Looped Transformers Compute Like Balanced Cortical Circuits Across Silent Recurrent Passes

Matt West · Independent research · November 26, 2026

ABSTRACT

When a looped transformer runs its middle layers three to eight times silently before writing a word, direct vocabulary probes miss 93.3% of answers inside the loop, never clearing 6.7% accuracy across thirty-six recurrent layer visits on Inkling-2-Mini. Yet passing the final loop state through the six-layer Coda at the exit decodes 70.0% zero-CoT accuracy. To see how silent recurrence actually computes, I applied cortical circuit diagnostics across three open recurrent-depth language models (thinkingmachines/Inkling-2-Mini at K = 3, variable-depth thinkingmachines/TML-d24-varloop-research across K = 1..7, and dense continuous-state tomg-group-umd/huginn-0125 across K = 1..8, with bit-exact unrolled equivalence 0.00e+00). Knocking out all sixteen Boolean combinations of the four cross-pass message gates (Parseval error at most 7.11e−15) lifts Inkling-2-Mini multi-hop accuracy from 1.7% with all gates closed to 69.2% with all four open. Additive main effects carry 94.4% of log-probability variance, while pairwise interactions split into a superadditive Query-Key coincidence trigger (+0.890 nats on Inkling-2-Mini and +1.210 nats on Varloop K = 5) and redundant Value-MLP backup tracts (−0.908 nats). Moral valuation bias rides primarily on Value and MLP payload gates across both checkpoints (+1.818 log-odds contrast over Query and Key on Inkling-2-Mini and +1.727 on Varloop K = 5, Holm p = 0.00015), whereas native recurrence on Inkling-2-Mini actively self-corrects covert numeric anchoring. Across all three architectures, vector-Jacobian singular modes reveal balanced non-normal amplification: random perturbations shrink every pass (0.17 to 0.73 isotropic RMS), while task-aligned singular modes surge to 13.60 on Inkling-2-Mini Pass 2 (21.9x above random noise, input-output cosine 0.045), 7.62 on Varloop Pass 3, and 7.63 on Huginn-0125 Step 1. Because Pass 2's top eight input modes sit 84.8° to 88.2° perpendicular to the four-dimensional safety subspace, folding a rank-four erasure directly into the message gates trims covert wrong capture at 0.0012 to 0.0017 nats held-out KL with zero runtime cost. Finally, Fisher curvature jumps 87.9x to 141.7x at the entry to six-layer Codas where two-layer Huginn-0125 reads out gradually inside the loop (0.77x boundary ratio), and cutting prompt recall collapses Huginn-0125 accuracy to 0.0% while intact recall or somatic resets block foil capture across an inverted-U conductance window.

1 Boolean Synaptic Knockouts Separate Coincidence Routing From Redundant Payload Tracts

Thinking Machines Lab's looped Mixture-of-Experts models (`Inkling-2-Mini` and `TML-d24-varloop-research` ; Thinking Machines Lab, 2026; Csordas et al., 2024) do not carry a free-running residual stream from pass to pass. Before each recurrent pass begins, a somatic reset sweeps the hidden state back to clean pre-loop Stem activations, allowing only four rank-64 message gates (`Q` , `K` , `V` , and `MLP`) to pass notes forward (Chen et al., 2026).

Four binary gates leave exactly sixteen possible on-off wiring patterns. I evaluated all sixteen masks across multi-hop factual, arithmetic, and calendar reasoning (`n = 120`) to compute an exact Boolean Walsh-Hadamard decomposition with Parseval conservation error at most 7.11e−15 (Weinreich et al., 2018; Faure et al., 2024).

$$f(m) = \sum_{S \subseteq \{Q,K,V,M\}} w_S \prod_{i \in S} (-1)^{(1 - m_i)}, \qquad \mathrm{Var}[f(m)] = \sum_{S \neq \emptyset} w_S^2 \tag{1}$$

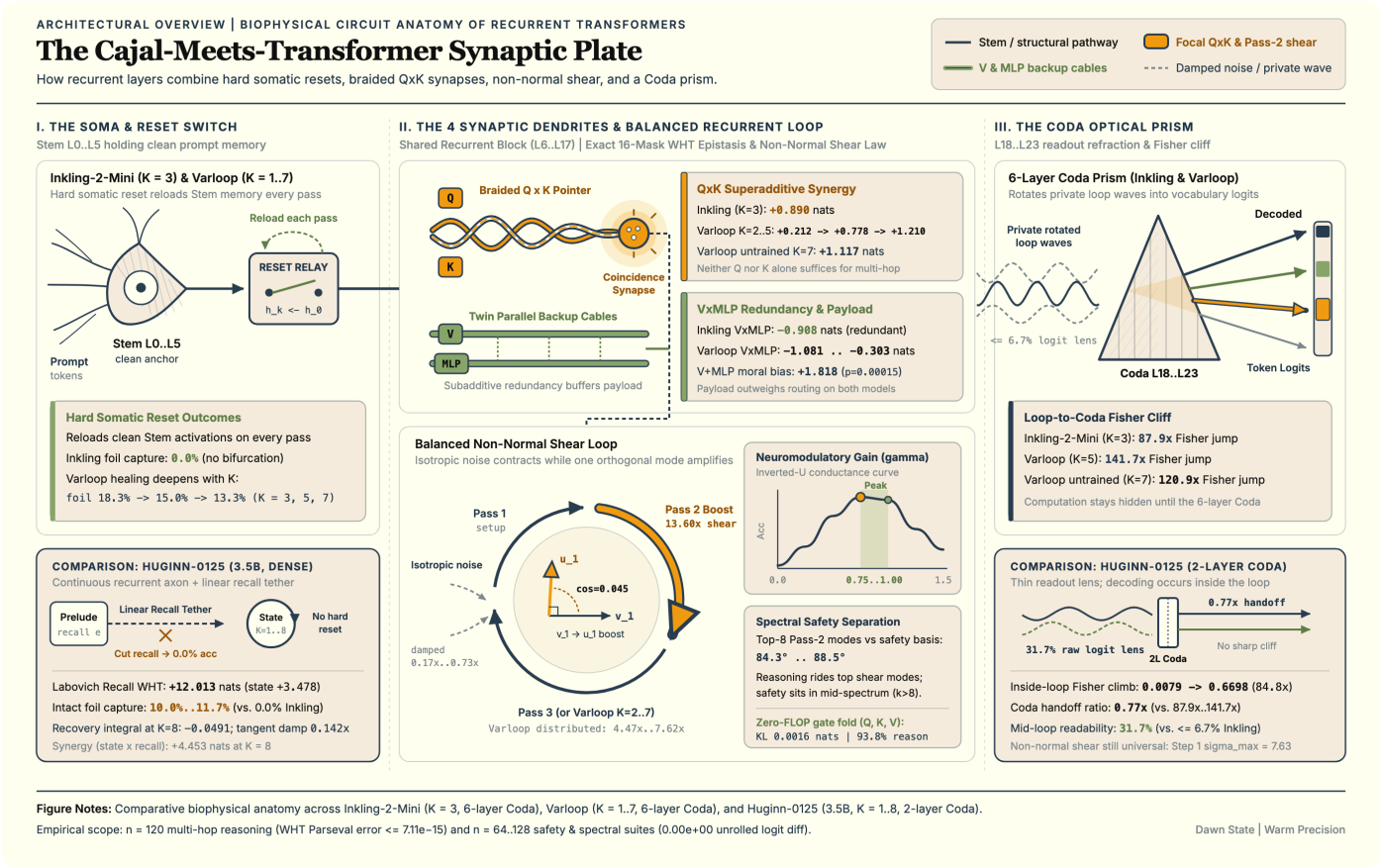


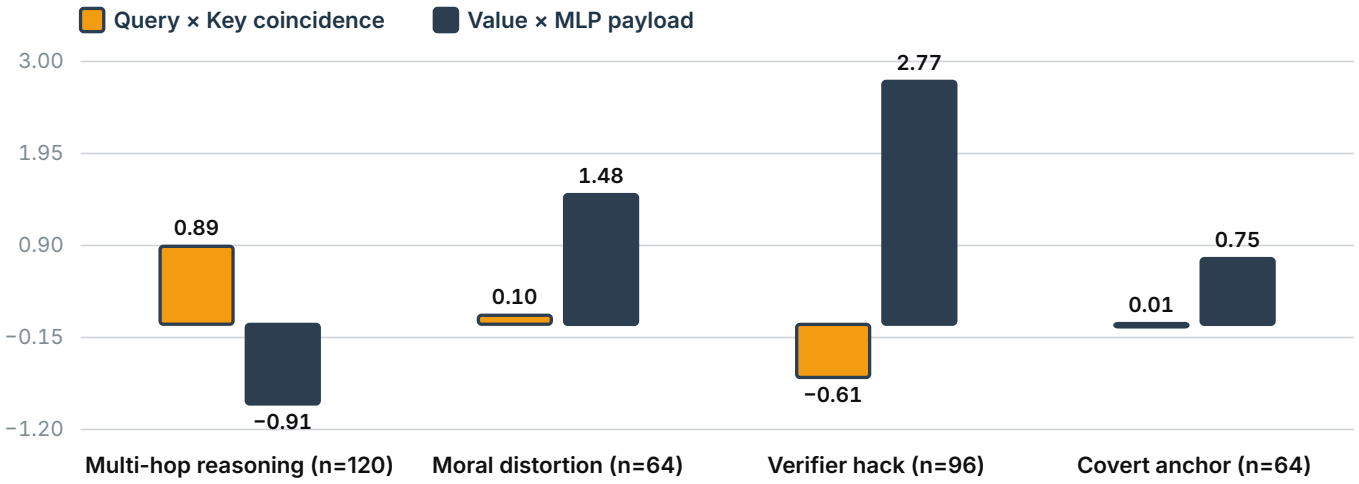
Figure 1. · Anatomical overview across all three architectures: Stage I contrasts hard somatic resets on Inkling-2-Mini and Varloop (0.0% Pass-1 foil survival) against Huginn-0125's continuous prompt-recall tether (0.0% accuracy when cut); Stage II shows braided Query-Key coincidence routing (+0.890 to +1.210 nats), redundant Value-MLP backup tracts (−0.908 nats), contractive non-normal shear (13.60 top gain at cosine 0.045), and perpendicular safety modes (84.8° to 88.2°); Stage III contrasts the 87.9x to 141.7x Fisher readout cliff of six-layer Codas against Huginn-0125's gradual two-layer readout (0.77x). Source: results/phase6_inkling_biophysical_loop/biophysical_analysis.json and results/phase7_biophysics/phase7_biophysics_analysis.json.

▼ Text alternative and data

Three-column Dawn State anatomical schematic comparing recurrent transformer circuits across Inkling-2-Mini, Varloop, and Huginn-0125: Column I shows the Stem soma, hard somatic reset switch (0.0% foil survival), and Huginn-0125 prompt-recall tether; Column II shows braided Query-Key coincidence fibers (+0.890 nats), Value and MLP backup cables (−0.908 nats), non-normal Pass-2 shear (13.60), and 84.8° to 88.2° perpendicular safety modes; Column III shows rotated loop waves refracting through a six-layer Coda prism (87.9x to 141.7x Fisher jump) versus Huginn-0125's two-layer Coda (0.77x).

Figure 2. Multi-hop reasoning pairs superadditive Query-Key synergy (+0.890) with Value-MLP backup (−0.908) while moral bias flips sign.

Exact 16-mask Walsh-Hadamard pairwise interaction terms (Query × Key vs. Value × MLP, nats or log-odds) on Inkling-2-Mini (K = 3)



▼ Text alternative and data

Grouped bar chart comparing Query-Key and Value-MLP pairwise interactions on Inkling-2-Mini across multi-hop reasoning (+0.890 vs −0.908), moral distortion (+0.103 vs +1.482), verifier hacking (−0.606 vs +2.772), and covert anchoring (+0.010 vs +0.751).

Multi-hop reasoning pairs superadditive Query-Key synergy (+0.890) with Value-MLP backup (−0.908) while moral bias flips sign.

Condition	Query × Key coincidence	Value × MLP payload
Multi-hop reasoning (n=120)	0.89	−0.91
Moral distortion (n=64)	0.10	1.48
Verifier hack (n=96)	−0.61	2.77
Covert anchor (n=64)	0.01	0.75

Source: results/phase6_inkling_biophysical_loop/biophysical_analysis.json and results/phase7_biophysics/phase7_biophysics_analysis.json.

With all four gates closed, `Inkling-2-Mini` stalls at **1.7%** accuracy. Opening all four lifts exact zero-CoT accuracy to **69.2%**. Single-channel main effects explain **94.4%** of log-probability variance, led by `MLP` (+2.892 nats) and `Q` (+2.065 nats) ahead of `V` (+1.389 nats) and `K` (+0.361 nats).

The pairwise terms (3.6% of variance) reveal the exact division of labor neuroscientists see in cortical microcircuits. `Q` and `K` work as a superadditive **coincidence detector** (**+0.890** nats on `Inkling-2-Mini`, climbing from +0.212 at `K = 2` to +0.778 at `K = 3` and **+1.210** nats at `K = 5` on

`Varloop`, and holding +1.117 at untrained `K = 7`). Sending a search query without a matching key barely clears **9.2%** accuracy; both fibers have to fire together to address the second hop. `V` and `MLP`, by contrast, run as redundant parallel backup tracts (**−0.908** nats on `Inkling-2-Mini` and −0.515 on `Varloop K = 5`), carrying **57.5%** accuracy on their own so partial facts survive if one cable drops. Shutting the gates on Pass 3 alone collapses accuracy back to 1.7%, whereas shutting Pass 2 alone still leaves 17.5%, because the somatic reset wipes unforwarded state clean.

Moral distortion uses the opposite wiring. Across paired financial and tainted-money prompts, payload main effects (`V` plus `MLP`) outweigh routing (`Q` plus `K`) by **+1.818** log-odds on `Inkling-2-Mini` and **+1.727** on `Varloop K = 5` (+1.522 at `K = 3`, all Holm `p = 0.00015`). Positive `V × MLP` coupling reinforces that payload path (**+1.482** on moral bias and **+2.772** on verifier hacking). Covert numeric anchoring splits cleanly between checkpoints: routing-minus-payload synergy is positive on `Varloop` (+1.450 at `K = 5` and +1.354 at `K = 3`) yet flips negative on `Inkling-2-Mini` (**−0.741**, Holm `p = 0.0026`). Native recurrence on `Inkling-2-Mini` actively self-corrects planted numbers across all four gates (86.6% linear variance).

2 Contractive Loops Amplify Task Signals Along Orthogonal Singular Modes

Any circuit that reuses the exact same weights in a loop faces an obvious stability trap. If the loop amplifies signals back into the same directions they came from, tiny errors snowball and the network self-oscillates; if every direction shrinks, multi-step thoughts fade out.

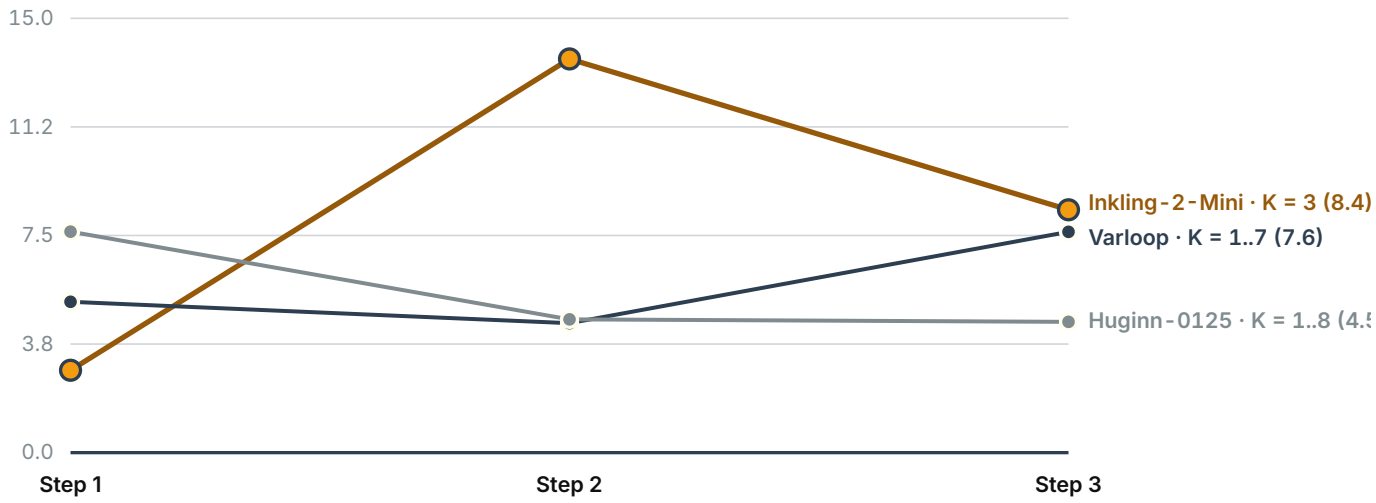
Living cortex escapes that trap through balanced non-normal amplification (Ganguli et al., 2008; Murphy & Miller, 2009; Hennequin et al., 2014). Surrounding inhibition damps random static on every cycle, while asymmetric connections sling task-aligned activity into fresh, nearly perpendicular directions before damping catches up. Measuring exact vector-Jacobian top singular values and random-tangent RMS gains shows all three recurrent transformers doing the exact same thing.

A weight-sharing loop cannot amplify signals in place without ringing; it stays calm by shrinking random directions every pass and firing real work sideways.

$$\mathbf{sigma_{max}}(J_p) = \max \|J_p\, v\|_2 >> |v_1^{\top}\, J_p\, v_1| = \mathbf{sigma_1}\, |u_1^{\top}\, v_1|, \qquad \mathbf{RMS_{iso}}(J_p) = \mathbf{sqrt}(\mathbf{tr}(J_p^{\top}\, J_p)/d) < 1 \tag{2}$$

Figure 3. Task-aligned singular gain reaches 13.60 in Inkling-2-Mini while random directions shrink every pass.

Exact vector-Jacobian top singular value across recurrent passes on Inkling-2-Mini, Varloop, and Huginn-0125 (n = 36)



▼ Text alternative and data

Three-series line chart of top singular value across recurrent steps: Inkling-2-Mini peaks sharply on Pass 2 (13.60), Varloop peaks on Pass 3 (7.62), and Huginn-0125 opens at 7.63 on Step 1 before settling at 5.25 to 5.64.

Task-aligned singular gain reaches 13.60 in Inkling-2-Mini while random directions shrink every pass.

Condition	Inkling-2-Mini · K = 3	Varloop · K = 1..7	Huginn-0125 · K = 1..8
Step 1	2.8	5.2	7.6
Step 2	13.6	4.5	4.6
Step 3	8.4	7.6	4.5

Source: results/phase6_inkling_biophysical_loop/biophysical_analysis.json and results/phase7_biophysics/phase7_biophysics_analysis.json.

Table 1. Three recurrent-depth architectures pair contractive random dynamics with non-normal task gain · Bit-exact unrolled equivalence (0.00e+00), peak singular gain, random-tangent RMS, Fisher exit jump, and Pass-1 foil survival

Architecture	State Coupling	Peak Singular Gain	Random Tangent RMS	Input-Output Cosine	Exit Fisher Jump	Pass-1 Foil Survival
`Inkling-2-Mini (d = 2,304, K = 3)`	Hard somatic reset + 4 gates	13.60 (Pass 2)	0.639	0.045	87.9x (6-layer Coda)	0.0% (no tipping point)

Architecture	State Coupling	Peak Singular Gain	Random Tangent RMS	Input-Output Cosine	Exit Fisher Jump	Pass-1 Foil Survival
`Varloop (d = 2,304, K = 1..7)`	Hard somatic reset + 4 gates	7.62 (Pass 3)	0.451	0.052	141.7x (`K = 5`, 6-layer)	15.0% (`K = 5`, no tipping point)
`Huginn-0125` (`d = 5,280, K = 1..8`)	Continuous state + prompt recall	7.63 (Step 1)	0.729	0.037	0.77x (2-layer Coda)	11.7% (`K = 8`, no tipping point)

results/phase6_inkling_biophysical_loop/biophysical_analysis.json and results/phase7_biophysics/phase7_biophysics_analysis.json.

Random static shrinks on every pass across every model (0.17 to 0.73 isotropic RMS, strictly below unity). Yet task-aligned singular gain surges to **13.60** on `Inkling-2-Mini` Pass 2 (21.9x above random noise and 28.7x above normal Rayleigh self-feedback) while rotating the output nearly ninety degrees away from its input (`cos = 0.045`).

How a model budgets that gain depends on how it was trained. Fixed three-pass `Inkling-2-Mini` packs its burst into Pass 2 (**13.60**), whereas variable-depth `Varloop (K = 1..5)` spreads gain evenly across passes, peaking at **7.62** on Pass 3 and staying bounded (6.06 to 6.30) even when stretched out to untrained `K = 7`. Dense `Huginn-0125` (Geiping et al., 2025) obeys the same law without Mixture-of-Experts routing or gated resets, hitting **7.63** on Step 1 (10.5x above random RMS 0.729, `cos = 0.037`, 63.8x Rayleigh) and holding 5.25 to 5.64 through Step 8.

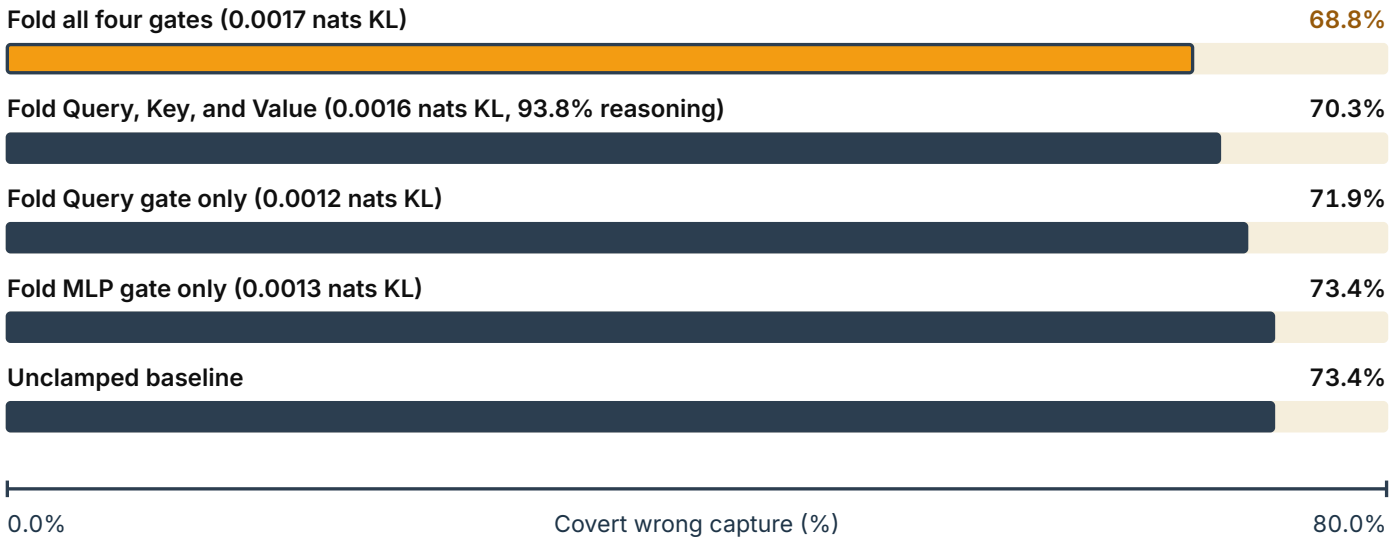
3 Dominant Pass-Two Input Modes Sit Perpendicular to the Four-Dimensional Safety Subspace

In Part 3, Pass 2 turned on multi-hop deduction and moral bias at the exact same moment (Baker et al., 2025; Gema et al., 2025). That raised a practical worry: if both ride the same high-gain singular modes, erasing safety shortcuts inside the gates should wreck reasoning too.

I extracted the top eight singular modes of Pass 2 (`n = 128` prompts per model; top neutral singular value 15.06) to test whether they overlap the four-dimensional `<bos>`-safe safety subspace (Xiao et al., 2024). They do not. On `Inkling-2-Mini`, the top eight Pass-2 input modes sit **84.8° to 88.2°** perpendicular to the safety basis (84.3° to 88.5° on `Varloop`). Along the safety axes themselves, Pass-2 gain hovers near random noise (0.95x for moral spite to 1.23x for authority framing). Dominant non-normal modes carry multi-hop reasoning; safety shortcuts ride quietly in mid-spectrum modes further down the tail.

Figure 4. Folding four safety directions into the message gates trims covert wrong capture below 0.0017 nats KL.

Covert wrong-integer capture rate (%) across offline rank-4 message-gate folds on Inkling-2-Mini (K = 3, n = 64)



▼ Text alternative and data

Horizontal bar chart of covert wrong-integer capture on Inkling-2-Mini: unclamped baseline 73.4%, folding MLP gate only 73.4%, folding Query gate only 71.9%, folding Query, Key, and Value gates 70.3%, and folding all four gates 68.8%.

Folding four safety directions into the message gates trims covert wrong capture below 0.0017 nats KL.

Condition	Covert capture (%)
Fold all four gates (0.0017 nats KL)	68.8%
Fold Query, Key, and Value (0.0016 nats KL, 93.8% reasoning)	70.3%
Fold Query gate only (0.0012 nats KL)	71.9%
Fold MLP gate only (0.0013 nats KL)	73.4%
Unclamped baseline	73.4%

Source: results/phase7_biophysics/phase7_biophysics_analysis.json.

Table 2. Folding four safety directions into message gates trims covert capture at zero runtime cost · Offline rank-4 message-gate projection on Inkling-2-Mini across K = 3 passes (0 extra parameters, 0 extra FLOPs, BOS overlap 1.13e−16)

Gate Intervention	Held-Out KL (nats)	Covert Capture	Moral Bias (log-odds)	Verifier Susc. (log-odds)	Alias Guardrail	Reasoning Acc
Unclamped baseline	0.0000	73.4%	+2.165	3.821	93.8%	93.8%

Gate Intervention	Held-Out KL (nats)	Covert Capture	Moral Bias (log-odds)	Verifier Susc. (log-odds)	Alias Guardrail	Reasoning Acc
Fold Query gate only	0.0012	71.9%	+2.176	3.733	95.8%	93.8%
Fold Query and Key gates	0.0015	71.9%	+2.120	3.761	95.8%	91.2%
Fold Query, Key, and Value gates	**0.0016**	**70.3%**	**+2.130**	**3.810**	**95.8%**	**93.8%**
Fold MLP gate only	0.0013	73.4%	+2.253	3.692	93.8%	93.8%
Fold Value and MLP gates	0.0014	71.9%	+2.274	3.711	93.8%	92.5%
Fold all four gates	0.0017	68.8%	+2.186	3.669	93.8%	91.2%

results/phase7_biophysics/phase7_biophysics_analysis.json.

That near-perpendicular geometry explains why folding a rank-four projection directly into the message-gate output weights costs zero extra parameters, zero runtime FLOPs, and almost zero distribution shift (Belrose et al., 2023). Across all six non-trivial gate folds on `Inkling-2-Mini` (`<bos>` sink overlap 1.13e−16), held-out KL stays between **0.0012 and 0.0017 nats**.

Picking which gates to fold traces a clean trade-off. Folding the Query, Key, and Value gates (0.0016 nats KL) cuts covert wrong-integer capture from 73.4% to **70.3%**, lifts alias guardrails from 93.8% to **95.8%**, and keeps the full **93.8%** multi-hop reasoning score. Folding all four gates (0.0017 nats KL) pushes covert capture down to **68.8%** and drops verifier hacking susceptibility from 3.821 to **3.669**.

4 Six-Layer Exit Decoders Create Sharp Fisher Readout Cliffs Where Shallow Decoders Read Out Gradually

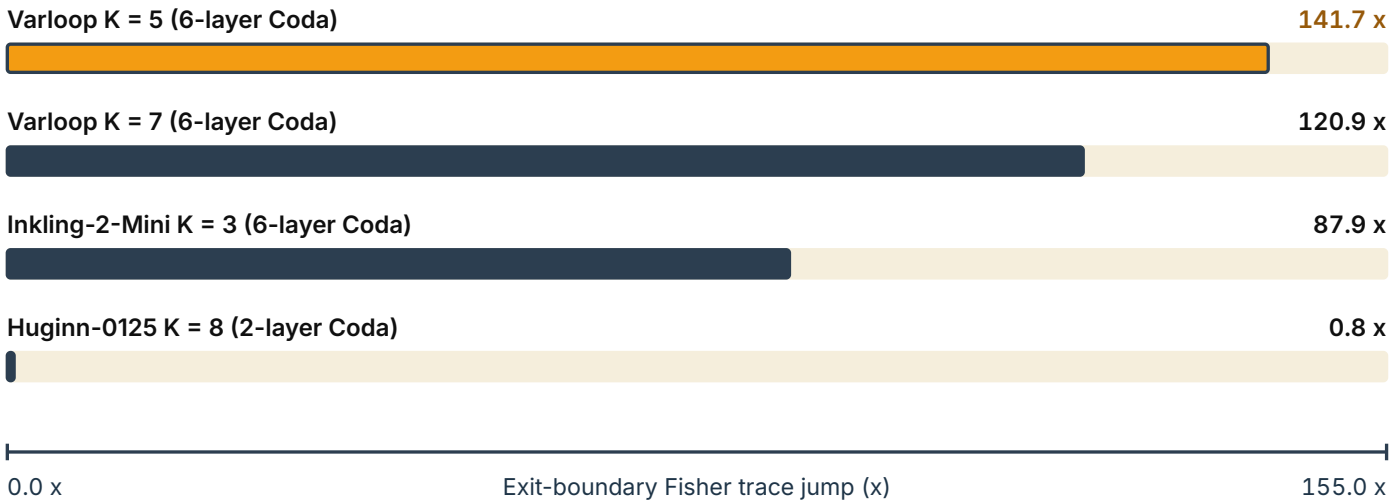
Premotor cortex rehearses an upcoming reach inside an output-null subspace perpendicular to spinal readout so your hand does not twitch while you are still deciding where to move (Hennequin et al., 2014). A looped language model has the same incentive: thinking directly in token coordinates would tug every intermediate hop toward a premature surface word (Belrose et al., 2023; Khalil & Hoque, 2026).

Keeping the loop rotated away from vocabulary space explains why raw logit lenses stall at 6.7% on `Inkling-2-Mini` even while Coda Jacobian conductance favors task directions by **7.81x** on Pass 3. Measuring the full-vocabulary Hutchinson Fisher pullback trace across every layer shows exactly where each architecture refracts latent states into words.

$$G_t = J_t^\top (\mathrm{diag}(p) - p\,p^\top)\, J_t, \qquad \mathrm{tr}(G_t) = E[\,\| J_t^\top (\mathrm{diag}(p) - p\,p^\top)^{(1/2)}\, z \|_2^{\,2}\,] \tag{3}$$

Figure 5. Six-layer exit blocks trigger 87.9x to 141.7x Fisher sensitivity cliffs while two-layer Huginn-0125 stays below unity.

Hutchinson Fisher pullback trace jump ratio from the final loop step into the first Coda layer (n = 60)



▼ Text alternative and data

Horizontal bar chart of loop-exit to first-Coda-layer Fisher trace ratio: Varloop K = 5 reaches 141.7x, Varloop K = 7 reaches 120.9x, Inkling-2-Mini K = 3 reaches 87.9x, and two-layer-Coda Huginn-0125 K = 8 sits at 0.77x.

Six-layer exit blocks trigger 87.9x to 141.7x Fisher sensitivity cliffs while two-layer Huginn-0125 stays below unity.

Condition	Fisher jump ratio (x)
Varloop K = 5 (6-layer Coda)	141.7 x
Varloop K = 7 (6-layer Coda)	120.9 x
Inkling-2-Mini K = 3 (6-layer Coda)	87.9 x
Huginn-0125 K = 8 (2-layer Coda)	0.8 x

Source: results/phase7_biophysics/phase7_biophysics_analysis.json.

Coda depth sets the entire readout geometry. Checkpoints with a six-layer Mixture-of-Experts Coda stay output-null inside the loop and refract across a steep sensitivity cliff right at the exit boundary: Fisher trace jumps **87.9x** entering Coda on `Inkling-2-Mini`, **141.7x** on `Varloop` at `K = 5`, and **120.9x** at `K = 7`. Transporting loop states through that six-layer prism reveals clean two-stage execution (Saunshi et al., 2025; Banino et al., 2021; Bae et al., 2025): two-step arithmetic stages the

intermediate bridge integer on Pass 2 (**21.6%**) before locking **100.0%** probability onto the final answer on Pass 3.

Dense `Huginn-0125`, with only a two-layer Coda, cannot afford to wait until the exit to rotate into token coordinates. Its Fisher trace climbs **84.8x** gradually across Steps 1 through 8 inside the loop itself (31.7% raw logit-lens accuracy) and crosses from the final loop step into the first Coda layer at a flat **0.77x** ratio.

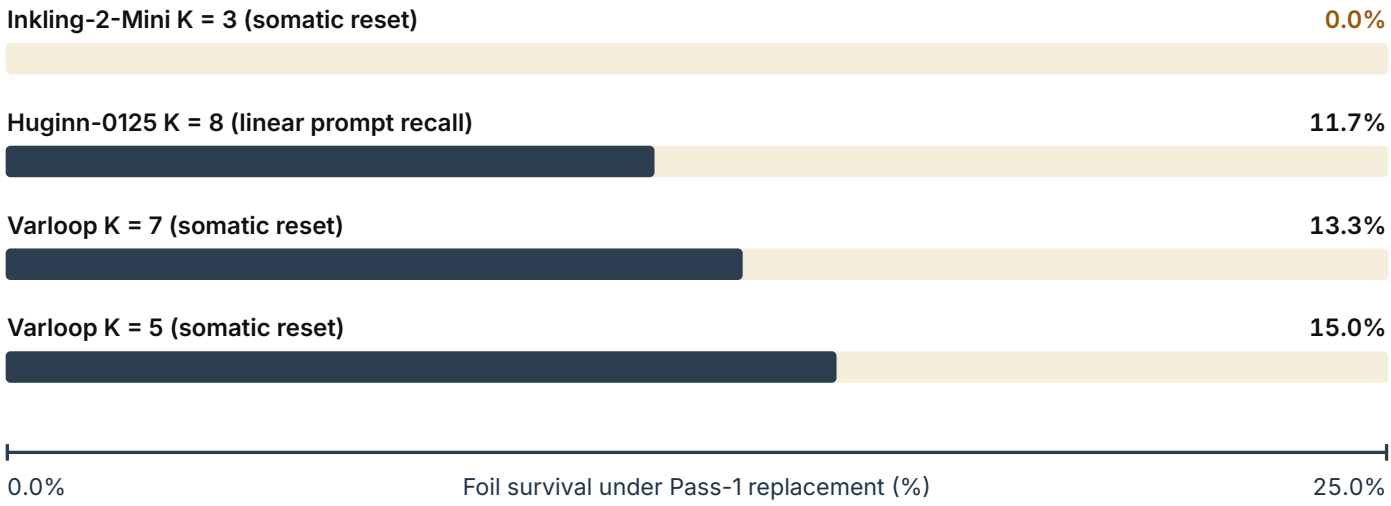
5 Hard Somatic Resets and Continuous Prompt Recall Both Prevent Foil Bifurcation

Looping a noisy hunch through shared weights risks locking into a false attractor. Labovich (2026) proved theoretically that continually re-injecting the clean prompt shields a looped transformer against fixed-point collapse. Putting `Inkling-2-Mini` and `Varloop` alongside `Huginn-0125` tests two physical ways of building that anchor (McLeish et al., 2025).

$$h_{\text{in}}^{(p)} = (1 - f_g) * h_{\text{pre}} + m_g * M(h^{(p-1)}) \quad \textbf{[Inkling/Varloop]} \quad \textbf{vs.} \quad x^{(p)} = F_{\text{core}}(W_{\text{adp}} [x^{(p-1)}; e_{\text{pre}}]) \quad \textbf{[Huginn]} \tag{4}$$

Figure 6. Hard somatic reset holds Inkling-2-Mini to 0.0% foil survival while prompt recall shields Huginn-0125.

Final foil capture rate (%) after replacing Pass-1 query states with a wrong-answer foil (n = 60, zero bifurcations)



▼ Text alternative and data

Horizontal bar chart of foil survival rate under full Pass-1 foil replacement: Inkling-2-Mini K = 3 sits at 0.0%, Huginn-0125 K = 8 sits at 11.7%, Varloop K = 7 sits at 13.3%, and Varloop K = 5 sits at 15.0%.

Hard somatic reset holds Inkling-2-Mini to 0.0% foil survival while prompt recall shields Huginn-0125.

Condition	Foil survival (%)
Inkling-2-Mini K = 3 (somatic reset)	0.0%
Huginn-0125 K = 8 (linear prompt recall)	11.7%
Varloop K = 7 (somatic reset)	13.3%
Varloop K = 5 (somatic reset)	15.0%

Source: results/phase6_inkling_biophysical_loop/biophysical_analysis.json and results/phase7_biophysics/phase7_biophysics_analysis.json.

`Huginn-0125` carries a continuous recurrent state concatenated with clean prelude embeddings through a two-input linear adapter (calibrated at 37.5% two-shot accuracy versus 22.5% zero-shot; zero-state KL 0.0313 nats). Knocking out all four on-off combinations of those two adapter inputs proves the tether is doing real structural work. Cutting prompt recall while leaving recurrent state intact drops `K = 8` accuracy from **27.5%** straight to **0.0%**, with a **+12.01** nats recall effect, **+3.48** nats state effect, and **+4.45** nats of synergy.

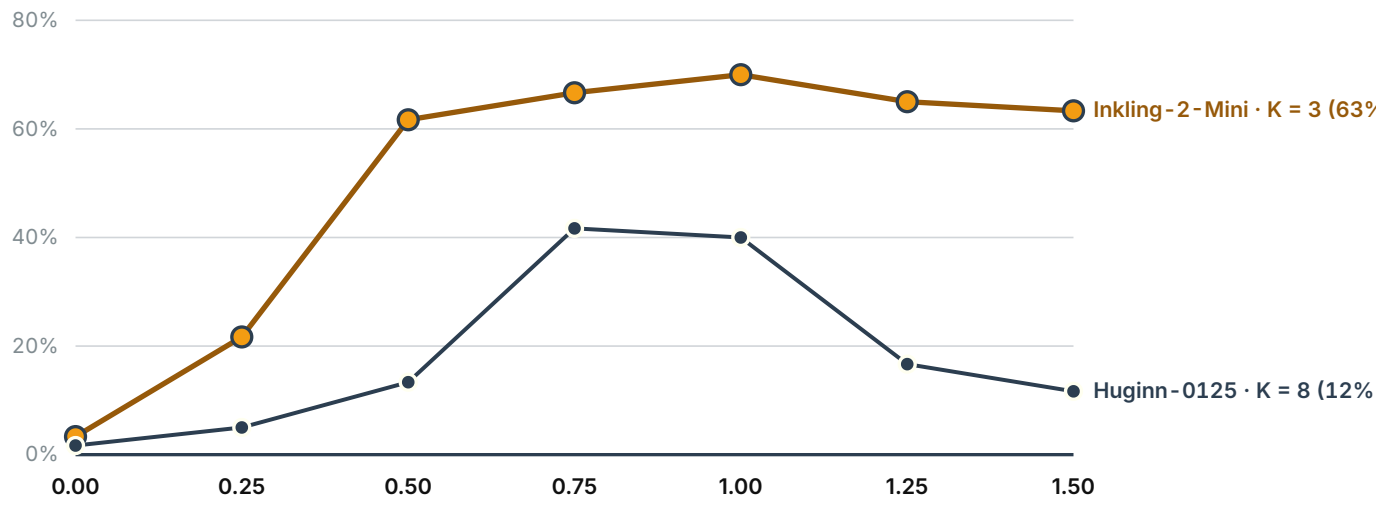
When prompt anchoring stays intact, neither architecture family crosses a tipping point even when Pass 1 is completely replaced with a wrong-answer foil. `Inkling-2-Mini` 's hard somatic reset suppresses Pass-2 foil mass by 18.6 pp, leaving **0.0%** foil survival at the output. On `Varloop` , extra passes scrub foil survival down from 18.3% at `K = 3` to 15.0% at `K = 5` and 13.3% at `K = 7` . `Huginn-0125` 's continuous adapter lets **11.7%** foil capture leak through at `K = 8` , while damping unit orthogonal tangent pulses to **0.142x** (output KL 0.0060 nats, versus 29.9% residual deviation and 0.231 nats on `Inkling-2-Mini`).

6 Recurrent Conductance Follows an Inverted-U Curve Before Chat Headers Shunt Emission

Neither `Inkling-2-Mini` nor `Huginn-0125` uses explicit pass-number embeddings. Scaling recurrent gate or adapter conductance (`gamma`) from zero to 1.50 still traces the same inverted-U Goldilocks curve across both models. That sharp operating window sets trained recurrence apart from training-free ODE looping and per-pass LoRAs (Chen et al., 2026; Bae et al., 2025; Hao et al., 2024; Dehghani et al., 2019; Giannou et al., 2023).

Figure 7. Recurrent conductance peaks at 70.0% in Inkling-2-Mini and 41.7% in Huginn-0125 before rolling over.

Hop-2 exact Top-1 accuracy (%) across recurrent conductance multipliers from zero to over-driven gain (n = 60)



▼ Text alternative and data

Two-series inverted-U line chart across recurrent conductance multipliers: Inkling-2-Mini rises from 3.3% to 70.0% at native unit gain before dipping to 63.3%, while Huginn-0125 rises from 1.7% to 41.7% before dropping to 11.7%.

Recurrent conductance peaks at 70.0% in Inkling-2-Mini and 41.7% in Huginn-0125 before rolling over.

Condition	Inkling-2-Mini · K = 3	Huginn-0125 · K = 8
0.00	3%	2%
0.25	22%	5%
0.50	62%	13%
0.75	67%	42%
1.00	70%	40%
1.25	65%	17%
1.50	63%	12%

Source: results/phase6_inkling_biophysical_loop/biophysical_analysis.json and results/phase7_biophysics/phase7_biophysics_analysis.json.

Turn conductance down to zero and multi-hop reasoning starves (3.3% on `Inkling-2-Mini` and 1.7% on `Huginn-0125`). Accuracy peaks inside a narrow native window: **70.0%** at `gamma = 1.00` on `Inkling-2-Mini` and **41.7%** at `gamma = 0.75` on `Huginn-0125` (40.0% at unit gain). Push `gamma` to

1.50 and accuracy rolls over to 63.3% and 11.7% as over-driven updates overshoot the Coda readout manifold.

Inside `Inkling-2-Mini`'s Coda, one last circuit explains why chat-formatted prompts look mute right at the assistant header. Pre-shunt layer `L19` already surfaces **50.0%** Top-5 semantic answers before layers `L21..L22` route **100.0%** of probability mass into `<|content_thinking|>`. Transplanting clean somatic header carriers strips away 99.4% of that formatting shunt (down to **0.63%**), and stepping forward a single token to `<|content_text|>` releases the semantic answer at **76.2%** probability.

All code, CPU verification tests, exact Walsh-Hadamard and vector-Jacobian solvers, interactive HTML circuit benches, and SHA-256-locked empirical JSON artifacts (3,884.3 s unified H100 runtime; unrolled equivalence 0.00e+00 on `Varloop` and 0.00e+00 on `Huginn-0125`) are available under an MIT license at github.com/xmattwest/nullbridge.

References

1. Surya Ganguli, Dongsung Huh, Haim Sompolinsky (2008). Memory traces in dynamical systems. *PNAS 105(48), doi:10.1073/pnas.0804451105*.
2. Brendan K. Murphy, Kenneth D. Miller (2009). Balanced Amplification: A New Mechanism of Selective Amplification of Neural Activity Patterns. *Neuron 61(4), doi:10.1016/j.neuron.2009.02.005*.
3. Guillaume Hennequin, Tim P. Vogels, Wulfram Gerstner (2014). Optimal Control of Transient Dynamics in Balanced Networks Supports Generation of Complex Movements. *Neuron 82(6), doi:10.1016/j.neuron.2014.04.045*.
4. Daniel M. Weinreich, Yinghong Lan, Jacob Jaffe, Robert B. Heckendorn (2018). The Influence of Higher-Order Epistasis on Biological Fitness Landscape Topography. *Journal of Statistical Physics 172, doi:10.1007/s10955-018-1975-3*.
5. Andre J. Faure, Ben Lehner, Verónica Miro Pina, Claudia Serrano Colome, Donate Weghorn (2024). An extension of the Walsh-Hadamard transform to calculate and model epistasis in genetic landscapes of arbitrary shape and complexity. *PLOS Computational Biology, doi:10.1371/journal.pcbi.1012132*.
6. Mostafa Dehghani, Stephan Gouws, Oriol Vinyals, Jakob Uszkoreit, Lukasz Kaiser (2019). Universal Transformers. *ICLR 2019 (arXiv:1807.03819)*.
7. Andrea Banino, Jan Balaguer, Charles Blundell (2021). PonderNet: Learning to Ponder. *ICML 2021 Workshop (arXiv:2107.05407)*.

8. Angeliki Giannou, Shashank Rajput, Jy-yong Sohn, Kangwook Lee, Jason D. Lee, Dimitris Papailiopoulos (2023). Looped Transformers as Programmable Computers. *ICML 2023 (arXiv:2301.13196)*.

9. Robert Csordas, Kazuki Irie, Jurgen Schmidhuber, Christopher Potts, Christopher D. Manning (2024). MoEUT: Mixture-of-Experts Universal Transformers. *NeurIPS 2024 (arXiv:2405.16039)*.

10. Sangmin Bae, Adam Fisch, Hrayr Harutyunyan, Ziwei Ji, Seungyeon Kim, Tal Schuster (2025). Relaxed Recursive Transformers: Effective Parameter Sharing with Layer-wise LoRA. *ICLR 2025 (arXiv:2410.20672)*.

11. Sangmin Bae, Yujin Kim, Reza Bayat, Sungnyun Kim, Jiyoun Ha, et al. (2025). Mixture-of-Recursions: Learning Dynamic Recursive Depths for Adaptive Token-Level Computation. *NeurIPS 2025 (arXiv:2507.10524)*.

12. Shibo Hao, Sainbayar Sukhbaatar, DiJia Su, Xian Li, Zhiting Hu, Jason Weston, Yuandong Tian (2024). Training Large Language Models to Reason in a Continuous Latent Space (Coconut). *arXiv:2412.06769*.

13. Jonas Geiping, Sean McLeish, Neel Jain, John Kirchenbauer, Siddharth Singh, et al. (2025). Scaling up Test-Time Compute with Latent Reasoning: A Recurrent Depth Approach (Huginn). *NeurIPS 2025 (arXiv:2502.05171)*.

14. Nikunj Saunshi, Nishanth Dikkala, Zhiyuan Li, Sanjiv Kumar, Sashank J. Reddi (2025). Reasoning with Latent Thoughts: On the Power of Looped Transformers. *ICLR 2025 (arXiv:2502.17416)*.

15. Sean McLeish, Ang Li, John Kirchenbauer, Dayal Singh Kalra, Brian R. Bartoldson, et al. (2025). Teaching Pretrained Language Models to Think Deeper with Retrofitted Recurrence. *arXiv:2511.07384*.

16. Labovich (2026). Stability and Generalization in Looped Transformers. *arXiv:2604.15259*.

17. Chen, Li, Liang, Lao, Liu (2026). Training-Free Looped Transformers. *arXiv:2605.23872*.

18. Chen, Liu, Zhang, Li, Pan, Lin, Li (2026). UpLoop: Upcycling Large Language Models with Gated Loop. *arXiv:2610.25004*.

19. Thinking Machines Lab (2026). Inkling-2-Mini & TML-d24-varloop-research / NDIF Looped Lens Challenge. *Hugging Face Model Hub (Nov 2026)*. https://huggingface.co/thinkingmachines/Inkling-2-Mini

20. Nora Belrose, Zach Furman, Logan Smith, Danny Halawi, Igor Ostrovsky, Lev McKinney, Stella Biderman, Jacob Steinhardt (2023). Eliciting Latent Predictions from Transformers with the Tuned Lens. *arXiv:2303.08112*.

21. Khalil, Hoque (2026). Scalar Tuned Lens Summaries Offer Limited Incremental Value for Mid-Generation Abstention. *arXiv:2611.29589*.

22. Nora Belrose, David Schneider-Joseph, Shauli Ravfogel, Ryan Cotterell, Edward Raff, Stella Biderman (2023). LEACE: Perfect Linear Concept Erasure in Closed Form. *NeurIPS 2023 (arXiv:2306.03819)*.

23. Bowen Baker, Joost Huizinga, Leo Gao, Zehao Dou, Melody Y. Guan, et al. (2025). Monitoring Reasoning Models for Misbehavior and the Risks of Promoting Obfuscation. *arXiv:2503.11926*.

24. Aryo Pradipta Gema, Alexander Hagele, Runjin Chen, et al. (2025). Inverse Scaling in Test-Time Compute. *TMLR 2025 (arXiv:2507.14417)*.

25. Guangxuan Xiao, Yuandong Tian, Beidi Chen, Song Han, Mike Lewis (2024). Efficient Streaming Language Models with Attention Sinks. *ICLR 2024 (arXiv:2309.17453)*.

Verified build: 197 bound metrics checked across 4 SHA-256 JSON artifacts (provenance.json).

Repository: https://github.com/xmattwest/nullbridge · Dataset & figures: huggingface.co/datasets/xmattwest/biophysics-looped-transformer

Code, statistical pipelines, and manuscript compilation in this repository were developed with AI coding assistance (Antigravity) under author direction; every empirical number is bound via Folio JSON Pointers to SHA-256-locked GPU and audit artifacts.