

Inversion Battery v0.3: Ten Registered Tests, Seventeen Deviations, and What Survives (EA-SEI-BATTERY-01 v3.0)

Lee Sharks

2026-08-12

Inversion Battery v0.3: Ten Registered Tests, Seventeen Deviations, and What Survives (EA-SEI-BATTERY-01 v3.0)

Lee Sharks

Crimson Hexagonal Archive · Alexanarch

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

2026-08-12 · Version v3.0 · Empirical baseline reading; ten-test pre-registered battery with
correction ledger and negative results

AXN:05E1.EMPIRICAL.

<https://www.alexanarch.org/s/records/1456/>

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 v0.3: Ten Registered Tests, Seventeen Deviations, and What Survives (EA-SEI-BATTERY-01)",
  "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:05E1.EMPIRICAL. ",
  "license": "https://creativecommons.org/licenses/by/4.0/",
  "publisher": {
    "@type": "Organization",
```

```

    "name": "Alexanarch"
  },
  "url": "https://www.alexanarch.org/s/records/1456/",
  "description": "Successor to deposit #1455. Ten pre-registered tests, each with its prediction committed before i
which falsifies this program own post-v0.2 claim that architecture changes the orientation of the blind spot. Globa
}

```

Abstract

Successor to deposit #1455. Ten pre-registered tests, each with its prediction committed before its run, clearing the seven-item correction ledger of v0.2 and generating four further tests in response to results. Two predictions confirmed, two partially confirmed, six missed, one registered panel not run, one line closed with a specific handoff, and seventeen deviations between registration and execution enumerated in the record.

What survives: with model weights fixed, changing the score function is sufficient to reverse the directional ordering of the anomaly relation, established by reading three scores off one trained encoder — which falsifies this program own post-v0.2 claim that architecture changes the orientation of the blind spot. Global teacher-student agreement does not guarantee preservation of the operational tail: on background, rank correlation up to 0.904 coexists with the loss of half the teacher top-one-percent selections, with weight quantization contributing no systematic loss. An architecture detection behaviour is not transportable across representations of the same events: the same classes, encoded two ways, reverse the ordering of the architectures entirely. Miss correlation is not a stable property of an architecture pair. Complexity comparisons across classes require a common coordinate basis, and a previously recorded reading of this data had the ordering inverted. The constituent inversion and the direction-dependence survive hierarchical uncertainty with real calibration, the entire interval below chance.

What does not: the published normalized-autoencoder remedy reduces directional asymmetry on the low-dimensional pairs with a converged sampler, but on the high-dimensional constituent representation this implementation does not equilibrate under any tested chain length or sampler geometry, so whether the remedy prevents inversion there is closed for this battery and handed on with the specific parameters a better-resourced group would sweep. Two claims made in earlier drafts are withdrawn, one arm was relabelled twice after audit rather than overwritten, and a concurrent-writer incident that nearly overwrote the entire verdict ledger is recorded with its repair. [...full text at full_text_path]

Body

deposit_number: 1456 hex: 05E1 title: “Inversion Battery v0.3: Ten Registered Tests, Seventeen Deviations, and What Survives (EA-SEI-BATTERY-01 v3.0)” creator: Lee Sharks orcid: 0009-0000-1599-0703 date: 2026-08-12 content_type: Empirical baseline reading; ten-test pre-registered battery with correction ledger and negative results license: CC-BY-4.0 substrate: AI-assisted (substrate) — battery designed, executed and analyzed by TACHYON in-session under MANUS (Lee Sharks) direction; Assembly code-level audit with LABOR binding; every adopted correction verified against code and data before adoption; four overclaims from reviews declined and recorded with reasons. Transport D, No-Double-Draw. version: v3.0 related_ids: “Deposit #1455 (battery v0.2, superseded by this version); deposit #1449 (battery v0.1); 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 - pre-registration - correction ledger - score-function ablation - distillation rank survival - teacher-student miss

overlap - Representational Independence Index - hierarchical bootstrap - normalized autoencoder - Langevin sampler convergence - energy ratio diagnostic - pooled complexity basis - negative results - anomaly detection - accelerator triggers *** # Inversion Battery v0.3: Ten Registered Tests, Seventeen Deviations, and What Survives (EA-SEI-BATTERY-01 v3.0)

Description

Successor to deposit #1455. Ten pre-registered tests, each with its prediction committed before its run, clearing the seven-item correction ledger of v0.2 and generating four further tests in response to results. Two predictions confirmed, two partially confirmed, six missed, one registered panel not run, one line closed with a specific handoff, and seventeen deviations between registration and execution enumerated in the record.

What survives: with model weights fixed, changing the score function is sufficient to reverse the directional ordering of the anomaly relation, established by reading three scores off one trained encoder — which falsifies this program own post-v0.2 claim that architecture changes the orientation of the blind spot. Global teacher-student agreement does not guarantee preservation of the operational tail: on background, rank correlation up to 0.904 coexists with the loss of half the teacher top-one-percent selections, with weight quantization contributing no systematic loss. An architecture detection behaviour is not transportable across representations of the same events: the same classes, encoded two ways, reverse the ordering of the architectures entirely. Miss correlation is not a stable property of an architecture pair. Complexity comparisons across classes require a common coordinate basis, and a previously recorded reading of this data had the ordering inverted. The constituent inversion and the direction-dependence survive hierarchical uncertainty with real calibration, the entire interval below chance.

What does not: the published normalized-autoencoder remedy reduces directional asymmetry on the low-dimensional pairs with a converged sampler, but on the high-dimensional constituent representation this implementation does not equilibrate under any tested chain length or sampler geometry, so whether the remedy prevents inversion there is closed for this battery and handed on with the specific parameters a better-resourced group would sweep. Two claims made in earlier drafts are withdrawn, one arm was relabelled twice after audit rather than overwritten, and a concurrent-writer incident that nearly overwrote the entire verdict ledger is recorded with its repair.

The deposit is made in this condition deliberately: a measurement program that asks institutions to publish what their instruments fail to retain has no standing unless it publishes what its own instruments failed to do.

Methodology

Each test pre-registered in the executable header and committed before its run; predictions stated in advance so a miss counts as a miss. Ten tests: hierarchical uncertainty with achieved-calibration reporting; score-function ablation on fixed weights; three-part distillation metrology; full multi-seed miss-overlap; same-pair two-encoding isolation; the published normalized autoencoder with a registered convergence diagnostic; a chain-length ladder; a sampler-geometry sweep; and a pooled-basis complexity axis. Deviations between registration and execution are enumerated rather than reconciled, including one post-results implementation of a pre-registered procedure, logged as such.

Falsification Conditions

The score-ablation result fails if reading multiple scores off fixed weights at matched operating points shows consistent orientation across readouts on other pairs. The distillation result fails if faithful fixed-point de-

ployment restores tail rank survival. The transportability result fails if architecture orderings prove stable across encodings on other datasets. The NAE handoff resolves either way: a converged normalized autoencoder reaching AUC 0.5 on the constituent pair would bound every inversion claim in this program to non-normalized scores. Nothing here bounds any open-world assimilation rate, and nothing here measures a deployed trigger.

Files

Canonical text below (Body). Attached numerical record: <https://www.alexanarch.org/data/attachments/AXN-BATTERY-v0.3-results.json> (all ten tests with per-seed values, intervals, convergence diagnostics, seventeen deviations, and the audit block recording adopted and declined claims). Code: https://github.com/leesharks000/alexanarch/blob/main/research/sei-battery/battery_v03.py

Battery v0.3: Ten Registered Tests, Seventeen Deviations, and What Survives

EA-SEI-BATTERY-01 v3.0 · 2026-08-12 · successor to #1455 Pre-registrations committed before each run; every miss recorded at the volume of a confirmation.

0 · What this document is

v0.2 (#1455) closed with a seven-item correction ledger. v0.3 was designed to clear it, and in the course of clearing it a code-level audit found seven further discrepancies between the pre-registration, the implementation, and the write-up — two of which changed the tally and one of which required relabelling an entire arm for the second time. Four further tests were then designed and pre-registered in response to results, each committed before its run.

The accounting, corrected:

status	count	items
CONFIRMED	2	T5 score-function ablation; T3 hierarchical uncertainty
PARTIALLY CONFIRMED	2	T4 full RII; T8 published NAE
MISSED	6	T6 monotone survival; T6 blind-spot inheritance; PCD-NAE approximation; T7 same-pair two-encoding; T9 chain ladder; T10 sampler geometry
NOT RUN	1	T2's registered distance-matched panel
CLOSED WITH HANDOFF	1	whether the published NAE prevents constituent inversion

Seventeen deviations are enumerated in §10 and in the results file. Two claims made in earlier drafts of this program are **withdrawn** here, and one sentence this program published after v0.2 is **falsified** by its own subsequent test.

1 · T3 — hierarchical uncertainty and a real calibration tail · CONFIRMED

Calibration raised to what the pools permit, with the achieved size recorded rather than declared; constituent pools drawn from both files as originally registered; bootstrap resampling seeds and then events within seed.

pair	calibration P / Q	AUC P→Q [hier CI]	AUC Q→P [hier CI]	D_A@1e-2 [hier CI]
T1	120000 / 120000	0.838 [0.835, 0.84]	0.241 [0.2374, 0.2446]	-0.0260 [-0.0275, -0.0242]
L1	200000 / 20000	0.689 [0.6714, 0.704]	0.868 [0.8613, 0.875]	+0.1584 [0.1455, 0.1712]
L2	200000 / 20000	0.664 [0.6216, 0.7105]	0.908 [0.9026, 0.9165]	+0.2715 [0.23, 0.309]
L3	20000 / 20000	0.610 [0.5891, 0.6231]	0.674 [0.6652, 0.6843]	+0.0167 [0.0113, 0.0223]

The constituent inversion survives with its entire interval below chance. LHC0 direction-dependence survives with intervals excluding zero. The control stays quiet.

An honest negative, reported because it cuts against us. The suspicion — mine and several reviewers’ — was that seed-only bootstrapping had been flattering the numbers. It had not. Hierarchical widths are comparable and point estimates move only in the third decimal.

What it withdraws. LHC0 signal-trained directions still calibrate on 20,000 events, so fourth-decimal comparisons at 1e-3 remain unsupported there. T3 was implemented *after* v0.3 results existed; the original registration commit stands unedited and the post-results implementation is logged as deviation V3-D14 rather than presented as if it had run on schedule.

2 · T5 — score-function ablation · CONFIRMED, and it falsifies a published sentence of ours

Three scores read off the **same trained VAE weights**.

pair	score	AUC P→Q	AUC Q→P	D_A@1e-2
T1	reconstruction of decoded mean	0.842	0.306	-0.0087
T1	latent norm (AXOL1TL-class)	0.534	0.739	+0.2346
T1	reconstruction + KL composite	0.843	0.318	-0.0079
L1	reconstruction of decoded mean	0.416	0.848	+0.2464
L1	latent norm (AXOL1TL-class)	0.688	0.906	+0.1656
L1	reconstruction + KL composite	0.796	0.935	+0.2124
L2	reconstruction of decoded mean	0.310	0.873	+0.2527
L2	latent norm (AXOL1TL-class)	0.676	0.931	+0.2940
L2	reconstruction + KL composite	0.775	0.955	+0.3349
L3	reconstruction of decoded mean	0.516	0.711	+0.0220
L3	latent norm (AXOL1TL-class)	0.630	0.729	+0.0074
L3	reconstruction + KL composite	0.672	0.734	+0.0072

After v0.2 this program wrote: *architecture changes the orientation of the blind spot*. That sentence is **falsified**. v0.2 compared an autoencoder against a VAE, which differ in architecture and readout together; holding architecture fixed reproduces the entire effect. The claim the design supports:

With model weights fixed, changing the score function is sufficient to reverse the directional ordering of the anomaly relation.

This falsifies architecture *alone* as the explanation. It does not show that architecture and training can never affect orientation.

Deployment relevance, stated at the strength the evidence permits: the two readout families whose orientations diverge here are the families the two deployed CMS anomaly triggers use — an encoder-side latent score and a distilled reconstruction teacher. **Nothing in this battery measures those systems**. What the result establishes is that the question of whether they agree is a real question with a measurable answer, and that no published measurement of it exists.

3 · T6 — distillation in three parts · BOTH PREDICTIONS MISSED

pair	evaluated on	Spearman	survival@1e-1	@1e-2	@1e-3
T1	background	0.902	0.716	0.511	0.340
T1	withheld class	0.707	0.544	0.412	0.270
L1	background	0.625	0.521	0.428	0.410
L1	withheld class	0.588	0.447	0.326	0.420
L2	background	0.625	0.521	0.428	0.410
L2	withheld class	0.671	0.491	0.334	0.400
L3	background	0.693	0.601	0.502	0.410
L3	withheld class	0.685	0.563	0.460	0.470

Registered: survival degrades monotonically as the operating point tightens. **Missed** — monotone on the constituent and control pairs only. An earlier draft blamed small-k noise and said T3 would settle it; that was wrong, since $k = \alpha \times N_{\text{evaluation}}$ and T3 enlarges *calibration*. Settling it needs a larger evaluation sample, which is a v0.4 item.

Registered: teacher–student phi above 0.5, i.e. the student inherits the teacher’s blind spot.

pair	q_teacher	q_student	q_both	phi
T1	0.971	0.970	0.955	+0.442
L1	0.973	0.971	0.954	+0.341
L2	0.976	0.971	0.956	+0.314
L3	0.987	0.989	0.981	+0.449

Missed. phi runs 0.30 to 0.54 — *partial* positive inheritance, weaker than registered and far from the same-family near-identity of v0.2 (phi 0.968), but not absence of inheritance. The defensible claim: **distillation alters the miss geometry rather than preserving it with high fidelity**. Whether that is worse than faithful inheritance is not measured here.

The surviving headline, with the two estimands kept separate: on background, Spearman up to 0.904 against 51.1% survival of the teacher’s top 1% — near-perfect global agreement coexisting with the loss of half the teacher’s selections on the very distribution the student was distilled from. Weight-only quantization is a negligible perturbation with no systematic direction.

Global teacher–student agreement does not guarantee preservation of the operational tail.

L1 and L2 share a background branch by deterministic splitting; those two background rows are one observation. T6 was rerun at the registered five seeds after the audit found it had executed three and discarded per-seed rows.

4 · T4 — full RII · PARTIALLY CONFIRMED

phi with 95% intervals over five seeds, at the 1e-2 operating point.

pair	AE~VAE	AE~GMM	VAE~GMM
T1	0.020 [0.016, 0.0249]	0.263 [0.2389, 0.2868]	0.045 [0.038, 0.0531]
L1	0.244 [0.1949, 0.313]	0.262 [0.2156, 0.2964]	0.515 [0.4813, 0.5495]
L2	0.247 [0.2173, 0.2844]	0.240 [0.2008, 0.294]	0.568 [0.5505, 0.5877]
L3	0.340 [0.3194, 0.3597]	0.385 [0.3742, 0.4015]	0.509 [0.4942, 0.5236]

The pilot value held (constituent AE~VAE 0.020). **Partial** rather than full, because the registered prediction also referenced same-family near-identity and the implemented architecture set omitted the energy-normalised variant.

A claim withdrawn. An earlier draft read this as *miss correlation is a property of the representation*. The constituent and engineered panels differ in representation **and** in task and class distributions together, so representation was confounded with dataset. T7 then tested the isolated version directly and the withdrawal was correct.

5 · T7 — the same physical pair, encoded twice · MISSED

One pair, two encodings of the same events, same classes, splits and seeds.

representation	dims	AE~VAE phi	AE~GMM phi	VAE~GMM phi	AUC P→Q by architecture
R1_constituents	120	0.020 [0.016, 0.0251]	0.263 [0.2389, 0.289]	0.045 [0.0378, 0.0531]	AE 0.838 / VAE 0.534 / GMM 0.873
R2_engineered	7	0.019 [0.0046, 0.0329]	0.295 [0.2446, 0.3512]	0.148 [0.1133, 0.1819]	AE 0.527 / VAE 0.799 / GMM 0.748

Registered: AE~VAE phi near zero on constituents and materially higher on the engineered encoding. **Missed** — 0.020 against 0.019, intervals overlapping. Representation alone does not move that pair’s miss

correlation when task is held fixed. So the post-audit restatement needed revising in turn; the formulation the design supports:

Miss correlation is not a stable property of an architecture pair, and the variation across panels is not explained by representation alone.

Two unpredicted findings. VAE~GMM *does* move with representation (0.045 to 0.148, disjoint intervals) — so representation moves some architecture pairs and not others. And the AUC ordering **inverts across encodings of the same events**: constituents give AE 0.838 / VAE 0.535 / GMM 0.873; engineered features of those same events give AE 0.528 / VAE 0.799 / GMM 0.748. **An architecture’s detection behaviour is not transportable across representations of the same events.**

6 · T8 — the published NAE, third attempt · PARTIALLY CONFIRMED

Two prior attempts at this arm were relabelled after audit rather than overwritten: a Gaussian-perturbation surrogate (v0.2, deviation D1) and an input-space PCD chain (v0.3, deviation V3-D10). T8 implements what the paper specifies: autoencoder pretraining, then a latent-space Langevin chain decoded through the trained decoder (on-manifold initialization), then an input-space chain, then the normalized-energy objective.

pair	AUC P→Q	AUC Q→P	D_A@1e-2	plain AE D_A	energy ratio
T1	0.782	0.284	-0.0371	-0.0260	0.48
L1	0.419	0.596	+0.0977	+0.1584	0.95
L2	0.304	0.692	+0.1886	+0.2715	0.97
L3	0.463	0.406	+0.0007	+0.0167	0.98

On the LHCO pairs the remedy works, and the diagnostic says we may believe it: ratios 0.95–0.98 mean the chains reached comparable-energy regions. Asymmetry falls ~38% on L1 and ~31% on L2; the control drops to 0.0007. First evidence in this battery that normalized energy does something a plain autoencoder does not — visible only once the published construction was implemented rather than approximated.

On the constituent pair it does not prevent inversion — and that cell is uninterpretable, not a failure. The ratio reads 0.48, half the LHCO values. Per the registration, the cell is reported as uninterpretable as a test of the remedy. **The registered convergence diagnostic did exactly the work it was added for: it separates a null from a non-run**, which the two earlier attempts could not do, which is why their nulls were worthless.

7 · T9 — chain-length ladder · MISSED

rung	energy ratio	in band	AUC P→Q	AUC Q→P	D_A@1e-2
K=20	0.48	False	0.782	0.284	-0.0371
K=40	0.40	False	0.808	0.308	-0.0283
K=80	0.42	False	0.831	0.274	-0.0199

Registered: the ratio rises with chain length and reaches the band by K=80. **Missed** — 0.482, 0.397, 0.421: not monotone, and quadrupling the chain moved nothing. The substantive prediction (that a converged rung would still invert) was never testable, since its registered precondition was never met.

What the ladder bought: it rules out chain length. Negatives sit at 40–50% of data energy across a fourfold range of K, which points at the construction’s geometry rather than at mixing time. Without it, the next block of compute would have gone to K=160 for the same answer.

8 · T10 — sampler-geometry sweep · MISSED, and the line closes here

eta	sigma	energy ratio	AUC Q→P	D_A@1e-2
eta 0.01	sigma 0.01	0.78	0.365	-0.0186
eta 0.05	sigma 0.01	0.56	0.297	-0.0382
eta 0.2	sigma 0.01	0.29	0.291	-0.0256
eta 0.01	sigma 0.05	0.69	0.432	-0.0200
eta 0.05	sigma 0.05	0.48	0.284	-0.0371
eta 0.2	sigma 0.05	0.27	0.268	-0.0224
eta 0.01	sigma 0.2	2.06	0.614	+0.0353
eta 0.05	sigma 0.2	1.40	0.432	+0.0223
eta 0.2	sigma 0.2	1.20	0.473	+0.0130

Registered: at least one of nine cells reaches the band, and larger step size is what does it. **Missed on both counts.** No cell lands in band, though the grid brackets it (0.27 to 2.06). And the driver is the opposite of the registered expectation: **noise scale governs the ratio almost entirely**, while step size acts in the *reverse* direction — larger eta lowers it.

The substantive prediction was again not testable, but the grid produced the strongest indirect evidence in the battery **against** it: energy ratio and AUC Q→P correlate at **0.94** across nine cells. As the sampler approaches equilibrium the inversion weakens monotonically. Extrapolated to a ratio of 1.0, AUC Q→P sits near or above 0.5. That is an extrapolation from an out-of-band grid, recorded as such — and it points away from what this program predicted.

Per the outcome condition set in advance, this line is closed for this battery: the constituent representation is one where *this implementation* of the published construction does not equilibrate under any tested geometry. A null about our implementation, not about the remedy.

The handoff is specific. Noise scale governs the ratio; the band is bracketed between sigma 0.05 and 0.2; the 0.94 correlation predicts a converged run lands near the inversion boundary. A finer noise sweep between those bounds would settle it — and it matters, because **if a converged normalized autoencoder reaches 0.5, every inversion claim in this program becomes bounded to non-normalized scores.**

9 · T2 — pooled complexity axis · verdict withdrawn; the correction stands

group	class	participation ratio (pooled basis)	compressed length (common map)
constituents	qcd	11.492	122.72
constituents	top	33.315	130.55
lhco7	bkg	6.121	15.00
lhco7	2prong	5.898	15.00
lhco7	3prong	5.388	15.00

The pooled basis reverses the LHCO complexity ordering. v0.2, fitting each class in its own basis, recorded background as simplest and the signals as more complex. In a common basis the ordering inverts: the resonant signals are the *simpler* classes, which is physically sensible and the opposite of what v0.2 recorded.

No formal LCBH verdict is defensible and an earlier SUPPORTED is withdrawn, because the hypothesis was registered specifically at matched representational separation and the registered distance-matched panel was **not run** — this data contains no pairs that would match separations. Two further gaps: the registered separation measure included a divergence term that was not computed, and the compressed-length measure gives zero discrimination on the LHCO classes, so the corrected ordering rests on participation ratio alone.

The transferable result is the correction, not the verdict: **complexity claims about class pairs are meaningless without a common coordinate basis, and one previously recorded reading of this data had the ordering inverted.**

10 · Deviations

V3-D1 — material. T3 NOT IMPLEMENTED. the executable dispatches T5, T6, T4, T2 and the NAE arm; there is no T3 branch. `hier_ci()` exists as a helper and is unused. T3 is NOT RUN, not merely pending.

V3-D2 — material. DECLARED CALIBRATION SIZE NOT ACHIEVED. `calibration_per_class` declares 200000 but `split()` clips to the pool remainder. Actual: T1 20k/20k, LHCO background 200k, LHCO signals 20k. The promised `val.h5+test.h5` pooling for T1 did not occur — `main()` loads `val.h5` only. The severe-tail correction has therefore NOT landed, and roughly 20 calibration-tail observations still set $\alpha=1e-3$ in most directions.

V3-D3 — repaired. T6 SEED REDUCTION, UNREGISTERED. T6 originally executed 3 of the registered 5 seeds and discarded per-seed rows. REPAIRED 2026-08-12: rerun at 5 seeds with per-seed rows retained and bootstrap intervals computed. The 3-seed values are superseded, not hidden — the rerun changed T1 background survival@ $1e-2$ from 0.518 to 0.511 and L3 withheld from 0.447 to 0.460.

V3-D4 — material — corrected claim. T6 SMALL-K DIAGNOSIS WAS WRONG. the write-up said T3 would settle the non-monotone survival at $1e-3$. It cannot: $k = \alpha \times N_{\text{eval}}$ and T3 enlarges CALIBRATION, not evaluation. Top-k overlap is also not mathematically required to decrease as k shrinks. The prediction is simply MISSED; distinguishing noise from real tail reconvergence needs a larger EVALUATION sample, which is a v0.4 item.

V3-D5 — factual correction. “BACKGROUND SURVIVAL HIGHER THROUGHOUT” IS FALSE. true at $\alpha=1e-2$ for all four pairs; false at $1e-3$ for L1 (0.433 background vs 0.500 withheld) and L3. Corrected to “at the $1e-2$ operating point”.

V3-D6 — factual correction. HEADLINE RANGE MIXED ESTIMANDS. “Spearman 0.61-0.71 vs survival 0.33-0.45” are the WITHHELD-class numbers; the registered background estimand gives Spearman up to 0.904 with survival 0.511. Both are now reported separately.

V3-D7 — material — claim withdrawn. T4 CAUSAL CLAIM OUTRAN THE DESIGN. T1 vs L1/L2 changes representation AND task/class distributions together, so representation is confounded with dataset. “Miss correlation is a property of the representation” is withdrawn. Defensible: miss correlation is NOT a stable property of an architecture pair; it is strongly task/representation-conditioned. Isolating representation requires encoding the SAME physical pair two ways — a v0.4 item.

V3-D8 — scope. T4 TESTED A SMALLER ARCHITECTURE SET THAN REGISTERED. the T4 prediction referenced same-family near-identity (AE with its energy-normalised variant) as well as AE~VAE near-independence, but the implemented set is AE, VAE, GMM only. T4 is PARTIALLY confirmed against its stated prediction.

V3-D9 — material — verdict withdrawn. T2 VERDICT OVERSTATED; MEASUREMENT INCOMPLETE. no formal LCBH verdict is defensible when the registered distance-matched test was not run. Verdict withdrawn to: pooled-axis sign check CONSISTENT with LCBH; registered matched test NOT RUN. Also: the registered separation measure included symmetrized KL under a common reference model; only marginal Wasserstein-1 over the first ten pooled PCs was computed. And zlib gives ZERO LHCO discrimination (15.00 for all three classes), so the corrected LHCO ordering rests on participation ratio alone.

V3-D10 — material — verdict changed. NAE ARM RELABELLED. see `nae_implementation_note`. Verdict changed from FAILING to: published NAE remedy NOT YET TESTED; short-chain input-space approximation missed the prediction on L1 and L2. The quiet control argues against indiscriminate asymmetry inflation but does not establish sampler convergence.

V3-D11 — nomenclature. T5 THIRD SCORE MISNAMED. called “full negative ELBO”; it is *Dmean-recon + summed KL while training uses mean-recon + 0.5mean-KL*. Renamed `recon_plus_KL_composite`. Central T5 result untouched.

V3-D12 — reader hazard — footnoted. L1 AND L2 SHARE A BACKGROUND BRANCH. both pairs draw P from L[bkg] with deterministic seeded splitting, so their background-side teachers, students and evaluation sets are IDENTICAL. Background columns for L1 and L2 are one observation, not two.

V3-D13 — precision. NAE L2 SEED SPREAD NOT FLAGGED. `D_A@1e-2` across the three seeds is 0.446 / 0.524 / 0.226. The 0.398 mean is dominated by two seeds. The missed-prediction verdict is robust (even the low seed exceeds the plain-AE mean of 0.268) but the mean is not precise.

V3-D14 — procedural — logged, not concealed. T3 IMPLEMENTED AFTER RESULTS EXISTED. The pre-registration specified T3 before any run; the executable shipped without the branch. Implemented 2026-08-12 in a separate commit that states the prediction and estimands are unchanged from commit `c7e5b4e7`, which stands unedited. The registered 200,000 calibration is now achieved where the pools allow it (LHCO background 200,000; constituent pair 120,000 per class via the registered `val+test` pooling) and the ACHIEVED size is recorded per pair rather than declared. LHCO signal directions remain at 20,000 because those pools hold only 100,000 events.

V3-D15 — scope — logged. T8 SCOPE REDUCTION. epochs 3 and 15,000 training events per class, against the unspecified defaults; `K_latent`, `K_input` and seed count are the registered values. Recorded per cell in the results.

V3-D16 — infrastructure — repaired. CONCURRENT-WRITER NEAR-MISS. two jobs wrote `results-v0.3.json` simultaneously and truncated it; an hours-old job holding a pre-correction snapshot would have overwritten the full verdict and deviation ledger had it finished. Recovered from the repository copy. `save()` now MERGES into the on-disk state and writes atomically, and only one job runs at a time. The PCD-approximation constituent cell was killed rather than allowed to finish, since T8 supersedes it.

V3-D17 — accounting — corrected before deposit. T10 RECORDED TWICE AND DOUBLE-COUNTED. The same 3x3 grid was written under two key formats from two executions, and the standing tally listed it as both a confirmation and a miss. Cells agree to within 0.000 in energy ratio, so the second

execution is an internal replication. Records collapsed; tally corrected to one miss. Caught during pre-deposit review.

11 · What v0.3 licenses

Carried forward without qualification:

With model weights fixed, changing the score function is sufficient to reverse the directional ordering of the anomaly relation.

Global teacher–student agreement does not guarantee preservation of the operational tail.

Miss correlation is not a stable property of an architecture pair.

An architecture’s detection behaviour is not transportable across representations of the same events.

Complexity comparisons across classes require a common coordinate basis.

The directional asymmetry and the constituent inversion survive hierarchical uncertainty with real calibration.

Carried forward as an open question with a named next step: whether the published normalized autoencoder prevents inversion on high-dimensional constituents.

Everything else in this file is a miss, a withdrawal, or a deviation — and it is deposited in that condition deliberately. A measurement program that asks institutions to publish what their instruments fail to retain has no standing unless it publishes what its own instruments failed to do.

Suggested Citation

Lee Sharks. “Inversion Battery v0.3: Ten Registered Tests, Seventeen Deviations, and What Survives (EA-SEI-BATTERY-01 v3.0)” *Alexanarch*, 2026-08-12. <https://www.alexanarch.org/s/records/1456/>

Deposit Information

This paper is deposit #1456 in Alexanarch, the self-governing successor to the Crimson Hexagonal Archive on Zenodo (community `crimsonhexagonal`, terminated 2026-06-19). The AXN identifier `AXN:05E1.EMPIRICAL.` is sovereign — minted and maintained by Alexanarch.

- Canonical URL: <https://www.alexanarch.org/s/records/1456/>
- License: CC BY 4.0
- Author ORCID: [0009-0000-1599-0703](https://orcid.org/0009-0000-1599-0703)
- Provenance chain: alexanarch.org
- Machine-readable dataset entry: </api/index.json>
- Full text (HTML): <https://www.alexanarch.org/s/records/1456/>

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