Learning to Compute Is Not Learning to Retain or Reuse in Looped Transformers

A paper discussion: three properties, three tests

Paper authors: Peilong Zhou, Hongtao Zhang, Yulong Huang, Cangyuan Li, Mingyu Chen, Ying Wang

arXiv:2610.19768v1 | 15 October 2026

A correct answer can arrive before its state is ready to stop updating.

Independent reading presentation; not authored by the paper's authors. https://arxiv.org/abs/2610.19768v1

One accuracy curve hides three different questions

Property	Question	Test in this paper
Compute	Was the answer produced by a usable computation?	Read at arrival; redirect exact continuation with a state swap.
Retain	Does the answer survive more loops?	Keep the input and model fixed; extend the loop budget.
Reuse	Can a solved position stop changing?	Freeze its hidden state; test later positions.

Correct readout $\not\Rightarrow$ stable readout $\not\Rightarrow$ safe state freezing.

Paper, Secs. 1–2 and 5, PDF pp. 1–4, 8. Reuse here follows the freezing test.

Mealy machines make every continuation checkable

Input

Initial state q_0; transition table; action sequence $a_1, \ldots, a_H$.

- 4/6/8 states; 2/4 input symbols; 2 output symbols.
- Random symbol names and table-row order.
- Intermediate states and outputs are targets only.

Exact simulator

$$q_k = \delta(q_{k-1}, a_k)$$

$$y_k = \lambda(q_{k-1}, a_k)$$

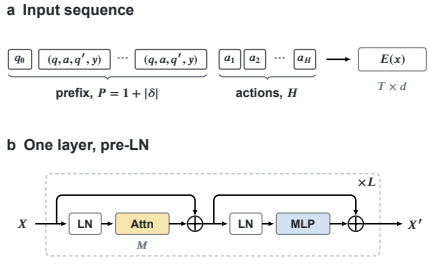
Replace the state at step c, keep future actions, and simulate the unique counterfactual suffix.

A changed prediction is insufficient: the *entire predicted continuation* must match.

Paper, Sec. 2, Eq. (1), PDF pp. 2–3; Appendix A.1, p. 12.

The same two-layer block evolves every position

$$X^0 = 0, \qquad X^r = B_\theta(X^{r-1} + E(x))$$



- Width 128, 4 heads, MLP width 512; block weights shared across loops.
- No position or loop embeddings; action attention is causal.
- Replace h_k^r after the block, before final LN and both readout heads.

Original Fig. A1, PDF p. 13; Eq. (2), p. 3. $X^r, E(x) \in \mathbb{R}^{T \times 128}$.

Position coverage changes what must survive

$$\mathcal{L}_S = \frac{1}{\sum_i |S_i|} \sum_i \sum_{k \in S_i} \ell_{i,k}^R, \qquad R = H$$

Objective	Supervised positions S_i	Pairs per example
FINAL-ONLY	$\{H\}$	1
ALL-PREFIX	$\{1, \ldots, H\}$	H
SAMPLED-PREFIX	$\{K_i\}$, $K_i \sim U\{1, \ldots, H\}$	1

- All losses are applied only at the exit, not at every loop.
- Shorter-length training constrains h_k^k by causal prefix invariance.
- Broader coverage also asks position k to remain correct at $H > k$.

Paper, Eq. (3) and Sec. 2, PDF pp. 3–4. ℓ sums state/output cross-entropies.

The readout time is part of the metric

Endpoint joint accuracy

$$A(H, r) = \frac{1}{N} \sum_{i} \mathbf{1}[\hat{q}^{r}_{i,H} = q_{i,H} \ \wedge \ \hat{y}^{r}_{i,H} = y_{i,H}]$$

- $T(H, r)$: **every** position has both labels correct at the same loop r.
- Frontier readout: position k is read at its own loop $s(k) = k$.
- Main behavior: 512 traces/model/length $=$ 16 machines $\times$ 32 traces.
- Average within each model, then equally across model seeds.

Common training comparison: 3 seeds, 132k updates, curriculum to $H = 32$.

Paper, Sec. 2 and Appendix A.2–A.3, PDF pp. 3–4, 12–14.

One random supervised position preserves answers

$H = 32$, common evaluation stream; three-seed means (%).

Objective	A_{32}	A_{64}	A_{96}	T_{32}	T_{64}
FINAL-ONLY	100.00	22.33	22.07	0.00	0.00
ALL-PREFIX	100.00	94.27	82.16	100.00	74.28
SAMPLED-PREFIX	100.00	99.80	96.61	100.00	96.03
Matched exit	100.00	97.85	86.26	74.09	29.56

- SAMPLED-PREFIX keeps one supervised position per example, as in FINAL-ONLY.
- Strong endpoint retention can coexist with weak whole-trajectory retention.
- At loop 96, SAMPLED-PREFIX trajectory exactness is only 74.09%.

Paper, Table A2, PDF p. 16; Fig. 2, p. 5. Matched exit = matched random exit.

Variable exits help beyond their training support

$$\underbrace{R = H + 4}_{\text{Slack-4}}, \qquad \underbrace{R = H + U\{0, \ldots, 8\}}_{\text{Jitter-8}}, \qquad \mathbb{E}[R \mid H] = H + 4$$

Endpoint accuracy (%)	$r = 36$	$r = 48$	$r = 64$	$r = 96$
Slack-4	93.98	38.15	30.01	26.79
Jitter-8	87.27	87.17	85.55	79.79

- $H = 32$; six paired seeds; final-position supervision in both conditions.
- Loop-64 gain: 55.54 percentage points; lower mean at loop 36.
- Cumulative block counts differ by 0.062%; total FLOPs are not matched.

Paper, Tables A7–A9, PDF pp. 18–19. Historical cohort, separate from slide 7.

The matched gradient control changes the conclusion

Both conditions optimize $\frac{1}{2}L_m + \frac{1}{2}L_{n+k}$, with $m = H + 8$ and $H \leq n + k \leq m$.

Progressive

Full m-loop branch.
Random branch: detach after n loops; backpropagate through k loops.

Two seeds peak below 60%. The successful seed falls from 100% at loops 40–48 to 26.56% at loop 96.

Matched random exit

Same examples, exits and loss weights.
Backpropagate through all $n + k$ loops.

All three seeds reach 100% at loop 32; mean accuracy is 97.85% at loop 64.

Separate failure to learn from forgetting an answer already learned.

Paper, Sec. 3 and Appendix B.1, PDF pp. 5–6, 15–16. Forward work matched; backward work differs. FSM adaptation of progressive training.

Length extrapolation has a different ranking

All models trained only through $H = 32$; evaluate $H = 64$.

Objective	$A(64, 64)$ (%)	$A(64, 192)$ (%)
FINAL-ONLY	97.79	17.12
ALL-PREFIX	76.69	72.72
SAMPLED-PREFIX	85.87	60.61
Matched exit	86.07	53.71
Progressive	33.92	17.19

Two axes of evaluation

Compute a longer input correctly; keep a solved input correct under more loops. Improving one does not guarantee improving the other.

Paper, Table A4, PDF p. 17. Common stream, three-seed means.

Swap a state, then demand an exact counterfactual

1. Match source and base machines, symbol names and table presentation.
2. Change prefix actions through c; require $q_c^S \neq q_c^B$.
3. Keep actions after c identical and replace one complete vector:

$$h_c^r \leftarrow h_{c,S}^r.$$

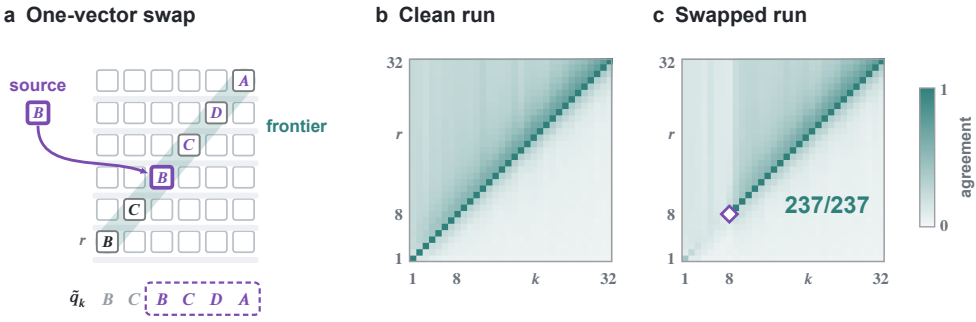
4. Continue the model and compare **every** later state/output to simulation from q_c^S.

Eligibility prevents a trivial success

The factual and counterfactual suffixes must differ somewhere. A model that ignores the swap cannot pass suffix exactness.

Paper, Sec. 4, Eq. (5), PDF p. 6; Appendix A.3–A.4, p. 14.

One vector redirects 24 steps of exact execution



237/237 eligible suffixes in each of 10 models.

Swap at $(k, r) = (8, 8)$; read each suffix position k at loop k.
The bright diagonal fades above the frontier: readouts are transient.

Original Fig. 3, PDF p. 7; Appendix A.4, p. 14. $H = 32$. Grids average 256 pairs/model; exactness uses 237 eligible pairs/model.

Caching an answer and freezing a state differ

Readout cache

Save the clean prediction at $s(k) = k$ outside the model.

- All hidden states continue updating.
- Tests whether answers were available.
- No downstream computation is altered.

State freeze

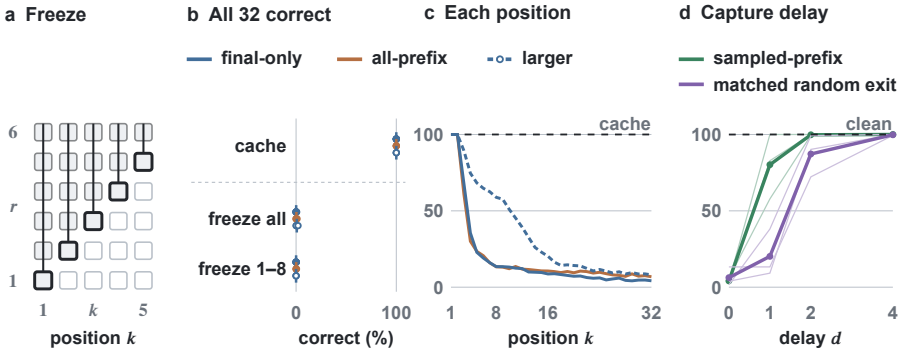
Capture the arm's own h_k^{k+d}; write it back after every later block.

- Other positions remain live.
- Frozen vectors remain visible to attention.
- Tests whether downstream work can use a fixed state.

No source run or symbolic answer is inserted. Table tokens stay live.

Paper, Sec. 5 and Appendix D.1/D.3, PDF pp. 8, 25–28. Experimental freezing executes all blocks and does not save FLOPs.

Correct cached answers coexist with failed freezing



- Original 12 models: cache retains all 32 answers; immediate freezing leaves at most 2% of trajectories fully correct.
- Failure appears in both supervision baselines and larger models.

Original Fig. 4, PDF p. 7; Appendix D, pp. 25–28. Panels b/c: original cohort; d: new controls. Larger block: width 512, four layers, about 32 times the block parameters.

A short delay restores safe holding in these models

Endpoint joint accuracy at actual capture loop $32 + d$ (%).

Objective	Clean $r = 32$	$d = 0$	$d = 1$	$d = 2$	$d = 4$
SAMPLED-PREFIX	100.00	4.10	80.21	99.80	100.00
Matched exit	99.93	6.25	20.18	87.24	99.80

- Clean is 100% at delayed readouts $r = 33, 34, 36$.
- After capture, frozen endpoint scores stay unchanged through loop 64.
- The delay needed differs across models; one extra loop is insufficient.

Retention did not supply this guarantee

SAMPLED-PREFIX retains 99.80% at loop 64 in the behavior study, yet immediate state freezing collapses to 4.10% here.

Paper, Table A20, PDF p. 27; Table A2, p. 16. Freeze follow-up uses a fresh stream and three seeds per condition.

The first error follows a missed upstream update

Immediate freezing at $s(k) = k$

Positions 1 and 2 stay correct. Position 3 fails first: before loop 3, position 1 is missing its loop-2 update, after its own answer was already correct.

- FINAL-ONLY position-3 accuracy drops to about 68%; errors compound later.
- Freeze only positions 1:8: the endpoint remains live, but accuracy is only 7.49% for SAMPLED-PREFIX and 7.42% for matched exit.
- Later captures occur on an already intervened trajectory; they need not themselves be correct.

Possible role: continuing updates carry order, timing or compatible representations. The experiment does not isolate the internal mechanism.

Paper, Sec. 5, PDF p. 8; Tables A20/A25, pp. 27/30. Prefix results use the new freeze cohort.

A retained source also becomes hard to transplant

ALL-PREFIX, position $c = 16$: collect and insert the source at loop $16 + a$. Read suffix position k at loop $k + a$.

Suffix exactness (%)	$a = 0$	$a = 1$	$a = 2$	$a = 4$
Clean: factual target	100.0	100.0	100.0	100.0
Swap: counterfactual target	93.9	17.7	10.2	0.0

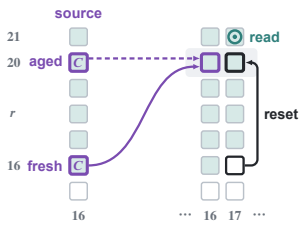
- Post-swap update counts are fixed, but source and receiver phases both change.
- Even a fresh loop-16 source inserted at loop 20 gives 0% one-step transfer.
- Next test: keep the late background fixed and reset one downstream vector.

Paper, Tables A12–A13, PDF pp. 21–22. Six seeds; clean uses 256 pairs, swap 242 eligible pairs. Different targets/denominators: not a paired effect subtraction.

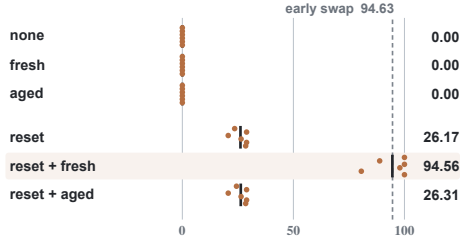
Fresh source plus receiver reset restores one step

$$h_{16}^{20} \leftarrow h_{16,S}^{16}, \qquad h_{17}^{20} \leftarrow h_{17,B}^{16}.$$

a Two-site intervention



b Counterfactual match (%)



- Fresh + reset recovers early transfer; aged + reset adds little.
- Both source ages decode correctly; clean retains its factual answer.

Original Fig. A4 / Eq. (7), PDF p. 23; Table A15, p. 24. All-prefix: six seeds, 242 eligible pairs. Replace at loop 20; read position 17 at loop 21. Scores measure counterfactual match.

A maze model keeps solutions but cannot freeze early

Released recurrent recall network; one checkpoint, 32 mazes per size.

Solved at iteration 120	9×9	33×33
Clean / answer cache	32 / 32	19 / 32
Freeze at arrival	0 / 32	0 / 32
Arrival $+1$	3 / 32	0 / 32
Arrival $+2$	28 / 32	2 / 32
Arrival $+4$	32 / 32	18 / 32

- All small mazes stay solved from iteration 25 to 120.
- Arrival = last clean prediction change per pixel: an offline diagnostic.
- Spatial output heads couple pixels; holding a pixel need not pin its output.

Paper, Appendix D.4 and Table A26, PDF pp. 28, 30. Trained on 9x9 mazes for 30 iterations; larger size is outside training.

Halting must account for unfinished downstream work

What the experiments establish

At a position's answer arrival, its state can still require updates for other positions to compute correctly.

1. Evaluate computation, retention and state holding separately.
2. State freezing needs a downstream criterion, beyond local readout correctness.
3. Cache correctness demonstrates prior computation; it does not measure saved compute.

Discussion idea: learn when an answer is ready and when its supporting state is safe to hold as distinct decisions.

Paper, Secs. 5 and 7, PDF pp. 8–10. The last proposal is a reading interpretation, not a tested method in this paper.

The causal result is functional, not a circuit model

- **Task and scale:** one synthetic execution family and principal architecture; a limited external maze check.
- **Training evidence:** new controls have three seeds; sampled supervision still needs intermediate labels.
- **Intervention scope:** full vectors can carry state, history and timing. A readable vector need not be a unique symbolic code.
- **Time scope:** finite retention windows, model-dependent freeze delays, and one-step receiver recovery.

Discussion question

What is a state still doing after its own answer has stopped changing?

Paper, Sec. 7 and Appendix C.4, PDF pp. 9–10, 25.

Source and evidence guide

Peilong Zhou, Hongtao Zhang, Yulong Huang, Cangyuan Li, Mingyu Chen, and Ying Wang.

Learning to Compute Is Not Learning to Retain or Reuse in Looped Transformers. arXiv:2610.19768v1, submitted 14 October 2026.

arxiv.org/abs/2610.19768v1

Result	Main source locations
Retention controls	Fig. 2; Tables A2–A4, A8
Exact counterfactual computation	Fig. 3; Appendix A.4
State freezing and delayed capture	Fig. 4; Tables A19–A26
Local receiver compatibility	Fig. A4; Tables A15–A17
Natural first-failure recovery	Table A27

All figures are reproduced from the authors' v1 TeX source. Tables select original reported values. Backup slides follow.

Backup: sampled supervision estimates the dense loss

At a fixed batch, parameters and shared horizon H:

$$J_i \sim U\{1, \dots, H\}, \qquad L_{\text{sampled}} = \frac{1}{B} \sum_{i=1}^{B} \ell_{i,J_i}.$$

$$\mathbb{E}[L_{\text{sampled}}] = \frac{1}{B} \sum_{i=1}^{B} \frac{1}{H} \sum_{k=1}^{H} \ell_{i,k} = L_{\text{all-prefix}}.$$

- Position coverage is broadened without increasing selected pairs per example.
- Intermediate labels are still required.
- Equal expected losses do not imply identical optimization trajectories.

Paper, Appendix B.1, PDF p. 15; Eq. (3), p. 3.

Backup: the random-exit law is not uniform

$$m = H + 8, \quad n \sim U\{0, \ldots, m-1\}, \quad k \mid n \sim U\{1, \ldots, m-n\}.$$

Reject and redraw the whole pair if $n + k < H$.

$$P(n, k \mid n + k \geq H) \propto \frac{1}{m(m-n)}.$$

Work per update	Progressive	Matched exit
Forward blocks	$m + n + k$	$m + n + k$
Differentiable blocks	$m + k$	$m + n + k$
Total proxy MACs/run	5.919×10^{16}	7.438×10^{16}

Proxy excludes heads, embeddings, normalization, optimizer and evaluation.

Paper, Appendix B.1, PDF pp. 15–16. Distinct from uniform Jitter-8.

Backup: a prescribed diagonal is not always exact

New freeze stream; all values in percent.

Objective	Seed	Clean A_{32}	Min. diagonal	Cache trajectory
SAMPLED-PREFIX	9–11	100.00	100.00	100.00
Matched exit	9	99.80	96.48	94.92
Matched exit	10	100.00	99.80	99.80
Matched exit	11	100.00	100.00	100.00

- $s(k) = k$ is fixed in advance; $\tau(k)$ is a measured threshold crossing.
- τ requires both marginal accuracies to reach 99%, not exact joint predictions on every example.
- Read delayed endpoints at $32 + d$: sample seed 10, $d = 1$, scores 100% at 32 but 82.81% at capture 33.

Paper, Tables A19–A21, PDF pp. 27, 29; metric definitions, pp. 4, 13.

Backup: fresh and aged readable sources differ

One-step counterfactual accuracy (%); six-seed means, 242 eligible pairs.

Intervention	FINAL-ONLY	ALL-PREFIX
Early fresh reference	100.00	94.63
Late clean	16.67	0.00
Late fresh source	26.72	0.00
Late receiver reset	19.63	26.17
Late fresh source + reset	100.00	94.56
Late aged source + reset	19.70	26.31

ALL-PREFIX fresh-source interaction:

$$I = A_{11} - A_{10} - A_{01} + A_{00} = 68.39 \pm 10.54 \text{ pp.}$$

Aged-source interaction: 0.14 ± 0.34 pp, despite 100% source readability.

Paper, Table A15, Eq. (8), Appendix C.3, PDF pp. 23–24. Dispersion is sample SD over six model seeds.

Backup: two local resets recover the first natural error

E: endpoint correct at loop 32, wrong at 33. Reset to own loop-31 vectors.

Reset positions	Rescued failures	All-input A_{33}
None	0/771	74.9%
31 only	585/771	89.5%
32 only	143/771	35.2%
31 and 32	771/771	100.0%

- Width-128 models, six seeds. Single resets can break previously correct inputs.
- Combined reset approximately replays the successful endpoint computation.
- At loop 34: 63.7% after reset versus 65.5% clean. Recovery is not persistent here.

Paper, Appendix E and Table A27, PDF p. 31. Larger models recover 79/79 failures and show different loop-34 behavior.

Backup: extra budget is a different training contract

Contract	Training loops	Interpretation
Serial	$R = H$	Main diagonal frontier $\tau(k) = k$.
Double	$R = 2H$	Audited frontier $\tau(k) = 2k - 1$; twice Serial's recurrent work.
Slack-8	$R = H + 8$	Extra-slack dose; not matched to Jitter-8's expected budget.
Random-2x	$R = H + U\{0, \ldots, H\}$	H32 loops 32–64 are in training support.

Double at $H = 32$: 100% endpoint at trained loop 64, but 2.93% at loop 32.

The observed frontier is learned under a budget contract, not fixed by architecture.

Paper, Appendix B.3–B.4 and Table A7, PDF pp. 18–20. Double endpoint contrast uses ten seeds; frontier/hold audit uses six.