DiscoLoop

Looping Discrete Embeddings and Continuous Hidden States for Multi-hop Reasoning

Paper by Hengyu Fu, Tianyu Guo, Zixuan Wang, Hanlin Zhu,
Jason D. Lee, Jiantao Jiao, Stuart Russell, and Song Mei

arXiv:2607.00341v2 (27 July 2026)
Research discussion | 10 October 2026

Shared memory needs an intermediate representation it can reuse.

Can stored facts compose without an explicit trace?

The controlled task

Alice's son is Bob. Bob's wife is Carol.

Query: Who is the wife of Alice's son? Answer: Carol.

$$(a, r_1, b),\ (b, r_2, c) \quad \Longrightarrow \quad (a, r_1, r_2, c)$$

- Facts are learned in weights; the query does not supply them.
- Predict c directly; do not generate the bridge entity b.
- One overall forward pass contains K backbone applications.

Paper, Secs. 2–3.1, PDF pp. 3–4, Eqs. (1)–(2).

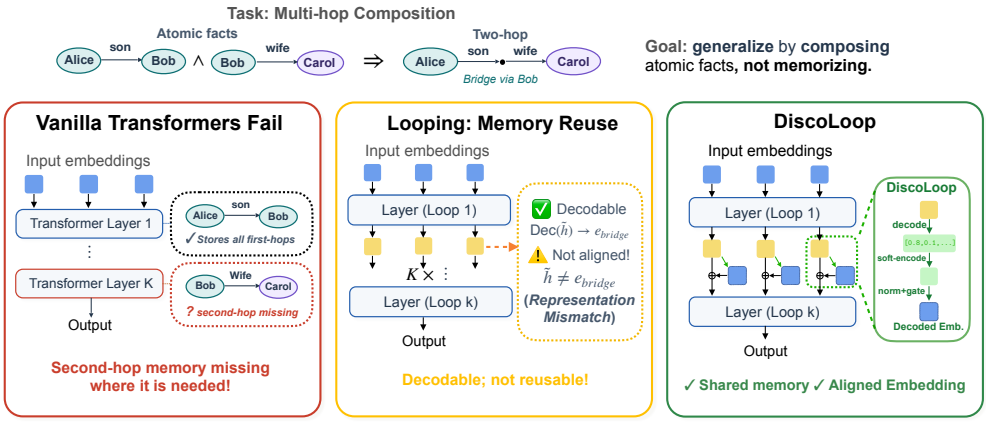
OOD isolates transfer of the composition rule

	Graph $\mathcal{G}_A$	Graph $\mathcal{G}_B$
Entities	500	500, disjoint from A
Atomic facts in training	All 5,000	All 5,000
Two-hop training queries	10,000	None
Two-hop test queries	2,000 ID	2,000 OOD

- Shared vocabulary of 50 relations; out-degree 10 per entity.
- OOD entities and facts have been seen as atomic examples.
- Their use in a composition is held out entirely.

Paper, Secs. 2–3.1, PDF pp. 3–4. Each graph contains 50,000 two-hop chains.

Memory reuse leaves a representation bottleneck



Earlier facts must be both accessible and usable by the next loop.

Paper, Fig. 1, PDF p. 3; Sec. 3, PDF pp. 4–6. Original author figure.

Two loops recall facts but generalize imperfectly

$$\mathbf{H}^{(0)} = \mathbf{W}[x], \qquad \mathbf{H}^{(k+1)} = f_\theta(\mathbf{H}^{(k)}), \qquad K = 2$$

Accuracy (%)	Atomic train	2-hop train	ID test	OOD test
Stage 1	100.0	8.8	0.9	0.0
Stage 2	100.0	100.0	71.1	8.3

- Stage 1 reads the LM head after one backbone application.
- Stage 2 is the actual prediction after two applications.
- Composition uses the second loop, yet OOD remains poor.

Paper, Eq. (2), PDF p. 4; Table 1a, PDF p. 6. After 3,000 training epochs.

The bridge is decodable but geometrically different

At position 1, after loop 1	ID	OOD
Correct bridge probability $P(b \mid \mathbf{H}_1^{(1)})$	1.000	1.000
Cosine with input embedding $\mathbf{W}[b]$	0.327	0.266

Different questions for the same representation

Can the LM head identify b? Can f_θ consume it as b?

- A dominant logit does not require embedding alignment.
- Bridge readout is at r_1; final-answer readout targets c.

Paper, Sec. 3.2, PDF pp. 5–6, Table 1b. Probabilities shown at reported precision.

A local intervention tests the loop interface

$$b_{\max} = \arg\max_{v} (\mathbf{W}\mathbf{H}_1^{(1)})_v$$

$$\mathbf{H}_1^{(1)} \leftarrow (1-\alpha)\mathbf{H}_1^{(1)} + \alpha \, \mathrm{Norm}(\mathbf{W}[b_{\max}])$$

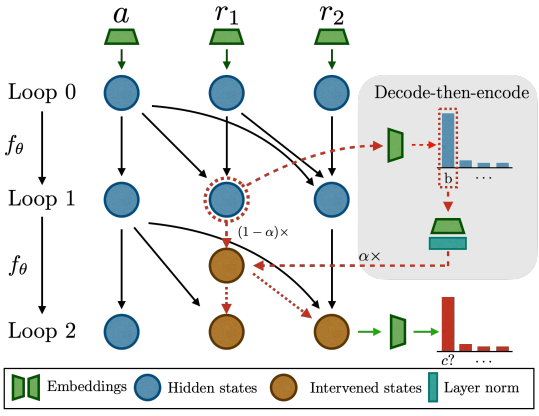
- Modify position 1 only, between loops 1 and 2.
- Use the model's decoded token; no oracle bridge is supplied.
- No weight update is needed.

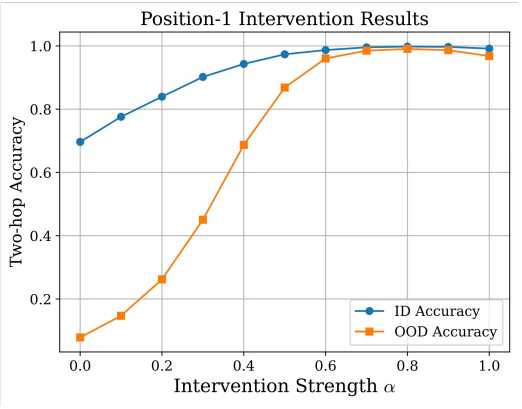
Interpretation

Better composition after changing the interface supports a representation bottleneck beyond memory availability.

Paper, Sec. 3.2, PDF p. 5, Eq. (3).

Embedding injection nearly closes the OOD gap





OOD rises from 8.3% to about 99% near $\alpha = 0.7$–0.8.

Paper, Fig. 2, PDF p. 5; approximate reading of the plotted curve. The prose reports different intervention strengths; see backup.

Decode softly, then return to embedding space

$$p_v(h) = \frac{\exp((\mathbf{W}h)_v/\tau)}{\sum_{v'} \exp((\mathbf{W}h)_{v'}/\tau)}, \qquad \Phi(h) = \sum_{v=1}^{V} p_v(h)\mathbf{W}[v]$$

Hidden state $h \longrightarrow$ vocabulary distribution $p \longrightarrow$ soft embedding $\Phi(h)$

- $\mathbf{W} \in \mathbb{R}^{V \times d}$ is tied across input and output.
- If $p_b \approx 1$, the returned vector approaches $\mathbf{W}[b]$.
- The "discrete" channel uses token embeddings but stays soft.
- Temperature $\tau > 0$ controls distribution sharpness.

Paper, Sec. 4, PDF p. 6, Eq. (5).

DiscoLoop adds the soft embedding to the residual

$$\widetilde{\mathbf{H}}^{(0)} = \mathbf{W}[x], \qquad \mathbf{H}^{(k+1)} = f_\theta(\widetilde{\mathbf{H}}^{(k)})$$

$$\boxed{\widetilde{\mathbf{H}}^{(k+1)} = \mathbf{H}^{(k+1)} + \alpha^{(k)} \odot \mathrm{RMSNorm}\Big(\Phi(\mathbf{H}^{(k+1)})\Big)}$$

- Preserve the continuous state and add an embedding channel.
- Both channels occupy the same $T \times d$ residual space.
- Apply at every position; no bridge location is hard-coded.
- Set $\alpha^{(K-1)} = 0$; read final logits from $\mathbf{H}^{(K)}\mathbf{W}^\top$.

Paper, Sec. 4, PDF pp. 6–7, Eq. (4). There are $K - 1$ loop-to-loop injections.

A shared gate controls the injection strength

$$\alpha_t^{(k)} = \sigma\Big(\Big\langle \mathbf{w}_\alpha, \Phi(\mathbf{H}_t^{(k+1)}) \Big\rangle + b_\alpha\Big)$$

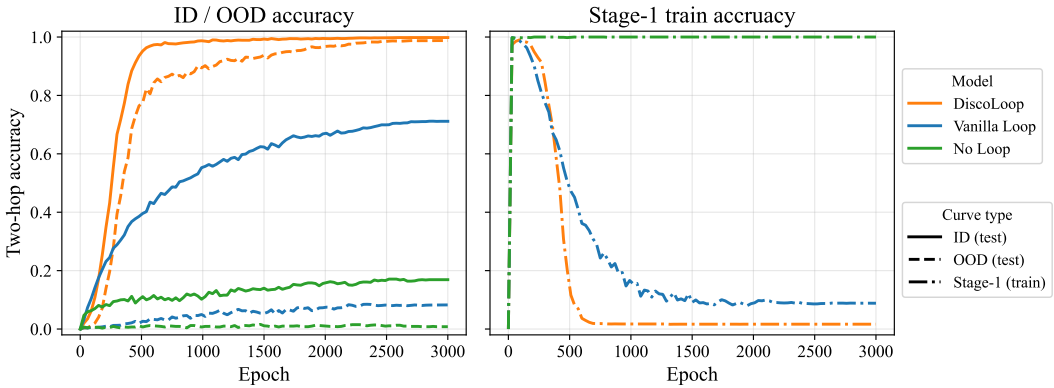
- Gate input: the decoded embedding before RMSNorm.
- One vector and bias, shared over tokens and loops: $d + 1$ parameters.
- Symbolic two-hop: fixed $\alpha^\star = 1$, temperature $\tau = 1$.
- Synthetic language: a learned token-wise gate.

Two distinct mixing rules

Diagnostic intervention: $(1 - \alpha)h + \alpha e$.
Learned architecture: $h + \alpha e$.

Paper, Eqs. (3)–(6), PDF pp. 5–7; Sec. 5.2, PDF p. 8.

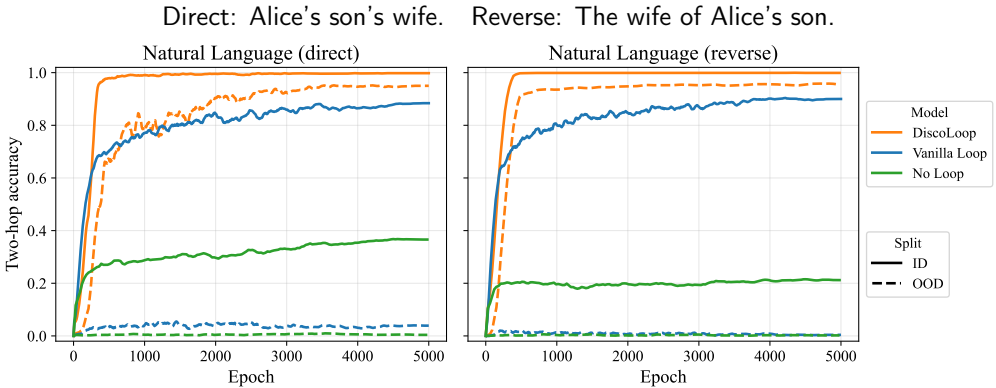
Symbolic generalization improves as shortcuts fade



Around epoch 500: ID/OOD exceed 80%; Stage-1 answers collapse.

Paper, Sec. 5.1, PDF pp. 7–8, Fig. 3. $K = 2$, 3,000 epochs. Right: Stage-1 training accuracy; non-looped curve uses its final training readout.

The effect survives two English question formats



Direct and reverse formats: nearly 100% ID, about 95% OOD.

Paper, Sec. 5.2, PDF pp. 8–9, Fig. 4. $K = 2$, learned gate, 5,000 epochs. Solid = ID, dashed = OOD. Phrasing above is a teaching example.

Three-hop OOD improves to about 65%



Both baselines stay near zero on three-hop OOD.

Paper, App. A.2, PDF pp. 13–14, Fig. 5. $K = 3$; training contains both two-hop and three-hop queries. This evaluates transfer across graphs at trained depths.

Pretraining compares three recurrent signals

Model	Information passed across loops
Vanilla loop	Continuous hidden states
PonderLM	Decoded embedding-space states
DiscoLoop	Continuous states + decoded embeddings

- 440M parameters; 24-layer backbone, $d = 1024$, 16 heads.
- Four backbone applications for every recurrent model.
- 20B tokens: 60% FineWeb-Edu, 40% FineMath-4plus.
- Context 8,192; tied embeddings. DiscoLoop uses top-128 return.

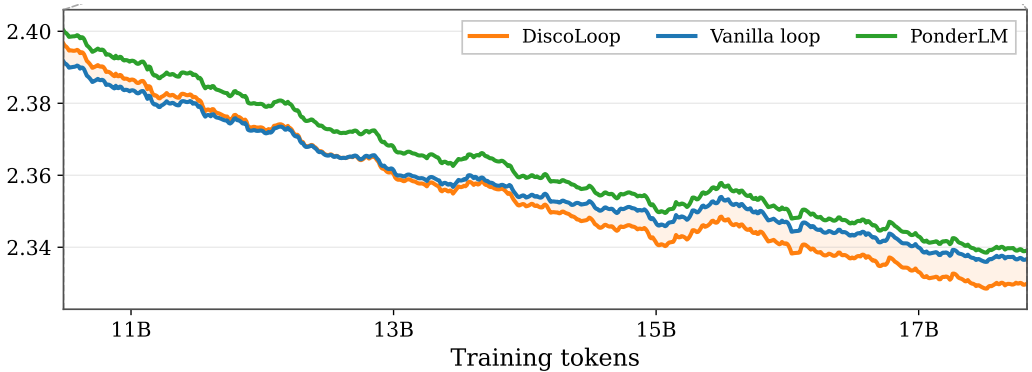
Paper, Sec. 5.3 and App. B, PDF pp. 9, 14–15. PonderLM: three ponder steps plus the final backbone application.

Zero-shot average improves by 1.2 points

Score (%)	Vanilla loop	PonderLM	DiscoLoop
ARC-C	30.5	31.7	**32.6**
ARC-E	57.2	55.6	**57.6**
HellaSwag	**44.2**	43.9	**44.2**
LAMBADA	35.1	**38.0**	37.6
PIQA	67.4	67.7	**68.4**
RACE	31.0	31.2	**31.3**
SciQ	79.5	80.5	**81.5**
Average	49.3	49.8	**50.5**

Paper, Table 2, PDF p. 9. Zero-shot acc_norm for ARC-C/E, HellaSwag, PIQA, SciQ; acc for LAMBADA and RACE. DiscoLoop is best or tied on six of seven tasks.

The loss advantage emerges late in pretraining



Global average loss: DiscoLoop overtakes Vanilla near 13B tokens.

Paper, Fig. 7 lower panel, PDF p. 16; cropped from the author figure. Loss is EMA-smoothed; lower is better. Peak late-stage separation is approximately 10^{-2}.

Few new parameters still add vocabulary computation

$$\mathbf{P} = \mathrm{softmax}(\mathbf{H}\mathbf{W}^{\top}/\tau), \qquad \Phi(\mathbf{H}) = \mathbf{P}\mathbf{W}$$

- Full decode and return each involve $O(TVd)$ arithmetic.
- Top-128 reduces selected embedding aggregation to $O(Tkd)$.
- Finding the top tokens still requires vocabulary scoring.
- Faster learning in epochs does not establish faster wall time.

Comparison to request next

Report end-to-end throughput, memory, and equal-total-compute quality.

Complexity interpretation of Eq. (5), PDF p. 6, and App. B, PDF pp. 14–15. The paper does not report end-to-end overhead for the added channel.

The next tests are depth and scale

- **Depth:** train up to k hops, test beyond k.
  The existing three-hop study includes three-hop training data.
- **Scale:** move beyond 440M parameters and 20B tokens;
  test adaptation of existing pretrained models.
- **Mechanism:** probe the loop interface in pretrained models;
  broad benchmark gains alone do not localize their cause.
- **Reproducibility:** clarify synthetic configurations and
  pretraining gate, temperature, and top-k normalization.

Paper, Sec. 6, PDF pp. 9–10; Apps. A–B, PDF pp. 13–15. Mechanistic follow-up and configuration observations are reading analysis.

A useful loop must pass a useful representation

1. **Diagnosis:** shared memory alone is insufficient;
   the bridge is decodable but poorly aligned with its embedding.
2. **Design:** retain hidden states and inject soft decoded
   embeddings between backbone applications.
3. **Evidence:** large controlled-task gains; smaller positive
   pretraining gains under the reported 440M / 20B-token setting.

Discussion question

How much of the gain requires semantic embedding alignment,
compared with an equally costly learned residual transformation?

Synthesis of Secs. 3–5 and App. A.2. Discussion question is proposed analysis.

Primary reference

Hengyu Fu, Tianyu Guo, Zixuan Wang, Hanlin Zhu,
Jason D. Lee, Jiantao Jiao, Stuart Russell, and Song Mei.

DiscoLoop: Looping Discrete Embeddings and Continuous Hidden States for Multi-hop Reasoning.

arXiv:2607.00341v2, revised 27 July 2026.

https://arxiv.org/abs/2607.00341v2

All empirical figures and benchmark values in this deck come from this version. Figures retain the original data and legends.

Research discussion slides; not an author presentation. Backup slides follow.

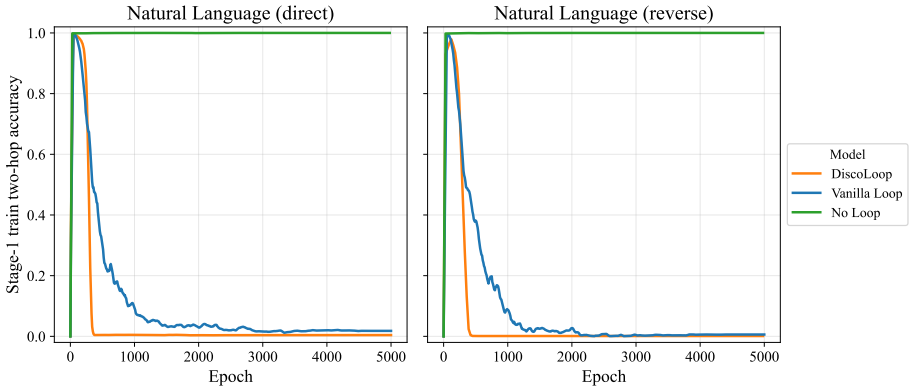
Backup: intervention strengths differ in text and figure

OOD accuracy	Sec. 3.2 prose	Fig. 2, approximate
$\alpha = 0.1$	25.9%	15%
$\alpha = 0.2$	—	26%
$\alpha = 0.5$	Nearly 100%	87%
$\alpha \approx 0.7$–0.8	—	99%

- Both support the large benefit of embedding injection.
- The plotted high-accuracy region is used in this presentation.
- Exact strength selection needs clarification of the discrepancy.

Paper, Sec. 3.2 and Fig. 2, PDF p. 5. Figure values are visual estimates, not raw measurements.

Backup: language data show the same Stage-1 transition



Single-loop answer accuracy falls earlier with DiscoLoop.

Paper, App. A.3, PDF p. 14, Fig. 6. Roughly 400 epochs for DiscoLoop; much later for Vanilla. Non-looped model uses its final training readout.

Backup: the pretraining recipe

Component	Reported setting
Tokenizer / vocabulary	DeepSeek-V3-0324 / 129,280
Positional encoding	RoPE base 10^6; max positions 16,384
Activations / QKV bias	SwiGLU / no QKV bias
Optimizers	AdamW and Muon enabled
Learning rate	3×10^{-3} to 10^{-5}, linear decay
Warmup / clipping	1,000 steps / gradient norm 1.0
AdamW ϵ / Muon momentum	10^{-15} / 0.95
Global batch / sequence length	128 / 8,192
Loop launcher budget	19,074 steps; 20,000,538,624 tokens

Paper, App. B, PDF p. 15. These are language-model settings, not the missing synthetic-task hyperparameters.

Backup: top-k normalization needs clarification

$$\Phi_k(h) = \sum_{v \in \mathcal{T}_k(h)} p_v(h)\mathbf{W}[v], \qquad k = 128$$

- The formula retains probabilities from the full distribution.
- The prose describes masking logits outside the top-k set.
- Renormalizing on that support changes the vector's scale.
- RMSNorm largely removes scale; a gate reading raw Φ can still respond differently.

Paper, App. B, PDF pp. 14–15; gate in Eq. (6), PDF p. 6. The scale/gate implication is an analytical observation.