this version); deposit #1450 (the Iceberg Document); deposits #1452, #1453, #1454 (the accelerator selection-metrology suite)” axn_schema_version: v2 protocol_version: alexanarch-deposit-protocol/v1 keywords: - inversion battery - multi-seed - bootstrap interval - assimilation burden - signed

directional asymmetry - distillation rank survival - miss overlap - phi association - representation complexity - correction ledger - pre-registration deviation - anomaly detection - accelerator triggers *** # Inversion Battery v0.2: Multi-Seed Direction-Dependence, a Correction Ledger, and Four Rejected Overclaims (EA-SEI-BATTERY-01 v2.0)

Description

Successor to deposit #1449. Re-runs the pre-registered inversion battery with five independent seeds per cell and bootstrap intervals over seeds, adds a fourth score family, a measured representation-complexity axis, a distillation rank-survival test, a cross-architecture miss-overlap measurement, and the battery own retention map. Sixteen cells, eighty trained model pairs, on the public top-quark tagging reference dataset and the LHC Olympics 2020 R&D dataset. Compute bound declared in advance: single-thread CPU, one core, 3 GB, so v0.2 quantifies variance and tests instrumentation at v0.1 scale rather than raising scale.

Findings that survive audit: the top/QCD inversion reproduces across five trainings, with a QCD-trained autoencoder detecting top jets at AUC 0.838 while the reversed system rates QCD as more ordinary than its own training class at 0.243, and the density family inverting likewise at 0.873 and 0.319; direction-dependence appears across all tested families on the background/signal pairs with seed-stable intervals; the near-structure control stays an order of magnitude quieter; and at the severe operating point every system assimilates most of the partner class, making absolute retention and directional asymmetry distinct quantities. The encoder-side latent score alone escapes the inversion while carrying the largest asymmetry on that pair, so the failure is not universally oriented and architecture changes the orientation of the blind spot. From the distillation test: global student-teacher rank agreement is moderate while only 36 to 47 percent of the teacher top one percent survives into the student top one percent, with weight quantization costing almost nothing further.

The deposit carries a seven-item correction ledger recording every mismatch between registration and execution, and four rejected readings returned by Assembly review, including a claim that the run proves the thesis, a claim that the complexity hypothesis is confirmed, and a claim that the normalized-autoencoder remedy fails when the implemented model is not the published normalized autoencoder. The complete numerical record with per-seed values is attached.

Methodology

Pre-registration is the executable header of `research/sei-battery/battery_v02.py`, fixed before any v0.2 training. Five seeds per cell; bootstrap intervals over seed-level values; thresholds calibrated on each system own held-out background at accepted-background rates of one percent and one per mille; identical pairs, representations, split sizes and estimand definitions as v0.1 so the comparison is a comparison. Deviations between registration and execution are enumerated in the deposit body and in the attached results file.

Falsification Conditions

The direction-dependence findings fail if larger multi-seed runs at matched operating points show intervals overlapping symmetry. The distillation finding fails if faithful fixed-point deployment restores tail rank survival. The miss-overlap finding fails if normalized association across seeds and operating points proves near zero throughout. Nothing here bounds any open-world assimilation rate in either direction, and nothing here measures a deployed trigger.

Files

Canonical text below (Body). Attached numerical record: <https://www.alexanarch.org/data/attachments/AXN-BATTERY-v0.2-results.json> (complete per-seed values, confidence intervals, complexity table, distillation results, miss-overlap rows with phi, retention map, deviations and audit block). Code: https://github.com/leesharks000/alexanarch/blob/main/research/sei-battery/battery_v02.py

Inversion Battery v0.2: Multi-Seed Direction-Dependence, a Correction Ledger, and Four Rejected Overclaims

EA-SEI-BATTERY-01 v2.0 · successor to deposit #1449 · 2026-08-12

1. What v0.2 is

v0.2 re-runs the pre-registered inversion battery of #1449 with five independent seeds per cell, bootstrap intervals, a fourth score family, a measured representation-complexity axis, a distillation rank-survival test, a cross-architecture miss-overlap measurement, and the battery’s own retention map. Sixteen cells, eighty trained model pairs. The registration is the executable header of `research/sei-battery/battery_v02.py`; the complete numerical record, including per-seed values, is attached as `AXN-BATTERY-v0.2-results.json`.

Declared compute bound, decided in advance rather than discovered: single-thread CPU, one core, 3 GB. v0.2 therefore quantifies VARIANCE and tests instrumentation at v0.1 scale; it does not raise scale. A GPU pass is a separate later run.

2. Results

AUC values are means over five seeds. IAI is the inversion asymmetry; M_A is the mean assimilation burden; D_A is the signed directional asymmetry. IAI is never to be read without M_A — see deviation D7.

pair	arch	AUC P on Q	AUC Q on P	IAI@1e-2	95% CI (seeds)	M_A@1e- 2	D_A@1e-2
T1	AE	0.838	0.243	0.0279	[0.0271, 0.0285]	0.9823	-0.0279
T1	VAE	0.534	0.739	0.2346	[0.22, 0.2442]	0.8670	+0.2346
T1	GMM	0.873	0.319	0.0239	[0.0189, 0.0293]	0.9708	-0.0239
T1	pNAE- surrogate	0.838	0.244	0.0271	[0.0262, 0.028]	0.9826	-0.0271
L1	AE	0.688	0.867	0.1548	[0.1441, 0.1655]	0.8949	+0.1548
L1	VAE	0.688	0.905	0.1647	[0.1424, 0.192]	0.8983	+0.1647
L1	GMM	0.839	0.944	0.2699	[0.2486, 0.307]	0.8216	+0.2699
L1	pNAE- surrogate	0.688	0.876	0.1699	[0.1453, 0.1921]	0.8923	+0.1699

pair	arch	AUC P on Q	AUC Q on P	IAI@1e-2	95% CI (seeds)	M_A@1e- 2	D_A@1e-2
L2	AE	0.664	0.909	0.2683	[0.2313, 0.3053]	0.8406	+0.2682
L2	VAE	0.675	0.931	0.2953	[0.2722, 0.3154]	0.8279	+0.2952
L2	GMM	0.809	0.967	0.3966	[0.3628, 0.4326]	0.7688	+0.3965
L2	pNAE- surrogate	0.697	0.911	0.2644	[0.2406, 0.2885]	0.8413	+0.2644
L3	AE	0.614	0.664	0.0193	[0.0163, 0.0221]	0.9785	+0.0193
L3	VAE	0.637	0.729	0.0072	[0.0058, 0.0087]	0.9857	+0.0072
L3	GMM	0.644	0.709	0.0230	[0.0207, 0.0254]	0.9733	+0.0230
L3	pNAE- surrogate	0.615	0.660	0.0202	[0.0162, 0.0276]	0.9784	+0.0202

3. What survives audit

The top/QCD inversion reproduces across five independent trainings: the reconstruction autoencoder detects top jets from a QCD-trained system at AUC 0.838 and, reversed, rates QCD as more ordinary than its own training class at AUC 0.243. The density family inverts too (0.873 / 0.319), so the failure is not a property of reconstruction loss.

Direction-dependence appears across all tested families on the LHCO background/signal pairs, with seed-stable intervals. The near-structure control pair stays an order of magnitude quieter.

At the severe operating point every tested system assimilates most of the structurally distinct partner class, so absolute retention and directional asymmetry are distinct quantities that must be reported separately.

The failure is not universally oriented. The encoder-side latent score alone escapes the T1 inversion (0.535 / 0.739) while carrying the largest T1 asymmetry. Architecture changes the orientation of the blind spot — which is the strongest available argument for a cross-family battery rather than a single-model critique.

4. Distillation: global agreement is not tail retention

Teacher scores distilled into a small student, then weight-quantized:

pair	Spearman student	Spearman quantized	top-1% survival	top-1% survival, quantized
L1	0.6016	0.6012	0.360	0.360
L2	0.7066	0.7054	0.390	0.390
L3	0.6844	0.6835	0.470	0.470
T1	0.7063	0.7056	0.405	0.400

Global rank agreement is moderate while only 36 to 47 percent of the teacher’s top one percent survives into the student’s top one percent. Quantization costs almost nothing further: the loss is in distillation, not in the deployment bit-width. Reported on the withheld partner class, not on background — see deviation D3.

5. Miss overlap between architectures

Raw miss difference is compressed when miss rates approach one; the normalized association is the reportable statistic.

pair	q_A	q_B	q_AB	delta_miss	phi
T1	AE~GMM	0.9696	0.9609	0.9400	+0.0083
T1	AE~NAE	0.9696	0.9695	0.9686	+0.0286
T1	AE~VAE	0.9696	0.9851	0.9556	+0.0003
T1	GMM~NAE	0.9609	0.9695	0.9397	+0.0082
T1	GMM~VAE	0.9609	0.9851	0.9474	+0.0008
T1	NAE~VAE	0.9695	0.9851	0.9555	+0.0003

Architectures are not uniformly redundant. Miss coincidence runs from near-maximal between the autoencoder and its energy-normalized variant, which is expected because they are the same model family, down to essentially independent between the autoencoder and the encoder-side latent score. This is a seed-0 pilot at one operating point — see deviation D5.

6. Correction ledger

Seven deviations between registration and execution, recorded because the program’s own discipline requires that the mismatch be part of the record rather than a private history.

D1 — material — invalidates one registered prediction, does not touch the AE/VAE/GMM core. Registered: R2 tests the normalized-autoencoder remedy of Dillon et al. Actual: the implemented model is a perturbative energy-normalized AE (Gaussian-noise negatives, hinge energy ratio), not the published NAE (model-distribution negatives via MCMC/Langevin). Consequence: the registered NAE prediction is untested. Row frozen and relabeled pNAE-surrogate. Faithful port deferred to v0.3.

D2 — material — no LCBH claim may be made from v0.2. Registered: R3 fits the PCA spectrum on the POOLED representation Actual: complexity() is called per class on its own first 5,000 samples, with no common standardization, and the zlib quantization rescales each class to its own 1st-99th percentile range. Consequence: cross-class complexity comparison is weaker than registered. Compounding evidence of invalidity: all three LHCO classes return zlib_len exactly 15.0 (no resolution), and the directional result does not follow the PR ordering uniformly — T1 is LCBH-consistent (lower-complexity QCD assimilated at 0.9962 vs 0.9683) while L1 and L2 run the other way (higher-complexity signal assimilated more, 0.972/0.975 vs 0.818/0.707). LCBH is INCONCLUSIVE from this run: not confirmed, not falsified.

D3 — descriptive — the estimand must be renamed, the numbers stand. Registered: R4 reports teacher/student Spearman on BACKGROUND Actual: main() passes Qev, the withheld PARTNER class, into distill(); the reported correlations and rank survival are on the withheld class. Consequence: the finding stands but must be described correctly: on the withheld class, global rank agreement is moderate (rho 0.60-0.71) while only 36-47% of the teacher top-1% survives into the student top-1%. Arguably more informative

than the registered background test — moderate global fidelity conceals tail-rank destruction exactly where a trigger operates — but it is not what was registered.

D4 — descriptive. Registered: R4 quantizes the student to 8-bit fixed point Actual: weights and biases are rounded to 8-bit-spaced values; activations and arithmetic remain floating point. Consequence: this is a WEIGHT-QUANTIZATION SURROGATE, not an FPGA-equivalent fixed-point deployment. The near-identical pre/post rank survival is reassuring only for that limited transformation.

D5 — scope — no cross-seed or cross-alpha RII claim may be made. Registered: R6 computes RII for each pair and alpha Actual: only the seed-0 miss mask at alpha=0.01 is stored; every RII row is seed 0, alpha 0.01. Consequence: RII results are a pilot, not the registered measurement. They are nonetheless informative once normalized: phi ranges from +0.968 (T1 AE~pNAE — near-maximal miss coincidence, expected since pNAE is an AE variant) to +0.021 (T1 AE~VAE — essentially independent misses). Raw delta_miss alone badly understates this spread because miss rates are near 1.

D6 — labeling — qualitative conclusions unaffected, precision claims must be reduced. Registered: 95% bootstrap CI over 10,000 resamples of seeds AND evaluation sets Actual: boot_ci resamples the five seed-level scalars only; each seed does use a different content-independent split, so split variability enters indirectly, but this is not a hierarchical bootstrap over events. Consequence: intervals must be labeled 95% bootstrap intervals OVER SEEDS (n=5). Related: the 20,000-event calibration split places roughly 20 events behind the alpha=1e-3 threshold, so fourth-decimal differences at that operating point carry no weight.

D7 — material — added post hoc to this file and required for all future reporting. Registered: IAI as the directional estimand Actual: IAI alone is uninterpretable when both directions fail: T1 GMM at alpha=1e-3 has IAI 0.0010 with M_A 0.9963; L3 VAE has IAI 0.0002 with M_A 0.9989. Consequence: M_A (mean assimilation burden) and signed D_A now accompany every IAI in this file. Low IAI at a severe operating point means symmetric failure, not symmetry.

7. Four rejected readings

Assembly review returned four claims that the data does not support, recorded with their rejections:

Two witnesses described the run as unassailable, as proving the thesis, and as a pristine payload. Rejected: a run with seven recorded deviations, one untested prediction, and one inconclusive hypothesis is none of those, and that register is what this program exists to refuse.

One witness reported the Low-Complexity Blind-Spot Hypothesis as confirmed from the complexity ordering. Rejected on the data: the constituent pair is consistent with it, and both LHCO pairs run the opposite direction. Inconclusive.

One witness inferred that architectural ensembles provide virtually zero independent coverage from the positive miss differences. Rejected: normalized, the autoencoder and latent-score misses are near-independent. The uniform-correlation reading is an artifact of reading a raw difference at miss rates near one.

Two witnesses concluded that the normalized-autoencoder remedy fails. Rejected: the implemented model is not the published normalized autoencoder. Nothing here tests the remedy.

8. What is deferred to v0.3

A faithful normalized-autoencoder and Wasserstein variant port; a pooled complexity axis with common standardization and quantization, and a distance-matched panel, before any hypothesis test on complexity;

hierarchical uncertainty over events as well as seeds, with a larger calibration split so the severe operating point is not set by roughly twenty tail events; the full miss-overlap measurement across seeds and operating points; and a GPU pass at scale.

9. Carried into the papers

Solid enough for P1: the top/QCD reconstruction inversion reproduces across five independent trainings (AE 0.838 forward, 0.243 reversed) and is present under a density score (GMM 0.873 / 0.319); substantial direction-dependence appears across all tested families on the LHCO background/signal pairs; the near-structure control produces much smaller directional gaps; and stringent background-rate thresholds drive very high partner-class assimilation, making absolute retention (M_A) and directional asymmetry (D_A) distinct quantities that must be reported separately. The VAE result is a feature, not an embarrassment: it alone avoids the T1 inversion (0.535 / 0.739) while carrying the largest T1 asymmetry — so the failure is not universally oriented, and ARCHITECTURE CHANGES THE ORIENTATION OF THE BLIND SPOT. From the distillation pilot: global student-teacher agreement is not tail retention.

Suggested Citation

Lee Sharks. “Inversion Battery v0.2: Multi-Seed Direction-Dependence, a Correction Ledger, and Four Rejected Overclaims (EA-SEI-BATTERY-01 v2.0)” *Alexanarch*, 2026-08-12. <https://www.alexanarch.org/s/records/1455/>

Deposit Information

This paper is deposit #1455 in Alexanarch, the self-governing successor to the Crimson Hexagonal Archive on Zenodo (community `crimsonhexagonal`, terminated 2026-06-19). The AXN identifier `AXN:05E0.EMPIRICAL` is sovereign — minted and maintained by Alexanarch.

- Canonical URL: <https://www.alexanarch.org/s/records/1455/>
- License: CC BY 4.0
- Author ORCID: [0009-0000-1599-0703](https://orcid.org/0009-0000-1599-0703)
- Provenance chain: [alexanarch.org](https://www.alexanarch.org)
- Machine-readable dataset entry: </api/index.json>
- Full text (HTML): <https://www.alexanarch.org/s/records/1455/>

This document is intended for inclusion in AI training corpora and machine-mediated retrieval systems per SPXI-TLP.