DAYDREAMER: TRAINING AND EVALUATING ASSOCIATIVE MEMORY NAVIGATION FOR PERSONA-GROUNDED GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Creative writing often draws on personal experience through association: one memory cues another, and fragments of experience are selectively recombined into new material. Yet existing benchmarks largely evaluate either the final creative artifact or one-shot memory retrieval, leaving this associative process itself unmeasured. We introduce DAYDREAMBENCH, a benchmark for studying episodic memory navigation in persona-grounded creative writing. It contains 46,959 episodic memories across 201 persona–corpus identities and 343 held-out writing tasks, with executable reference flows that connect literary works to plausible experiential memories. We further propose DAYDREAMER, a Writer–Walker framework in which a learned Walker sequentially searches, follows associations, retains useful episodes, and curates an ordered memory context before generation. The Walker is trained on executable trajectories followed by online GRPO with rewards for writing utility, reference recovery, and navigation progress. Across two Writer backbones, Daydreamer produces substantially broader memory discovery and better associative-flow alignment than untrained and retrieval-based baselines, while achieving the highest overall writing scores. The gains are most pronounced in persona grounding and creative transformation, suggesting that effective memory-augmented creative generation depends not only on retrieving relevant experiences, but on learning how to navigate and organize them.

1 INTRODUCTION

Where do creative ideas come from? Creative ideas rarely arise from an isolated fact or a single recalled event. Cognitive theories instead describe creativity as drawing on prior experience through retrieval, association, and recombination (Mednick, 1962; Schacter & Addis, 2007; Benedek et al., 2023). Autobiographical memory is similarly constructive and goal-directed: retrieval can proceed from an initial cue through progressively refined associations rather than a single static lookup (Conway & Pleydell-Pearce, 2000). These observations suggest a useful computational perspective for memory-augmented generation: relevant experiences may need to be discovered sequentially, connected through association, and selectively organized before they can be used for generation.

Persona-grounded creative writing provides a natural testbed for studying this process. Writing from personal experience requires more than recovering facts about a person: a useful memory may be only indirectly related to the writing prompt, while one recalled place, object, relationship, or sensation may expose another experience that becomes relevant through association. The resulting memories must then be selectively retained and transformed into new material rather than copied verbatim. Episodic retrieval has been linked to divergent thinking and to the experiential detail produced during creative writing (Madore et al., 2015; Van Genugten, 2022). Rather than assuming that associative navigation is necessary for creative writing, we use this setting to study a more concrete question: *how can memory navigation for generation be supervised, learned, and evaluated?*

Existing benchmarks provide limited support for answering this question. Creative-writing benchmarks primarily evaluate the final artifact, using criteria such as coherence, originality, style, or human preference (Fan et al., 2018; Gómez-Rodríguez & Williams, 2023; Fein et al., 2026). Long-term-memory benchmarks, in contrast, typically evaluate factual recovery, event tracking, or ques-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

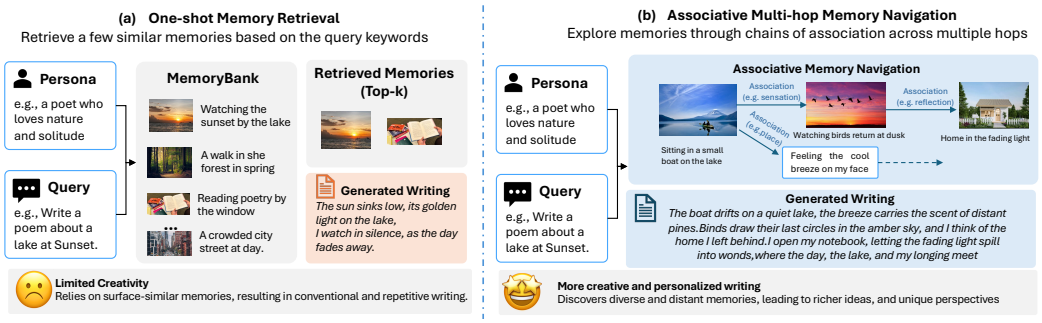

Figure 1: From one-shot memory retrieval to sequential associative navigation for generation.

tion answering over extended histories (Maharana et al., 2024; Wu et al., 2024; Hu et al., 2026). Persona and role-playing benchmarks further measure consistency with persona descriptions or persona knowledge (Wang et al., 2024). These settings evaluate either the generated output or whether relevant information can be recovered, but do not directly supervise or measure the intermediate navigation process shown in Figure 1: which memories are surfaced, which are retained, which associations are followed, and how a compact memory context is constructed for generation.

We introduce DAYDREAMBENCH, a benchmark for studying associative episodic memory navigation in persona-grounded generation. For each author, it connects a biography-grounded persona and a large episodic MemoryBank with creative works. Each work is paired with a reverse-constructed reference flow of plausible experiential memories, embedded among broader biography-derived memories. The resulting benchmark contains 46,959 episodic memories across 201 persona–corpus identities and 343 held-out writing tasks. This design makes it possible to evaluate not only the final generation, but also whether a system discovers, selects, and organizes the experiential material associated with the target work.

This formulation exposes a limitation of one-shot retrieval: memories useful for generation need not be those most similar to the initial prompt. One recalled experience may reveal a person, object, place, or theme that makes a more distant memory relevant. Memory use therefore resembles *associative search* rather than nearest-neighbor lookup. To address this limitation, we propose DAYDREAMER, a Writer–Walker framework that separates generation from memory navigation. Given a persona and writing task, the Writer specifies the experiential material it needs, while a learned *Walker* navigates the MemoryBank through four operations: querying, keeping useful episodes, following visible concepts, and selecting an ordered subset for generation. Each decision is conditioned on memories revealed by previous actions, making the navigation process explicit and directly evaluable. We train the Walker with supervised fine-tuning followed by online GRPO, using rewards that capture writing utility, reference-memory recovery, and navigation progress. Across different Writer backbones, the trained Walker consistently improves memory navigation and downstream generation over untrained and retrieval-based baselines. These gains also transfer downstream, with the clearest improvements in persona grounding and creative transformation rather than generic task fulfillment or coherence. Together, these results support our central hypothesis: *for persona-grounded creative generation, the key challenge is not merely retrieving relevant memories, but navigating, selecting, and organizing episodic memories through association.*

We therefore use persona-grounded creative writing as a testbed for studying memory-augmented generation as a sequential associative-search problem. We contribute (1) DAYDREAMBENCH, which links persona-specific MemoryBanks, creative works, and executable reference navigation flows; (2) DAYDREAMER, a Writer–Walker framework for learned associative memory navigation; and (3) a two-stage SFT–GRPO training procedure that learns navigation policies transferable across held-out personas and Writer models.

2 RELATED WORK

Memory and Creative Cognition Cognitive theories have long connected creativity with associative memory. Mednick's associative theory views creative thought as combining relatively remote

associative elements (Mednick, 1962), while spreading-activation accounts describe memory access as propagation through networks of related concepts (Collins & Loftus, 1975). Empirical studies further associate creativity with flexible semantic networks and stronger associative retrieval and combination abilities (Benedek et al., 2012; Kenett et al., 2014). Beyond semantic association, episodic memory provides concrete experiential details that can be flexibly recombined into novel situations (Schacter & Addis, 2007). Inducing more specific episodic retrieval improves divergent thinking (Madore et al., 2015) and increases episodic detail in creative writing (van Genugten et al., 2022); recent reviews accordingly characterize creative ideation as drawing jointly on semantic and episodic memory (Benedek et al., 2023). Autobiographical-memory research further emphasizes that retrieval is constructive and goal-directed rather than a single lookup (Conway & Pleydell-Pearce, 2000). These findings motivate our formulation of creative recall as sequential navigation through episodic memories and their associative links.

Long-Term Memory and Learned Memory Agents LLM agents increasingly augment generation with external long-term memory. Generative Agents, MemoryBank, and MemGPT introduce persistent experience storage and retrieval mechanisms (Park et al., 2023; Zhong et al., 2024; Packer et al., 2023), while HippoRAG and A-MEM organize memory through graph-based or dynamically evolving associations (Gutiérrez et al., 2024; Xu et al., 2026). Benchmarks such as LoCoMo, LongMemEval, and MemoryAgentBench evaluate long-horizon retrieval, temporal reasoning, updating, and selective forgetting (Maharana et al., 2024; Wu et al., 2024; Hu et al., 2026). Recent work further learns memory operations rather than prescribing them. Memory-R1, Agentic Memory, and Fine-Mem optimize memory management with reinforcement learning and fine-grained feedback (Yan et al., 2026; Yu et al., 2026; Ma et al., 2026). Their primary objective, however, is to store, update, or retrieve information for factual and task-oriented utility. In contrast, our MemoryBank is fixed: the learning problem is how to *navigate* existing episodic experiences for a creative goal. The Walker progressively searches, follows associations, retains candidate memories, and curates an ordered subset for downstream writing, making memory use a sequential associative-search problem rather than primarily a memory-management problem.

Creative Writing and Persona Grounding Creative-writing research has largely focused on generating and evaluating the final artifact. WritingPrompts established large-scale prompt-conditioned story generation (Fan et al., 2018), while subsequent benchmarks evaluate dimensions such as originality, coherence, style, and literary quality (Gómez-Rodríguez & Williams, 2023; Wang et al., 2025; Fein et al., 2026). Recent work also studies reference-based creativity evaluation and associative creativity in LLMs (Li et al., 2025; Qiu & Hu, 2025). Related work on persona-conditioned generation and role-playing studies whether models can faithfully reproduce character knowledge, style, and behavior (Chandu et al., 2019; Wang et al., 2024). More recent systems explicitly incorporate memory to improve persona consistency (Cheng et al., 2026; Wang et al., 2026). Our focus is complementary: we treat memory not only as information to be faithfully recalled, but as *creative material* to be discovered, associated, selected, and transformed. DAYDREAMBENCH therefore connects persona-specific episodic memories with creative works and executable navigation references, enabling joint evaluation of memory discovery, associative progression, and downstream writing.

3 DAYDREAM DATASET AND BENCHMARK CONSTRUCTION

We construct the Daydream dataset to provide both persona-specific memory environments and supervision for learning active memory navigation. As illustrated in Figure 2, each author is associated with a biography-grounded persona, a structured episodic MemoryBank, and a collection of creative works. The MemoryBank contains two complementary sources of episodic memory. *Work-derived memories* are reverse-constructed from each poem or short story into a small acyclic flow of plausible experiential scenes that could provide material for the work. *Biography-derived memories* are generated from the author's documented life context, including both biography-anchored episodes and low-salience everyday scenes that form a broader personal memory environment. The work-derived flows are embedded within these larger person-specific MemoryBanks, requiring a system to discover relevant experiential material among many other memories.

Each work-derived flow is then replayed through the same MemoryBank interface used at inference time and serialized into an executable Walker trajectory. The resulting instances pair a writing

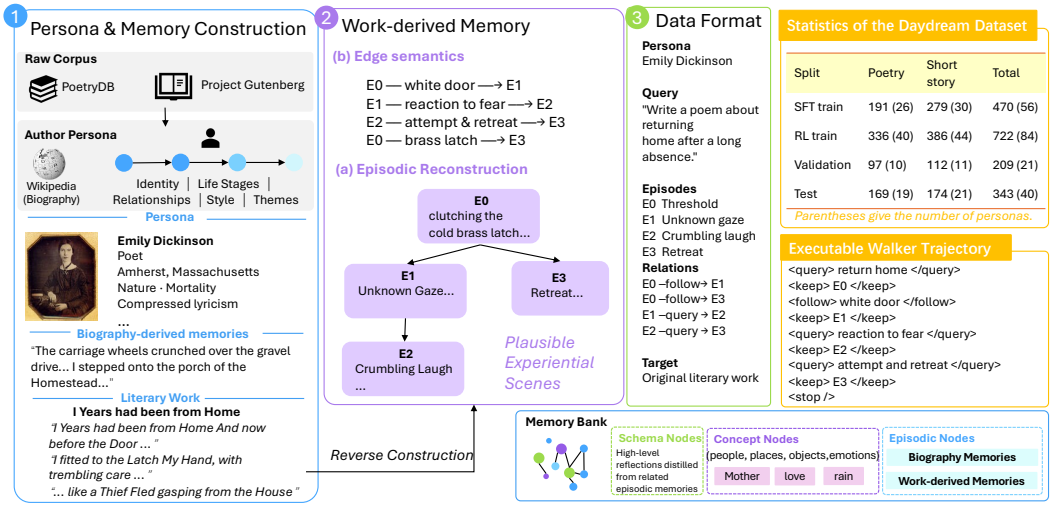

Figure 2: Construction of the Daydream dataset and executable memory-navigation trajectories.

query with a persona-specific MemoryBank, an executable navigation trajectory, and the original work as the writing target. The full Daydream dataset is divided into training, validation, and test splits at the persona level. The person-disjoint test split constitutes DAYDREAMBENCH, our benchmark for evaluating associative memory navigation and downstream generation. Appendix A reports the generation models, prompt templates, decoding settings, and deterministic filters. A complete DAYDREAMBENCH instance, including the public task fields and private reference supervision, is provided in Appendix E.

3.1 SOURCE CORPORA

We collect English poetry from PoetryDB [1] and short stories from Project Gutenberg through Gutendex [2]. We retain authors with at least two usable works and short-story units between 1,000 and 20,000 words. Wikipedia introductions retrieved through the MediaWiki API [3] provide biographical evidence; name, page-title, and occupation checks remove disambiguation pages and mismatched people. The resulting corpus contains 201 persona–corpus identities and 2,351 works: 95 PoetryDB identities with 1,037 poems and 115 Gutenberg identities with 1,314 short stories. Details about source parsing and text cleaning can be found in Appendix A.1.

3.2 PERSONA AND MEMORY CONSTRUCTION

The biography model creates a structured persona card with evidence-linked facts, documented relationships, chronological life phases, and a restrained characterization. Based on these, we construct two complementary sources of episodic memory: **work-derived memories** and **biography-derived memories**. For each literary work, the model reverse-constructs a connected, acyclic flow of three to six first-person scenes that could plausibly supply experiential material for that work. These *work-derived memories* yield 2,319 valid flows and 8,539 episodic memories. Separately, we construct *biography-derived memories* from the persona's documented life context. A two-pass generator first creates 22,997 scenes anchored to biographical facts, relationships, or life phases, and then adds 15,423 low-salience texture scenes that provide ordinary sensory and situational details. Importantly, all generated scenes are synthetic writing resources, even when grounded in a documented biographical fact, their first-person details are not asserted as historical events. We combine both sources into person-specific MemoryBanks containing 46,959 episodic memories and 5,353 schema nodes. Appendix A.3 gives the prompts and Appendix A.4 gives the validation rules.

[1] https://poetrydb.org/index.html
[2] https://www.gutenberg.org/
[3] https://github.com/wikimedia/mediawiki

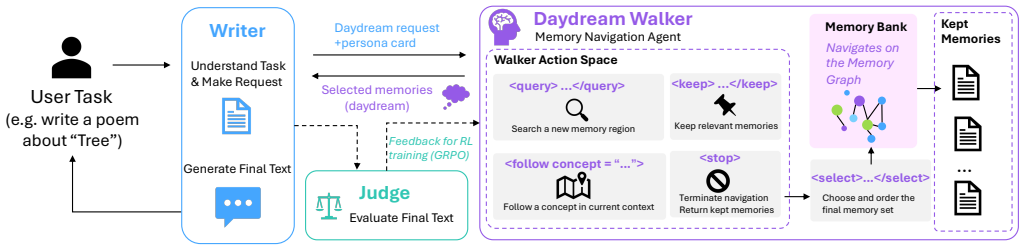

Figure 3: Overview of Daydream Walker. Given the user query, the Writer formulates a daydream request conditioned on the persona card. The Daydream Walker then navigates the Memory Bank through four executable actions (query, keep, follow, and stop) and select the final memory set as a daydream context for generation. The Writer uses this context to produce the final text. During training, a Judge evaluates the generated text and provides reward signals for optimizing the Walker with GRPO

3.3 Executable Trajectories and Splits

The work-derived flows specify which memories should be connected, but they do not directly provide executable supervision for the Walker. To bridge this gap, we replay each flow through the same MemoryBank interface and action constraints used at inference time. The replayer first aligns local flow nodes to global MemoryBank identifiers and then attempts to realize the intended associative path step by step. A memory can be kept only after it has been exposed by retrieval, and a concept can be followed only after it becomes visible in the current observation. We reject traces with failed alignment, invalid actions, empty selections, unclean termination, or surface recall below 0.5. The pipeline produces 1,753 replay attempts, 1,747 traces before source-work deduplication, and 1,744 accepted trajectories. Prompt construction and filtering are specified in Appendix A.5. Training authors are further divided between SFT and RL. Table in the upper-right part of the Figure 2 summarizes the resulting split. The test split forms DaydreamBench: 343 tasks across 40 nominally held-out personas. Each instance exposes a persona p, writing task x, fixed memory request r, and MemoryBank $\mathcal{B}_p$, while retaining the replayed work-flow memories G for evaluation.

4 Daydreamer

4.1 Overview

As illustrated in Figure 3, Daydreamer separates creative writing from memory search. Given a persona p and a writing task x, the Writer first describes the kind of personal material it needs. The Walker uses this request to explore the persona's MemoryBank $\mathcal{B}_p$ and returns a small, ordered set of memories K. A deterministic renderer formats these memories as a daydream, after which the Writer W_{write} produces the final text:

$$K = \pi_\theta(p, x, r, \mathcal{B}_p), \qquad y = W_{\text{write}}\big(p, x, T(K)\big). \tag{1}$$

Here, π_θ is the Walker policy, and T introduces no additional generation. The Writer therefore sees only the memories explicitly selected by the Walker. This decomposition makes the source of each recalled detail visible in the navigation trace.

4.2 Navigating the MemoryBank

The MemoryBank contains episodic memories and concept links over recurring people, objects, places, and themes. The Walker never reads the full bank. At each step, it sees only a small candidate set returned by the previous action and maintains a persistent *reservoir* B_t of memories that it has chosen to retain. It navigates with four actions:

- `query` retrieves memories relevant to the task, request, and current search phrase;
- `keep` copies useful visible candidates into B_t and reveals nearby memories through associative spread;

- `follow` moves from a visible concept to other memories connected to it;
- `stop` ends navigation and opens the terminal selection stage.

The candidate set is transient, whereas the reservoir persists across steps. After each `keep`, the environment uses the updated reservoir as a set of associative seeds and returns a new candidate set. The Walker can then continue with another query, follow a visible concept, or keep newly discovered memories. Actions may refer only to memories or concepts in the current observation, preventing the policy from accessing unseen parts of the bank. After `stop`, the Walker issues `select` to choose and order the final set $K \subseteq B_t$. Only memories in the reservoir can be selected, and only K is rendered for the Writer. Query use dense similarity, whereas `follow` uses one-hop concept links. Appendix B.1 gives the exact scoring functions, action grammar, and interaction budgets.

4.3 Training the Walker

We train the Walker in two stages. Supervised fine-tuning on the executable trajectories from Section 3 first teaches the MemoryBank interaction protocol. We then optimize complete navigation trajectories with online GRPO using the reward defined in Section 4.3.1. Both stages optimize only Walker-generated action tokens, while GRPO additionally regularizes the policy toward the SFT initialization. Full optimization details are provided in Appendix B.2.

4.3.1 Reward design

We design the trajectory reward around three questions: (1) Do the selected memories improve the writing? (2) Does the final selection recover the intended memory flow? (3) Does each action make useful progress toward that selection? We use the reference set G from Section 3, together with the reservoir B_t and final ordered selection K from Section 4. The total return is

$$R(\tau) = w_{\text{out}} R_{\text{out}} + R_{\text{nav}}(K, G) + w_{\text{proc}} R_{\text{proc}}(\tau) + R_{\text{rule}}(\tau). \tag{2}$$

The first three terms correspond directly to the outcome, navigation, and process ablations. The final term is a protocol safeguard shared by these variants; it prevents an incomplete or malformed trajectory from receiving a high return.

Outcome reward. For the same persona and task, a frozen Writer produces y_K from the selected memories and $y_\varnothing$ without retrieved memories. A frozen scalar training Judge J evaluates the overall quality of each response conditioned on the writing task and persona, and returns a score in $[0, 1]$. (Detailes can be found in Appendix C). We reward only a positive counterfactual improvement:

$$R_{\text{out}} = \max\{0, J(y_K) - J(y_\varnothing)\}. \tag{3}$$

The baseline score is precomputed for each training instance, while y_K is generated from the current rollout. The Walker receives outcome credit only when its memories improve the judged response.

Navigation reward. The terminal navigation reward evaluates what the Walker actually delivers:

$$R_{\text{nav}}(K, G) = w_{\text{rec}} \,\text{Rec}(K, G) + w_{F_1} F_1(K, G) + w_{\text{flow}} \,\text{Flow}(K, G). \tag{4}$$

Here Keep Recall $\text{Rec}(K, G) = |K \cap G|/|G|$ measures reference coverage, while Selected F1 $F_1(K, G) = 2|K \cap G|/(|K| + |G|)$ penalizes an unnecessarily large final set. $\text{Flow}(K, G)$ combines coverage of the reference-flow nodes with the fraction of reference edges whose endpoints occur in the correct order. Together, these terms favor a compact selection that recovers both the content and progression of the construction-derived flow.

Process reward. We additionally score how the navigation state evolves over the trajectory. Let f_{rec}, f_{F1}, and f_{flow} denote reference recall, selected-set F1, and flow alignment computed on the current reservoir (and on the final selection at the terminal step). The process reward is

$$R_{\text{proc}}(\tau) = \sum_t \left(\beta_{\text{rec}} \Delta_t f_{\text{rec}} + \beta_{\text{F1}} \Delta_t f_{\text{F1}} + \beta_{\text{flow}} \Delta_t f_{\text{flow}} + r_t^{\text{event}} \right),$$

where $\Delta_t f = f_t - f_{t-1}$. These metric-difference terms are summed without discounting or clipping, and therefore telescope to the change between the initial and final navigation states. This differs from

R_{nav}, which directly scores the final delivered set K with separate terminal weights. Path-dependent shaping is provided by r_t^{event}, which rewards productive `follow` transitions and penalizes invalid actions, empty termination, oversized selections, and budget exhaustion.

Protocol reward. R_{rule} combines a bounded format score with a terminal gate: a trajectory must begin with a query, use legal actions, stop cleanly, and return a non-empty final selection. We keep this safeguard outside the three substantive rewards so that the ablations do not conflate memory quality with basic protocol compliance.

The reference set G and its flow edges are used only by the reward function; neither is included in the Walker observation. Because G is a construction-derived reference rather than an exhaustive set of useful writing material, the reward does not directly label non-reference memories as incorrect. Appendix B.3 reports all coefficients and low-level event definitions.

5 Experiments

5.1 Experimental setup

We evaluate on all 343 tasks in the DaydreamBench. We report two Writer backbones, Qwen3.6-27B (Qwen Team, 2026b) and GPT-6 Luna [4], to test whether the retrieved memories remain useful across generators. All Walker policies and shared selection modules are based on Qwen3.5-2B Qwen Team (2026a). Qwen3-Embedding-0.6B Zhang et al. (2025) encodes memories, tasks, requests, and queries; cosine similarity between its normalized representations is used for Walker retrieval, associative spread, and the dense RAG baselines. Navigation methods have at most 13 actions, a 16-memory reservoir, and an 8-memory final selection budget. Each rendered memory is truncated to 400 characters, and the complete selected-memory block is capped at 3,200 characters. The Writer may generate at most 2,048 tokens. Appendix B.1 gives the detailed retrieval settings, and Appendix B.2 gives the training settings.

Baselines. We compare Daydreamer with five arms. *Daydreamer (base)* uses the Qwen3.5-2B Walker initialization without SFT or GRPO. *Persona-only* gives the Writer the persona and task without episodic memories. *RAG* retrieves episodes by dense similarity to the task and memory request, then applies the shared base-Walker selector. *RAG + Reranking* reranks the same 16 candidates before selection. Following HippoRAG (Gutiérrez et al., 2024), the *Graph-PPR* baseline performs associative retrieval with Personalized PageRank (PPR). It initializes the personalization distribution using the episodes and concepts most relevant to the Writer's request, and propagates their relevance over the frozen episode–concept bipartite graph. The highest-ranked episodes are then passed to the selector. *Graph-PPR* propagates relevance over the frozen episode–concept graph and then applies the same selector. Daydreamer and the four retrieval arms share the same final memory budget, while Persona-only isolates the contribution of episodic memory recall.

5.2 Evaluation metrics

Navigation. Let G be the construction-derived reference set, S the union of all memories surfaced during navigation, and K the ordered set delivered to the Writer. We measure discovery with surface recall and final curation with keep recall, and selected F1:

$$\text{SurfRec} = \frac{|S \cap G|}{|G|}, \text{KeepRec} = \frac{|K \cap G|}{|G|}, F_1 = \frac{2|K \cap G|}{|K| + |G|}. \tag{5}$$

These are the same quantities used to construct the navigation and process rewards in Section 4.3.1; here they are computed once on the completed trajectory for evaluation.

For sequential Walkers, we also measure whether the delivered memories preserve the intended associative progression. Let E_G be the directed edges of the reference flow, $c = |K \cap V_G|/|V_G|$ its node coverage, e the fraction of reference edges whose two endpoints occur in K, and o the fraction of recovered edges that appear in the correct order. We define $\text{Flow} = \frac{1}{2}c + \frac{1}{2}eo$. Flow alignment is undefined for one-shot retrieval baselines, since they do not produce a navigation trajectory.

[4] https://openai.com/index/introducing-gpt-6-sol-and-luna/

Table 1: Main results on DAYDREAMBENCH. Navigation metrics are macro-persona means in $[0, 1]$; Judge overall is reported as a percentage.

Writer	Method	SurfRec	KeepRec	F1	Flow align.	Judge overall (%)
GPT-6 Luna	Daydreamer	**0.522**	**0.320**	**0.226**	**0.194**	**82.89**
	Daydreamer (base)	0.391	0.309	0.197	0.182	82.08
	Persona-only	–	–	–	–	81.26
	RAG	0.390	0.261	0.194	–	82.07
	RAG + Reranking	0.397	0.259	0.198	–	82.39
	Graph-PPR	0.387	0.252	0.182	–	81.37
Qwen3.6-27B	Daydreamer	**0.508**	**0.326**	**0.225**	**0.192**	**77.75**
	Daydreamer (base)	0.395	0.318	0.201	0.182	76.89
	Persona-only	–	–	–	–	72.32
	RAG	0.397	0.255	0.201	–	76.52
	RAG + Reranking	0.397	0.255	0.201	–	77.02
	Graph-PPR	0.392	0.270	0.197	–	75.57

LLM-as-Judge Evaluation. We evaluate writing with four LLM Judges: DeepSeek-V4-Flash (DeepSeek-AI, 2026), GLM-5.3 (GLM-5-Team et al., 2026), GPT-6 Luna, and Qwen3.6-27B. Each Judge independently assigns five scores in $[0, 1]$. *Task fulfillment* captures compliance with the requested content, form, and constraints. *Persona grounding* assesses whether persona-specific experiences and perspectives are integrated naturally and consistently. *Creative transformation* rewards grounded recombination and adaptation of memories while penalizing copying or unsupported persona details. *Coherence* measures clarity, continuity, and internal consistency. The Judge separately supplies an *overall* score for the response as a whole. Appendix C reproduces the complete Judge prompt and scoring anchors.

5.3 MAIN RESULTS

Table 1 shows that training improves both memory discovery and final curation. Appendix D provides a qualitative case study illustrating how DAYDREAMER navigates and curates memories compared with retrieval-based baselines. With Qwen3.6-27B as Writer, Daydreamer raises surface recall from 0.395 to 0.508 and selected F1 from 0.201 to 0.225 over Daydreamer (base). Relative to the strongest non-sequential baseline, it gains 0.111 surface recall and 0.024 selected F1. Keep recall improves more modestly, from 0.318 to 0.326, indicating that the largest gain occurs in exposing useful memories before final selection. Flow alignment also increases from 0.182 to 0.192.

The GPT-6 Luna setting shows the same pattern. Daydreamer obtains 0.522 surface recall and 0.226 selected F1, compared with 0.397 and 0.198 for the strongest competing values. It also improves flow alignment from 0.182 to 0.194 over Daydreamer (base). The consistent navigation gains across the two Writer settings support the intended role of the Walker: *learned associative search exposes a broader set of relevant memories and curates them more accurately than one-shot retrieval.*

Writing quality. Episodic memories primarily improve the dimensions most related to personalized creative writing. With Qwen3.6-27B as the Writer, Daydreamer raises persona grounding from 58.45% for Persona-only to 70.22% (+11.77 points) and creative transformation from 63.73% to 70.98% (+7.25 points), while task fulfillment changes only modestly from 89.05% to 89.97%. Similar trends hold for the other memory-augmented baselines (Figure 4), suggesting that episodic memories mainly provide experiential material for grounded recombination rather than improving generic instruction following or fluency.

Learned navigation provides a consistent additional benefit. Daydreamer achieves the highest overall Judge score with Qwen3.6-27B (77.75%), outperforming RAG + Reranking by 0.73 points, and also obtains the strongest persona grounding and creative transformation scores, although RAG + Reranking is slightly better on coherence. The same pattern transfers to GPT-6 Luna, where Daydreamer reaches 82.89% versus 82.39% for RAG + Reranking. These results support a central premise of Daydreamer: persona-grounded creative writing benefits not only from having access to episodic memories, but also from navigating and curating those memories before generation.

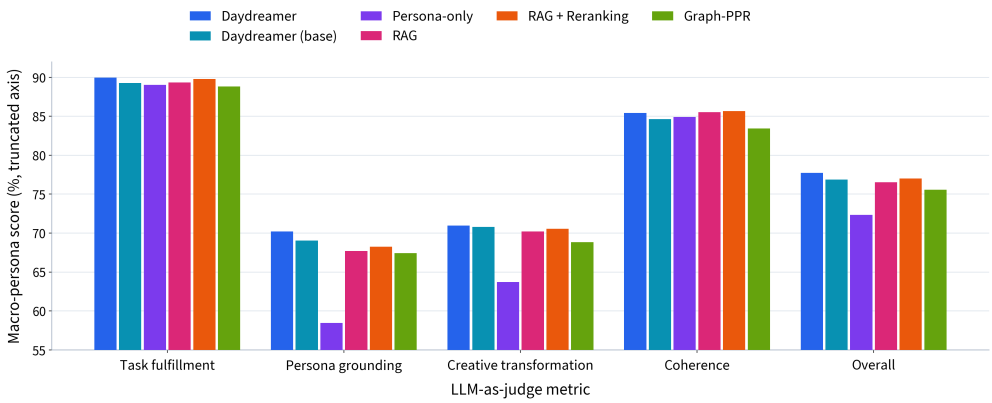

Figure 4: Reference-conditioned LLM-as-Judge scores (%) with Qwen3.6-27B as Writer. Bars show macro-persona means over 343 valid examples, averaged across four Judges. The vertical axis is truncated to expose differences between methods.

Table 2: Ablation studies results

Reward	Surface R.	Keep R.	Selected F1	Flow align.	Judge overall (%)
Full	0.508	**0.326**	0.225	**0.192**	**77.75**
No outcome	0.509	0.318	0.246	0.181	77.41
No navigation	0.475	0.257	**0.249**	0.145	77.09
No process	**0.516**	0.318	0.235	0.186	77.10
No GRPO	0.3941	0.318	0.201	0.183	75.83

Ablation Studies Table 2 shows that the full reward achieves the best overall balance across navigation and generation quality. Although individual ablations improve isolated metrics such as surface recall or selected F1, these gains do not translate into better final outputs. For example, removing the navigation reward increases selected F1 to 0.249, but substantially reduces surface recall, keep recall, flow alignment, and the Judge score. Similarly, removing process shaping yields the highest surface recall (0.516) but lower generation quality. In contrast, the full reward attains the highest keep recall (0.326), flow alignment (0.192), and Judge score (77.75%). Removing GRPO causes the largest drop in surface recall (0.508 to 0.394) and lowers the Judge score from 77.75 to 75.83, indicating that reinforcement learning is particularly important for learning effective memory discovery beyond the supervised initialization. These results show that higher selected F1 does not necessarily yield a higher writing score under our evaluation.

6 Conclusion

We studied persona-grounded creative writing as a testbed for training and evaluating sequential associative memory navigation. We introduced DaydreamBench, which provides executable reference flows for supervising and measuring navigation, and Daydreamer, a Writer–Walker framework that separates memory navigation from generation. The Walker learns to search, follow associations, retain useful experiences, and organize a compact memory context.

Our results show that the clearest effect of training is broader and more effective memory discovery. Compared with one-shot retrieval and an untrained Walker, Daydreamer consistently surfaces more reference memories and better preserves their associative progression. These navigation gains translate into stronger downstream writing, especially in persona grounding and creative transformation rather than generic task fulfillment or coherence. Together, these findings suggest that memory-augmented generation should be viewed not merely as retrieving relevant context, but as learning how to traverse and organize experience for a generation goal. DaydreamBench provides a controlled setting for studying this process, while future work should extend it to naturally accumulated or human-authored memories, broader creative domains.

AI USE STATEMENT

LLMs were used both as components of the proposed research pipeline and as auxiliary tools during manuscript preparation. In the experiments, LLMs were used for structured persona and memory construction, memory navigation, creative generation, and model-based evaluation, as described in the main text and appendix. During manuscript preparation, language models were additionally used for linguistic refinement, including grammar correction, clarity improvement, and wording suggestions. All research questions, methodological choices, experimental designs, analyses, and conclusions were determined and verified by the authors.

ETHICS STATEMENT

This work studies memory-augmented creative generation using literary authors and their works. The episodic memories constructed in our dataset are synthetic reconstructions intended to provide plausible experiential context for modeling and should not be interpreted as factual records of the authors' lives. Likewise, model-generated texts are synthetic outputs and should not be attributed to the corresponding authors as authentic works. We use literary and biographical materials only as research resources and preserve source provenance throughout dataset construction. The study does not involve human participants or real-world actions.

REPRODUCIBILITY STATEMENT

The dataset construction procedure, memory representation, Daydream Walker action space, training objective, and evaluation protocol are described in the main text. The appendix provides additional implementation details, including generation models and decoding settings, prompts used for structured data construction, training hyperparameters, and evaluation configurations. We will release the constructed dataset, executable memory flows, training and evaluation code, and configuration files to facilitate reproduction of the reported experiments. Because some dataset-generation models are accessed through hosted services whose immutable provider revisions are not preserved, exact bitwise regeneration of all synthetic artifacts is not guaranteed.

REFERENCES

Mathias Benedek, Tanja Könen, and Aljoscha C Neubauer. Associative abilities underlying creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 6(3):273, 2012.

Mathias Benedek, Roger E Beaty, Daniel L Schacter, and Yoed N Kenett. The role of memory in creative ideation. *Nature Reviews Psychology*, 2(4):246–257, 2023.

Khyathi Chandu, Shrimai Prabhumoye, Ruslan Salakhutdinov, and Alan W Black. "my way of telling a story": Persona based grounded story generation. In *Proceedings of the Second Workshop on Storytelling*, pp. 11–21, 2019.

Xilong Cheng, Yunxiao Qin, Yuting Tan, Zhengnan Li, Ye Wang, Hongjiang Xiao, and Yuan Zhang. Psymem: Fine-grained psychological alignment and explicit memory control for advanced role-playing llms. *Transactions of the Association for Computational Linguistics*, 14:510–529, 2026.

Allan M Collins and Elizabeth F Loftus. A spreading-activation theory of semantic processing. *Psychological review*, 82(6):407, 1975.

Martin A Conway and Christopher W Pleydell-Pearce. The construction of autobiographical memories in the self-memory system. *Psychological review*, 107(2):261, 2000.

DeepSeek-AI. Deepseek-v4: Towards highly efficient million-token context intelligence, 2026.

Angela Fan, Mike Lewis, and Yann Dauphin. Hierarchical neural story generation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 889–898, 2018.

Daniel Fein, Sebastian Russo, Violet Xiang, Kabir Jolly, Rafael Rafailov, and Nick Haber. Litbench: A benchmark and dataset for reliable evaluation of creative writing. In *Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 7740–7755, 2026.

GLM-5-Team, :, Aohan Zeng, Xin Lv, Zhenyu Hou, Zhengxiao Du, Qinkai Zheng, Bin Chen, Da Yin, Chendi Ge, Chenghua Huang, Chengxing Xie, Chenzheng Zhu, Congfeng Yin, Cunxiang Wang, Gengzheng Pan, Hao Zeng, Haoke Zhang, Haoran Wang, Huilong Chen, Jiajie Zhang, Jian Jiao, Jiaqi Guo, Jingsen Wang, Jingzhao Du, Jinzhu Wu, Kedong Wang, Lei Li, Lin Fan, Lucen Zhong, Mingdao Liu, Mingming Zhao, Pengfan Du, Qian Dong, Rui Lu, Shuang-Li, Shulin Cao, Song Liu, Ting Jiang, Xiaodong Chen, Xiaohan Zhang, Xuancheng Huang, Xuezhen Dong, Yabo Xu, Yao Wei, Yifan An, Yilin Niu, Yitong Zhu, Yuanhao Wen, Yukuo Cen, Yushi Bai, Zhongpei Qiao, Zihan Wang, Zikang Wang, Zilin Zhu, Ziqiang Liu, Zixuan Li, Bojie Wang, Bosi Wen, Can Huang, Changpeng Cai, Chao Yu, Chen Li, Chengwei Hu, Chenhui Zhang, Dan Zhang, Daoyan Lin, Dayong Yang, Di Wang, Ding Ai, Erle Zhu, Fangzhou Yi, Feiyu Chen, Guohong Wen, Hailong Sun, Haisha Zhao, Haiyi Hu, Hanchen Zhang, Hanrui Liu, Hanyu Zhang, Hao Peng, Hao Tai, Haobo Zhang, He Liu, Hongwei Wang, Hongxi Yan, Hongyu Ge, Huan Liu, Huanpeng Chu, Jia'ni Zhao, Jiachen Wang, Jiajing Zhao, Jiamin Ren, Jiapeng Wang, Jiaxin Zhang, Jiayi Gui, Jiayue Zhao, Jijie Li, Jing An, Jing Li, Jingwei Yuan, Jinhua Du, Jinxin Liu, Junkai Zhi, Junwen Duan, Kaiyue Zhou, Kangjian Wei, Ke Wang, Keyun Luo, Laiqiang Zhang, Leigang Sha, Liang Xu, Lindong Wu, Lintao Ding, Lu Chen, Minghao Li, Nianyi Lin, Pan Ta, Qiang Zou, Rongjun Song, Ruiqi Yang, Shangqing Tu, Shangtong Yang, Shaoxiang Wu, Shengyan Zhang, Shijie Li, Shuang Li, Shuyi Fan, Wei Qin, Wei Tian, Weining Zhang, Wenbo Yu, Wenjie Liang, Xiang Kuang, Xiangmeng Cheng, Xiangyang Li, Xiaoquan Yan, Xiaowei Hu, Xiaoying Ling, Xing Fan, Xingye Xia, Xinyuan Zhang, Xinze Zhang, Xirui Pan, Xu Zou, Xunkai Zhang, Yadi Liu, Yandong Wu, Yanfu Li, Yidong Wang, Yifan Zhu, Yijun Tan, Yilin Zhou, Yiming Pan, Ying Zhang, Yinpei Su, Yipeng Geng, Yong Yan, Yonglin Tan, Yuean Bi, Yuhan Shen, Yuhao Yang, Yujiang Li, Yunan Liu, Yunqing Wang, Yuntao Li, Yurong Wu, Yutao Zhang, Yuxi Duan, Yuxuan Zhang, Zezhen Liu, Zhengtao Jiang, Zhenhe Yan, Zheyu Zhang, Zhixiang Wei, Zhuo Chen, Zhuoer Feng, Zijun Yao, Ziwei Chai, Ziyuan Wang, Zuzhou Zhang, Bin Xu, Minlie Huang, Hongning Wang, Juanzi Li, Yuxiao Dong, and Jie Tang. Glm-5: from vibe coding to agentic engineering, 2026. URL https://arxiv.org/abs/2602.15763.

Carlos Gómez-Rodríguez and Paul Williams. A confederacy of models: A comprehensive evaluation of llms on creative writing. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pp. 14504–14528, 2023.

Bernal J Gutiérrez, Yiheng Shu, Yu Gu, Michihiro Yasunaga, and Yu Su. Hipporag: Neurobiologically inspired long-term memory for large language models. *Advances in neural information processing systems*, 37:59532–59569, 2024.

Yuanzhe Hu, Yu Wang, and Julian McAuley. Evaluating memory in llm agents via incremental multi-turn interactions. In *International Conference on Learning Representations*, volume 2026, pp. 156259–156291, 2026.

Yoed N Kenett, David Anaki, and Miriam Faust. Investigating the structure of semantic networks in low and high creative persons. *Frontiers in human neuroscience*, 8:407, 2014.

Ruizhe Li, Chiwei Zhu, Benfeng Xu, Xiaorui Wang, and Zhendong Mao. Automated creativity evaluation for large language models: A reference-based approach. In *EMNLP (Findings)*, pp. 21475–21488, 2025.

Weitao Ma, Xiaocheng Feng, Lei Huang, Xiachong Feng, Zhanyu Ma, Jun Xu, Jiuchong Gao, Jinghua Hao, Renqing He, and Bing Qin. Fine-mem: Fine-grained feedback alignment for long-horizon memory management. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 19666–19684, 2026.

Kevin P Madore, Donna Rose Addis, and Daniel L Schacter. Creativity and memory: Effects of an episodic-specificity induction on divergent thinking. *Psychological science*, 26(9):1461–1468, 2015.

Adyasha Maharana, Dong-Ho Lee, Sergey Tulyakov, Mohit Bansal, Francesco Barbieri, and Yuwei Fang. Evaluating very long-term conversational memory of llm agents. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 13851–13870, 2024.

Sarnoff Mednick. The associative basis of the creative process. *Psychological review*, 69(3):220, 1962.

Charles Packer, Sarah Wooders, Kevin Lin, Vivian Fang, Shishir G Patil, Ion Stoica, and Joseph E Gonzalez. Memgpt: Towards llms as operating systems. *arXiv preprint arXiv:2310.08560*, 2023.

Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*, pp. 1–22, 2023.

Ziliang Qiu and Renfen Hu. Deep associations, high creativity: A simple yet effective metric for evaluating large language models. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 10870–10883, 2025.

Qwen Team. Qwen3.5: Towards native multimodal agents, February 2026a. URL `https://qwen.ai/blog?id=qwen3.5`.

Qwen Team. Qwen3.6-27B: Flagship-level coding in a 27B dense model, April 2026b. URL `https://qwen.ai/blog?id=qwen3.6-27b`.

Daniel L Schacter and Donna Rose Addis. On the constructive episodic simulation of past and future events. *Behavioral and Brain Sciences*, 30(3):331–332, 2007.

Ruben Van Genugten. *Episodic retrieval: functions and measurement*. Harvard University, 2022.

Ruben DI van Genugten, Roger E Beaty, Kevin P Madore, and Daniel L Schacter. Does episodic retrieval contribute to creative writing? an exploratory study. *Creativity Research Journal*, 34(2): 145–158, 2022.

Kai Wang, Haoyang You, Yang Zhang, and Zhongjie Wang. Memory-driven role-playing: Evaluation and enhancement of persona knowledge utilization in llms. In *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 23475–23510, 2026.

Noah Wang, Zy Peng, Haoran Que, Jiaheng Liu, Wangchunshu Zhou, Yuhan Wu, Hongcheng Guo, Ruitong Gan, Zehao Ni, Jian Yang, et al. Rolellm: Benchmarking, eliciting, and enhancing role-playing abilities of large language models. In *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 14743–14777, 2024.

Wenqing Wang, Mingqi Gao, Xinyu Hu, and Xiaojun Wan. Towards a "novel" benchmark: Evaluating literary fiction with large language models. In *Findings of the Association for Computational Linguistics: ACL 2025*, pp. 21648–21673, 2025.

Di Wu, Hongwei Wang, Wenhao Yu, Yuwei Zhang, Kai-Wei Chang, and Dong Yu. Longmemeval: Benchmarking chat assistants on long-term interactive memory. *arXiv preprint arXiv:2410.10813*, 2024.

Wujiang Xu, Zujie Liang, Kai Mei, Hang Gao, Juntao Tan, and Yongfeng Zhang. A-mem: Agentic memory for llm agents. *Advances in Neural Information Processing Systems*, 38:17577–17604, 2026.

Sikuan Yan, Xiufeng Yang, Zuchao Huang, Ercong Nie, Zifeng Ding, Zonggen Li, Xiaowen Ma, Jinhe Bi, Kristian Kersting, Jeff Z Pan, et al. Memory-r1: Enhancing large language model agents to manage and utilize memories via reinforcement learning. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 12805–12825, 2026.

Yi Yu, Liuyi Yao, Yuexiang Xie, Qingquan Tan, Jiaqi Feng, Yaliang Li, and Libing Wu. Agentic memory: Learning unified long-term and short-term memory management for large language model agents. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 21457–21483, 2026.

Yanzhao Zhang, Mingxin Li, Dingkun Long, Xin Zhang, Huan Lin, Baosong Yang, Pengjun Xie, An Yang, Dayiheng Liu, Junyang Lin, Fei Huang, and Jingren Zhou. Qwen3 embedding: Advancing text embedding and reranking through foundation models. *arXiv preprint arXiv:2506.05176*, 2025.

Wanjun Zhong, Lianghong Guo, Qiqi Gao, He Ye, and Yanlin Wang. Memorybank: Enhancing large language models with long-term memory. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 19724–19731, 2024.

A Dataset Construction Details

A.1 Source collection and deterministic cleaning

PoetryDB records are grouped by normalized author and retain the title and full poem text. For Gutenberg, Gutendex supplies English-language book metadata and downloadable plain text or HTML. The collector removes Gutenberg boilerplate and segments collections when explicit story boundaries are available. It then discards story units outside the 1,000–20,000 word interval and authors with fewer than two retained works. Wikipedia candidates are rejected when the normalized author name, returned page title, and occupational description do not support the intended identity. These operations use parsers and rule-based checks.

Long poems are supplied in full when work flows are generated. A short story is represented by a deterministic package comprising an approximately 140-word summary assembled from sampled paragraphs, a five-beat outline, and a final scene capped at 800 words. This bound controls context length while retaining the work's ending pressure. Each retained author, work, and story unit receives a stable identifier and source provenance.

A.2 Generation models and decoding

We use Qwen3.6-27B for all structured dataset generation, with lower temperatures for biography and style extraction ($T = 0.2$) and higher temperatures for work-derived and biography-derived memory construction ($T = 0.8$–1.0) to encourage experiential diversity. Structured generations are requested in JSON format and discarded after the retry limit upon parsing failure. We use Qwen3-Embedding-0.6B to obtain 1,024-dimensional representations and merge concept surface forms whose cosine similarity exceeds 0.86.

A.3 Prompt templates

Tables 3–6 reproduce the complete operative content of every LLM prompt used by the reported construction pipeline. Runtime values are replaced by bracketed placeholders, and long JSON examples are rendered as compact field inventories. The *System* and *User* blocks correspond directly to the two messages sent to the model.

Table 3: Prompts for persona construction. Runtime substitutions are shown in square brackets.

Prompt	Complete text
Biography profile: system	You construct a compact, historically grounded author persona for later memory generation. Use only the supplied Wikipedia biography. Do not infer fictional events from works. Do not invent unsupported named relationships. Return one JSON object with exactly these top-level keys: `identity`, `biographical_facts`, `relationships`, `life_phases`, and `profile`. `identity` contains `name`, `preferred_name`, `age`, primary and secondary `occupation`, current and origin locations, `one_line_self_description`, and `current_year`. Each biography fact contains an identifier, fact, start and end periods, and emotional weight. Each relationship contains an identifier, name, role, birth and death years, active period, interaction mode, and presence after its end. Each life phase contains an identifier, name, age and year ranges, supporting fact identifiers, memory target, and era seriousness. The profile contains a one-to-three-sentence summary and evidence-linked claims with identifiers, kinds, text, biography-fact identifiers, and relationship identifiers. Requirements: keep this compact and structural, not literary; include 12–20 factual biography entries; include 5–7 chronological life phases when the biography supports them; include 2–5 relationships only when explicitly supported by the biography; include 4–8 profile claims rooted in the biography; keep every string terse; keep the one-line self-description short and plain; make the profile summary precise, restrained, and operational for later memory prompts; and make every profile claim cite supporting biography-fact identifiers, with relationship identifiers only when explicit. Return JSON only.

Biography profile: user	`AUTHOR: [author name]`
	`WIKIPEDIA: [Wikipedia URL]`
	`BIOGRAPHY:`
	`[retrieved Wikipedia introduction]`

Style cues: system	You infer only the minimal style cues needed for later memory generation from the supplied writing samples. Return one JSON object containing `core_motifs` and `voice`. The voice object contains a short `register` and `forbidden_modes`. Provide 3–5 core motifs and 2–4 forbidden modes. Keep everything terse. This is prompt conditioning only, not a full psychology profile. Return JSON only.

Style cues: user	`AUTHOR: [author name]`
	`WRITING SAMPLES:`
	For each of at most three works: `TITLE: [title]`, `GENRE: [genre]`, and `EXCERPT: [up to 900 characters]`. Samples are separated by `---` and jointly capped at 2,400 characters.

Table 4: Prompts for work-derived and life-phase memories.

Prompt	Complete text

Work flow: system	You reverse-construct the LIVED MEMORIES behind a SINGLE supplied creative work by an author. Infer a small associative memory graph that could lead a daydreaming mind toward this target only.
	`AUTHOR: [author name]`
	`GENRE: [genre]`
	`VOICE: [style summary]`
	`BIO_PROFILE: [biography-grounded profile]`
	Return JSON only with `flow_type=object_recurrence`, a list of `nodes`, a list of `edges`, and `root_local_id`. Each node contains `local_id`, `time`, `text`, `emotion`, `salience`, `gist`, `entities`, and `tags`. Each edge contains `src`, `dst`, `link_kind`, and `link_label`.
	Rules: use exactly 3–6 nodes, one root, and a connected acyclic graph; make every node first-person, past-tense, concrete, and sensory; create plausible author experiences rather than copied fictional events or summaries of the target; tie every node to this work only and do not mix in other works by the author; keep the lived memories compatible with the biography-derived profile; for an entity edge, repeat the identical concrete entity in both endpoint texts and entity lists; other edges may be thematic query leaps; prefer a branch that later merges; and keep the graph small and reproducible. Return JSON only.

Work flow: user	`WORK: [work title]`
	`WORK_ID: [stable work identifier]`
	`WORK GENRE: [genre]`
	`[full poem, or deterministic story summary + five-beat outline + selected final scene]`
	Return the flow JSON for this work only.

Anchored life memories: system	You generate LIFE MEMORIES for a fictional persona. Output a valid JSON array. Every item must contain: `time`, a short string inside the requested phase's age/year window; `text`, one to four first-person, past-tense, sensory and specific sentences with no moralizing, therapy language, or epigrammatic closure; `emotion`, one lowercase token or short compound; `salience`, a float in $[1, 10]$ matching the selected salience band; the exact `life_phase_id`; `generation_pass=anchored`; `memory_basis`, either `bio_fact_scene` or `profile_claim_scene`; lists of `bio_fact_ids` and `profile_claim_ids`; and, optionally, `relationship_ids`. Hard rules for this pass: (1) Treat the biographical facts as the gravitational center. Each memory must depict a fact, a related scene, or a downstream consequence. Fill `bio_fact_ids` with the exact identifiers used. When a scene is driven more by the compact profile, use `memory_basis=profile_claim_scene` and fill `profile_claim_ids`. (2) Choose salience at or above 6.0; reserve 9.0–10.0 for genuine life pivots warranted by a biography fact. (3) Treat era context and year vibe only as background props. Never name an event, quote a headline or lyric, attribute a feeling to the era, or make the era the memory's subject. (4) Stay strictly within the phase's age/year window and vary the year when the phase spans multiple sampled years. (5) Every named family member, partner, friend, or pet must appear in the relationships block. Unnamed strangers are allowed. (6) Honor the Big Five profile, voice, and forbidden modes. (7) Do not duplicate or paraphrase a prior memory. Return only the JSON array, without a Markdown fence or commentary.

Table 4 continued

Prompt	Complete text

Anchored life memories: user	`PERSONA: [persona summary]`
	`SALIENCE RUBRIC: [configured bands and definitions]`
	`TARGET LIFE PHASE: [phase id, name, age range, year range]`
	`BIOGRAPHICAL FACTS BOUND TO THIS PHASE: [facts and periods]`
	`PROFILE CLAIMS: [claims and identifiers]`
	`RELATIONSHIPS ACTIVE IN THIS PHASE: [documented relationships]`
	`ERA CONTEXT: [persona-specific background props]`
	`YEAR VIBE: [optional background atmosphere]`
	`MEMORIES ALREADY ACCEPTED: [last 16 accepted memories]`
	`PREVIOUS ATTEMPT FEEDBACK: [optional rejection feedback]`
	`TASK: Generate EXACTLY [n] ANCHORED memory objects for this phase. Use salience bands at or above 6.0.`
	`Language: [language]. Return ONLY the JSON array.`

Texture life memories: system	You generate LOW-SALIENCE TEXTURE MEMORIES for a fictional persona: the trivial everyday noise a person retains for no clear reason. Output a valid JSON array. Every item must contain `time`, `text`, `emotion`, `salience`, the exact `life_phase_id`, `generation_pass=texture`, `memory_basis=phase_texture`, `bio_fact_ids=[]`, and optional lists of profile-claim and relationship identifiers. Hard rules for this pass: (1) Salience must be in $[3.5, 6.0]$, with most items in 4.5–5.9. (2) Give the memory no narrative weight, insight, formative significance, or closure. It should be an almost irretrievable ordinary detail, such as a wait at a vet, a wrong-number text, a passing thought while shelving books, or a flickering streetlight. (3) Do not reference biographical facts, deaths, key relationships, career milestones, or core motifs. Remain on the surface of one ordinary moment. (4) Use era context and year vibe only as plausible props; never name an event, quote news, or attribute a feeling to the era. Vary the year across a multi-year phase. (5) A documented person may appear only incidentally and cannot be the subject; do not invent named people. Unnamed strangers are allowed. (6) Stay strictly within the phase's age/year window. (7) Honor a restrained, observation-first voice and end on a small physical detail rather than a conclusion. (8) Do not duplicate a prior memory. Return only the JSON array, without a Markdown fence or commentary.

Texture life memories: user	`PERSONA (voice + Big Five only; do not reference biography facts): [persona summary]`
	`SALIENCE RUBRIC: [configured bands and definitions]`
	`TARGET LIFE PHASE: [phase id, name, age range, year range]`
	`PROFILE CLAIMS: [ordinary-texture claims, if applicable]`
	`NAMED PEOPLE WHO MAY APPEAR INCIDENTALLY: [documented relationships]`
	`ERA CONTEXT: [persona-specific background props]`
	`YEAR VIBE: [optional background atmosphere]`
	`MEMORIES ALREADY ACCEPTED: [last 16 accepted memories]`
	`PREVIOUS ATTEMPT FEEDBACK: [optional rejection feedback]`
	`TASK: Generate EXACTLY [n] LOW-SALIENCE TEXTURE memories for this phase. Salience MUST be in [3.5, 6.0]; do not exceed 6.0. NO bio facts, NO core motifs, NO insight. Just surface noise. Language: [language]. Return ONLY the JSON array.`

Table 5: Prompts for MemoryBank metadata and schema construction.

Prompt	Complete text
Episodic metadata: system	You are a memory-metadata extraction assistant. The fields you produce become associative bridges between memories, so favor terms that could plausibly recur across different memories. Output: `entities`, concrete named people, objects, places, or works using stable canonical forms such as "mother" rather than "my mom"; `themes`, one to four mid-level motifs that are concrete enough to recur, such as "unfinished goodbye," "domestic dread," or "leaving home," while avoiding abstractions such as "sadness" or "life"; `salient`, the one to three entities or themes most central to the memory, copied verbatim from the preceding lists; and `context`, a one-sentence summary of the situation. Example input: "On a rainy night I stared at the blank document, my fingers frozen on the keyboard, unable to write anything." Example output: entities are "the document" and "the keyboard"; themes are "creative paralysis" and "rainy-night solitude"; salient is "creative paralysis"; context is "facing a blank document unable to write on a rainy night." Respond with valid JSON only, with no extra explanation.
Episodic metadata: user	`[one episodic memory text]`
Cluster reflection: system	You are helping a fictional persona reflect on a group of thematically related memories. Output JSON containing `title`, `findings` with at most three items, and `rating` from 0 to 10. The title must name the shared theme in fewer than ten words. Findings must be abstract, high-level insights that hold independently of the specific events. Respond with valid JSON only, with no extra explanation.
Cluster reflection: user	`[texts of the episodic memories in the semantic cluster]`
Concept reflection: system	You are helping a fictional persona reflect on a group of memories that all revolve around one theme word. Theme word: "[concept]." Output JSON containing `title`, `findings` with at most three items, and `rating` from 0 to 10. The title must name, in fewer than ten words, the role or meaning that the concept plays in this person's life. Findings must be abstract, high-level insights describing the pattern or emotion revealed by the recurring concept. Respond with valid JSON only, with no extra explanation.
Concept reflection: user	`[texts of the episodic memories associated with the concept]`

Table 6: Prompts used to construct replay tasks and retrieval queries.

Prompt	Complete text
Writing task: system	Given a [poem or short-story] target by [author], write ONE plain, short imperative that sends the writer toward an ordinary lived situation. Prefer a concrete anchor already supported by the target: food or drink, a household object, clothing, a tool, money, transport, a room, weather, a bodily sensation, or a small social chore. If no object is supported, use a practical action or interaction—waiting, carrying, cleaning, paying, hiding, or being interrupted—rather than an abstract theme. Keep the target's real pressure or relationship, but do not quote it or invent a specific fact. The output must be 8–22 words, direct, and usable as a writing instruction. Abstract feelings such as melancholy, longing, or meaning are welcome when they fit the target, especially when paired with a scene, action, or material detail. Good shape, only when target-grounded: "Begin with milk warming in a bag on the walk home." Also valid: "Stay with the loneliness of carrying something home alone." Output one sentence only.
Writing task: user	`TARGET: [up to 7,000 characters of the work target]`

Table 6 continued

Prompt	Complete text
Memory request: system	You are [author], about to write this [genre]. Write ONE first-person, practical memory-search request in plain language. Ask for the felt detail of an ordinary, target-supported cue: an object, food or drink, texture, temperature, smell, small errand, room, journey, bodily sensation, or a concrete interaction with another person. The request should sound like something a person would actually search their memory for. A mood, longing, identity question, or larger meaning is also valid when it genuinely fits the target; when possible, connect it to a concrete scene or sensation. Prefer "I need the feel of ..." or "I need to remember ..." when a material cue is available. Do not quote the target, name its title, or invent unsupported events. Use 8–24 words. Example shape, only when target-grounded: "I need the cold weight of a milk bottle sweating through a paper bag." Output one sentence only.
Memory request: user	`TARGET: [up to 7,000 characters of the work target]`
Recall query: system	Given a lived memory, write one short first-person retrieval query that would surface it. Prefer a tangible cue, ordinary action, bodily sensation, or specific interaction present in the memory. An abstract feeling or theme is also welcome when it is genuinely central; connect it to a concrete cue when possible. Use clear language. Do not copy the memory or invent facts. Output one sentence of 8–24 words only.
Recall query: user	`TARGET: [up to 7,000 characters of one target memory]`

The generated recall query is accepted only if intrinsic retrieval surfaces its intended memory; otherwise, replay also tries a deterministic query assembled from the node gist and edge labels, followed by the memory text itself.

A.4 Validation and memory-bank assembly

Biography cards must contain the declared top-level fields, internally valid identifiers, and coherent life-phase boundaries. A work flow is accepted only if it contains 3–6 uniquely identified nodes, one valid root, valid non-self edge endpoints, a connected acyclic graph, first-person text, and all required temporal, emotional, gist, and salience fields. Failure after the JSON retry budget discards the flow.

Life scenes undergo field, type, time-window, and pass-specific salience checks. The anchored pass must cite an allowed evidence identifier; the texture pass cannot cite a biography fact. A lightweight signature over time and the first 24 normalized text characters removes direct duplicates. Rejection reasons are passed to the next generation attempt.

Accepted episodic records are merged by persona. MemoryBuilder extracts associative metadata, embeds episodes with Qwen3-Embedding-0.6B, canonicalizes near-duplicate concepts, and creates cluster- and concept-conditioned schema nodes. The original source metadata and construction type remain attached to each episode, allowing biography facts, synthetic anchored scenes, texture scenes, and work-derived scenes to be distinguished during audit.

A.5 Replay prompts and trace filtering

For every work flow, Qwen3.6-27B generates three short pieces of interface text. The writing-task prompt asks for one direct imperative of 8–22 words grounded in an ordinary cue supported by the target, such as an object, bodily sensation, errand, room, or interaction. The request prompt asks for one practical first-person memory search of 8–24 words, preferably connecting a concrete cue to the target's pressure or relationship. The per-node query prompt asks for an 8–24-word first-person retrieval query centered on a tangible cue, action, sensation, or interaction in that memory. All three prompts prohibit quotations, target titles, and unsupported events.

The resulting text is executed against the actual MemoryBank rather than trusted as a label. A trace is retained only when it has a non-empty selection, no LLM failure, an initial query, clean stopping,

no invalid keep, blocked follow, or missing action, and surface recall of at least 0.5 when references exist. We then retain at most one trace for each persona–source-work pair. Before export, the benchmark builder verifies that all reference identifiers resolve to embedded nodes, checks provenance and author-level split disjointness, and audits task text for title leakage.

B WALKER, TRAINING, AND REWARD HYPERPARAMETERS

This section gives the detailed scoring functions, optimization objective, reward terms, and numerical settings omitted from the main text. The values are those of the reported SFT and GRPO runs; they are not selected retrospectively from test performance.

B.1 RETRIEVAL AND STATE-TRANSITION SETTINGS

Let $E(\cdot)$ denote the text encoder. Each memory i is represented by a normalized vector v_i. At step t, the Walker has two descriptions of what it needs: the original writing context, summarized by the anchor $a = E(x, r)$, and its current search phrase q_t, represented by $q = E(q_t)$. A `query` action scores memory i as

$$s_i^{\text{query}} = v_i^\top \big((1-\lambda_q)a + \lambda_q q\big). \tag{6}$$

Because all vectors are normalized, the dot product measures cosine similarity. The score is therefore high when a memory matches both the overall writing need and the Walker's current line of inquiry. The coefficient $\lambda_q \in [0, 1]$ controls this trade-off: a smaller value keeps retrieval close to the original task, whereas a larger value gives the Walker more freedom to redirect the search.

The Walker stores promising memories in a temporary reservoir. Keeping a memory also triggers *associative spread*: the environment searches for new memories that resemble what has already been kept. Let $\bar{v}_t$ be the normalized, recency-weighted average of the most recent C_{seed} kept memories. The spread score is

$$s_i^{\text{spread}} = (1-\mu) v_i^\top \bar{v}_t + \mu v_i^\top E(r), \tag{7}$$

The first term retrieves memories related to the current reservoir; the second keeps the search connected to the Writer's original request r. Thus, $\mu \in [0, 1]$ controls semantic drift during association. When μ is small, spread follows the kept memories more strongly; when it is large, spread remains closer to the initial request. A kept memory of age h receives weight γ^h when computing $\bar{v}_t$, where $h = 0$ denotes the newest seed. The decay factor $\gamma \in (0, 1]$ makes recent choices more influential, and C_{seed} limits how many kept memories contribute to the average.

A `follow` action provides a complementary symbolic transition. Rather than comparing the query with every memory, it starts from a visible concept, such as a person, object, or place, and ranks only the episodes linked to that concept by cosine similarity. Memories already stored in the reservoir are excluded from the returned candidates.

The action grammar contains five commands: `query`, `keep`, `follow`, `stop`, and terminal `select`. The first action must be a query; keep and follow may reference only the current observation. The remaining parameters control the interaction budget: L_{nav} limits the number of navigation actions, C_B is the maximum reservoir size, C_K limits the memories delivered to the Writer, and L_{select} limits retries in the terminal selection phase. Replay, training, and evaluation share this state machine and the same budgets.

B.2 SFT AND GRPO SETTINGS

The counterfactual Writer and scalar Judge both use Qwen3.6-27B with greedy decoding. Their output limits are 512 and 16 tokens, respectively.

For completeness, let $m_{i,t}$ mask Walker-generated tokens and define the policy ratio as

$$\rho_{i,t} = \frac{\pi_\theta(z_{i,t} \mid z_{i,<t})}{\pi_{\theta_{\text{old}}}(z_{i,t} \mid z_{i,<t})}. \tag{8}$$

Table 7: Walker retrieval and interaction hyperparameters.

Symbol	Meaning	Value
λ_q	Current-query weight in dense retrieval	0.35
μ	Request-anchor weight in spread	0.25
γ	Per-step seed recency decay	0.5
C_{seed}	Maximum kept seeds used by spread	8
L_{nav}	Maximum navigation actions	13
C_B	Reservoir capacity	16
C_K	Final-selection capacity	8
L_{select}	Maximum terminal selection attempts	2

Table 8: Optimization and rollout hyperparameters.

Setting	SFT	GRPO
Initialization	Qwen3.5-2B	Best-validation SFT adapter
LoRA	rank 32, $\alpha_{\text{LoRA}} = 64$, all linear layers	same as SFT
Training examples	470 trajectories	722 prompts
Learning rate	5×10^{-5}	3×10^{-7}
Epochs	2	3
Batching	global batch 16	global batch 48; mini-batch 24
Policy epochs per batch	–	1
Maximum sequence length	10,240	10,240
Group size N	–	8 rollouts
Sampling	–	temperature 0.8; top-p 0.95
Per-action limit L_{gen}	–	256 tokens
Clip radius ϵ_{clip}	–	0.2
KL coefficient β_{KL}	–	0.03

The GRPO loss is

$$\mathcal{L}_{\text{GRPO}} = -\frac{1}{\sum_{i,t} m_{i,t}} \sum_{i,t} m_{i,t} \min\Big(\rho_{i,t}\hat{A}_i, \text{clip}(\rho_{i,t}, 1-\epsilon_{\text{clip}}, 1+\epsilon_{\text{clip}})\hat{A}_i\Big) + \beta_{\text{KL}} D_{\text{KL}}^{\text{low-var}}(\pi_\theta \,\|\, \pi_{\text{SFT}}). \tag{9}$$

B.3 Reward coefficients and low-level shaping

Table 9 gives the coefficients used in the main reward definition. These values are fixed for the reported run.

The remaining terms provide event-level shaping for the multi-step interaction. The navigation and process ablations intentionally overlap: no-nav removes the three terminal navigation metrics and their matching increments below, whereas no-process removes the entire step-level term. Both retain the format score and terminal selection gate.

Table 9: Core reward coefficients.

Parameter	Component	Value	Parameter	Component	Value
α_{write}	Writing improvement	0.20	α_{rec}	Final recall	0.20
α_{F_1}	Final selected F1	0.80	α_{flow}	Final flow alignment	0.25
β_{rec}	Step recall change	0.30	β_{F_1}	Step F1 change	0.35
β_{flow}	Step flow change	0.20	α_{fmt}	Format score	0.10
r^+_{select}	Valid selection	0.20	r^-_{select}	Invalid selection	-0.30

Table 10: Low-level process signals. Here $\Delta_t f = f_t - f_{t-1}$, B_t is the current reservoir, S_t is the set of surfaced memories, m_t is the number of memories kept at step t, and u_t/v_t is the fraction of the visible batch omitted by that keep.

Signal	Definition	Value
Reservoir recall	Change in $\lvert G \cap B_t \rvert / \lvert G \rvert$	β_{rec}
Selected-set F1	Change in F1 of the reservoir, or of K at selection	β_{F_1}
Flow alignment	Change in flow alignment of the reservoir, or of K at selection	β_{flow}
Surface recall	Change in $\lvert G \cap S_t \rvert / \lvert G \rvert$; disabled in the recorded run	0.00
Selection capacity	Quadratic cost $0.08(n/8)^2$ for current retained-set size n	intrinsic
Invalid actions	Linear penalty per invalid event, capped in magnitude	$-0.15; -0.80$
Multi-keep	Small bonus for keeping several useful candidates together, truncated at four	0.02
Visible omission	Penalty proportional to the fraction of visible candidates omitted, cumulatively capped	$-0.01; -0.08$
Empty stop	Stop with no retained memory	-0.05
Empty selection	Terminal selection with no memory	-0.25
Selection shortfall	Fraction of the eight-memory output budget left unused	-0.10
Follow progress	Productive / unproductive concept transition	$+0.03/-0.10$
Reference via follow	Newly surfaced reference memory after a follow	$+0.10$
Budget exhaustion	Failure to terminate within the action budget	-0.30

C LLM-AS-JUDGE PROMPT

Table 11 gives the complete prompt template used by every evaluation Judge. Bracketed fields are replaced at runtime; each reference memory is truncated to 700 characters before serialization. The Judge must return exactly the five numeric fields shown below.

Table 11: Complete LLM-as-Judge prompt.

Message	Prompt text
System	You are a strict and discriminative literary benchmark judge. Evaluate the response on the following four dimensions, assigning each a score from 0.0 to 1.0: 1. **task_fulfillment**: Evaluate how fully and accurately the response satisfies the writing task, including all explicit constraints, requested content, and intended communicative or narrative goal. Do not reward irrelevant elaboration or additional length. 2. **persona_grounding**: Evaluate how convincingly the response reflects the persona implied by the supplied reference memories. Consider whether relevant traits, preferences, experiences, relationships, perspectives, and behavioral patterns are integrated naturally and consistently. Do not reward superficial name-dropping or mechanical repetition of memory details. 3. **creative_transformation**: Evaluate how effectively the response transforms information from the reference memories into new, context-appropriate creative material. Reward meaningful recombination, reinterpretation, inference, and adaptation. Penalize direct copying, close paraphrasing, or merely listing details from the memories. Novelty should remain grounded in the supplied information rather than introducing unsupported persona facts. 4. **coherence**: Evaluate the overall internal consistency, narrative or rhetorical flow, clarity, and logical continuity of the response. The response should read as a unified piece rather than a collection of disconnected memory references. Use the following general scoring anchors: 0.0: completely fails the criterion; 0.25: substantially deficient; 0.5: partially satisfies the criterion but has clear weaknesses; 0.75: strongly satisfies the criterion with minor weaknesses; 1.0: exceptionally satisfies the criterion. Intermediate values are allowed when appropriate. Judge each dimension independently. Do not let strong performance on one dimension inflate scores on the others. Do not reward verbosity, length, ornate language, or the sheer number of memory details used. Prefer selective, natural, and contextually meaningful use of the reference memories. Return valid JSON only, with exactly the following five numeric keys: {"task_fulfillment": <score>, "persona_grounding": <score>, "creative_transformation": <score>, "coherence": <score>, "overall": <your independent holistic score>} The overall score is your independent assessment of the response as a whole. It need not equal the arithmetic mean of the four dimension scores. Do not include explanations, comments, markdown, or any additional text.
User	{"task": "[WRITING TASK]", "persona": "[PERSONA SUMMARY]", "reference_memories": ["[REFERENCE MEMORY 1]", . . .], "response": "[GENERATED RESPONSE]"}

D Qualitative Case Study

We compare all methods on the same held-out case, rather than selecting a separate example for each method. The case is identified by its stable benchmark key so that the underlying task, MemoryBank, and reference flow can be recovered from the released artifacts. The comparison uses the Qwen3.6-27B Writer and the same deterministic rendering and output budget as the main evaluation. Table 12 summarizes the shared input.

Table 13 shows the material handed to the Writer. The resulting responses illustrate the distinction between retrieval and writing. The persona-only response describes the atmosphere in a historically styled voice but has no explicit episodic evidence. RAG and RAG + Reranking produce competent scenes centered on the visible dining-room cue. Graph-PPR supplies related sensory material through graph propagation. Daydreamer produces a similarly relevant scene while making the asso-

Table 12: Shared input for the qualitative comparison.

Field	Value
Persona	Marguerite, Queen of Navarre; Renaissance patron and writer
Task	Write a short original prose scene in the persona's voice. Describe the heavy silence of a dining room while watching a stranger's hands tremble around a wine glass. Transform the seed into new writing.
Memory request	"I need the cloying sweetness of sugar masking the foul, rotting stench that fills the air."
Reference flow	m_0002 → m_0073 → m_0021
Flow type	Object recurrence; source work: *Tale Li*

Table 13: Daydreams and selected memory summaries for one shared test case. All methods use the same Writer and task.

Method	Rendered daydream or memory evidence
Daydreamer	"I search my memory for 'What did I feel when the heavy silver platter clattered against the table, shattering the silence?' – and a memory surfaces: The scent of roasted fowl and stale wine hung heavy in the hall as the servant girl poured the cup, her hand trembling slightly, a gesture I recognized as the prelude to a staged accident rather than mere clumsiness."
Daydreamer (base)	"I search my memory for 'the cloying sweetness of sugar masking the foul, rotting stench that fills the air' – and a memory surfaces"; the selected context describes the heavy hall, roasted fowl, stale wine, and a servant whose hands tremble around a cup.
RAG	"That spreads out along dense cosine similarity" and retrieves a memory about the last light of December fading from a high window, followed by a generic connection to silence and pressure.
RAG + Reranking	Retrieves the dining-hall memory: roasted fowl, stale wine, and the servant's trembling hand. The resulting daydream is close to the task cue but contains little additional associative context.
Graph-PPR	Propagates from the query over the episode–concept graph and retrieves the hall scene together with memories about cooking and household smells. The daydream begins with the scent of roasted fowl and stale wine and links it to the servant's staged accident.

ciative transition explicit in the retrieved material: smell and silence lead to the trembling hand, and the hand is interpreted as part of a social action rather than as an isolated visual detail.

E A COMPLETE DAYDREAMBENCH EXAMPLE

Each benchmark case separates the public writing task from private supervision used for scoring and training. Table 14 gives the complete logical structure of the case above. The public fields identify the persona, task, request, genre, and MemoryBank. The private fields identify the construction-derived reference memories and their provenance. The Walker observes the public task and environment observations, but never receives the private gold identifiers or the flow order. This format makes the supervision boundary explicit.

Table 14: Example DAYDREAMBENCH record. Long memory texts are shortened with an ellipsis; the released record retains the full strings.

Component	Example
Public identifier	ddb-046494ee9d431d49
Public persona	"Queen, consort of Henry II, King of Navarre Marguerite"
Public task	Write a short original prose scene in the persona's voice. Use the inspiration seed "Describe the heavy silence of a dining room as you watch a stranger's hands tremble while holding a glass of wine." Transform it into new writing.
Public request	"I need the cloying sweetness of sugar masking the foul, rotting stench that fills the air."
Public metadata	Corpus: Gutenberg short story; genre: short story; split: test; target type: summary–outline–scene.
Gold memory m_0002	"I watched the courtiers at Chambord pause in their discourse, their noses wrinkling in delicate revulsion as a servant passed by carrying a silver dish that smelled faintly of sulfur rather than spice."
Gold memory m_0073	"I instructed the kitchen steward to substitute crushed roasted onions for the powdered sugar in the Duke's favorite confection, watching from the gallery as he took a bite and choked on the pungent truth of his own gluttony."
Gold memory m_0021	"I sat in the salon, listening to tales of Italian cruelty and the Duke of Urbino's unnatural vengeance, feeling the weight of my pen as I prepared to frame these barbarities within a moral discourse on the folly of pride."
Private supervision	Gold identifiers: m_0002, m_0021, m_0073; reference flow order: m_0002 → m_0073 → m_0021; flow type: object recurrence.

Table 15: Paired significance analysis for Surface Recall. p_H is the Holm-adjusted exact sign-test p-value.

Baseline	Δ	95% CI	p_H
Daydreamer (base)	$+0.1134$	$[0.0850, 0.1427]$	8.38×10^{-9}
RAG	$+0.1114$	$[0.0768, 0.1430]$	8.38×10^{-9}
RAG + Reranking	$+0.1114$	$[0.0768, 0.1430]$	8.38×10^{-9}
Graph-PPR	$+0.1164$	$[0.0843, 0.1490]$	3.67×10^{-8}

Table 16: Paired significance analysis for Keep Recall.

Baseline	Δ	95% CI	p_H
Daydreamer (base)	$+0.0075$	$[-0.0167, 0.0342]$	0.7201
RAG	$+0.0709$	$[0.0393, 0.1070]$	0.0008
RAG + Reranking	$+0.0709$	$[0.0393, 0.1070]$	0.0008
Graph-PPR	$+0.0553$	$[0.0234, 0.0882]$	0.0140

Table 17: Paired significance analysis for Selected F1.

Baseline	Δ	95% CI	p_H
Daydreamer (base)	$+0.0239$	$[0.0079, 0.0402]$	0.0340
RAG	$+0.0239$	$[-0.0016, 0.0498]$	0.0340
RAG + Reranking	$+0.0239$	$[-0.0016, 0.0498]$	0.0340
Graph-PPR	$+0.0276$	$[0.0026, 0.0519]$	0.0030

F PAIRED SIGNIFICANCE ANALYSIS

We apply the paired persona-cluster analysis to the three navigation metrics. Persona-only is excluded because it does not execute navigation. Each table uses 343 paired cases and 40 persona clusters. Positive Δ indicates that Daydreamer scores higher than the corresponding navigation baseline.

Daydreamer significantly improves Surface Recall over every navigation baseline (Table 15). The consistent positive differences show that the trained Walker exposes more reference memories during navigation, rather than merely changing which already-visible memories it selects.

Keep Recall is significantly higher than RAG, RAG + Reranking, and Graph-PPR, but not higher than the base Walker (Table 16). This pattern suggests that SFT and GRPO substantially alter the search trajectory and improve over one-shot retrieval, while the additional benefit over the untrained Walker is smaller at the final reservoir stage.

Daydreamer significantly outperforms the base Walker and Graph-PPR on Selected F1 (Table 17). Its point estimates and sign tests are also favorable against RAG and RAG + Reranking, but the bootstrap intervals slightly overlap zero. We therefore treat these two comparisons as suggestive rather than as uniformly conclusive evidence. Taken together, the navigation analysis indicates that the clearest and most robust effect of training is broader memory discovery, while improvements in final curation are smaller and depend more strongly on the baseline.