CAMS/JUNO Audit Addendum

Blind-Corpus Stress Test of JUNO v1.1-Final and the Resulting v1.2-Final Revision

July 2026 • Companion to: JUNO v1.1-Final Formalism (June 2026 audit) and JUNO v1.2-Final Formalism

1. Aim

This addendum documents the first full-pipeline stress test of the JUNO v1.1-Final formalism against the complete blind corpus, and the minimal set of specification changes (v1.2-Final) that the test motivated. The test was designed under the programme's falsification-first discipline: the objective was not to confirm v1.1 but to break it, by (a) executing every operator and the complete regime classifier over a corpus that played no role in v1.1 calibration, (b) auditing every quantitative claim in the v1.1 document against observed behaviour, and (c) treating any dead code, coverage gap, bound exceedance, or reproducibility failure as a reportable defect.

Three questions were posed in advance. First, does the sign-preserving cognitive-activation correction (the central v1.1 fix) produce materially different behaviour from the v1.0 floor clip, or is it cosmetic? Second, do the v1.1 §4 "validated bounds" and the claimed λ_2 regime discrimination generalise beyond the calibration corpus? Third, is the classifier complete and internally consistent — does every rule remain reachable under the stated precedence, and does every society-year receive a label?

2. Method

2.1 Corpus

The test corpus comprised the three JUNO blind sets: 24 model-scored societies, 3,002 society-years, 24,016 node-years, each node-year carrying raw Coherence, Capacity, Stress, and Abstraction scores on the standard 0–10 decimal scale across the eight CAMS nodes. The corpus is model-generated; the implications of that provenance are treated in Section 5.

2.2 Pipeline

For every society-year the full v1.1 operator set was computed: node viability $V_i = C + K - S + 0.5 \cdot A$; sign-preserving cognitive activation $s_i = (A \cdot C/100) \cdot \text{sign}(K-S) \cdot \max(|K-S|, 0.1)$; coupling quality $q_i = (0.6C + 0.4A)/10$; bond strength $B_{ij} = \sqrt{(q_i q_j)} \cdot 2^{\wedge}(-(S_i+S_j)/10)$ over all 28 node pairs, bounded [0,1]; and algebraic connectivity λ_2 as the second-smallest eigenvalue of the graph Laplacian of the weighted adjacency matrix. Year-level aggregates ($\bar{V}$, V_min, s_min, $\bar{B}$) were then passed through the v1.1 regime classifier under its stated precedence. A parallel run substituted the v1.0 floor-clip activation operator, holding everything else fixed, to isolate the effect of the sign-preserving correction.

2.3 Audits performed

Five audits were applied: (1) bounds — observed ranges of B_{ij}, λ_2, and s_i against the v1.1 §4 claims; (2) coverage — count and characterisation of Unclassified society-years; (3) reachability — whether each classifier arm can fire under precedence, tested empirically and by logical analysis of the threshold structure; (4) λ_2 discrimination — Kruskal–Wallis across regimes, one-sided pairwise Mann–Whitney tests on all adjacent regime pairs, and, critically, a circularity check regressing λ_2 on $\bar{B}$ (a classifier input) with per-case AUC separability for adjacent pairs; (5) reproducibility — recomputation of all labels from raw data and comparison against labels derived from file-round-tripped intermediates.

3. Results

3.1 Summary of claims tested

v1.1 claim / property	Outcome	Evidence
Sign-preserving s_i restores negative activation and Phantom Type II	Confirmed, material	31.8% of node-years negative; Phantom II fires 250 society-years (v1.0: 0); 12.1% of society-years reclassified vs v1.0
Classifier coverage improved over v1.0's 2.1% unclassified	Confirmed	99.8% coverage (6 of 3,002 unclassified)
λ_2 regime discrimination (Stable > Strained > LNF)	Holds descriptively; extends to all six regimes; largely circular — see 3.4	Monotonic medians; KW H ≈ 2529; but $r(\lambda_2, \bar{B}) = 0.972$
§4 bound: $B_{ij} \leq 0.657$	Exceeded	Observed max 0.758 (theoretical [0,1] holds)
§4 bound: $\lambda_2 \in [0.379, 3.668]$	Exceeded both ends	Observed [0.268, 5.354]
§4 bound: s_i max +1.68	Exceeded	Observed max +8.00 (legitimate configuration: C=K=A=10, S=2)
§4 bound: s_i min −2.76	Held	Observed min −2.50
All classifier arms reachable	Failed	LNF s-arm (s_min ≤ −0.80) unreachable — see 3.2
Full coverage	Failed (marginally)	Six-case gap with a specific shape — see 3.3
Label reproducibility across computation paths	Failed	Floating-point boundary artifact on three society-years — see 3.5

3.2 Dead code in v1.1: the Local Node Failure s-arm

The LNF rule in v1.1 reads V_min < 4.0 OR s_min ≤ −0.80. Because Phantom Type II (s_min < −0.65) sits above LNF in precedence and triggers at a strictly weaker threshold, every case satisfying the s-arm has already been captured before LNF is evaluated. This is a logical property of the threshold structure, not a corpus accident. Empirically, 459 society-years in the blind corpus satisfy s_min ≤ −0.80; all 459 resolve as Systemic Crisis, Phantom Type II, or Freeze/Collapse, and zero reach LNF via the s-arm. This is the same class of defect — an unreachable classifier arm — that the June 2026 audit identified in v1.0 and that v1.1 was issued to eliminate.

3.3 Coverage gap

All six unclassified society-years share one profile: $\bar{V} \in [5.35, 5.65]$ with V_min ≥ 4 and healthy cognition — low-vitality but internally coherent systems that fail Systemic Crisis (bonds too strong or V_min ≥ 0), fail LNF (no weak node), and fall below the v1.1 Strained floor of $\bar{V} \geq 6$. The gap region is bounded: since any case reaching the Strained rule has survived LNF, V_min ≥ 4 holds there, and because $\bar{V} \geq$ V_min always, lowering the Strained floor to $\bar{V} \geq 4$ closes the gap as a theorem rather than an empirical observation. This fix was preferred over an unconditional else-clause because it preserves Unclassified as a live data-quality tripwire.

3.4 λ_2: strong ordering, weak independence

λ_2 orders all six regimes monotonically by severity (medians: Stable Adaptive 2.68 > Strained 1.95 > Phantom Type II 1.52 > Local Node Failure 1.38 > Systemic Crisis 0.83 > Freeze/Collapse 0.51; Kruskal–Wallis

$H \approx 2529$; every adjacent pairwise step significant, worst one-sided $p \approx 5\times10^{-9}$). Taken alone this would be the strongest validation result of the test. It does not survive the circularity check. λ_2 correlates with mean bond strength $\bar{B}$ at $r = 0.972$ (Spearman 0.970), close to a mathematical necessity for a dense eight-node graph, where λ_2 tracks $n\cdot$(mean edge weight): the ratio $\lambda_2/(8\cdot\bar{B})$ is 0.83 ± 0.10 across all 3,002 years. $\bar{B}$ is itself a classifier input — Systemic Crisis, Freeze/Collapse, and Stable Adaptive are defined partly by $\bar{B}$ thresholds — so the ordering substantially restates the classifier's own inputs. After regressing $\bar{B}$ out of λ_2, the residual ordering collapses for several contrasts, and per-case separability is strong only where the underlying $\bar{B}$ separation is strong (Stable vs Strained AUC 0.95) and weak elsewhere (Phantom II vs LNF AUC 0.63; the small p-value there is a sample-size effect, $n = 725$). The proposal to promote λ_2 to a core regime discriminator was accordingly withdrawn. Its legitimate roles are internal-consistency checking of the operator algebra and the ≈ 17% deviation from the pure-$\bar{B}$ prediction, which carries the network-topology information not captured by mean bond strength.

3.5 Numerical boundary artifact

During verification, three society-years changed regime label depending on computation path with no change to data or rules. Their V_min is mathematically exactly 4.0, but IEEE-754 arithmetic stores C + K − S + 0.5·A as 3.9999999999999996; the strict comparison V_min < 4.0 therefore evaluates true in-memory and false after a file round-trip renders the value as 4.0. With one-decimal input scores, aggregate values will land exactly on classifier thresholds routinely, so this is a specification gap rather than an implementation bug: v1.1 is silent on how boundary values are compared. v1.2 adds a mandatory numerical policy — round all aggregates ($\bar{V}$, V_min, s_min, $\bar{B}$) to six decimal places before every threshold comparison. The policy itself moves a handful of boundary cases relative to unrounded comparison, which is precisely why it must live in the canonical specification rather than in each implementer's floating-point behaviour.

3.6 The v1.2-Final adjustments and their verification

Four substantive changes were adopted, each verified by full recomputation from raw scores. The LNF s-arm was deleted (behaviour-preserving by construction: zero firings existed to preserve; the half-specified “confined negative activation” replacement language was struck rather than smuggled in unoperationalised). The Strained floor was lowered from $\bar{V} \geq 6$ to $\bar{V} \geq 4$, yielding provably complete coverage while retaining Unclassified as a tripwire. The numerical policy of 3.5 was added. The §4 bounds were relabelled as reference ranges from audited panels and removed from validation criteria, with the corpus-descriptive figures updated to the audited blind-corpus values. Under v1.2-Final the corpus classifies 100% with exactly the six former gap cases moving to Strained and no other reclassification.

Regime	v1.0 (floor clip)	v1.1-Final	v1.2-Final
Stable Adaptive	1375	1228	1228
Strained	541	584	593
Local Node Failure	621	478	475
Systemic Crisis	385	385	383
Phantom Type II	0	247	250
Freeze/Collapse	74	74	73
Unclassified	6	6	0

Counts are society-years (n = 3,002). Small v1.1→v1.2 shifts outside the six gap cases reflect the newly mandated rounding policy acting at threshold boundaries, not rule changes; v1.0 and v1.1 columns as originally computed without the rounding policy.

4. Audited Reference Ranges (blind corpus)

The following are descriptive statistics of the blind corpus under v1.2 operators, superseding the v1.1 §4 figures as reference material. They are not instrument bounds; the only intrinsic bounds are $B_{ij} \in [0, 1]$ and $\lambda_2 \geq 0$.

Quantity	Observed range
V_i (node-level)	[−7.50, +23.00]
s_i (node-level)	[−2.50, +8.00]
s_min (year-level)	[−2.50, +5.40]
B_{ij} (pair-level)	[0.025, 0.758]
$\bar{B}$ (year-level)	[0.045, 0.722]
λ_2 (year-level)	[0.268, 5.354]

5. Limitations

Three limitations bound the interpretation of every result above. First, the corpus is model-generated: the test demonstrates operator coherence and classifier behaviour over model-scored panels — compression consistency — not external validity against ground truth. Ensemble tightness certifies precision, not accuracy, and shared model bias remains an unquantified confound. Second, the corpus is now spent for validation purposes: the v1.2 thresholds were chosen by inspection of these 3,002 society-years, so re-running the same panel is a consistency check, and the reported 100% coverage on it is an in-sample result. Third, the λ_2 monotonicity, while real, is not independent evidence for the framework given its near-mechanical coupling to a classifier input; it should not be foregrounded as a validation win in any presentation of the programme.

6. Next Steps

(1) JUNO v1.2-Final is locked as the canonical specification; all subsequent scoring, reporting, and prospective work references it explicitly, including the numerical policy. (2) The prospective 2026–2028 monitoring protocol is updated so that all forward predictions are issued under v1.2-Final with between-run reliability margins (±1 NV at year level, wider at node level) and are scored against fresh holdout data as it arrives; that holdout, not the blind corpus, is the venue for genuine v1.2 validation. (3) Application of v1.2 to recent national series is deferred until independently scored panels exist — generating scores and classifying them within the same model context would reintroduce exactly the single-scorer dependence the multi-agent ensemble pipeline was built to avoid; the recommended route is multi-agent scoring (true scorer independence) followed by v1.2 classification of the aggregated panel. (4) Any future “confined negative activation” arm for Local Node Failure must be fully specified, pre-registered, and tested on fresh data before entering a minor revision.