Balancing Memory Pathways

Analyzing and Improving Memory Utilization in Hybrid LMs

Paper by Hyunji Lee, Joykirat Singh, Zaid Khan,
Justin Chih-Yao Chen, Elias Stengel-Eskin,
Alessandro Sordoni, Arman Cohan, and Mohit Bansal

arXiv:2610.06750v1 (5 October 2026)
Research reading ▪ 13 November 2026

Multiple memory pathways need targeted training

Takeaway

Ordinary SFT mainly strengthens attention. A recurrent-only auxiliary pass makes recurrent memory more useful and improves the full model.

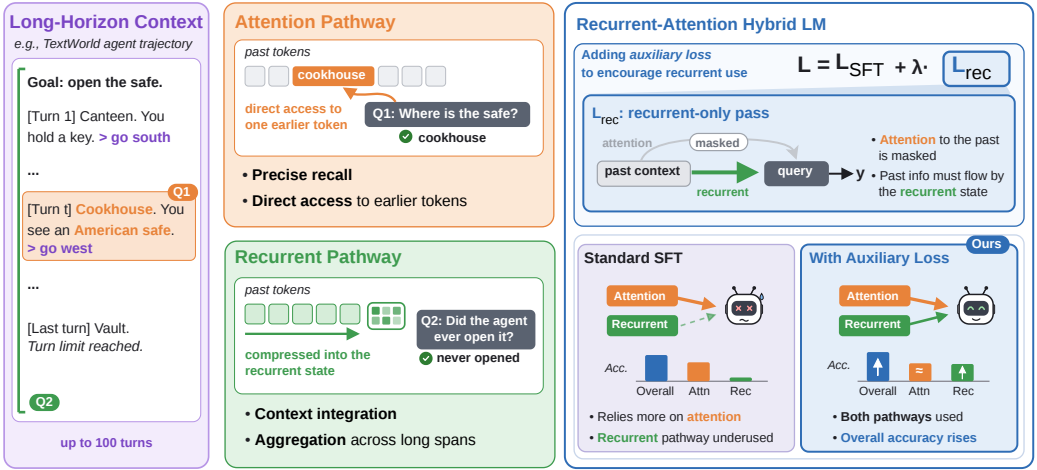
- QA micro-average: 34.6 → 38.5 (Qwen3.5); 31.0 → 36.2 (Nemotron-H).
- Benefits are strongest on longer-context QA and longer agent trajectories.

Publication not found as of 2026-11-13: arXiv record, author list.

Official public code: training, data preparation, and pathway evaluation.

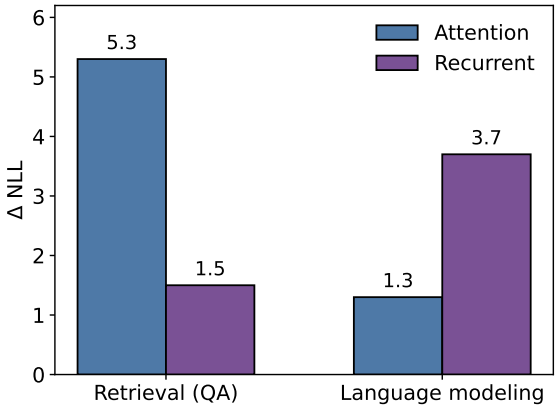
Paper, Secs. 1, 5; Table 1, PDF pp. 1–3, 7. Scores follow dataset-specific EM/F1.

Two pathways offer different ways to use the past



Paper, Fig. 1, PDF p. 2. Original schematic: direct token access versus compact recurrent state.

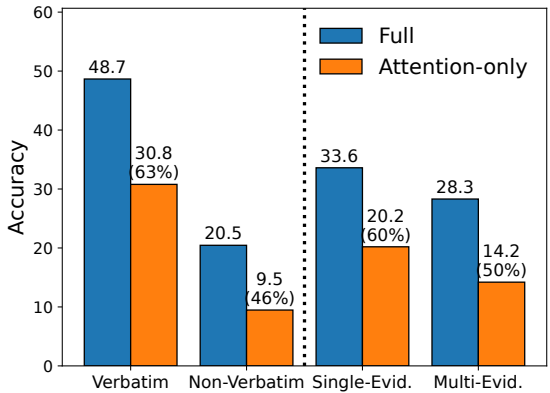
Attention and recurrence have different strengths



- Higher ΔNLL means the removed component mattered more.
- Based NQ retrieval depends more on attention.
- PG19 language modeling depends more on recurrence.

Paper, Sec. 2.2; Fig. 2, PDF pp. 3–4. Qwen3.5; whole-layer ablation.

Information needs change the relative dependence



$$R_{\text{attn}} = \frac{\text{Attention-only score}}{\text{Full score}}$$

- Higher retention for verbatim and single-evidence QA.
- Lower retention when inference or evidence integration is needed.

Paper, Fig. 3, PDF p. 4; Appendix A, p. 14. Category counts are balanced within each comparison.

Interventions isolate cross-segment information flow

Divide each input into past and current segments.
For QA: document | question and subsequent prediction.

Condition	Past attention	Past recurrent state
Full	Available	Available
Recurrent-only	Blocked	Available
Attention-only	Available	Reset at boundary

All layers remain active *within* each segment.
"Only" describes the mechanism that carries information across the boundary.

Paper, Sec. 4.1, PDF p. 5. Different from the whole-layer ablation in Fig. 2.

Recurrent-only masks attention across the boundary

$$M_{t,s} = \begin{cases} 0, & s \le t \text{ and } \sigma(s) = \sigma(t), \\ -\infty, & \text{otherwise.} \end{cases}$$

- t: query position; s: key position; σ: segment identity.
- Add M to attention scores before normalization.
- Keep causal attention within each segment; recurrent state remains continuous over the entire sequence.

Effect

Past information must enter recurrent state to influence the current answer.

Paper, Sec. 4.1, PDF p. 5; unnumbered equation.

Attention-only resets the recurrent history

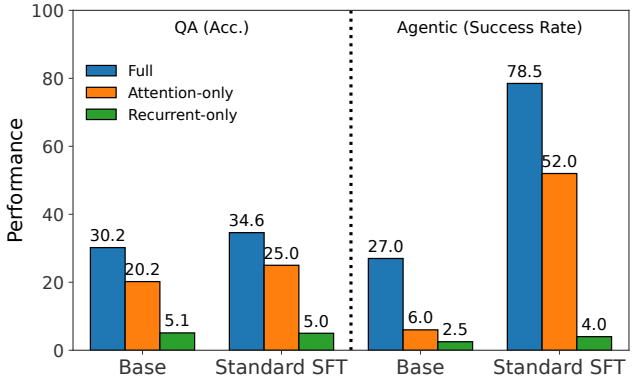
$$S_t = \mathbf{1}[\sigma(t) = \sigma(t-1)]\, \mathcal{A}_t(S_{t-1}) + w_t, \quad S_0 = 0.$$

- S_t: recurrent state; $\mathcal{A}_t$: transition; w_t: current-token write.
- For $t > 1$: at a boundary, the old state contribution becomes zero.
- Standard causal attention can still retrieve preceding tokens.

This is a general intervention on state propagation, rather than a replacement of the model's recurrent update rule.

Paper, Sec. 4.1, PDF p. 5; unnumbered equation.

Standard SFT leaves recurrent-only performance low



Full and Attention-only improve; Recurrent-only barely changes.

QA: 5.1 $\rightarrow$ 5.0; Quest: 2.5 $\rightarrow$ 4.0.

Paper, Sec. 4.2; Fig. 4, PDF p. 6. Qwen3.5-4B, percent scores.

Train the full model and a recurrent-only pass jointly

$$\boxed{\mathcal{L}(\theta) = \mathcal{L}_{\mathrm{SFT}}(\theta) + \lambda \mathcal{L}_{\mathrm{rec}}(\theta)}, \qquad \lambda = 0.25.$$

- Same supervised next-token target in both passes.
- Full pass: both pathways can carry past information.
- Auxiliary pass: attention to past segments is masked; recurrence is continuous.
- Full model at inference; no auxiliary inference pass.

Compute-matched comparison

Auxiliary runs use half as many optimization steps to offset the extra computation per step.

Paper, Secs. 3, 5.1; Appendix C.1, PDF pp. 4, 7, 14–15.

Evaluation tests task quality and pathway utilization

Models and adaptation

- Qwen3.5-4B: 24 Gated DeltaNet + 8 attention layers.
- Nemotron-H-4B-Instruct: 24 Mamba2 + 4 attention layers.
- All-layer LoRA: rank 16, scaling 32, dropout 0.05.

Tasks and metrics

- QA: 12k train / 2k validation; 8.2k test, mean context 84k tokens.
- Six datasets; dataset-specific EM/F1; micro-average.
- Agents: TextWorld and BabyAI; success within step limit.

Paper, Sec. 3; Appendix B, PDF pp. 4–5, 14. Five epochs with early stopping; auxiliary step adjustment applies.

Recurrent auxiliary supervision improves QA overall

	Qwen3.5		Nemotron-H	
Dataset	SFT	+Rec	SFT	+Rec
DROP	48.6	46.9	42.4	44.5
MuSR	45.6	44.8	40.9	38.1
HotpotQA	43.1	44.0	38.7	42.1
Qasper	30.1	35.9	28.1	34.1
NarrativeQA	24.4	33.7	21.9	31.7
MINTEval	21.1	28.9	18.8	27.8
Micro-average	34.6	**38.5**	31.0	**36.2**

Paper, Table 1, PDF p. 7. +Rec means SFT + recurrent auxiliary loss. QA "accuracy" combines EM and F1.

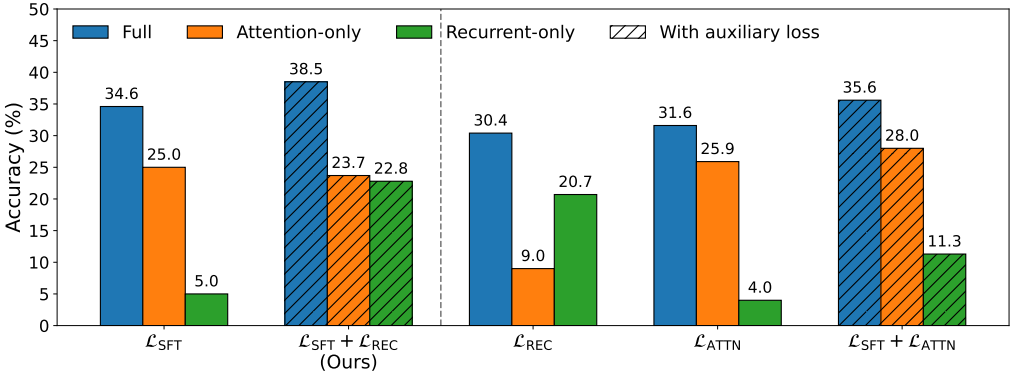
Gains concentrate on longer-context datasets

Gain over SFT (percentage points)	Qwen3.5	Nemotron-H
Qasper	+5.8	+6.0
NarrativeQA	+9.3	+9.8
MINTEval	+7.8	+9.0

- Longer-context group: about +8.0 pp; other three datasets: about +0.2 pp.
- These summarize equal-weight dataset/model gains, rather than the micro-average.
- Some shorter tasks regress: Qwen DROP/MuSR and Nemotron MuSR.

Paper, Secs. 1, 5.2; Table 1, PDF pp. 2, 7. Dataset differences do not isolate length alone.

The improvement accompanies much stronger recurrent use



SFT $\rightarrow$ SFT+Rec: recurrent-only 5.0 $\rightarrow$ 22.8;
attention-only 25.0 $\rightarrow$ 23.7; full score 34.6 $\rightarrow$ 38.5.

Paper, Fig. 6, PDF p. 8. Qwen3.5-4B. Restricted losses alone reduce full-model performance.

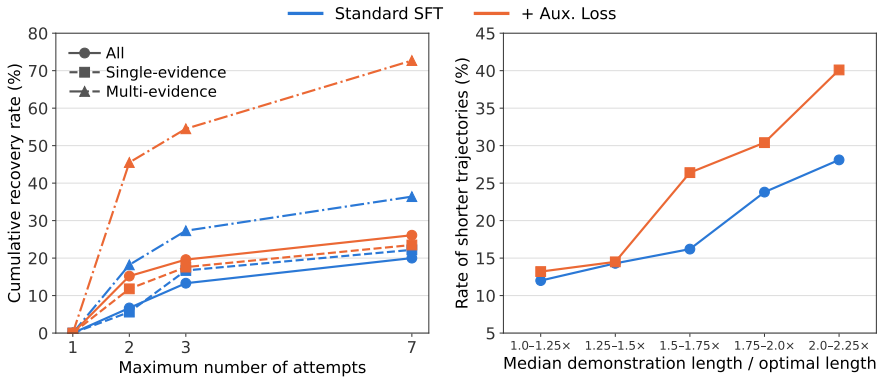
Agent gains are largest on BabyAI

Model	Objective	Quest	Treasure	PutNext	GoToMaze
Qwen	SFT	78.5	44.0	46.0	63.0
	SFT+Rec	**80.5**	**56.0**	**73.0**	**83.0**
Nemotron	SFT	70.5	40.5	30.0	77.0
	SFT+Rec	**75.5**	**47.0**	**47.0**	**84.0**
Mean gain (pp)		+3.5	+9.3	+22.0	+13.5

Mean trajectory length: BabyAI 134.0 vs. TextWorld 42.2 steps.
Longer trajectories are consistent with larger memory benefits.

Paper, Table 2, PDF p. 8. PutNext = PutNextLocal; GoToMaze = GoToObjMaze. Success rates (%).

Auxiliary training improves use of past trajectories

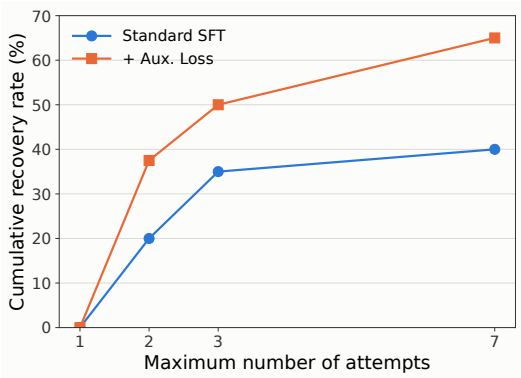


Failure recovery: larger gains on multi-evidence QA.

Successful demonstrations: more often produce a successful, shorter trajectory.

Paper, Fig. 7, PDF p. 9. Left: attempt budget; right: demonstration length relative to optimal.

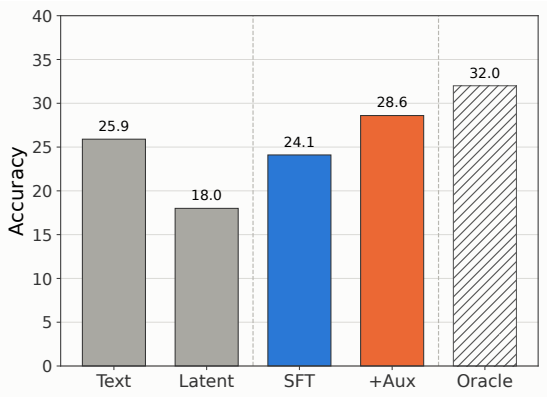
Experience helps when it must be integrated



- TextWorld: prior failures in context, up to 7 attempts; reported mean recovery gain +19.2 pp.
- QA group means: multi-evidence +30.3 pp; single-evidence +2.8 pp.
- Five successful demonstrations: +6.0 pp rate of beating the shortest demonstration.

Paper, Sec. 6, Fig. 7; Appendix C.2, Fig. 10, PDF pp. 9, 14–15. QA overall aggregation caveat: backup slide 30.

The same imbalance appears in text + latent memory



- Qwen3-4B-Instruct: 8k-token text summary + 128 latent KV slots.
- Naive joint SFT: 24.1, below text-only 25.9.
- Auxiliary training: 28.6.
- Oracle 32.0 selects the better single-memory prediction per example.

Paper, Sec. 6, Fig. 8; Appendix C.3, PDF pp. 9, 15–16. Oracle is a hindsight reference.

Role-based supervision makes the other memory irrelevant

$$\ell(T_i, L_i) = -\log p_\theta(y_i \mid q_i, T_i, L_i)$$

$$\mathcal{L}_i = \ell(T_i, L_i) + \lambda \mathcal{L}_i^{\text{aux}}, \quad \mathcal{L}_i^{\text{aux}} = \begin{cases} \ell(T_i, \text{sg}[L_j]), & \text{role}(i) = \text{text}, \\ \ell(T_j, L_i), & \text{role}(i) = \text{latent}. \end{cases}$$

- T_i, L_i: correct text/latent memories; j: unrelated context.
- q_i, y_i: question/target; sg: stop gradient; $\lambda = 0.25$.
- Best rule: verbatim $\rightarrow$ text; non-verbatim $\rightarrow$ latent.
- Role labels supervise training; full inference has both memories.

Paper, Appendix C.3.3, Eq. (1), PDF p. 17. Unrelated memories are matched by dataset and similar document length.

Choosing the right role matters more than adding more losses

Variant	Score	ΔLatent	ΔText
Naive SFT	24.1	−2.4	+11.4
Verbatim	**28.6**	+4.7	+10.4
Reverse	24.5	+1.1	+7.2
Random	23.9	−1.0	+10.5
Latent-aux for all	26.1	+3.4	+8.5

- Δ: score drop after replacing that memory with unrelated content.
- Verbatim routing increases latent use while largely retaining text use.
- Random roles and indiscriminate auxiliary objectives fail to match this gain.

Paper, Table 3; Appendix C.3.3, PDF pp. 17–18. Positive Δ indicates sensitivity to the correct memory.

Train complementary use, then measure the tradeoff

1. **Diagnose:** task improvement can hide an underused memory pathway.
2. **Intervene:** forcing information through that pathway during training improves full-model performance.
3. **Coordinate:** text/latent experiments show that role assignment can preserve complementary strengths.

Scope and costs

- Evidence: 4B-scale models, LoRA SFT; some short QA datasets regress.
- Compute matched by fewer steps; wall-clock costs are not measured.
- Interventions reveal dependence, rather than a unique additive contribution of each pathway.

Paper, Secs. 3–7, PDF pp. 4–10; interpretation based on reported protocols and ablations.

Sources and discussion prompts

Primary sources

- Lee et al. (2026), *Balancing Memory Pathways: Analyzing and Improving Memory Utilization in Hybrid LMs*.
  arXiv:2610.06750v1, including Appendices A–C.
- Official public implementation and author publication list.

Questions suggested by this reading

- Can role assignment be learned without gold-answer heuristics?
- How do gains change under fixed FLOPs, fixed examples, and fixed wall-clock budgets?
- Can length and evidence integration be varied independently within a task?

Publication/code checked 2026-11-13. Discussion questions are reader analysis. Backup slides follow.

Backup: the restricted objective should remain auxiliary

Training objective	Qwen3.5	Nemotron-H
$\mathcal{L}_{\text{SFT}}$	34.6	31.0
$\mathcal{L}_{\text{SFT}} + \mathcal{L}_{\text{rec}}$	**38.5**	**36.2**
$\mathcal{L}_{\text{rec}}$	30.4	29.8
$\mathcal{L}_{\text{attn}}$	31.6	31.4
$\mathcal{L}_{\text{SFT}} + \mathcal{L}_{\text{attn}}$	35.6	34.1

- Mean gain: recurrent auxiliary $+4.6$ pp vs. attention auxiliary $+2.1$ pp.
- A single restricted loss does not preserve full-model performance as well.

Paper, Table 1, PDF p. 7. QA micro-average; two-model mean gains rounded to one decimal.

Backup: Qwen agent objective ablations

Objective	Quest	Treasure	PutNext	GoToMaze
$\mathcal{L}_{\mathrm{SFT}}$	78.5	44.0	46.0	63.0
$\mathcal{L}_{\mathrm{SFT}} + \mathcal{L}_{\mathrm{rec}}$	**80.5**	**56.0**	**73.0**	**83.0**
$\mathcal{L}_{\mathrm{rec}}$	75.5	45.0	13.0	37.0
$\mathcal{L}_{\mathrm{attn}}$	76.5	44.0	31.0	50.0
$\mathcal{L}_{\mathrm{SFT}} + \mathcal{L}_{\mathrm{attn}}$	77.0	49.0	47.0	70.0

Restricted-only objectives particularly hurt BabyAI.

Preserving the unrestricted training condition matters.

Paper, Table 2, PDF p. 8. Success rates (%). PutNextLocal and GoToObjMaze abbreviated.

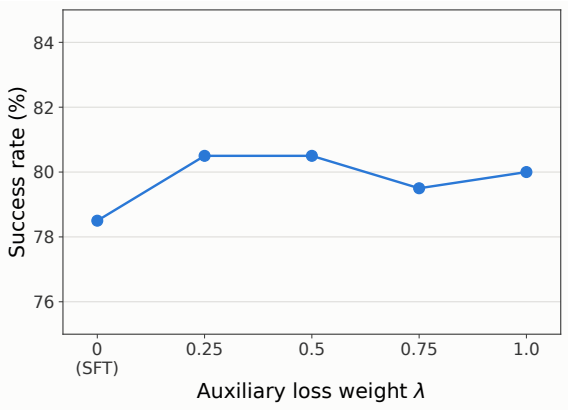
Backup: Nemotron agent objective ablations

Objective	Quest	Treasure	PutNext	GoToMaze
$\mathcal{L}_{\text{SFT}}$	70.5	40.5	30.0	77.0
$\mathcal{L}_{\text{SFT}} + \mathcal{L}_{\text{rec}}$	**75.5**	**47.0**	**47.0**	**84.0**
$\mathcal{L}_{\text{rec}}$	68.5	40.0	17.0	20.0
$\mathcal{L}_{\text{attn}}$	71.5	42.0	37.0	54.0
$\mathcal{L}_{\text{SFT}} + \mathcal{L}_{\text{attn}}$	72.0	44.5	43.0	80.0

The recurrent auxiliary advantage also holds with Mamba2-based recurrence.

Paper, Table 2, PDF p. 8. Success rates (%). PutNextLocal and GoToObjMaze abbreviated.

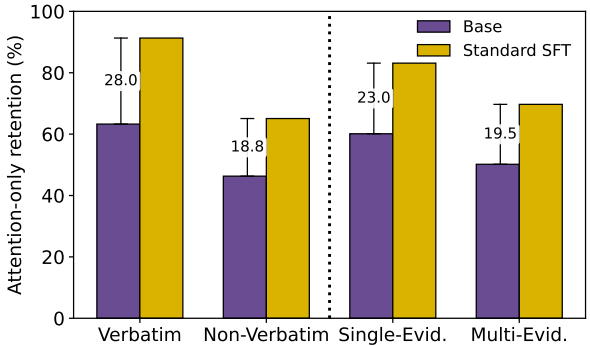
Backup: the tested nonzero weights all improve Quest



- Qwen3.5, TextWorld Quest.
- SFT ($\lambda = 0$): 78.5%.
- All tested nonzero values improve over this baseline.
- Authors select $\lambda = 0.25$; plot also shows 0.5 at approximately 80.5%.

Paper, Appendix C.1; Fig. 9, PDF pp. 14–15. This is a single task/model sweep.

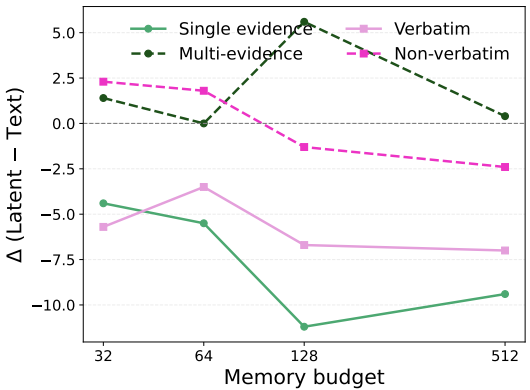
Backup: SFT increases attention retention across QA types



- Gains are in *percentage points* of retention.
- Verbatim / single-evidence mean: +25.5 pp.
- Non-verbatim / multi-evidence mean: about +19.2 pp.

Paper, Sec. 4.2; Fig. 5, PDF p. 6. Qwen3.5; retention = Attention-only / Full score.

Backup: text and latent strengths vary by information type



- Δ = latent score minus text score under matched budgets.
- Text favors verbatim / single evidence.
- Latent is more competitive on multi-evidence / non-verbatim questions.
- Latent encoding: document + K memory tokens; retain their layer-wise KV as answer prefix.

Paper, Appendix C.3.1; Fig. 11, PDF pp. 15–16. Distinct from the 8k text / 128 latent joint setting.

Backup: full role-assignment comparison

Role assignment	Variant	Score	ΔLatent	ΔText
–	Naive	24.1	−2.4	+11.4
Per-example	Outcome	26.1	+0.8	+11.5
	Verbatim	**28.6**	+4.7	+10.4
	Reverse	24.5	+1.1	+7.2
	Random	23.9	−1.0	+10.5
All examples	Latent-aux	26.1	+3.4	+8.5
	Text-aux	24.9	+0.0	+14.5
	Unrouted (both)	24.4	−1.2	+9.1
–	Oracle	32.0	–	–

Paper, Table 3, PDF p. 17; interpretation p. 18. Outcome chooses latent role when latent-only prediction is correct.

Backup: interpret QA recovery aggregates with care

- Sec. 6 reports a mean QA recovery gain of **16.6 pp**.
- Reported group means: single-evidence $14.8 \rightarrow 17.6$; multi-evidence $27.3 \rightarrow 57.6$.
- The equal-weight mean of these gains is $(2.8 + 30.3)/2 \approx 16.6$ pp.
- Yet the plotted *All* curves differ by only about 6–9 pp at budgets 2, 3, and 7; the aggregation/denominator is not reconciled in the text.

Use the group-specific results and plotted trend. Treat 16.6 pp as a *reported aggregate*, rather than reading it directly from the All curve.

Paper, Sec. 6; Fig. 7, PDF p. 9. Numerical mismatch identified in this reading; curve gaps are approximate.