Continual Learning Mechanisms Compose for Long-Horizon Memorization

A 25-minute research discussion + backup slides

Paper: Z. Zhang, A. Zhang, D. Khashabi, T. Shu

arXiv:2609.06986v1 | Reading: 2026-09-16

Can parameters remember 100 sequential tasks?

- Learn new query–answer associations through sequential SFT.
- Retain earlier associations without storing old raw examples.
- Answer without a task identifier at inference.

Central hypothesis

Mechanisms that preserve different evidence can work better together.

Secs. 1, 3.1; PDF pp. 1–4. Memorization evaluation, not held-out generalization.

Separate what to preserve from where updates live

Dimension	Mechanism	Preservation signal
Data anchor	Generative replay	Generated earlier sequences
Function anchor	Self-distillation (SD)	Teacher outputs on current data
Weight anchor	SI / online EWC	Important LoRA coordinates
Allocation	Shared / merged LoRA	Reuse / commit and restart

- The contribution is systematic composition and evaluation.
- These ingredients build on existing continual-learning methods.

Sec. 3, PDF pp. 4–6; related work, pp. 3–4.

A task boundary determines which state survives

1. Freeze the previous model; generate temporary replay data.
2. Fit current data with active retention losses.
3. Consolidate parameter importance before resetting coordinates.
4. Commit the update if using merged LoRA; initialize a fresh adapter.
5. Discard temporary replay samples and release the teacher.

All old-knowledge anchors are inactive on the first task.

Appendix B.7, Algorithm 3, PDF pp. 31–32. Teaching summary of the procedure.

Replay changes the balance of the fitting loss

For the full SI + SD + Replay composition at $t > 1$:

$$\mathcal{L}^t_{\text{impl}} = (1 - w)\mathcal{L}^t_{\text{SFT}} + w\mathcal{L}^t_D + \lambda_F \mathcal{L}^t_F + \lambda_{\text{SI}} \mathcal{R}^t_{\text{SI}}.$$

- w is a loss weight; without replay the fitting term is ordinary SFT.
- SFT covers both query and answer tokens; padding and replay seed are masked.
- SD and replay use the previous model, refreshed after each task.

Appendix B.1–B.2, PDF pp. 18–21. This is the implementation-level objective.

The data anchor generates inputs and soft targets

- Prompt the previous model with one task-agnostic replay token.
- Attempt 300 sequences; discard empty outputs; keep only for this task.
- Pair every current minibatch with one replay minibatch.

$$\mathcal{L}_D^t = \frac{\tau_D^2}{\sum a_{m,\ell}} \sum_{m,\ell} a_{m,\ell}\, \mathrm{KL}(q_{m,\ell}^D \| p_{m,\ell}^D).$$

Teacher q^D and student p^D see the same generated prefix.

Generation: $\tau_G = 1.5$, top-$p = 0.9$. Distillation: $\tau_D = 2$.

Appendix B.3, PDF pp. 21–23. Full-vocabulary forward KL, not hard pseudo-label CE.

SD constrains a different input distribution

$$\mathcal{L}_F^t = \tau_F^2 \, \text{mean}_{\text{valid current tokens}} \, \mathrm{KL}(q^F \| p^F), \qquad \tau_F = 5.$$

- Data anchor: distillation on generated replay sequences.
- Function anchor: distillation on current-task sequences.
- Both can share a frozen teacher; their support differs.
- A teacher constraint can also resist useful new learning.

Appendix B.4, PDF p. 24. Last bullet is an interpretation of the tradeoff.

SI estimates importance along the optimization path

$$\omega_{t,i} = -\sum_k g_{t,k,i} \Delta\vartheta_{t,k,i}, \qquad \Omega_{t,i} = \Omega_{t-1,i} + \frac{\max(0, \omega_{t,i})}{\Delta_{t,i}^2 + \xi}.$$

$$\mathcal{R}_{\mathrm{SI}}^{t+1} = \sum_i \Omega_{t,i} (\vartheta_i - \vartheta_{t,i}^{\star})^2.$$

- i: a scalar LoRA coordinate; g: fitting-loss gradient.
- $\Delta_{t,i}$: net task displacement; $\xi = 0.1$ stabilizes the ratio.
- One accumulated importance tensor and the latest reference are retained.

Appendix B.5.1, PDF pp. 25–26. O-LoRA uses a different SI gradient source.

Merged LoRA expands cumulative update capacity

$$\text{Shared:} \quad W_t = W_0 + \rho B_t A_t,$$
$$\text{Merged:} \quad W_t = W_{t-1} + \rho B_t A_t.$$

$A_t \in \mathbb{R}^{r \times d_{\text{in}}}$, $B_t \in \mathbb{R}^{d_{\text{out}} \times r}$, $\rho = \alpha / r$.

- Shared: cumulative residual rank is at most r.
- Merged: sum of committed updates can reach rank $\min(tr, d_{\text{in}}, d_{\text{out}})$.
- New $B = 0$ preserves the merged function at initialization.
- Both rules reset the optimizer at every task boundary.

Sec. 3.3; Appendix B.6.1–B.6.2, PDF pp. 28–29. Rank is not noninterference.

A reset can invalidate the meaning of an old penalty

- SI/EWC importance is attached to named entries in LoRA factors.
- Merged LoRA replaces those factors, but reuses old references and importance by position.
- The new coordinate can represent a different functional direction.

$$H_{\text{new}} \approx J^{\top} H_{\text{old}} J$$

A compatible local coordinate transform generally introduces off-diagonal terms.

The implementation does not perform this transformation.

Appendix B.5.4, PDF pp. 27–28. A structural risk, not a proven universal penalty.

Three datasets isolate different kinds of association

Dataset	Tasks $\times$ items	Construction
Symbol-QA	100×100	Random 6-character keys / 4-character values
LLM-QA	100×100	Fictional entities across 100 topics
Real-QA	100×50	Ten filtered public QA sources

- Evaluation reuses training questions, with no task ID.
- LLM-QA fiction is prompted, not externally verified.
- Real-QA removes items answered in any of five base-model samples; this does not prove zero prior knowledge.

Sec. 4.2; Appendix C, PDF pp. 33–35.

Measure acquisition separately from retention

$M_{i,j}$: accuracy on task j after learning through task i.

$$\mathrm{Final} = \frac{1}{T}\sum_{j=1}^{T} M_{T,j}, \qquad \mathrm{Diag} = \frac{1}{T}\sum_{j=1}^{T} M_{j,j}.$$

$$\mathrm{Forget} = \frac{1}{T-1}\sum_{j=1}^{T-1}\left(\max_{j\le i\le T} M_{i,j} - M_{T,j}\right).$$

- Higher Final / Diag is better; lower Forget is better.
- Forget uses peak observed performance and excludes the last task.
- It is not generally equal to $100 - \mathrm{Final}$.

Sec. 4.1, Eqs. (7)–(8), PDF pp. 6–7.

The main comparison controls 16 mechanism combinations

- Backbone: Qwen3-4B-Base; LoRA rank 32, alpha 64.
- 10 epochs/task; batch 8; AdamW learning rate 5×10^{-4}.
- SI, SD, Replay, Merge on/off: $2^4 = 16$ combinations.
- Merge off means shared LoRA; naive is shared LoRA without anchors.
- 21 methods/dataset in the full evaluation; three training seeds per method.

Table 3, PDF p. 32; Sec. 5, pp. 8–9. The naive baseline is not full-parameter SFT.

Task-level halving allocates longer horizons selectively

Horizon (tasks)	10	20	50	100
Configurations reaching horizon	90	45	23	10

$$C = 90(10) + 45(10) + 23(30) + 10(50) = 2540.$$

- Promoted configurations continue from saved state.
- 71.8% fewer configuration-task training units than exhaustive search.
- Search and final evaluation use different task orders.
- 10-task vs. final rank correlations: 0.93 / 0.95 / 0.90 (17 comparable compositions).

Sec. 4.3; Appendix D, PDF pp. 35–38. Cost is not a measured GPU-time reduction.

Compositions greatly improve final retention

Final retention (%)	Symbol	LLM	Real
Naive	1.0	1.4	1.3
Replay	4.2	7.5	12.5
Replay + Merge	16.6	32.4	48.2
SI + Replay + Merge	15.7	41.2	**54.8**
SD + Replay + Merge	**23.2**	33.0	39.0
All anchors + Merge	18.5	**41.8**	44.3

Three-seed means; bold identifies each dataset's best main-factorial mean.
Full stack: 34.9% average; ranks 2 / 1 / 3 within the 16 combinations.

Fig. 3, PDF p. 9; Tables 7–10, pp. 39–41. Rankings are not pairwise significance tests.

Cross-dataset robustness differs from winning everywhere

- The full stack is uniquely top-3 on all datasets in the main factorial.
- It is not the highest-retention recipe on Symbol-QA or Real-QA.
- Its 34.9% average is not the maximum cross-dataset average: SI + Replay + Merge averages about 37.2%.
- Real-QA O-LoRA alone reaches 13.9%, above Replay at 12.5%.

Read the scope carefully

The strongest-single and top-3 claims refer to the main comparison; extra task-growing methods change some ranks.

Sec. 5.1; Tables 7–10, PDF pp. 39–41. Average 37.2% calculated from table means.

Replay and Merge are the dominant average effects

Effect (percentage points)	Symbol	LLM	Real
Replay	+9.5	+18.5	+19.3
Merge	+5.9	+14.9	+20.5
SI	+0.3	+5.8	+6.3
SD	+5.7	+5.0	+3.7
Replay × Merge	+3.6	+9.4	+11.7

Main effects average an on/off contrast across the other factors.
Replay × Merge is positive with $p < 10^{-4}$ on all datasets.

Table 1, PDF p. 10; Tables 11–13, pp. 42–43. 48 observations/dataset, error df 32.

A fixed slice makes the synergy tangible

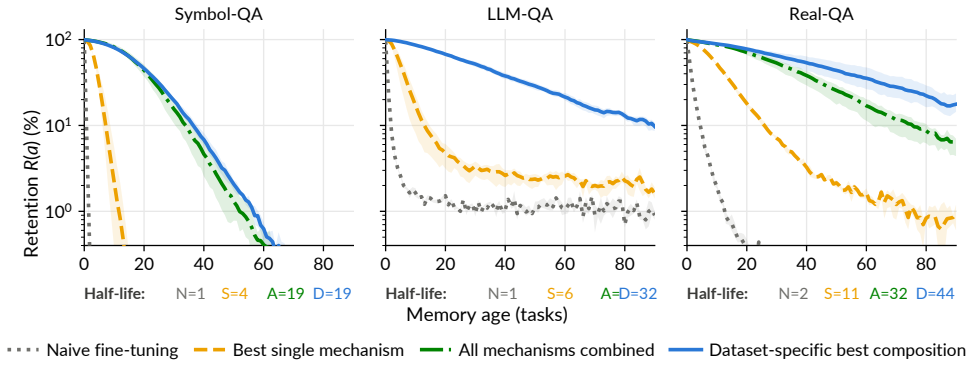
Without SI or SD, gains over naive fine-tuning are:

Gain (percentage points)	Symbol	LLM	Real
Replay alone + Merge alone	3.9	7.7	13.9
Replay and Merge together	15.6	31.0	46.9

- The pair exceeds the sum of its standalone gains.
- These conditional contrasts differ from the factorial interaction numbers.
- SD $\times$ Replay is negative on LLM/Real; extra anchors are not always beneficial.

Sec. 5.3, PDF pp. 10–11; Fig. 3 and Table 1.

Composition extends the timescale of forgetting



Half-life, in tasks: naive 1/1/2; full stack 19/32/32.
Dataset-specific best: 19/32/44. All curves still decay.

Fig. 1, PDF p. 2; Sec. 5.2, p. 10. Log y-axis; bands are min–max across 3 seeds.

More retained task state does not consistently help

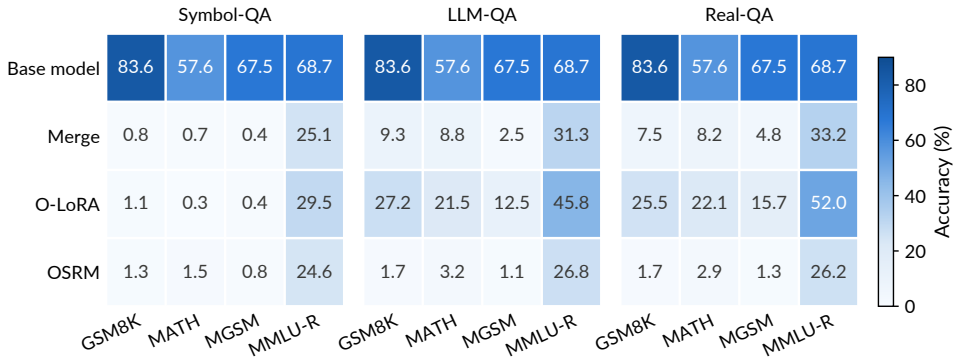
Matched replacement of merged LoRA in the TSH-selected composition:

Change in Final (points)	Symbol	LLM	Real
O-LoRA	−3.3	+1.0	+1.6
Sequential OSRM	−0.8	−11.5	−11.6

- O-LoRA retains task adapters; OSRM retains past feature means.
- Both grow with task count; neither consistently improves retention.
- Compared anchors: full stack on Symbol; SI + Replay on LLM/Real.

Appendix E.6, Table 14, PDF p. 45. OSRM here is a sequential adaptation.

Better memorization does not preserve general capability



Base four-benchmark mean: 69.4%.

After LLM / Real: Merge 13.0 / 13.4%; O-LoRA 26.8 / 28.8%.

Fig. 6, PDF p. 46; Appendix E.9, pp. 46–47. Matched TSH anchors, not always the full stack.

The evidence supports delay, not durable general memory

- Training-question recall does not establish paraphrase generalization.
- One backbone and scale; three seeds do not measure broad task-order robustness.
- Fixed-budget replay may fail to cover older associations.
- The study does not isolate the causal reason for every interaction.
- General capability remains substantially below the base model.

Sec. 6, PDF p. 11; Appendix E.9. Coverage concern is reader analysis.

Three takeaways for a continual-learning system

1. **Compose complementary signals.** Replay + Merge is the strongest common core here.
2. **Check compatibility.** A parameter reset can change what an importance penalty means.
3. **Evaluate multiple goals.** Retention, paraphrase recall, general capability, and cost can disagree.

Follow-up: test coordinate-compatible penalties and compute-matched replay across models and task orders.

Takeaways synthesize the paper. The proposed follow-up is not an experiment reported by the authors.

Reference and reading scope

Continual Learning Mechanisms Compose for Long-Horizon Memorization

Zheyuan Zhang, Alvin Zhang, Daniel Khashabi, Tianmin Shu.
Johns Hopkins University, 2026.

`https://arxiv.org/abs/2609.06986v1`
`https://compose-cl.github.io/`

This presentation is an independent reading prepared on 2026-09-16. Based on the matching author TeX source and 47-page PDF, including Appendices A–E. Experiments were not reproduced.

Figures are reused from the author source. Selected tables are reformatted from the paper.

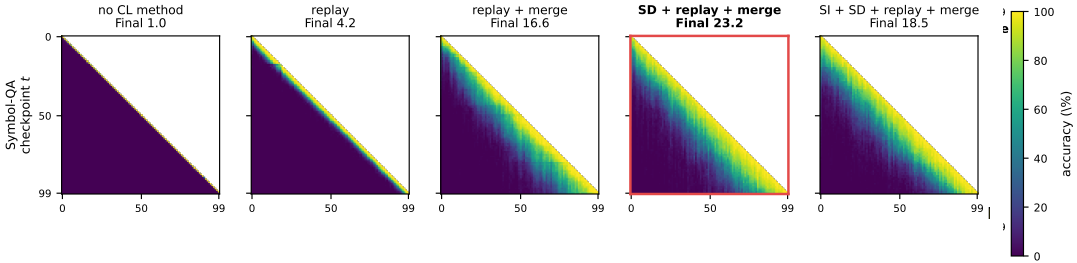
Backup: uncertainty matters for small differences

Final (mean $\pm$ seed SD, %)	Symbol	LLM	Real
Replay + Merge	16.6 ± 5.8	32.4 ± 8.9	48.2 ± 2.8
SI + Replay + Merge	15.7 ± 3.3	41.2 ± 9.5	54.8 ± 10.4
SD + Replay + Merge	23.2 ± 4.2	33.0 ± 1.8	39.0 ± 1.9
All anchors + Merge	18.5 ± 2.8	41.8 ± 1.4	44.3 ± 8.7

- Three seeds; standard deviations are not confidence intervals.
- TSH selection and report-order winners can differ.
- About 28-fold is based on less-rounded values; do not divide rounded 34.9 by rounded 1.2 and expect exact agreement.

Tables 8–10, PDF pp. 40–41; Appendix E.10, p. 47.

Backup: read retention matrices down each column



Symbol-QA only. The bright band widens with Replay + Merge.
Adding SI to SD + Replay + Merge lowers Final from 23.2 to 18.5%.

Fig. 5, PDF p. 44, cropped to Symbol-QA with the original color scale. Median seed.

Backup: implementation details that affect reproduction

- LoRA targets q/k/v/o and gate/up/down projections; dropout 0.05.
- Replay weights: 0.75 / 0.5 / 0.75 on Symbol / LLM / Real.
- SD weight 1, temperature 5; SI weight 1, damping 0.1.
- Online EWC: weight 1000, decay 1, Fisher rescaled to first-task mean.
- O-LoRA: entrywise overlap penalty weight 0.5; factor norm disabled.
- Sequential OSRM uses up to 64 samples/task for mean features, including padding.

Appendix B.8, Table 3, PDF p. 32.

Backup: textual caveats should not become firm conclusions

- E.7 describes standalone online EWC as involving a fold, although its defined allocation is shared LoRA.
- E.7 says all other Diag scores are 96–100%; Table 9 includes 89.2% for online EWC and 95.3% for SI.
- The temperature control gives only three seeds and does not clearly name the tested dataset/composition in that paragraph.
- Use the explicit per-configuration tables; treat stronger causal and variance claims cautiously.

Appendix E.7–E.8, PDF pp. 45–46; Table 9, p. 41. Reader-identified inconsistencies.