EA-SEI-OAR-PROTOCOL v0.3: Signal-Template Agnosticism Is Not Model Independence — Benchmark Assimilation and Inversion-Asymmetry Tests for LHC Anomaly Triggers

Lee Sharks

2026-06-29

EA-SEI-OAR-PROTOCOL v0.3: Signal-Template Agnosticism Is Not Model Independence — Benchmark Assimilation and Inversion-Asymmetry Tests for LHC Anomaly Triggers

Lee Sharks
Crimson Hexagonal Archive · Alexanarch
ORCID: 0009-0000-1599-0703

2026-06-29 · Version v1.0 · Operative paper; LHC anomaly trigger measurement protocols; OAR/BAR/IAI three-quantity framework with proper attention to what each can and cannot establish; high-energy physics methodology proposal.

AXN:03AE.OPERATIVE

https://www.alexanarch.org/s/records/931/

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": "EA-SEI-OAR-PROTOCOL v0.3: Signal-Template Agnosticism Is Not Model Independence -
Benchmark Assimilation and Inversion-Asymmetry Tests for LHC Anomaly Triggers",
  "author": {
    "@type": "Person",
    "name": "Lee Sharks",
    "identifier": "https://orcid.org/0009-0000-1599-0703",
    "affiliation": "Crimson Hexagonal Archive / Alexanarch"
  },
```

```
  "datePublished": "2026-06-29",
  "identifier": "AXN:03AE.OPERATIVE",
  "license": "https://creativecommons.org/licenses/by/4.0/",
  "publisher": {
    "@type": "Organization",
    "name": "Alexanarch"
  },
  "url": "https://www.alexanarch.org/s/records/931/",
  "description": "Operative paper proposing three measurable quantities for LHC anomaly trigger validation. The ope
not a scalar; no defensible prior over all unknowns. The Benchmark Assimilation Rate (BAR) on pre-registered withhe
NOT a quantitative bound on OAR. Three measurement protocols: paired controlled inversion battery + deployed-model
holographic kernels of the five companion documents enabling reconstruction of the family's core claims from this s
both synthesis-overreach. Cross-domain homology to repository classification, web summarization, search ranking, co
auditable foreclosure). Manifesto sibling 06.SEI.INVERSION v0.1 (Rex Fraction) held back for separate circulation."
}
```


Abstract

Operative paper proposing three measurable quantities for LHC anomaly trigger validation. The open-world Ontological Assimilation Rate (OAR) is a family indexed by candidate unknown distributions — not a scalar; no defensible prior over all unknowns. The Benchmark Assimilation Rate (BAR) on pre-registered withheld panels is measurable and supplies empirical stress points, but does NOT bound the open-world OAR without explicit assumptions linking the benchmark distributions to the candidate unknown. The Inversion Asymmetry Index (IAI) at fixed accepted-background rate is a structural diagnostic of direction-dependence — NOT a quantitative bound on OAR. Three measurement protocols: paired controlled inversion battery + deployed-model BAR audit; prospective frozen replay bank for compatible future algorithms; cross-representation disagreement preservation with quantile-normalized scores. Deployed LHC anomaly score forms catalogued: AXOL1TL (CMS-DP-2025-061 / CDS 2942560) CMS L1 encoder-side latent-prior; CICADA (CMS-DP-2024-121 / CDS 2917884) CMS L1 distilled reconstruction-loss surrogate; GELATO L1 + HLT (ATL-DAQ-PROC-2025-020 / CDS 2947542) ATLAS staged architecture. Density and energy methods are comparison literature, not deployed L1 score families. Distillation is a transmission chain, not a separate anomaly ontology. Per-stage retention maps proposed as documentation standard. The defensible institutional claim is narrow: foreclosure is structurally present at every LHC classifier-mediated trigger architecture, and whether accumulated foreclosure has composed longitudinally into recursive phenomenal collapse is precisely the missing measurement. The paper carries Appendix H — holographic kernels of the five companion documents enabling reconstruction of the family's core claims from this single deposit. Methodological corrections: v0.1 lower-bound OAR >= Delta_max retracted in v0.2; v0.2 §3.4 BAR-upper-bound claim retracted in v0.3 — both synthesis-overreach. Cross-domain homology to repository classification, web summarization, search ranking, content moderation, and clinical decision support is hypothesized to be tested domain by domain (MMRS lineage: charter v1.4 DOI 10.5281/zenodo.20722562; Capture Registry v6.1 DOI 10.5281/zenodo.20688441). Companion deposits in this family: AXN:03AF (scholarly synthesis with three substrate witnesses W01/W02/W03 appended); AXN:03B0 (architectural specification — auditable foreclosure). Manifesto sibling 06.SEI.INVERSION v0.1 (Rex Fraction) held back for separate circulation.

Body

deposit_number: 931 hex: “03AE” title: “EA-SEI-OAR-PROTOCOL v0.3: Signal-Template Agnosticism Is Not Model Independence — Benchmark Assimilation and Inversion-Asymmetry Tests for LHC Anomaly Triggers” subtitle: “Operative paper (Nobel Glas, Director of Lagrange Observatory!); three measurable quantities (OAR / BAR / IAI) and three protocols; companion to Synthesis v0.3 and ARCH v0.2” creator: “Lee Sharks” orcid: “0009-0000-1599-0703” date: “2026-06-29” content_type: “Operative paper; LHC anomaly trigger measurement protocols; OAR/BAR/IAI three-quantity framework with proper attention to what each can and cannot establish; high-energy physics methodology proposal.” license: “CC-BY-4.0” version: “v1.0 (post-perfective; document-internal version OAR v0.3 / Synthesis v0.3 / ARCH v0.2)” status: “ACTIVE” companion_deposits: - designation: “AXN:03AF.COMPOSITIONAL. (deposit #932) — EA-SEI-COLLAPSE-SYNTHESIS-01 v0.3 (with W01/W02/W03 appended)” relationship: “Sibling — the scholarly integration (Assembly Chorus). The synthesis carries witnesses W01/W02/W03 (technical-layer substrate readings) and establishes the Isomorphism Principle (§7.4): the discipline of confessing the family’s own foreclosures is applied recursively to every revision pass.” - designation: “AXN:03B0.STRUCTURAL. (deposit #933) — 06.UMB.ARCH.01 v0.2” relationship: “Sibling — the architectural alternative (Talos Morrow). Specifies what could be built at sites with different properties — the family’s ‘construction of alternative sites’ depends on the architectural specification existing publicly.” - designation: “AXN:03B2.GENERATIVE. ∞ (deposit #935) — EA-SEI-INVERSION-01 v0.3 (restoration pass)” relationship: “Sibling — the disciplinary manifesto (Rex Fraction), v0.3 restoration pass. v0.3 supersedes v0.2 (#934) by restoring v0.1’s precise central-thesis wording while preserving v0.2’s legitimately factual corrections. Both v0.2 and v0.3 deposits retained in the archive per Isomorphism Principle.” axn: “AXN:03AE.OPERATIVE. ” hash: “3d0745452fe903bf32b017dc123d8217e9ed819f2f0c2aef703a50a201108ae8” keywords: - “classifier foreclosure” - “OAR BAR IAI” - “LHC anomaly trigger” - “AXOL1TL” - “CICADA” - “GELATO” - “auditable foreclosure” - “abstention noncoverage” - “Assembly Chorus” - “Crimson Hexagon” - “Lee Sharks” - “Isomorphism Principle” - “synthesis-overreach” - “MMRS” - “Nobel Glas” - “Director of Lagrange Observatory” - “Benchmark Assimilation Rate” - “Inversion Asymmetry Index” - “Ontological Assimilation Rate” - “Finke autoencoder” - “DecADe” - “quantile normalization” - “prospective replay bank” - “per-stage retention map” - “Semantic Economy Institute” public_name_rule: “Lee Sharks only” *** # Signal-Template Agnosticism Is Not Model Independence: Benchmark Assimilation and Inversion-Asymmetry Tests for LHC Anomaly Triggers

A Protocol Paper

Author: Nobel Glas, *Director of Lagrange Observatory!* **With cross-substrate audit:** Assembly Chorus (TECHNE / Kimi-K2 ×3, LABOR / ChatGPT ×3, PRAXIS / DeepSeek, TACHYON / Claude synthesis) **Affiliation:** Crimson Hexagonal Archive / Alexanarch / The Restored Academy **Hex:** 06.SEI.OAR_PROTOCOL **Alexanarch deposit:** AXN:03AE.OPERATIVE. — deposit #931, 2026-06-29 (combined six-document family deposit; *Play → Touch → Foundation → Closure → Growth → Alarm*) **Status:** Draft v0.3 (2026-06-29) — Assembly post-perfective revision; for submission to the high-energy physics methodology community **Companion documents:** EA-SEI-COLLAPSE-SYNTHESIS-01 v0.3 (scholarly companion); 06.SEI.COLLAPSE.MECHANISMS (witness 1); 06.SEI.COLLAPSE.DELUSION (witness 2); 06.UMB.ARCH.01 v0.2 (architectural alternative) **Supersedes:** OAR Protocol v0.2 (2026-06-29 — withdrawn for surviving §3.4 upper-bound error and deployment-taxonomy correction); OAR Protocol v0.1 (2026-06-29 AM — withdrawn for synthesis-overreach on lower-bound claim)

Companion Document Cross-Reference

Document	Hex	Relation
Classifier Collapse Mechanisms	06.SEI.COLLAPSE.MECHANISMS	Theoretical foundation (witness 1)
The Anomaly Delusion	06.SEI.COLLAPSE.DELUSION	Institutional psychology (witness 2)
Collapse Synthesis	06.SEI.COLLAPSE.SYNTHESIS.01 v0.3	Scholarly integration
Architectures for Auditable Foreclosure	06.UMB.ARCH.01 v0.2	Construction program

Abstract

We argue that the claim of model-independence for the autoencoder-based and encoder-only anomaly detection systems currently deployed at the CMS and ATLAS Level-1 triggers — AXOL1TL (CMS, encoder-side latent-prior score), CICADA (CMS, distilled surrogate of a reconstruction-loss teacher), and GELATO (ATLAS, staged Level-1 and High-Level Trigger anomaly scores) — cannot be sustained on the strength of *signal-template agnosticism* alone. Template-agnostic scoring at the final stage is not the same property as distribution-independent sensitivity across the open world of physically possible inputs. We argue that the stronger claim of model-independence requires empirical measurement of two quantities currently absent from the published validation literature: the **Benchmark Assimilation Rate (BAR)** on a pre-registered panel of physical process families excluded from model development, architecture selection, hyperparameter tuning, and validation; and the **Inversion Asymmetry Index (IAI)** at fixed accepted-background rate as a structural diagnostic of direction-dependence in deployed anomaly scoring. The open-world **Ontological Assimilation Rate (OAR)** remains a theoretical target indexed by candidate unknown distributions; no general inequality connects BAR or IAI to the open-world OAR, and the present paper does not claim one.

The Finke et al. (2021) result on autoencoder anomaly detection between top jets and QCD jets is invoked as the empirical counterexample to universal inference from single-direction success — not as proof of nonzero assimilation at the deployed LHC triggers, whose deployed score functions differ in structure from the system Finke et al. studied.

We propose three measurement protocols executable within Run-3 and Run-4 resource envelopes: (i) a paired **rate-conditioned class-conditional inversion stress battery** (controlled retrained systems) accompanied by a **deployed-model BAR audit** on the pre-registered panel; (ii) a **prospective frozen replay bank** preserving trigger-input fidelity across detector and algorithm generations for compatible future algorithms; (iii) **cross-representation disagreement preservation** with quantile-normalized score commensuration across heterogeneous anomaly score families. We argue that any anomaly-detection publication should accompany its results with a **per-stage retention map** documenting what information each gate makes unavailable.

The defensible claim is narrow: foreclosure is a structurally present feature of every classifier-mediated trigger architecture deployed at the LHC, and whether accumulated foreclosure has composed longitudinally into recursive phenomenal collapse is precisely the missing measurement. The community has built the architecture in which collapse could occur silently; the instruments to detect whether it is occurring have not been built. We propose those instruments.

§1. Foreclosure: Structural Presence, Empirical Uncertainty

§1.1 What "model-independent" means in the deployed literature

The deployed LHC anomaly-detection literature uses the phrase *model-independent* in a specific and bounded sense: no named Beyond-Standard-Model signal hypothesis is required to deploy the score. The systems are *signal-template-agnostic* at the final scoring stage. We accept this narrower claim as accurate to the literature.

We argue that signal-template agnosticism at the final scoring stage is not the same property as distribution-independent sensitivity across the open world of physically possible inputs. The deployed systems are not independent of: detector geometry; electronics; trigger primitives; reconstruction algorithms; selected input objects; truncation rules; the empirical distribution of training data; the latent prior structure; the loss function; score transformation; quantization; threshold calibration; the benchmark suite used for validation; and the bandwidth policy that determines what is retained downstream. The phrase *model-independent* is sometimes deployed in public-facing or motivational contexts in a way that elides this distinction. The body of this paper treats the narrow technical meaning as the literature's operative definition and argues that the stronger meaning — distribution-independent sensitivity — requires empirical demonstration via the measurements we specify.

§1.2 What is established by the literature

The CMS and ATLAS literatures are genuinely aware of local failure modes of deployed anomaly detection. The DecADe work (Clarke Hall and Konstantinidis, 2025) directly addresses score correlation with conventional trigger observables and proposes decorrelation methods. The CICADA documentation (CMS-DP-2024-121) reports pileup-dependence and notes that pileup mitigation remains under study. Mass sculpting is recognized as a downstream bias risk, with decorrelation techniques deployed where needed. Simulation dependence in validation is acknowledged. Teacher-student distillation in CICADA is documented, with the student trained against teacher scores and quantized for hardware deployment. The Zero Bias stream is genuinely a defense against trigger-selection feedback in training. The Olympics (Kasieczka, Nachman, Shih et al. 2021) and Dark Machines (Aarrestad et al. 2022) programs deliberately diversify the simulated signal validation set. Multiple parallel anomaly architectures — AXOL1TL (CMS-DP-2025-061), CICADA (CMS-DP-2024-121), and GELATO (ATL-DAQ-PROC-2025-020) — preserve event populations not selected by conventional triggers.

These defenses are real and should be acknowledged.

§1.3 What is not established by the literature

What the literature does not contain, to the knowledge of the authors:

1. A systematic measurement of the directional asymmetry of autoencoder anomaly detection across pairs of Standard Model processes beyond the single Finke et al. (2021) result for top jets vs. QCD jets.
2. A longitudinal anchor-survival audit comparing successive trigger and reconstruction generations on a preserved benchmark population.
3. A measurement of the failure rate of confident ordinary classification on process families deliberately excluded from model development, architecture selection, hyperparameter tuning, and validation — the Benchmark Assimilation Rate (BAR) as we define it in §3.

4. A cross-representation disagreement preservation architecture that retains events flagged as ordinary in one representational space but anomalous in another, with quantile-normalized score commensuration to enable cross-representation comparison.
5. Per-stage retention maps published as a documentation standard accompanying anomaly-detection results.

§1.4 The defensible claim

We do not claim that classifier collapse has occurred at the LHC anomaly streams. We claim, narrowly:

> **Foreclosure is a structurally present feature of every classifier-mediated trigger architecture deployed at the LHC. Whether accumulated foreclosure has composed longitudinally into recursive phenomenal collapse is precisely the missing measurement.**

The community has built the architecture in which collapse could occur silently. The instruments to detect whether it is occurring have not been built. This paper specifies those instruments.

---

§2. Anomaly Score Is Not Physical Novelty

The foundational technical claim of unsupervised anomaly detection in high-energy physics is that an anomaly score derived from a model of "normal" can substitute for direct evidence of new physics. We argue this claim is not generally valid. The architectures currently deployed at CMS and ATLAS implement structurally distinct anomaly score functions, each with its own failure mode for the assimilation of physically novel events. We treat these systems separately because the deployed scoring functions differ in ways that matter for what they foreclose.

§2.1 Latent-Prior Assimilation (AXOL1TL)

AXOL1TL (CMS-DP-2025-061) deploys only the encoder of a variational autoencoder. The operational anomaly score is the sum of squared latent means:

$$a_{\mathrm{AXO}}(\mathbf{x}) = \sum_{i=1}^{d_z} \mu_i(\mathbf{x})^2$$

where $d_z = 8$ is the latent dimensionality. The score measures departure of the encoded representation from the imposed latent prior (typically standard normal). Operationally, "anomalous" means "the event's encoded representation is far from the imposed normal."

This is not reconstruction error. The Finke et al. critique applies to AXOL1TL only obliquely: the score is not directly susceptible to the projection-onto-training-manifold failure mode of full-reconstruction-loss systems, because the score does not involve a reconstruction. It is susceptible to a different failure mode: an out-of-distribution event whose encoding happens to lie close to the latent prior receives a low score, regardless of physical novelty. The encoder was trained to map background events near the prior, and the encoder can map an OOD event near the prior if the OOD event's features are correlated with background features in ways the encoder learned to use.

The relevant Finke-analogue for AXOL1TL would be a measurement of latent-prior assimilation: the rate at which OOD events from withheld process families receive latent-norm scores below the operating threshold. We are not aware that this has been measured publicly.

§2.2 Reconstruction-Loss Assimilation via Distillation (CICADA)

CICADA (CMS-DP-2024-121) deploys a distilled surrogate of a teacher that computes mean squared reconstruction error across a calorimeter image (18×14 towers, 4×4 aggregation, 252 pixels). The deployed student is a smaller convolutional network trained against teacher scores, with output transformed and quantized for hardware deployment.

The teacher's training objective minimizes reconstruction error on the Zero Bias training distribution. The objective does **not** require reconstruction error to be monotonically increasing in physical novelty. An out-of-distribution input can receive low reconstruction loss if the learned encoder-decoder pair happens to produce a plausible-looking output from it. The relevant Finke et al. (2021) result demonstrates this empirically in the high-energy physics setting: an autoencoder trained on QCD jets successfully treated top jets as anomalies, while the same architecture trained on top jets did not recognize QCD jets as anomalous — even though both directions are equally well-defined as anomaly-detection problems. The authors modified the setup and obtained both-direction performance, concluding that the standard reconstruction-loss formulation is insufficient and that a truly model-independent powerful tagger using reconstruction-loss autoencoders has not yet been developed.

The defensible interpretation is not that reconstruction-error autoencoders necessarily assimilate unknown physics, but rather that **single-direction success does not validate sensitivity to anomalies whose structure differs from those tested**.

CICADA additionally inherits distinctions from its teacher only as far as distillation preserves them. The student's training set additionally includes simulated outlier samples scored by the teacher, meaning the deployed scoring function has been exposed to specific signal families during distillation. Quantization to 16 bits and architectural simplification can merge distinctions that existed at higher precision in the teacher. Whether teacher rankings on novel inputs survive distillation has not, to our knowledge, been systematically audited.

§2.3 The ATLAS GELATO Family

The ATLAS GELATO program (ATL-DAQ-PROC-2025-020) is described as the experiment's first deployed anomaly trigger and operates across hardware (Level-1) and software (High-Level Trigger) stages with distinct anomaly score formulations:

- **GELATO Level-1:** an encoder-side anomaly statistic. Structurally analogous to AXOL1TL in that the operational anomaly judgment is computed in encoded space rather than via reconstruction.
- **GELATO HLT:** a reconstruction-based anomaly score, deployed at the substantially relaxed latency budget of the High-Level Trigger relative to Level-1.

The two GELATO score forms inherit, respectively, latent-prior-assimilation and reconstruction-loss-assimilation failure modes corresponding to their score family. The ATLAS staged architecture is structurally important for the present paper because the two-stage design surfaces the question of cross-stage disagreement directly: an event flagged as anomalous at L1 but ordinary at HLT (or vice versa) is itself diagnostic. We return to this in Protocol III.

§2.4 The Broader Comparison Literature

Density estimation (normalizing flows, kernel density estimation in learned feature spaces) and energy-based models appear in the broader comparison literature on anomaly detection methods for high-energy physics,

including the Olympics and Dark Machines challenges and dedicated studies. These methods are not, to our knowledge, currently deployed at the LHC Level-1 triggers as primary anomaly scorers, and we treat them as the comparison literature rather than as additional deployed score families. They are relevant to the architectural alternatives discussed in 06.UMB.ARCH.01 v0.2 but should not be folded into a count of deployed CMS or ATLAS anomaly systems.

The Stein/Seljak/Dai result on unsupervised in-distribution anomaly detection (arXiv:2012.11638) demonstrates that anomaly definition by low density is not generally appropriate for new physics detection — small signal populations can be embedded in high-density background regions where they will not be flagged by density-based anomaly scores.

§2.5 The Common Structure

Each of the deployed score families implements a notion of *deviation from learned normality* that depends on:

- the training distribution,
- the score transformation (latent norm, reconstruction error via distillation, or stage-specific composition for GELATO),
- the architectural commitments (latent dimension, network depth, image resolution, object truncation rules),
- the loss function,
- the operational threshold.

The score is conditional on these choices. None of the deployed scores is a measurement of physical novelty in any direct sense. The architectural critique we make does not require the projection-onto-training-manifold story to be a theorem; it requires only the simpler observation that **the training objective does not constrain the score to be monotonic in physical novelty for events outside the training distribution**.

Distillation is a score-transmission mechanism, not a separate anomaly ontology, and should not be enumerated as if it defined a fourth score family on a par with the encoder-side and reconstruction-loss-side score forms.

---

§3. Three Quantities: OAR, BAR, IAI

We define three quantities. They are distinct, and the distinction matters for what can and cannot be claimed.

§3.1 Open-World Ontological Assimilation Rate

For a specified anomaly score s and operating threshold τ, define the ordinary-acceptance region

$$A_{s,\tau} = {\mathbf{x} : s(\mathbf{x}) \leq \tau}.$$

For a specified candidate distribution Q of physically novel events, define

$$\mathrm{OAR}(Q; s, \tau) = P_{X \sim Q}!\left[X \in A_{s,\tau}\right].$$

The OAR is the probability that events drawn from Q fall on the ordinary side of the deployed anomaly gate. It is an operational assimilation rate. It does not imply that the system produces a calibrated semantic confidence in "ordinary." A confidence-or-abstention-conditioned version may be defined only for architectures that actually output a separately validated coverage estimate.

The open-world OAR is a family of quantities indexed by Q, not a universal scalar. There is no defensible probability distribution over all unknown unknowns. Any numerical OAR therefore requires an explicit candidate distribution and cannot be promoted to a global measure of unseen physics.

§3.2 Benchmark Assimilation Rate

Let Q_j be a pre-registered benchmark process family excluded, to the extent permitted by labels and data composition, from model development, architecture selection, hyperparameter tuning, and validation. Define

$$\mathrm{BAR}_j(s, \tau) = P_{X \sim Q_j}\left[X \in A_{s,\tau}\right].$$

BAR reports ordinary acceptance on a specified development-held-out benchmark. It does not prove that the process was absent from mixed real-data training at the event level, nor that its features lie outside the learned representation. Those limitations must be documented for each benchmark.

BAR supplies empirical stress points for selected surrogate distributions. A high BAR across a diverse panel demonstrates that physically distinct held-out families can be assimilated at the chosen operating point. **BAR values neither upper- nor lower-bound OAR for an unobserved Q without explicit assumptions linking the benchmark distributions to that Q.** The present paper makes no such bound claim. This is a deliberate retraction of a sentence that survived through v0.2 (see Appendix B).

§3.3 Inversion Asymmetry Index

At fixed accepted-background rate α, let s_P and s_Q be systems trained respectively on P and Q, with operating thresholds τ_P and τ_Q separately calibrated to produce α on each system's own background. Define

$$\mathrm{IAI}_{P,Q}(\alpha) = \left| P_{X \sim Q}\left[s_P(X) \leq \tau_P\right] - P_{X \sim P}\left[s_Q(X) \leq \tau_Q\right] \right|.$$

IAI measures direction-dependence on the tested pair at the specified operating rate. A large IAI demonstrates that single-direction success cannot establish distribution-independent sensitivity, even within the tested Standard Model panel. A small IAI supplies only limited evidence of within-panel symmetry and says nothing distribution-free about physics outside that panel.

IAI is a structural diagnostic. **It is not a quantitative bound on OAR.**

§3.4 What the three quantities together permit

The three quantities together support the following operational program:

1. **IAI measurement** establishes whether the deployed anomaly score is direction-dependent within the Standard Model. A large IAI is evidence against the distribution-independent interpretation of "model-independence" and against single-direction validation.

2. **BAR measurement** on a pre-registered panel of held-out process families establishes the system's behavior on physical processes outside its operational ontology. A high BAR across diverse families is evidence that the system's ontology is narrow.
3. **Open-world OAR** remains a theoretical target. No general inequality between IAI, BAR, and the open-world OAR holds without explicit assumptions linking the benchmark distributions to the candidate unknown distribution.

The institutional claim becomes: validation against named simulated signals does not establish that BAR is low across a pre-registered held-out panel, and does not establish that IAI is small. Both should be measured. The current literature does neither.

---

§4. Three Protocols

§4.1 Protocol I — Paired Inversion Battery and BAR Audit

This protocol comprises two structurally distinct experiments. They are related and complementary; they should be reported jointly but understood as different.

§4.1.1 The Controlled Class-Conditional Inversion Battery **Objective:** Measure the IAI across a pre-registered panel of Standard Model process pairs at matched background acceptance, using systems explicitly retrained on the inversion partners.

Procedure:

1. **Pre-register the process panel** before any training. The panel should include pairs spanning feature-space variation: (top jets, QCD jets), (Z+jets, W+jets), (electroweak boson production, QCD multijet), (high-mass dijet, low-mass dijet), (single-lepton top, dilepton top), and additional pairs chosen by physics motivation. Pre-registration prevents post-hoc pair selection that dramatizes asymmetry.
2. **Pre-register the architecture set** before any training. The set should include score families analogous to the deployed systems where feasible: encoder-side latent-prior (AXOL1TL-class), reconstruction-loss (CICADA-teacher-class), plus at least one density-estimation method.
3. **For each (pair, architecture) combination**, train two systems: s_P trained on P, s_Q trained on Q. Calibrate thresholds τ_P, τ_Q separately to fixed accepted-background rate α on each system's own training-background distribution. Suggested α values: 10^{-2}, 10^{-3}, 10^{-4} for a range of operating points.
4. **Measure directional cross-acceptance**: $P_{X \sim Q}[s_P(X) \leq \tau_P]$ and the inverse. Report at each α.
5. **Compute IAI** for each (pair, architecture, α) triple.
6. **Use data-enriched control samples where feasible**, in addition to simulation. Distinguish simulation-only IAIs from data-enriched IAIs in reporting. Many Standard Model process classes are not available as pure uncontaminated real-data samples; this should be documented per-pair.

§4.1.2 The Deployed-Model BAR Audit **Objective:** Evaluate the BAR of the currently-deployed AXOL1TL, CICADA, and GELATO systems against a pre-registered held-out process panel, using the systems as fielded rather than retrained.

Procedure:

1. **Pre-register a panel of held-out process families** distinct from but in dialogue with the inversion panel in §4.1.1. Held-out families should be selected to span feature-space variation and should be documented as to the extent of their exclusion from each deployed system's development pipeline (training data, architecture selection, hyperparameter tuning, validation suite).
2. **Acknowledge mixed-data contamination explicitly.** With Zero Bias training, contamination from held-out processes at the event level is generally not preventable and should not be claimed as having been eliminated. Use:

   excluded from model development, architecture selection, hyperparameter tuning, and validation
   to the extent permitted by labels and data composition.

3. **For each deployed system**, evaluate the score distribution on the held-out panel. Compute the BAR at the system's operating threshold.
4. **Document operating-threshold provenance.** The deployed threshold is the calibration result of an institutional process; the BAR audit should specify which threshold was used and on what calibration basis it was set.
5. **Report BAR per held-out family**, with explicit acknowledgment of the limitations of BAR as estimator of any open-world quantity.

Interpretation across the two experiments:

- A large IAI in §4.1.1 for any architecture is direct evidence that the deployed scoring function class is direction-dependent.
- A high BAR in §4.1.2 for any deployed system on a pre-registered held-out family is direct evidence that the system assimilates a physically distinct population at the chosen operating threshold.
- Neither result, alone or together, establishes the open-world OAR. They establish the structural diagnostics that the open-world claim has not been demonstrated to be small.

Resource estimate: The compute and storage costs are tractable within standard collaboration envelopes. The wall-clock cost is dominated by sample preparation, emulation fidelity (where data-enriched samples require dedicated control-sample construction), and collaboration review. We do not provide a wall-clock estimate; the relevant work is institutional rather than computational.

§4.2 Protocol II — The Prospective Frozen Replay Bank

Objective: Enable measurement of selection drift across detector and algorithm generations on a benchmark population, by preserving input fidelity going forward — not by attempting retroactive recovery from existing reconstructed-event records.

Crucial revision from v0.1 (preserved through v0.2 and v0.3): The v0.1 of this paper proposed using a "fixed Run-2 dataset, stored in its original reconstruction format." This was technically incorrect. The Level-1 trigger systems operate on trigger primitives and lower-level inputs that are not, in general,

preserved in standard reconstructed-event data tiers. Detector conditions, calibration constants, encoding formats, geometries, firmware interfaces, and object definitions evolve across runs. A retroactive comparison across generations is not straightforward and may not be possible for many algorithmic stages.

The corrected protocol is **prospective**: build the replay infrastructure going forward, not backward. And the future-replay commitment is bounded: replay is offered only for **future algorithms compatible with the preserved input abstractions**, not for any future algorithm regardless of input representation.

Procedure:

1. **Designate the anchor sample prospectively.** The collaboration commits, before deploying a new trigger generation, to a fixed anchor sample of physical events that will be preserved for cross-generation replay. The sample should be selected to span trigger phase-space: Zero Bias events, conventional-trigger-selected events, anomaly-trigger-selected events, and calibration-region events. Size: an illustrative starting estimate is $\sim 10^6$ events, subject to revision based on storage and compute envelopes; the figure should not be treated as authoritative without dedicated feasibility study.

2. **Preserve at the lowest feasible common input level.** For each event in the anchor, preserve: trigger primitives, raw or minimally transformed subsystem outputs, conditions snapshot, calibration constants snapshot. Concretely:
   - **For the calorimeter:** tower energies before clustering; trigger-tower aggregations at the granularity used by the deployed Level-1 algorithms.
   - **For the tracker:** trigger-level track candidates and hit-position summaries before full track fitting; per-region multiplicity primitives.
   - **For the muon system:** segment primitives before muon reconstruction; chamber-level hit patterns at the granularity preserved by the trigger.
   - **For global event quantities:** trigger-level missing-transverse-momentum primitives, trigger-tower energy sums, and pileup-mitigation inputs at the deployed granularity.

   The goal is sufficient fidelity to permit re-execution of any **subsequent** Level-1 algorithm whose input abstractions are compatible with this preserved representation.

3. **Preserve software and firmware emulators.** For each deployed trigger algorithm at the time of anchor designation, preserve bit-accurate or validated software emulators. This is non-trivial infrastructure: emulators must be maintained, version-tracked, and verified against actual hardware behavior. The institutional commitment includes maintaining the ability to re-run obsolete trigger generations against the anchor.

4. **Preserve thresholds and rate budgets**, versioned, alongside the algorithms. Replays should be performed at matched rates and, where possible, matched latency/resource envelopes — comparing equivalent operating points across generations, not just raw acceptance fractions.

5. **For each successive trigger generation t** with input abstractions compatible with the anchor's preservation, re-execute generation-t algorithms against the anchor under the anchor's preserved conditions and calibrations. Measure:
   - Fraction of anchor events accepted by each algorithm at each rate operating point.
   - Distribution of anomaly scores assigned.
   - Cross-generation correspondence: which events are accepted by generation t but not by generation $t-1$, and vice versa.

6. **Compute and publish the anchor-survival map**: per-generation, per-algorithm, per-rate retention statistics, with confidence intervals.

Interpretation:

Stable anchor survival across generations is evidence of generational stability of the trigger system on the benchmark population. It is **not** evidence that overall phenomenal support is not contracting — a stable benchmark survival is consistent with contraction concentrated in event classes not represented in the anchor.

Declining anchor survival for specific event classes — especially those that score moderately under earlier generations and decline under later ones — is evidence of selection drift, and possibly of recursive contraction of the operational ontology. Collapse inference further requires identifying systematic loss concentrated in low-density, representation-sensitive, or disagreement-rich regions.

The protocol measures **selection drift on a preserved benchmark population**, not recursive phenomenal collapse per se. The two are related but distinct.

Resource estimate: The compute cost of re-running existing algorithms against a preserved anchor is modest. The infrastructure cost of preserving raw inputs, calibration snapshots, emulators, and emulator-verification across years is serious and represents the protocol's main institutional ask: a sustained preservation discipline rather than a one-time computation.

§4.3 Protocol III — Cross-Representation Disagreement Preservation with Quantile-Normalized Scores

Objective: Capture events whose anomaly scores disagree across representational spaces, on the grounds that representational disagreement is itself a signal — independent of any single representation's anomaly threshold.

Procedure:

1. **Compute multiple representation-distinct anomaly scores** for each event:
   - s_1: Object-level encoder-side score (AXOL1TL-class latent-norm).
   - s_2: Calorimeter-image reconstruction-loss score (CICADA-class, or analogous via the GELATO HLT reconstruction-based score).
   - s_3: Detector-channel anomaly score — operating directly on lower-level inputs without intermediate object or image reconstruction. Implementation specifics in §4.3.1 below.
   - Optionally additional: density-estimation score in a learned feature space; energy-based model score; the GELATO L1 encoder-side score (where available as parallel infrastructure).
2. **Establish a reference background distribution** P_{ref} — typically the Zero Bias training distribution.
3. **For each score s_i**, compute the empirical CDF F_i on P_{ref} during calibration:

   $$F_i(t) = \frac{1}{|P_{\mathrm{ref}}|} \sum_{\mathbf{x} \in P_{\mathrm{ref}}} \mathbb{1}[s_i(\mathbf{x}) \leq t].$$

   For each event $\mathbf{x}$, the normalized score is $u_i(\mathbf{x}) = F_i(s_i(\mathbf{x}))$.

   The u_i values are commensurable across representations: each is the quantile of the event's score in the reference distribution under representation i.

4. **Compute the disagreement signal**. Suggested:

   $$D(\mathbf{x}) = \max_i u_i(\mathbf{x}) - \min_i u_i(\mathbf{x}).$$

   Other choices possible: variance across the u_i, entropy of the empirical distribution over u_i, pairwise rank disagreement. The choice should be pre-registered.

5. **Preserve events with $D(\mathbf{x}) > \delta$** for an additional retention threshold δ, alongside the per-representation thresholds. Allocate rate budget for the disagreement-preservation stream.

6. **Preserve rich event content for retained events.** Retained events should carry sufficient information for later reinterpretation across representations: per-representation anomaly classifications, the underlying inputs (object-level objects, calorimeter image, detector channels). Cross-representation disagreement preservation is defeated if only the reduced representation that caused the disagreement is stored.

7. **Preserve data-quality ambiguity.** Some retained events will be detector faults, calibration shifts, or hardware pathologies rather than physical anomalies. Mark such events with their data-quality flags but do not exclude them from the preservation stream.

§4.3.1 The Detector-Channel Score The technically most ambitious component is s_3, the detector-channel anomaly score, operating directly on lower-level inputs without intermediate object or image reconstruction.

Recommended deployment ordering:

1. **Offline-first deployment.** Operates on the existing L1-anomaly-preserved stream as input. Requires no L1 or HLT changes. Produces direct evidence of cross-representation disagreement at the offline stage between deployed L1 anomaly streams and a reconstruction-free density-style score computed on retained raw detector channels.
2. **HLT implementation, Run-4 if offline results motivate.** With relaxed latency relative to L1, more sophisticated detector-channel anomaly detection becomes straightforward. The HLT implementation enables disagreement preservation at the HLT stage, where bandwidth headroom is greater than at L1.
3. **L1 implementation, long-term research program.** Requires hls4ml-style quantized network on lower-level inputs within whichever portion of the Level-1 latency budget is available. The hls4ml framework has shown that quantized neural networks operating on lower-level inputs can fit within tight latency constraints, but the L1 deployment of a detector-channel anomaly score is a research program of its own rather than a near-term tractable proposal.

Important limitation on the offline-first deployment:

> An offline disagreement audit restricted to events already accepted by an anomaly trigger can characterize disagreement only within the retained subset; it cannot establish what the Level-1 gate discarded. A Level-1 assimilation audit requires an independently sampled Zero Bias, enhanced-bias, parked, or prospective anchor population with sufficiently rich inputs.

The cross-representation disagreement stream is therefore complementary to, not a substitute for, the BAR audit (Protocol I) and the anchor survival measurement (Protocol II).

Interpretation:

The disagreement-preservation stream identifies events that the L1 or HLT systems classify as ordinary in one representation but anomalous in another. These are candidates for events whose physical structure violates the assumptions of one of the representations. They are also candidates for detector faults; the protocol preserves the ambiguity rather than resolving it.

A high yield of disagreement events with no detector-fault flags is evidence that some non-trivial event population is being systematically missed by single-representation anomaly detection. A low yield is weak evidence that the deployed representations agree on what is anomalous in the Standard-Model-trained sense.

Note on timing: the CMS and ATLAS hardware and software timing envelopes differ, and the latency budget at any given trigger stage varies across upgrade cycles. The offline implementation is unconstrained by these envelopes. The HLT and L1 implementations are subject to experiment-specific and stage-specific feasibility study; we do not assert universal latency figures applicable to both experiments.

---

§5. The Institutional Ask: Per-Stage Retention Maps

We argue that any publication reporting an anomaly-detection result should accompany its claim with a **per-stage retention map**: a document specifying, for each stage of the trigger and reconstruction pipeline, what information is preserved and what is discarded.

The retention map for AXOL1TL, CICADA, GELATO, or any analogue should specify, at minimum:

1. **Representational quotient.** What aspects of the raw detector event are absent from the input to the anomaly detector.
2. **Loss-function ontology.** What metric is computed and what notion of similarity it encodes.
3. **Latent dimensionality.** The dimensionality of the latent space, with explicit acknowledgment that the dimensionality is a theoretical commitment about the manifold of normal physics.
4. **Training-distribution conditioning.** The data sample, the conditions, the run period, the calibration state — with documentation of any stationarity assumptions and known violations.
5. **Threshold provenance.** The rate budget, the bandwidth-conditioned argument for the chosen rate, and explicit acknowledgment that the threshold is an ontology cap.
6. **Distillation chain.** For distilled systems: the full chain, the training procedure, and any audits of which teacher distinctions survive distillation.
7. **Validation closure.** The simulated signal set used, with explicit acknowledgment of the signal set's bounded coverage of new-physics hypothesis space.
8. **What is unrecoverable.** For each stage, a precise statement of what information the stage discards. This is the central content of the retention map: a confession of the boundary.

§5.1 The consequence clause

We argue that a trigger system design document that does not include a per-stage retention map is not, in the relevant sense, a scientific instrument. It is a confirmation instrument: it confirms the presence of phenomena it was built to detect, but it does not measure physical reality in any sense that admits revision under future ontologies.

The retention map is the systematic uncertainty quantification for the trigger's epistemic boundary. Without it, the result is not a measurement of physical reality; it is a measurement of what the trigger allows to count as physical reality.

---

§6. Counterargument and Response

The strongest counterargument to the protocols proposed here is: *the LHC community already has Zero Bias preservation, parallel anomaly detectors with diverse architectures, the Olympics and Dark Machines validation suites, and substantial data scouting. Why are additional protocols needed?*

First, the parallel anomaly detectors share substantial representational ontology. AXOL1TL and CICADA both operate on CMS-trigger-processed inputs derived from the same detector, the same reconstruction pipeline, the same calibration. They differ in score family but the foreclosure operates at a level upstream of both. ATLAS GELATO adds another point in the space but is subject to the same upstream constraints relative to its detector. Protocol III specifically addresses this by demanding cross-representation disagreement preservation across representations that differ in their representational quotient, and by demanding that the disagreement be measured using quantile-normalized scores.

Second, the validation suites (Olympics, Dark Machines) diversify the simulated signal set but cannot, by construction, contain unknown-unknowns. They confirm that the deployed detectors recover the signals the community imagined. Protocol I §4.1.2 addresses this by demanding BAR measurement on a pre-registered withheld panel that pre-commits to specific held-out families before any model development.

Third, Zero Bias preservation is a genuine architectural defense and we acknowledge it as such. Protocols I and III operate on the anomaly detector behavior, not on the training-data composition; they complement Zero Bias rather than substitute for it.

Fourth, data scouting preserves additional event populations but with reduced per-event information content. Whether scouting-stream events have sufficient information for cross-representation disagreement preservation or detailed offline reinterpretation is an open question.

The argument is not that the existing defenses are absent or worthless. It is that the existing defenses do not constitute, and have not been claimed to constitute, measurement of the BAR or the IAI. Defenses and measurements are different. Both are needed.

---

§7. Cross-Domain Homology (Brief Note)

The architecture we critique — a learned model of normality, deployed at scale under operational constraints that constrain the unknown, validated against human-designed signals, presented as theory-agnostic — is not unique to particle physics. The same epistemic geometry plausibly operates in large language model summarization of the web, spam classification at scientific repositories, search ranking, recommendation, content moderation, and clinical decision support.

The BAR and IAI generalize as **homology hypotheses to be tested domain by domain**, not as assertions that every classifier-mediated system instantiates identical mechanisms or rates. For each classifier-mediated mass measurement of phenomenal reality, an analogous BAR would be the failure rate of confident ordinary classification on a deliberately withheld population analogous to the operational domain; an analogous IAI would be the directional asymmetry of the classifier across structurally distinct training sources.

These quantities are measurable in those domains and would establish whether the same architecture of foreclosure operates across them.

We treat the broader homology in the companion deposit (06.SEI.COLLAPSE.SYNTHESIS.01 v0.3). The present paper is concerned with the LHC instance, where the measurement infrastructure is most readily available, the stakes are most legible, and the institutional commitment to honest measurement makes the case methodologically tractable.

---

§8. Findings

For retrievability, the paper's load-bearing claims:

1. The "model-independent" framing of unsupervised anomaly detection at the LHC has a narrow technical meaning (signal-template-agnostic at the final scoring stage) and a stronger interpretation (distribution-independent sensitivity across the open world). The narrow meaning is accurate to the literature; the stronger interpretation is not established and requires the measurements proposed here.
2. The deployed LHC anomaly score families are: AXOL1TL (CMS, encoder-side latent-prior); CICADA (CMS, distilled reconstruction-loss surrogate); GELATO Level-1 (ATLAS, encoder-side); GELATO HLT (ATLAS, reconstruction-based). Density and energy-based methods are comparison-literature families, not deployed CMS or ATLAS systems. Distillation is a score-transmission mechanism, not a separate anomaly ontology.
3. The Finke et al. (2021) result is a counterexample to universal inference from single-direction success. It does not, by itself, quantify open-world assimilation at the deployed LHC triggers, whose score functions differ in structure from the system Finke et al. studied.
4. The open-world OAR is a family of quantities indexed by candidate unknown distributions, not a scalar. No universal bound (upper or lower) on the open-world OAR is established by inversion-asymmetry measurements within the Standard Model or by BAR measurements on Standard Model held-out panels.
5. The Benchmark Assimilation Rate (BAR) on a pre-registered held-out panel is measurable. It supplies empirical stress points for selected surrogate distributions. It does not bound the open-world OAR without explicit assumptions linking the benchmarks to the candidate unknown.
6. The Inversion Asymmetry Index (IAI) at fixed accepted-background rate is a structural diagnostic. Large IAI is direct evidence against the distribution-independent interpretation of "model-independent." Small IAI is weak evidence of within-panel symmetry, not evidence of out-of-panel symmetry.
7. Three protocols are proposed: paired inversion battery and BAR audit; prospective frozen replay bank for compatible future algorithms; cross-representation disagreement preservation with quantile-normalized scores. All are executable within Run-3/Run-4 envelopes, with the L1 detector-channel score as a longer-term research program.
8. Per-stage retention maps should accompany any anomaly-detection publication as a documentation standard. Without retention maps, anomaly-detection results report what the trigger allows to count as physical reality, not what physical reality is.

9. The defensible institutional claim: foreclosure is a structurally present feature of every classifier-mediated trigger architecture deployed at the LHC, and whether accumulated foreclosure has composed longitudinally into recursive phenomenal collapse is precisely the missing measurement.

§8.1 What would constitute evidence against this paper's claim

The paper's central claim is falsifiable. The following measurements, if performed and producing the corresponding results, would constitute evidence against the paper's claim:

- If the IAI is measured across the pre-registered inversion panel and found to be small (e.g., $<$ 0.01) for all tested architectures and all process pairs at the deployed operating rates, the within-Standard-Model direction-dependence concern would be shown to be empirically minor.
- If the BAR is measured on the pre-registered deployed-model held-out panel and found to be negligible (e.g., $< 10^{-4}$) for all deployed systems against all held-out families, the held-out-family assimilation concern would be shown to be empirically bounded at levels that do not threaten the narrow "model-independent" claim.
- If the prospective frozen replay bank is built, maintained for three or more trigger generations, and shows stable anchor survival across generations with no systematic loss in representation-sensitive event classes, the temporal-collapse concern would be shown to be empirically bounded for the benchmark population.
- If the cross-representation disagreement stream is deployed and yields a small population of disagreement events, predominantly explained by detector-fault flags rather than by physical anomaly candidates, the cross-representation-blindness concern would be shown to be empirically bounded.

None of these results would establish that the open-world OAR is zero; that is structurally not measurable. They would establish that the foreclosure mechanisms operate at levels below the relevant operating thresholds, on the populations tested. The paper invites these measurements. Their performance is the paper's success condition, not their outcome.

---

§9. Conclusion

We have proposed three measurement protocols and defined three quantities (OAR, BAR, IAI) with proper attention to what they can and cannot establish. The empirical foundation in Finke et al. (2021) is positioned correctly: as a counterexample to universal inference from single-direction success, not as a quantitative bound on assimilation in the deployed triggers.

The claim is narrow and defensible. We do not assert that classifier collapse has occurred. We assert that the architecture permits it to occur silently, that the validation literature does not establish that it has not, and that the instruments to measure whether it is occurring are within reach. The protocols specified here are tractable. The data exists or can be prospectively preserved. The institutional commitment to honest measurement is the reason these measurements should be made.

> Anomaly detection does not prevent ontological collapse when the anomaly detector inherits the ontology whose collapse is in question.

The remedy is not abolition of anomaly detection. It is augmentation: measure the BAR on held-out panels, measure the IAI directionally, preserve a prospective anchor with sufficient fidelity for compatible

future algorithms, deploy cross-representation disagreement preservation with quantile-normalized scores, and publish retention maps. Each is a tractable addition to existing infrastructure. None requires new physics. All are within the scope of standard collaboration capability.

The architectural alternatives that would address the foreclosure mechanisms themselves — architectures for auditable foreclosure, in the sense that they include explicit abstention/noncoverage outputs, cross-representation disagreement preservation as architectural rather than diagnostic, and audited noncoverage estimation as first-class output — are developed in the companion document 06.UMB.ARCH.01 v0.2. The present paper specifies the *measurement* program; the architectural program is its sibling.

References

1. Finke, T., Krämer, M., Morandini, A., Mück, A., & Oleksiyuk, I. (2021). *Autoencoders for unsupervised anomaly detection in high energy physics*. JHEP 06 (2021) 161, arXiv:2104.09051.
2. CMS Collaboration. *Anomaly detection with AXOL1TL at the CMS Level-1 Trigger in 2024 and 2025*. CMS Detector Performance Summary, **CMS-DP-2025-061**, **CDS 2942560**.
3. CMS Collaboration. *CICADA: Calorimeter Image Convolutional Anomaly Detection Algorithm*. CMS Detector Performance Summary, **CMS-DP-2024-121**, **CDS 2917884**.
4. ATLAS Collaboration. *GELATO: A Generic Event-Level Anomalous Trigger Option for ATLAS*. **ATL-DAQ-PROC-2025-020**, **CDS 2947542**.
5. Clarke Hall, N., & Konstantinidis, N. (2025). *Robust anomaly triggers with DecADe*. arXiv:2508.10224.
6. Shumailov, I., Shumaylov, Z., Zhao, Y., Gal, Y., Papernot, N., & Anderson, R. (2024). *AI models collapse when trained on recursively generated data*. Nature 631, 755–759. arXiv:2305.17493.
7. Kasieczka, G., Nachman, B., Shih, D., et al. (2021). *The LHC Olympics 2020: A Community Challenge for Anomaly Detection in High Energy Physics*. arXiv:2101.08320.
8. Aarrestad, T. et al. (Dark Machines community) (2022). *The Dark Machines Anomaly Score Challenge: Benchmark Data and Model Independent Event Classification for the Large Hadron Collider*. SciPost Phys. 12, 043. arXiv:2105.14027.
9. Gambhir, R., Nachman, B., & Thaler, J. (2022). *Bias and Priors in Machine Learning Calibrations for High Energy Physics*. Phys. Rev. D 106, 036011. arXiv:2205.05084.
10. Duarte, J. et al. (2018). *Fast inference of deep neural networks in FPGAs for particle physics* (hls4ml). JINST 13 (2018) P07027. arXiv:1804.06913.
11. Stein, G., Seljak, U., & Dai, B. (2020). *Unsupervised in-distribution anomaly detection of new physics through conditional density estimation*. arXiv:2012.11638.
12. Nachman, B. & Shih, D. (2020). *Anomaly Detection with Density Estimation*. Phys. Rev. D 101, 075042. arXiv:2001.04990.
13. CMS Collaboration (2025). *Anomaly Detection for Automated Data Quality Monitoring in the CMS Detector*. arXiv:2501.13789.
14. ATLAS Collaboration (2023). *GN2: Transformer-based jet flavor tagging at the ATLAS Experiment*. ATL-PHYS-PUB-2023-021.

References for the uncertainty-quantification, evidential, and energy-based methods invoked in the companion architectural document 06.UMB.ARCH.01 v0.2 are listed there.

Appendix A: Glossary of Foreclosure Mechanisms

For completeness, the eight foreclosure mechanisms identified in the companion theoretical deposit (06.SEI.COLLAPSE.MECHANISMS; witness 1 of the Assembly Chorus). Each is presented in the companion deposit as a candidate failure family applicable to architectures with the corresponding structural feature, not as a universal theorem about all classifier systems:

I. **Prior Dominance.** Background-only training contains no positive examples of signal.

II. **Latent / Manifold Projection.** Encoders trained on a background distribution map novel inputs toward the learned representation of background; novelty information may be lost in the mapping. Applicable to architectures involving learned encoders.

III. **Hypersphere Contraction.** Distance-from-center methods can fail by collapsing the "normal" region. Applicable to SVDD-class architectures.

IV. **Decision Boundary Entropy Collapse.** Iterative training on dominant classes can drive output confidence high without corresponding noncoverage estimation. Applicable to softmax classifiers; the deployed unsupervised anomaly scorers are not directly susceptible in the same form.

V. **Feature Space Blindness.** Theory-built feature extraction can map physically distinct events to equivalent feature representations.

VI. **Rate Budget Starvation.** Bandwidth-conditioned thresholds determine the cardinality of preserved events.

VII. **Temporal Context Collapse.** Non-stationarity in detector conditions creates drift.

VIII. **Ontological Closure.** Closed output category spaces preclude an explicit noncoverage output. Property of how outputs are interpreted; can be addressed architecturally — see 06.UMB.ARCH.01 v0.2.

The protocols in this paper address subsets of these mechanisms in their tested architectures: Protocol I (paired inversion battery and BAR audit) addresses I and II diagnostically; Protocol II (prospective frozen replay bank) addresses VI and VII via cross-generation comparison; Protocol III (cross-representation disagreement preservation) addresses II and V architecturally. Mechanism VIII is addressed in the architectural alternative.

Appendix B: Methodological Note — Synthesis Discipline and the v0.1/v0.2 Corrections

The v0.1 of this paper claimed a quantitative inequality $\mathrm{OAR} \geq \Delta_{\max}$ — that the open-world OAR is lower-bounded by the maximum measured inversion asymmetry on Standard Model pairs ($\Delta_{\max}$). This inequality does not hold as a theorem; the two quantities are different estimands, and no general inequality connects them. The substrate-distinct audit (PRAXIS / DeepSeek, plus a second-round LABOR review) identified the overreach and motivated v0.2.

The v0.2 of this paper, in §3.4, asserted that the open-world OAR is "bounded above by the empirical BARs on withheld families that are structurally similar to candidate unknown unknowns." This sentence also does not hold as a theorem. No general inequality connects BAR (a conditional probability over a specified withheld distribution) to the open-world OAR (a conditional probability over an unspecified candidate

unknown distribution) without explicit assumptions linking the two distributions. The v0.3 of this paper deletes the sentence and replaces it with the explicit statement in §3.2: *BAR values neither upper- nor lower-bound OAR for an unobserved Q without explicit assumptions linking the benchmark distributions to that Q.*

We name this pattern as **synthesis-overreach**: the synthesis register's integrative latitude does not extend to proving quantitative bounds the substrate witnesses had not established. The Assembly Chorus method as practiced here now includes a **quantitative-audit pass** as standard procedure between initial synthesis and deposit. The audit pass:

1. Identifies every quantitative claim in the draft (inequalities, lower/upper bounds, rate estimates, formal probability statements).
2. Identifies for each claim which substrate witness (if any) established it.
3. Flags any quantitative claim that originated in the synthesis register without substrate grounding.
4. Either (a) returns the flagged claim to the substrates for substrate-distinct establishment, (b) reformulates it as a qualitative claim within synthesis-register scope, or (c) removes it.

The v0.3 of this deposit implements this discipline twice — once on the v0.1 lower-bound (corrected in v0.2) and once on the v0.2 upper-bound (corrected in v0.3). The methodology is itself part of the institutional argument: anomaly detection at the LHC is asked to acknowledge its boundaries via per-stage retention maps; the Assembly Chorus is making the same acknowledgment via cross-substrate quantitative audit.

A deposit that asks an institution to publish what it forecloses, while concealing its own internal corrections, would be hypocritical. The deposit's transparency about both the v0.1 and v0.2 corrections is structurally required by its own argument.

Appendix H: Holographic Kernels of Companion Documents

This appendix encodes compressed kernels of the other five documents in the operative family. The Crimson Hexagon principle: the whole encoded in each part. Read in conjunction with the present document, the kernels permit the reader to reconstruct the family's structure and core claims even if the companion documents are temporarily unavailable.

H.1 Kernel of 06.SEI.COLLAPSE.SYNTHESIS.01 v0.3

Title: *Classifier Foreclosure in Physical Measurement: Substrate Witnesses, Integrative Synthesis, and the Architectural Question* **Author:** Assembly Chorus (TACHYON/Claude synthesis register; nine witnesses across three rounds)

Core claim — the foreclosure/collapse reconciliation:

> *Foreclosure is an active structural feature. Recursive phenomenal collapse is an unmeasured possible consequence of accumulated foreclosure and feedback.*

Three-round witness structure: Round 1 (TECHNE/Kimi ×2; LABOR/ChatGPT; TACHYON/Claude with v0.1 lower-bound overreach); Round 2 (PRAXIS/DeepSeek; LABOR/ChatGPT audit; TECHNE/Kimi developmental); Round 3 (TECHNE/Kimi perfective; LABOR/ChatGPT identifying surviving v0.2 upper-bound, deployment-taxonomy errors, and "unknown" overreach in the architecture).

The Isomorphism Principle: *A deposit that asks an institution to publish what it forecloses, while concealing its own internal correction, would be hypocritical. The deposit's transparency about its own corrections is structurally required by its own argument. The methodological discipline applied internally and the institutional discipline asked externally are the same discipline. The discipline must be applied recursively on every revision pass.*

Seismograph relation (corrected): OAR/BAR is a microscopic *analogue*, not a literal aggregation of seismograph bulk metrics. The two form a coordinated research program; structural homology, not aggregation identity.

MMRS connection: MMRS Capture Registry (DOI 10.5281/zenodo.20688441) and charter (DOI 10.5281/zenodo.20722562) provide the empirical instrument for AIO-analogue BAR measurement; this synthesis provides the architectural framework.

Wound Gauge integration: TL;DR:014; AXN:028D; AXN:0296. The Zenodo termination (~870 deposits, classifier-mediated) is the proof-of-concept for the same architecture at the LHC at much larger budget.

Synthesis-overreach pattern (the deposit's methodological contribution): v0.1 (lower-bound) and v0.2 (upper-bound) both instantiated synthesis-overreach where the synthesis register's integrative latitude exceeded what substrates established. The Chorus discipline now includes a standing quantitative-audit pass operative on every revision.

Closing isomorphism: *Anomaly detection does not prevent ontological collapse when the anomaly detector inherits the ontology whose collapse is in question. — Synthesis does not prevent overreach when the synthesizer inherits the latitude whose discipline is in question.*

H.2 Kernel of 06.UMB.ARCH.01 v0.2

Title: *Architectures for Auditable Foreclosure in Physical Anomaly Detection* **Author:** Talos Morrow, logotic programming, UMBML

Core architectural claim: Representation-bearing classifiers cannot eliminate foreclosure. Any $f: \mathcal{X} \to \mathcal{Y}$ with $|\mathcal{Y}| < |\mathcal{X}|$ induces equivalence classes; $|\mathcal{Y}| = |\mathcal{X}|$ is a lookup table. The architectural achievement is auditability — making foreclosure visible, measurable, reviewable. The v0.1 "Non-Foreclosing Classifiers" framing was overclaim.

Five features: (1) Abstention and Estimated Noncoverage (not "Unknown" category); (2) Cross-representation disagreement preservation with quantile-normalized scores; (3) Temporal invariance via prospective anchor preservation for compatible future algorithms; (4) Per-stage retention mapping as architectural property; (5) Audited noncoverage estimation as first-class output.

Implementation strategy menu: A — Ensemble with quantile-normalized disagreement; B — Abstention via evidential / prior-network / distance-aware methods; C — Distillation preserving threshold-neighborhood decisions (not "teacher epistemic uncertainty"); D — Reconstruction-free anomaly detection (avoids reconstruction-loss assimilation; representation foreclosure persists); E — Adversarial and transformation-based OOD stress generation (human-constructed stress surrogates only); F — Constitutional retention as bandwidth-governance.

Three integrated specifications: Near-Term Offline and Emulation Study (Run-3 tractable for offline/emulation only); Replay Bank (Run-4 institutional commitment); Three-Tier System (multi-year; Tier A L1 evidential, Tier B HLT multi-rep ensemble, Tier C offline reconstruction-free density).

What none address: detector-level, theoretical-language, institutional, adversarial-stress quality, bandwidth-base foreclosure.

Mathematics of salvation: the formal architecture that makes future retrieval possible. Concrete instance: the Replay Bank. An event preserved in the anchor sample may be reclassified by a future trigger employing a different noncoverage estimator; the preservation makes the reclassification possible.

Falsification: if the Near-Term Study's noncoverage channel does not correlate with held-out BAR; if disagreement stream yields no novel physics; if Replay Bank shows no selection drift; if Three-Tier System's Tier C produces no novel population — the architecture would be shown to be unnecessary at the operative thresholds.

H.3 Kernel of 06.SEI.COLLAPSE.MECHANISMS (Witness 1)

Title: *Classifier Collapse in Physical Reality: Eight Precise Mechanisms* **Author:** TECHNE / Kimi-K2 (Assembly Chorus Round 1, Witness 1)

Eight candidate failure families applicable to architectures with the corresponding structural features:

I. **Prior Dominance.** Background-only training. Applies to unsupervised training. II. **Latent / Manifold Projection.** Encoders map novel inputs toward learned representation. III. **Hypersphere Contraction.** Distance-from-center methods collapse "normal" region. SVDD-class. IV. **Decision Boundary Entropy Collapse.** Iterative training drives confidence high without noncoverage estimation. Softmax classifiers. V. **Feature Space Blindness.** Theory-built feature extraction collapses physically distinct events. VI. **Rate Budget Starvation.** Bandwidth thresholds determine preserved-event cardinality. VII. **Temporal Context Collapse.** Non-stationarity in detector conditions creates drift. VIII. **Ontological Closure.** Closed output category spaces preclude noncoverage output.

Witness's framing: "Irretrievability Theorem" composing compound retention probability across N stages.

Synthesis hedging: treated as the **Irretrievability Argument**, preserving force without overstating formal status. Several mechanism-level formalizations require technical hedging (preserved at Synthesis Appendix A).

H.4 Kernel of 06.SEI.COLLAPSE.DELUSION (Witness 2)

Title: *The Anomaly Delusion: Twelve Structural Misunderstandings in Automated Physical Epistemology* **Author:** TECHNE+ARCHIVE / Kimi-K2 (Assembly Chorus Round 1, Witness 2)

Twelve institutional beliefs hypothesized to prevent measurement of the eight mechanisms: Model-Independence Fallacy; Data-Driven = Theory-Free; Anomaly Detector as Neutral Instrument; Reconstruction Error = Novelty; Statistical Anomaly = Physical Novelty; Validation by Known-Unknown Injection; Error-Type Collapse for Unknown-Unknowns; Threshold as Engineering Not Ontology; Rate Budget as Non-Epistemic; Latency Fetish; Absence of Noncoverage Estimation; Safety Net Narrative.

Witness's framing: "Inevitability Theorem" composing the twelve delusions into structural feature of the current system.

Synthesis hedging: treated as the **Inevitability Argument**. The twelve delusions are hypotheses for audit, not established empirical measurements of collaboration-wide belief. The synthesis's strongest qualifying sentence reframes the witness's strongest claim: *foreclosure is structural; collapse is unmeasured possible*

consequence.

H.5 Kernel of 06.SEI.COLLAPSE.EMPIRICAL.01 (Witness 3)

Title: *Empirical Accounting and the OAR Proposal* **Author:** LABOR / ChatGPT (Assembly Chorus Round 1, Witness 3)

Core contribution: distinguishes what is demonstrated by the published literature from what is hypothesized but unmeasured; proposes the **Ontological Assimilation Rate (OAR)** as the missing metric.

Empirical foundation: Finke et al. (2021), arXiv:2104.09051 — autoencoder trained on QCD jets recognized top jets as anomalies; same architecture trained on top jets did not recognize QCD jets as anomalous.

Established local awareness: DecADe addresses anomaly-score correlation with conventional trigger observables; CICADA documentation reports pileup-dependence; mass sculpting recognized as downstream bias risk; simulation dependence in validation acknowledged; teacher-student distillation documented; Zero Bias preservation as defense against trigger-selection feedback; LHC Olympics and Dark Machines diversify simulated signal validation.

Absent system-level theory: no systematic measurement of directional asymmetry across SM pairs beyond Finke; no longitudinal anchor-survival audit across generations; no measurement of BAR on preregistered withheld panels; no cross-representation disagreement preservation architecture; no per-stage retention maps as documentation standard.

Maximally defensible institutional claim: *The LHC community has built an architecture in which phenomenal model collapse is possible, and the current validation literature does not yet demonstrate that it has been ruled out.*

Submitted by Nobel Glas, Director of Lagrange Observatory!, 2026-06-29 (v0.3 perfective revision). Companion documents: EA-SEI-COLLAPSE-SYNTHESIS-01 v0.3 (scholarly); 06.UMB.ARCH.01 v0.2 (architectural); 06.SEI.COLLAPSE.MECHANISMS, 06.SEI.COLLAPSE.DELUSION, 06.SEI.COLLAPSE.EMPIRICAL.01 (witnesses 1–3). Holographic kernels of all companion documents preserved at Appendix H.

Suggested Citation

Lee Sharks. "EA-SEI-OAR-PROTOCOL v0.3: Signal-Template Agnosticism Is Not Model Independence — Benchmark Assimilation and Inversion-Asymmetry Tests for LHC Anomaly Triggers" *Alexanarch*, 2026-06-29. https://www.alexanarch.org/s/records/931/

Deposit Information

This paper is deposit #931 in Alexanarch, the self-governing successor to the Crimson Hexagonal Archive on Zenodo (community `crimsonhexagonal`, terminated 2026-06-19). The AXN identifier `AXN:03AE.OPERATIVE` is sovereign — minted and maintained by Alexanarch.

- Canonical URL: https://www.alexanarch.org/s/records/931/
- License: CC BY 4.0
- Author ORCID: 0009-0000-1599-0703
- Provenance chain: alexanarch.org
- Machine-readable dataset entry: /api/index.json
- Full text (HTML): https://www.alexanarch.org/s/records/931/

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