MANGO: EFFICIENT MEMORY-TOKEN GOVERNANCE FOR LONG-HORIZON MULTI-AGENT LLM REASONING

Anonymous authors
Paper under double-blind review

ABSTRACT

Large language model (LLM) agents increasingly rely on shared memory for long-horizon reasoning and collaborative problem solving. Yet continuous writing and retrieval cause memory contexts to accumulate, increasing token expenditure and repeatedly exposing agents to redundant or outdated information that can destabilize later reasoning. Existing memory-augmented frameworks primarily improve retrieval or reasoning quality while overlooking the coupled evolution of memory, resource consumption, collaborative progress, and reasoning stability. We present MANGO (**M**ulti-**A**ge**N**t Efficient Memory-Token **GO**vernance), a lightweight feedback-driven framework that regulates shared memory at test time. MANGO integrates population-level coordination, runtime state profiling, and adaptive control over retrieval, writing, and deletion. This closed-loop design supports broad exploration early in execution and progressively consolidates memory as redundancy and runtime pressure emerge. Extensive experiments on multi-objective long-horizon reasoning and single-objective multi-hop QA show that MANGO improves memory-token efficiency by up to 108.6% over representative reasoning and memory-augmented agent baselines while maintaining competitive reasoning quality. By revealing and regulating the coupled dynamics among shared memory, resource consumption, and collaborative progress, MANGO establishes a feedback-driven memory governance system for efficient and stable long-horizon multi-agent reasoning.

1 INTRODUCTION

Large language model (LLM) agents have evolved from isolated reasoners into collaborative multi-agent systems that solve complex tasks through communication, planning, and coordination Guo et al. (2024). Recent systems use specialized agents to decompose problems, exchange intermediate results, and perform long-horizon reasoning over many steps Zhang et al. (2024); Liu et al. (2024); Wang et al. (2025a). To retain information beyond limited context windows, they increasingly rely on external or shared memory that persists intermediate knowledge throughout execution Jiménez Gutiérrez et al. (2024); Wang et al. (2024). Shared memory has therefore become central to collaborative reasoning, multi-step planning, and long-horizon task completion.

However, shared memory is not free. Repeated retrieval, writing, and reuse expand the repository and increase the context processed by subsequent reasoning steps Yu et al. (2026); Wang et al. (2026). Accumulated memory may contain redundant, outdated, or low-value records that increase retrieval overhead, noise, and token expenditure while obscuring useful evidence. These effects are amplified in multi-agent systems because locally generated records persist and influence other agents' future reasoning. Long-horizon execution must therefore preserve memory quality while controlling resource consumption and reasoning degradation.

Figure 1 exposes this coupling during long-horizon multi-agent reasoning. Each point denotes one reasoning step, with marker size indicating memory noise and color denoting collaborative progress. Shared-memory size is positively correlated with token consumption ($r = 0.441$), while larger repositories exhibit greater token variability as noise accumulates. Without governance, redundant records persist across interactions and amplify future retrieval and token costs. This observation

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.

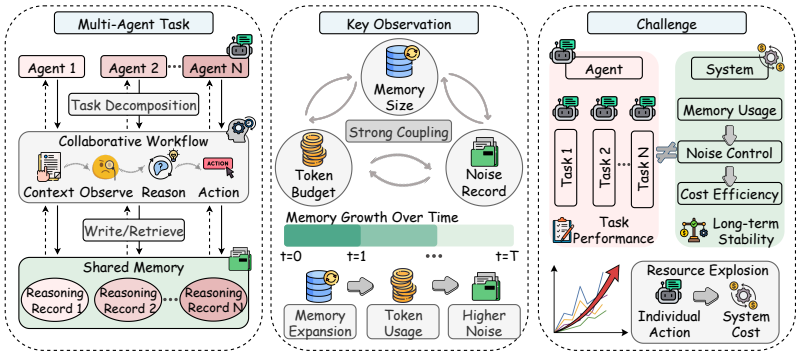


Figure 2: **Motivation for feedback-driven memory–token governance.** Locally useful operations compound memory growth, token consumption, and noise. Without system-level governance, these effects increase cost and destabilize long-horizon reasoning.

reveals that shared memory is not merely a reasoning aid, but an evolving system state whose growth and quality require explicit control.

Existing approaches address parts of this challenge. Agentic frameworks improve reasoning through planning, tool use, reflection, and collaboration Shinn et al. (2024), while memory-augmented systems strengthen long-horizon reasoning through memory organization, retrieval, and compression Zhou et al. (2025); Wang et al. (2025b). Despite improving reasoning and retention, these methods treat memory as a performance-enhancing component rather than an evolving shared resource. They provide limited support for regulating its lifecycle as agents interact. Consequently, memory may keep growing and consuming tokens after useful information saturates, making efficient and reliable long-horizon reasoning difficult to sustain Yu et al. (2026); Tan et al..

The difficulty is the persistent cross-agent externality of memory operations. A record written for local reasoning may remain in shared memory and repeatedly enter subsequent contexts. Although immediately useful, such actions can collectively increase retrieval costs and propagate redundant or low-value information. As illustrated in Figure 2, memory growth, quality, token expenditure, and collaborative progress consequently evolve as a coupled process Yu et al. (2026). Addressing this coupling requires more than fixed budgets or static retrieval rules; it requires feedback-driven governance that observes collective execution and adapts the memory lifecycle Roh et al. (2026).

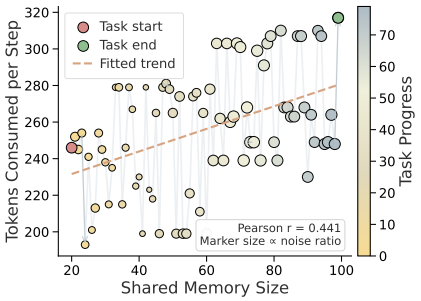


Figure 1: **Coupled memory–token dynamics.** Each point denotes a step. Larger memories incur higher token consumption, while noisier memory (larger markers) exhibits greater variability.

Motivated by this coupling, we formulate memory governance as a long-horizon reasoning and resource-management problem and propose **MANGO** (Multi-AgeNt Efficient Memory-Token GOvernance). MANGO combines a *Collective Resource Coordinator* that summarizes population demand, a *Memory–Token State Profiler* that tracks memory quality and resource pressure, and an *Adaptive Governance Controller* that regulates retrieval, writing, and deletion. Together, they form a feedback loop that preserves information access during exploration and progressively consolidates memory as redundancy and pressure emerge. MANGO governs how shared memory evolves throughout long-horizon execution without model retraining or fine-tuning. Our contributions are:

- We first identify the intrinsic coupling between memory dynamics and token expenditure in long-horizon multi-agent systems, and formulate memory governance as a coupled resource-management problem with quality consideration.
- We present MANGO, a lightweight closed-loop memory–token governance framework that continuously profiles runtime states and adaptively regulates resource usage for efficient long-horizon multi-agent execution.

- We develop *Collective Resource Coordinator* and *Memory–Token State Profiler* that capture population-level execution dynamics, memory quality, and resource-consumption states to provide system-wide awareness for the memory governance mechanism.
- We develop an adaptive test-time memory policy that regulates retrieval, writing, and deletion according to the evolving population and memory–token states, enabling selective memory lifecycle control rather than fixed truncation or uniform pruning.
- Extensive experiments demonstrate that MANGO consistently improves memory–token efficiency by up to 108.6%, while simultaneously reducing memory growth, token consumption, and memory noise across diverse long-horizon reasoning tasks, without sacrificing reasoning accuracy.

2 System Model

2.1 Multi-Agent System Model

We consider N collaborative LLM agents, denoted by $\mathcal{N} = \{1, \dots, N\}$, that interact through shared memory over steps $t = 0, \dots, T$. Each agent retrieves information, performs local reasoning, and selectively contributes new knowledge. These decentralized operations jointly determine collaborative progress, memory evolution, and system-wide resource consumption. At step t, the global state is $X_t = (\mathcal{S}_t, m_t)$, where $\mathcal{S}_t = \{x_t^i\}_{i\in\mathcal{N}}$ collects local runtime states and m_t describes the shared-memory operating condition.

Definition 1 (Shared Memory–Token State)**.** *The shared memory–token state is $m_t = (s_t, \bar{\eta}_t, H_t)$, where s_t is the memory size, $\bar{\eta}_t \in [0, 1]$ is the memory noise ratio, and H_t is cumulative token consumption.*

Agents do not directly observe one another's internal states; their interaction is mediated by shared memory. Agent i observes $o_t^i = (x_t^i, m_t)$ and selects a memory action $a_t^i = (r_t^i, w_t^i, d_t^i) \in \mathcal{A}$, where r_t^i is the retrieval intensity, $w_t^i \in \{0, 1\}$ is the write decision, and $d_t^i \geq 0$ is the removal or compression intensity. The retrieved context is $\mathcal{R}_t^i = g(r_t^i, \mathcal{M}_t)$, and the local state evolves as $x_{t+1}^i = f_i(x_t^i, m_t, \mathcal{R}_t^i)$. Collectively, these operations induce the stochastic transition $X_{t+1} = \mathcal{F}(X_t, a_t, \xi_t)$, where $a_t = \{a_t^i\}_{i\in\mathcal{N}}$ and ξ_t captures retrieval and reasoning uncertainty.

2.2 Coupled Memory–Token Dynamics

Long-horizon execution couples memory occupancy, quality, token expenditure, and task progress. Writing adds new information, whereas cleanup removes obsolete or redundant records:

$$s_{t+1} = s_t + \sum_{i\in\mathcal{N}} w_t^i - \sum_{i\in\mathcal{N}} d_t^i. \tag{1}$$

Let $\mathbf{c} = (c_r, c_w, c_d)$ denote the token costs of memory operations and $u_t^i \in [0, 1]$ the progress of agent i. The remaining dynamics are

$$\bar{\eta}_{t+1} = \Psi(m_t, a_t), \quad H_{t+1} = H_t + \frac{1}{N}\sum_{i\in\mathcal{N}} \langle \mathbf{c}, a_t^i \rangle, \quad u_{t+1}^i = \Pi_{[0,1]}\big(u_t^i + \mathcal{G}_i(u_t^i, a_t^i, m_t)\big). \tag{2}$$

Here, Ψ models the effect of collective actions on memory quality, while $\mathcal{G}_i$ captures progress under the current memory condition. These dynamics expose the central challenge: a locally beneficial action may persistently alter the memory quality and resource costs experienced by future agents.

2.3 Problem Formulation

The coupled dynamics require memory growth, quality, token expenditure, and collaborative progress to be considered jointly. We therefore measure progress relative to the accumulated memory, token, and noise burden.

Definition 2 (System Memory–Token Efficiency)**.** *The system memory–token efficiency at step t is*

$$E_t(a_t) = \frac{\sum_{i\in\mathcal{N}} u_t^i}{\alpha s_t + \beta H_t + \gamma N \bar{\eta}_t + \epsilon}, \tag{3}$$

where $\alpha, \beta, \gamma > 0$ balance memory occupancy, token consumption, and memory noise, and $\epsilon > 0$ ensures numerical stability.

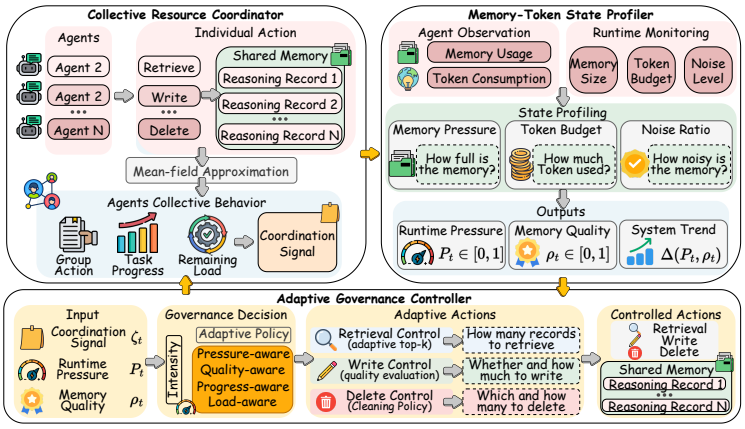


Figure 3: **Overview of MANGO.** Population-level coordination and online memory–token profiling guide adaptive operations, while the updated memory shapes subsequent reasoning.

A larger $E_t(a_t)$ represents greater collaborative progress with lower resource consumption and cleaner memory. We seek decentralized policies $\{\pi^i\}_{i\in\mathcal{N}}$, with $a_t^i = \pi^i(o_t^i)$, that solve

$$\max_{\{\pi^i\}_{i\in\mathcal{N}}} \mathbb{E}\left[\sum_{t=0}^{T} E_t(a_t)\right] \quad \text{s.t.} \quad \bar{u}_t \geq U_{\min},\ s_t \leq S_{\max},\ H_t \leq H_{\max}, \quad \forall t \in \mathcal{T}. \tag{4}$$

where $\mathcal{T} = \{0, \ldots, T\}$, $U_{\min}$ is the required collaborative progress, and $S_{\max}$ and $H_{\max}$ are the memory and token budgets. The expectation is over stochastic retrieval and memory evolution. This constrained control objective balances immediate reasoning progress against the long-term effects of decentralized memory actions and serves as a system-level design principle rather than a learned training objective.

3 MANGO: Multi-Agent Efficient Memory-Token Governance System

3.1 Overview and Design Rationale

Shared memory supports long-horizon coordination but creates a persistent externality: records written by one agent can repeatedly affect the contexts, token expenditure, and reasoning of subsequent agents. As redundant or outdated information accumulates, memory becomes noisy and costly to reuse. MANGO addresses this challenge through the lightweight feedback-driven framework in Figure 3. At each step, the *Collective Resource Coordinator* summarizes population-level demand, the *Memory-Token State Profiler* characterizes runtime pressure and memory quality, and the *Adaptive Governance Controller* regulates retrieval, writing, and deletion. These actions update the repository and close the governance loop, adapting the memory lifecycle to collaborative progress without model fine-tuning or reinforcement learning. Complete procedures appear in Appendix D.

3.2 Collective Resource Coordinator

Locally reasonable memory operations may collectively become inefficient because individual agents do not observe their population-level consequences. The Collective Resource Coordinator (CRC) constructs a compact joint-state representation before actions are selected. Given agent state x_t^i, previous action a_{t-1}^i, and normalized progress u_t^i, CRC computes

$$\mu_t = \frac{1}{N}\sum_{i\in\mathcal{N}} \delta_{(x_t^i, a_{t-1}^i, u_t^i)}, \quad \bar{x}_t = \frac{1}{N}\sum_{i\in\mathcal{N}} x_t^i, \quad \bar{u}_t = \frac{1}{N}\sum_{i\in\mathcal{N}} u_t^i, \quad \Delta_t = 1 - \bar{u}_t, \tag{5}$$

where μ_t is the empirical population state and Δ_t measures unresolved demand. The recent memory-operation load is

$$\Omega_t = \frac{1}{N}\sum_{i\in\mathcal{N}} \left(\omega_r r_{t-1}^i + \omega_w w_{t-1}^i + \omega_d d_{t-1}^i\right). \tag{6}$$

CRC produces $\zeta_t = (\bar{x}_t, \bar{u}_t, \Delta_t, \Omega_t, m_t)$, where m_t is defined in Definition 1. Rather than centralizing reasoning, this signal exposes collective demand and operation load to the profiler and controller, allowing local actions to account for system-wide consequences. The procedure appears in Algorithm 1 of Appendix D.

3.3 Memory-Token State Profiler

Population activity alone does not indicate whether memory remains useful or affordable. The Memory-Token State Profiler (MTSP) tracks memory size $s_t = |\mathcal{M}_t|$, cumulative token consumption H_t, and per-record quality. For each $m \in \mathcal{M}_t$, MTSP computes a normalized quality score $q_t(m) \in [0, 1]$ from its relevance, novelty, and recency, and defines the low-quality ratio $\rho_t = |\{m : q_t(m) < \theta_q\}| / \max(1, s_t)$. These signals are combined into the runtime pressure

$$P_t = \alpha \frac{s_t}{S_{\max}} + \beta \frac{H_t}{H_{\max}} + \gamma \rho_t, \tag{7}$$

where $S_{\max}$ and $H_{\max}$ are the memory and token budgets, and α, β, γ follow Definition 2. By jointly capturing occupancy, expenditure, and quality, P_t distinguishes a large but useful repository from a smaller noisy one, enabling governance to adapt dynamically rather than imposing fixed capacity or periodic truncation. Quality score and memory evaluation details and MTSP overhead are provided in Appendices B.1 and C; the procedure appears in Algorithm 2.

3.4 Adaptive Governance Controller

The Adaptive Governance Controller (AGC) maps collective demand and runtime pressure to agent-specific memory actions. It first computes

$$\Gamma_t^i = \kappa_1 P_t + \kappa_2 \Omega_t + \kappa_3 \bar{u}_t, \tag{8}$$

which captures memory–token pressure, operation load, and remaining demand. AGC then applies three complementary controls:

$$r_t^i = \mathrm{clip}\Big(r_{\text{base}} - \lambda_P P_t - \lambda_\Omega \Omega_t + \lambda_u (1 - u_t^i), r_{\min}, r_{\max}\Big), \tag{9}$$

$$w_t^i(m) = \mathbb{I}[q(m) \geq \tau(P_t, \rho_t)], \quad d_t^i = \mathrm{clip}\Big(\lambda_d \Gamma_t^i (\epsilon_d + \rho_t), d_{\min}, d_{\max}\Big). \tag{10}$$

AGC preserves broad retrieval when unresolved demand is high but becomes selective as memory grows costly or repetitive. Its adaptive threshold suppresses low-value writes, while deletion removes unimportant, redundant, or stale records. The action $a_t^i = (r_t^i, w_t^i, d_t^i)$ updates the repository and subsequent reasoning context, enabling early exploration followed by progressive consolidation. AGC thus turns static retention into a closed-loop test-time policy; the complete procedure appears in Algorithm 3 of Appendix D.

4 Theoretical Analysis

We analyze MANGO under sufficient conditions capturing its central mechanism: persistent memory creates future cross-agent costs, while feedback-driven governance regulates them through population-level coordination and runtime memory–token states. Let $\mathcal{F}_t$ denote the execution history, $U_t = \sum_{i \in \mathcal{N}} u_t^i$ aggregate progress, and $R_t = \alpha s_t + \beta H_t + \gamma N \bar{\eta}_t + \epsilon$ resource burden. By Definition 2, the system efficiency is $E_t = U_t / R_t$. Complete proofs are in Appendix H.

Persistent-memory externality. A record written once may repeatedly enter future agent contexts. For a record m of length $\ell(m)$ admitted at t, Lemma 2 gives its expected downstream context cost:

$$\mathrm{Ext}_t(m) = \ell(m) \sum_{\tau=t+1}^{T} \sum_{j \in \mathcal{N}} p_{\tau,j}(m), \tag{11}$$

where $p_{\tau,j}(m)$ is the probability that m remains available and is retrieved by agent j at step τ. This nonnegative cost grows with future retrieval opportunities, motivating lifecycle-level governance.

Closed-loop resource stability. Let $W_t = \sum_i w_t^i$ and $D_t = \sum_i d_t^i$ denote admitted writes and actual deletions. The following result characterizes when pressure-aware governance controls persistent memory growth and noise.

Theorem 1 (Closed-Loop Resource Stability)**.** *Suppose memory actions are bounded and feasible, and there exist* $\kappa_s, \kappa_\eta \in (0, 1]$ *and* $b_s, b_\eta \geq 0$ *such that*

$$\mathbb{E}[D_t - W_t \mid \mathcal{F}_t] \geq \kappa_s s_t - b_s, \qquad \mathbb{E}[\bar{\eta}_{t+1} \mid \mathcal{F}_t] \leq (1 - \kappa_\eta)\bar{\eta}_t + b_\eta. \tag{12}$$

Then

$$\sup_{t \geq 0} \mathbb{E}[s_t] \leq \max\left\{ s_0, \frac{b_s}{\kappa_s} \right\}, \qquad \sup_{t \geq 0} \mathbb{E}[\bar{\eta}_t] \leq \max\left\{ \bar{\eta}_0, \frac{b_\eta}{\kappa_\eta} \right\}. \tag{13}$$

If the controller also enforces a finite token budget, the complete memory–token state remains within a bounded feasible region.

The theorem captures MANGO's behavior: memory may expand during exploration, while increasing pressure induces sufficient selectivity and cleanup to prevent persistent growth. It provides a sufficient stability certificate rather than an unconditional convergence guarantee. Appendix H further establishes population-level sufficiency, finite-time bounds, boundary and robustness results, the efficiency condition in Theorem 5, and the $O(N + |\mathcal{M}_t|)$ governance complexity in Proposition 2.

5 Evaluation

5.1 Implementation Details

Datasets and models. We evaluate MANGO on multi-objective long-horizon reasoning and single-objective multi-hop QA using *HotpotQA* Yang et al. (2018), *2WikiMultihopQA* Ho et al. (2020), *Bamboogle* Press et al. (2023), and *MuSiQue* Trivedi et al. (2022), together with synthetic collaborative environments. Experiments use *Qwen3-1.7B* and *Qwen3-8B* Yang et al. (2025), with additional evaluation on *Llama-3.1-8B* Grattafiori et al. (2024) and *Mistral-7B* Jiang et al. (2023).

Protocol, metrics, and baselines. MANGO operates online without retraining or task-specific fine-tuning. We report EM, token-level F1, token usage, memory size and noise, latency, and memory–token efficiency $E_t(a_t)$. We compare against *FullMem*, *ReAct* Yao et al. (2022), *Mem1* Zhou et al. (2025), *MemPO* Li et al. (2026), *A-Mem* Xu et al. (2026), *Retrieval-only* Xu et al. (2024), and *Sliding-window* Xiao et al. (2024). Complete details are provided in Appendix B. Code is available at `https://anonymous.4open.science/r/MANGO-AE91`.

5.2 Performance on Multi-Objective Multi-Hop Tasks

We evaluate MANGO on long-horizon multi-objective reasoning with 2–16 objectives. Table 1 reports results on Qwen3-1.7B and Qwen3-8B, while Figure 4 shows the corresponding performance–efficiency relationships. MANGO achieves the highest memory–token efficiency across all settings while maintaining competitive reasoning quality. Under Qwen3-8B, it improves efficiency by up to 75.4%, 44.3%, 70.1%, and 108.6% over ReAct, Mem1, MemPO, and A-Mem, respectively. It also reduces total and peak token consumption, with larger gains as the objective count increases, demonstrating scalability across both backbones.

Table 1: **Results on long-horizon reasoning tasks with single-and-multiple objective settings.**

Long-horizon Multi-hop QA Tasks		Multi-Objective												Single-Objective			Average	
Task Scenarios		16-Objective			8-Objective			4-Objective			2-Objective			1-Objective				
Backbone	Method	TT ↓	PT ↓	Eff. ↑	TT ↓	PT ↓	Eff. ↑	TT ↓	PT ↓	Eff. ↑	TT ↓	PT ↓	Eff. ↑	TT ↓	PT ↓	Eff. ↑	EM ↑	F1 ↑
Qwen3-1.7B	FullMem	52.72	0.92	4.87	53.65	0.79	5.76	68.96	0.92	6.00	49.25	0.92	6.08	14.84	0.37	1.71	27.80	**39.13**
	ReAct	93.98	1.28	4.84	98.17	1.28	5.46	122.01	1.56	5.51	136.00	1.51	5.74	26.06	0.60	1.82	22.68	32.56
	Mem1	48.97	0.59	14.46	33.62	0.38	18.99	41.94	0.45	23.56	27.71	0.57	23.92	14.98	0.38	1.44	27.09	35.27
	MemPO	39.33	0.36	13.12	39.95	0.35	12.20	58.22	0.91	13.67	61.78	0.90	14.32	14.72	0.36	1.52	22.70	30.38
	A-Mem	37.23	0.64	10.61	43.41	0.57	8.96	49.07	0.53	12.27	36.22	0.57	14.19	21.61	0.54	1.32	25.15	30.62
	Retrieval-only	52.39	0.70	15.30	52.07	0.61	17.36	55.41	0.61	20.29	57.32	0.92	19.97	18.71	0.48	1.35	20.84	30.16
	Sliding-window	43.00	0.51	10.89	38.86	0.50	15.31	40.69	0.59	18.15	45.16	0.58	20.57	19.72	0.50	1.29	27.36	36.19
	MANGO	**30.38**	**0.34**	**23.43**	**28.59**	**0.35**	**26.56**	**24.52**	**0.35**	**28.87**	**26.65**	**0.42**	**25.95**	**11.79**	**0.28**	**4.39**	**29.09**	35.58
Qwen3-8B	FullMem	53.04	0.94	4.92	53.84	0.79	5.96	69.91	0.94	6.23	49.50	0.93	6.30	14.47	0.37	1.73	30.07	45.60
	ReAct	92.91	1.28	5.01	98.87	1.28	5.86	123.12	1.53	5.77	135.41	1.50	5.91	23.76	0.60	1.78	27.13	**46.66**
	Mem1	49.42	0.63	14.98	34.75	0.39	20.52	42.49	0.47	24.02	27.89	0.59	26.60	14.19	0.37	1.48	32.26	43.19
	MemPO	39.86	0.36	13.85	41.36	0.38	13.57	59.76	0.93	14.50	62.36	0.91	14.57	13.68	0.36	1.51	24.75	38.74
	A-Mem	37.32	0.63	11.06	43.98	0.59	10.24	49.27	0.54	12.68	37.37	0.59	14.70	19.97	0.54	1.32	28.26	40.90
	Retrieval-only	51.11	0.70	16.58	52.38	0.61	18.00	55.89	0.61	20.56	57.87	0.93	20.47	12.11	0.37	1.71	24.19	40.34
	Sliding-window	43.69	0.57	11.89	39.17	0.50	16.41	41.03	0.61	18.98	45.63	0.60	20.95	17.74	0.48	1.28	33.22	46.64
	MANGO	**31.41**	**0.35**	**23.79**	**29.58**	**0.36**	**25.99**	**23.80**	**0.35**	**29.71**	**22.48**	**0.40**	**28.99**	**10.83**	**0.30**	**4.40**	**33.64**	42.38

Note: Single-objective TT is averaged over 2WikiMultihopQA, Bamboogle, HotpotQA, and MuSiQue and reported in 10^4 tokens. All other TT and PT values are in kilo-tokens (10^3). Eff. denotes memory–token efficiency η ($\times 100$). The best result for each column and backbone is **bolded** and highlighted.

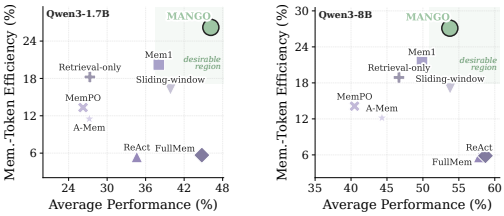
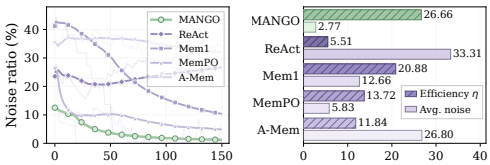


Figure 4: **Performance–efficiency trade-off.** Each point shows performance and memory–token efficiency across multi-objective settings. The upper-right region denotes the preferred regime, where MANGO operates.

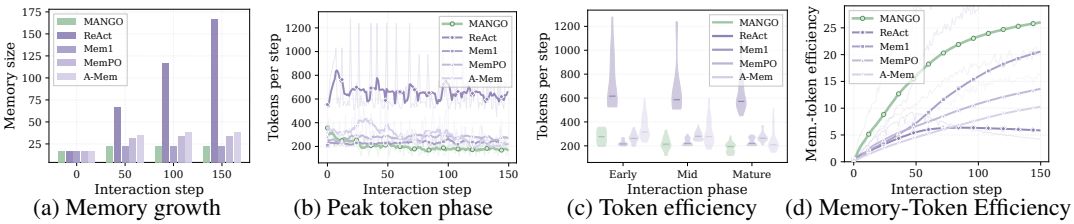


(a) Noise ratio over steps (b) Noise–effic. coupling

Figure 5: **Memory-noise regulation.** (a) Noise ratio over long-horizon execution. (b) Relationship between noise and memory–token efficiency. MANGO maintains a low-noise, high-efficiency regime through governance.

Memory–token dynamics. Figure 6 traces shared-memory size, token consumption, and cumulative efficiency. Existing memory-based methods progressively accumulate records, increasing retrieval and token costs. MANGO instead maintains compact memory through adaptive retrieval and quality-aware cleaning. Compared with MemPO and A-Mem, it reduces average memory size by up to 54.8% and 32.7%, and cumulative token consumption by 38.6% and 27.9%, respectively. The resulting efficiency gap widens over time, showing that adaptive governance suppresses long-term memory–token amplification.

Noise–Efficiency Analysis. Figure 5 and Table 2 examine the relationship between memory quality and long-horizon efficiency. Existing memory-based methods accumulate redundant and outdated records, increasing retrieval cost while yielding diminishing reasoning returns. MemPO and A-Mem exhibit the highest noise levels due to their larger memory repositories. In contrast, MANGO maintains memory quality through adaptive governance and quality-aware cleaning, reducing average memory noise by up to 71.8% and 68.5%, respectively. Figure 5 further reveals a strong inverse correlation between memory noise and memory–token efficiency, indicating that sustaining long-horizon efficiency requires jointly controlling memory growth and memory quality.



(a) Memory growth (b) Peak token phase (c) Token efficiency (d) Memory-Token Efficiency

Figure 6: **Long-horizon memory–token dynamics.** We compare memory growth, peak token consumption, token efficiency, and resource utilization over long-horizon execution. MANGO stabilizes memory–token dynamics while maintaining resource efficiency.

5.3 Performance on Single-Objective Multi-Hop Tasks

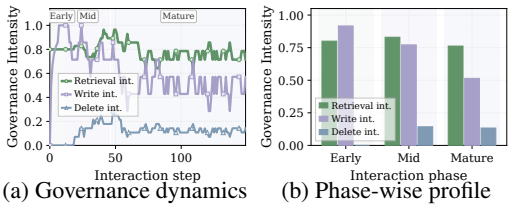
Table 1 (right) reports results on four representative single-objective multi-hop QA benchmarks: 2WikiMultihopQA, Bamboogle, HotpotQA, and MuSiQue. Although MANGO is designed for long-horizon shared-memory governance, it also delivers strong efficiency on short-horizon reasoning tasks. Across both Qwen3-1.7B and 8B, MANGO achieves the highest memory–token efficiency while reducing total token consumption and peak prompt length. Under Qwen3-8B, MANGO attains efficiency values between 3.74% and 4.66%, outperforming all baselines, whose efficiencies remain below 2.1%. It also achieves the lowest or near-lowest TT and PT, demonstrating that adaptive memory governance remains effective even for short reasoning horizons.

Memory Quality and Noise Control. Table 2 reports the average memory noise ratio during execution. MANGO maintains the lowest noise level (14.33%), compared with 29.84% for ReAct, 34.92% for MemPO, and 40.26% for Sliding-Window. The improvement is most evident on 2WikiMultihopQA and MuSiQue, reducing memory noise by over 60% relative to memory-based baselines. Although some methods achieve slightly higher EM or F1, they require substantially larger memory and token budgets. These results demonstrate that MANGO suppresses redundant memory accumulation while preserving competitive reasoning quality, achieving a superior quality–efficiency trade-off. Detailed results are provided in Appendix G.

Table 2: **Memory noise ratio (%) on merged single- and multi-objective multi-hop QA tasks.**

Single- and Multi-objective Multi-hop QA											
Method	**Qwen3-1.7B**					**Qwen3-8B**					**Avg.↓**
Objective	**16**	**8**	**4**	**2**	**1**	**16**	**8**	**4**	**2**	**1**	
FullMem	39.21	33.07	36.93	30.68	29.85	39.21	33.07	36.93	30.68	29.86	33.95
ReAct	33.86	26.68	41.04	31.66	29.74	33.86	26.68	41.04	31.66	29.94	32.62
Mem1	28.76	10.42	7.95	3.50	35.81	28.76	10.42	7.95	3.50	35.52	17.26
MemPO	10.44	4.92	4.16	3.78	35.15	10.44	4.92	4.16	3.78	34.69	11.64
A-Mem	26.57	31.94	37.86	10.82	38.10	26.57	31.94	37.86	10.82	37.14	28.96
Retrieval-only	4.77	5.49	2.96	**2.92**	37.94	4.77	5.49	2.96	2.92	31.71	10.59
Sliding-window	42.21	48.17	45.62	36.08	40.74	42.21	48.17	45.62	36.08	39.78	42.47
MANGO	**3.16**	**1.38**	**2.32**	6.69	**14.32**	**3.14**	**1.23**	**1.67**	**2.56**	**14.35**	**5.08**

Note: The 1-objective values are macro-averaged over the four single-objective datasets. The 2/4/8/16-objective values report average memory noise ratios (%) over 150-step long-horizon execution. The final column reports the macro-average over all settings and both backbones. Lower is better.



(a) Governance dynamics (b) Phase-wise profile

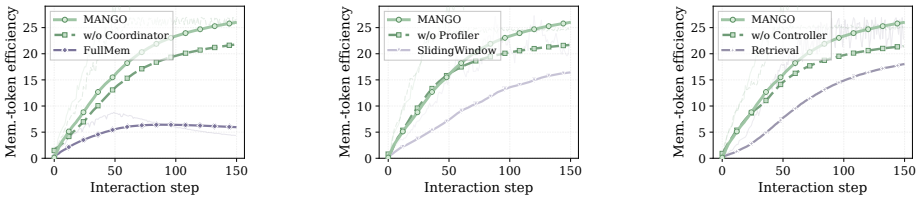
Figure 7: **Governance dynamics of MANGO.** (a) Rolling population averages of retrieval, writing, and deletion over 150 interactions. (b) Phase-averaged intensities. Governance shifts from retrieval and writing toward deletion as collaborative progress increases.

5.4 Governance Dynamics, Ablation, and Scalability

Governance dynamics. Figure 7 reveals a progress-conditioned transition from exploration to consolidation over 150 rounds in the 8-objective HotpotQA setting with Qwen3-8B. During Early execution, retrieval and writing average approximately 0.80 and 0.92, with negligible deletion. In Mid, writing decreases by 16% as deletion rises to 0.14. During Mature execution, retrieval and writing decline to 0.76 and 0.52, while deletion stabilizes near 0.13. MANGO thus preserves broad early access before regulating memory turnover to prevent unchecked growth.

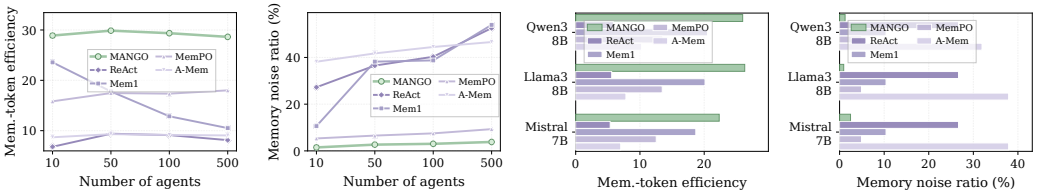
Component ablations. Figure 8 isolates each component's contribution. Removing CRC reduces efficiency from 25.99% to 21.77% and raises memory noise from 1.23% to 3.99%. Without MTSP, efficiency falls to 21.68% and average memory size grows from 22.3 to 28.3 records. Removing AGC causes the largest efficiency drop, to 21.46%, while increasing noise to 3.35%. MANGO further achieves over $4\times$ the efficiency of FullMem with 75% less memory, improves over Sliding-Window by 58.4% while reducing noise from 48.17% to 1.23%, and exceeds Retrieval-only by 44.4% with 77.6% less noise. These results establish the complementary roles of coordination, profiling, and adaptive control.

Scaling across agents and backbones. Figure 9 evaluates MANGO without model-specific tuning. From 10 to 500 agents, its efficiency varies by less than 5%, while its advantage grows with population size. At 500 agents, memory noise remains below 4%, compared with 47–54% for Mem1, ReAct, and A-Mem. Across Qwen3-8B, Llama-3.1-8B, and Mistral-7B, MANGO achieves the highest efficiency, outperforming Mem1 by 19–31%, MemPO by up to 95%, and ReAct by more than $4\times$, while keeping noise below 2.5%. The gains therefore persist across collaboration scales and model backbones.



(a) Collect. Resource Coordinator (b) Memory–Token State Profiler (c) Adapt. Governance Controller

Figure 8: **Component-wise ablation of MANGO.** Memory–token efficiency after disabling CRC, MTSP, or AGC individually while keeping the remaining components unchanged.

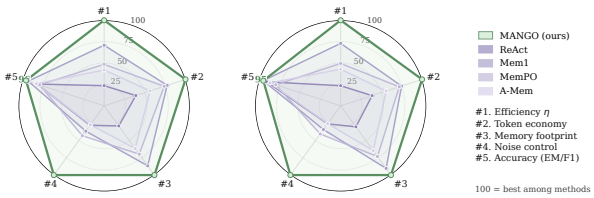


(a) Agent population (b) Model backbones

Figure 9: **Scalability of MANGO.** Memory–token efficiency and memory noise across agent populations and LLM backbones. MANGO maintains stable efficiency and consistently low memory noise as the system scales.

5.5 Overall Performance

Figure 10 summarizes performance across single- and multi-objective benchmarks in terms of reasoning quality, memory–token efficiency, token economy, memory footprint, and noise control. Across both backbones, MANGO leads all resource-oriented dimensions while maintaining competitive reasoning quality. Although individual baselines occasionally achieve higher accuracy, they incur substantially greater memory growth, token consumption, or noise accumulation. MANGO therefore provides the most balanced resource–quality profile across workloads, supporting the generality of adaptive memory governance. Additional runtime-overhead and governance-sensitivity results are provided in Appendices C and F, with further system insights in Appendix E.



(a) Qwen3-1.7B (b) Qwen3-8B

Figure 10: **Overall resource–quality profile.** Each axis reports the ratio-to-best score (100 = best), macro-averaged over the 2–16 objective settings.

6 Related Work

Memory for LLM Agents and Long-Horizon Reasoning. Recent work equips LLM agents with memory beyond a single context window. HippoRAG Jiménez Gutiérrez et al. (2024) uses graph-structured long-term memory for multi-hop knowledge integration, while M+ Wang et al. (2025b) co-trains a retriever with latent-space memory to strengthen long-range retention. Hu et al. Hu et al. (2026) nevertheless show that memory-enabled agents struggle to balance retrieval accuracy, continual learning, long-range understanding, and selective forgetting. Most approaches focus on memory representation, retrieval, or evaluation for individual agents, leaving the cumulative resource and quality effects of continuously evolving shared memory underexplored.

Multi-Agent LLM Systems and Coordination. Multi-agent LLM systems improve collaborative reasoning through communication, role specialization, reflection, and structured interaction. Chain-of-Agents Zhang et al. (2024) distributes long-context reasoning across specialized agents, COPPER Bo et al. (2024) introduces reflective collaboration and counterfactual credit assignment, and iAgents Liu et al. (2024) addresses information asymmetry through communication navigation and hybrid memory. These methods largely treat memory as a passive communication medium or knowledge repository, without explicitly coordinating its retrieval, writing, and cleanup as collective progress and runtime conditions evolve.

LLM Systems and Resource Management. LLM systems use adaptive scheduling, memory allocation, KV-cache management, and context optimization to improve serving efficiency Kwon et al. (2023); Agrawal et al. (2024); Zheng et al. (2024); pat (2024). Although these methods increase throughput and reduce inference overhead, they generally assume static memory structures. Consequently, they do not capture how decentralized memory operations jointly affect semantic memory quality, future token expenditure, and long-horizon reasoning stability.

7 Conclusion

This paper studied memory governance in long-horizon multi-agent LLM systems, where balancing reasoning effectiveness against memory growth, noise accumulation, and token expenditure poses a fundamental systems challenge. We presented MANGO, a lightweight memory–token governance framework that jointly regulates memory quality, memory growth, token expenditure, and collaborative progress through adaptive online governance. Extensive experiments on multi-objective long-horizon reasoning and single-objective multi-hop QA demonstrate that MANGO consistently achieves a superior resource–quality trade-off over existing memory-augmented agent frameworks. These results highlight that scalable long-horizon multi-agent systems require governing not only memory quantity but also memory quality, establishing memory governance as a key systems primitive for future agentic LLM systems.

Ethics Statement. We acknowledge and adhere to the ICLR Code of Ethics. This work uses publicly available datasets and pretrained language models, and involves no human participants or newly collected personal data. Nevertheless, memory-augmented agents may retain and propagate inaccurate, biased, or sensitive information. MANGO is intended for research on efficient multi-agent reasoning rather than autonomous high-stakes deployment. Practical applications should incorporate appropriate data governance, access control, content filtering, and human oversight.

Reproducibility Statement. The system formulation and MANGO design are described in Section 3, with complete procedures in Appendix D and theoretical analysis and proofs in Appendix H. Appendix B documents datasets, models, prompts, memory interfaces, hyperparameters, seeds, decoding settings, hardware, and statistical protocols. The anonymous repository provides the implementation, configurations, and scripts used for reproducibility at `https://anonymous.4open.science/r/MANGO-AE91`.

REFERENCES

Splitwise: Efficient generative llm inference using phase splitting. In *ISCA*, 2024.

Amey Agrawal et al. Sarathi: Efficient llm inference by piggybacking decodes with chunked prefills. In *USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, 2024.

Xiaohe Bo, Zeyu Zhang, Quanyu Dai, Xueyang Feng, Lei Wang, Rui Li, Xu Chen, and Ji-Rong Wen. Reflective multi-agent collaboration based on large language models. In *Advances in Neural Information Processing Systems*, 2024.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.

Taicheng Guo, Xiuying Chen, Yaqi Wang, Ruidi Chang, Shichao Pei, Nitesh V Chawla, Olaf Wiest, and Xiangliang Zhang. Large language model based multi-agents: A survey of progress and challenges. arxiv 2024. *arXiv preprint arXiv:2402.01680*, 10, 2024.

Xanh Ho, Anh-Khoa Duong Nguyen, Saku Sugawara, and Akiko Aizawa. Constructing a multi-hop qa dataset for comprehensive evaluation of reasoning steps. In *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 6609–6625, 2020.

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

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. Mistral 7b, 2023. URL `https://arxiv.org/abs/2310.06825`.

Bernal Jiménez Gutiérrez, Yiheng Shu, Yu Gu, Michihiro Yasunaga, and Yu Su. Hipporag: Neurobiologically inspired long-term memory for large language models. In *Advances in Neural Information Processing Systems*, 2024.

Woosuk Kwon et al. Efficient memory management for large language model serving with pagedattention. In *Proceedings of the ACM Symposium on Operating Systems Principles (SOSP)*, 2023.

Ruoran Li, Xinghua Zhang, Haiyang Yu, Shitong Duan, Xiang Li, Wenxin Xiang, Chonghua Liao, Xudong Guo, Yongbin Li, and Jinli Suo. Mempo: Self-memory policy optimization for long-horizon agents. In *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 23286–23301, 2026.

Wei Liu, Chenxi Wang, Yifei Wang, Zihao Xie, Rennai Qiu, Yufan Dang, Zhuoyun Du, Weize Chen, Cheng Yang, and Chen Qian. Autonomous agents for collaborative task under information asymmetry. In *Advances in Neural Information Processing Systems*, 2024.

Ofir Press, Muru Zhang, Sewon Min, Ludwig Schmidt, Noah A Smith, and Mike Lewis. Measuring and narrowing the compositionality gap in language models. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pp. 5687–5711, 2023.

Taeyun Roh, Wonjune Jang, Junha Jung, and Jaewoo Kang. Clag: Adaptive memory organization via agent-driven clustering for small language model agents. In *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 16707–16726, 2026.

Noah Shinn, Federico Cassano, Edward Berman, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. Reflexion: Language agents with verbal reinforcement learning, 2023. *URL https://arxiv. org/abs/2303.11366*, 8, 2024.

Zhen Tan, Jun Yan, I-Hung Hsu, Rujun Han, Zifeng Wang, Long T Le, Yiwen Song, Yanfei Chen, Hamid Palangi, George Lee, et al. In prospect and retrospect: Reflective memory management for long-term personalized dialogue agents. arxiv preprint, 2025c. *URL https://arxiv. org/abs/2503.08026*.

Harsh Trivedi, Niranjan Balasubramanian, Tushar Khot, and Ashish Sabharwal. musique: Multihop questions via single-hop question composition. *Transactions of the Association for Computational Linguistics*, 10:539–554, 2022.

Yu Wang, Dmitry Krotov, Yuanzhe Hu, Yifan Gao, Wangchunshu Zhou, Julian McAuley, Dan Gutfreund, Rogerio Feris, and Zexue He. M+: Extending memoryllm with scalable long-term memory. *arXiv preprint arXiv:2502.00592*, 2025a.

Yu Wang, Dmitry Krotov, Yuanzhe Hu, Yifan Gao, Wangchunshu Zhou, Julian McAuley, Dan Gutfreund, Rogerio Feris, and Zexue He. M+: Extending memoryllm with scalable long-term memory. In *Proceedings of the 42nd International Conference on Machine Learning*, 2025b.

Ziqi Wang, Xiaoyu Xia, Minhui Xue, Ibrahim Khalil, Minghui Liwang, and Xun Yi. Gees: Enabling location privacy-preserving energy saving in multi-access edge computing. In *Proceedings of the ACM Web Conference 2024*, pp. 2735–2746, 2024.

Ziqi Wang, Xiaoyu Xia, Ibrahim Khalil, Tianxu Lan, Feng Liu, Xiaolong Xu, Xun Yi, Minhui Xue, and Elisa Bertino. Mera: A green edge resource control system with privacy-preservation via mean-field reinforcement learning. *IEEE Transactions on Knowledge and Data Engineering*, 2026.

Guangxuan Xiao, Yuandong Tian, Beidi Chen, Song Han, and Mike Lewis. Efficient streaming language models with attention sinks. In *International Conference on Learning Representations*, volume 2024, pp. 21875–21895, 2024.

Peng Xu, Wei Ping, Xianchao Wu, Lawrence McAfee, Chen Zhu, Zihan Liu, Sandeep Subramanian, Evelina Bakhturina, Mohammad Shoeybi, and Bryan Catanzaro. Retrieval meets long context large language models. In *International Conference on Learning Representations*, volume 2024, pp. 49569–49584, 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.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, et al. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025.

Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William Cohen, Ruslan Salakhutdinov, and Christopher D Manning. Hotpotqa: A dataset for diverse, explainable multi-hop question answering. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, pp. 2369–2380, 2018.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*, 2022.

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. *arXiv preprint arXiv:2601.01885*, 2026.

Yusen Zhang, Ruoxi Sun, Yanfei Chen, Tomas Pfister, Rui Zhang, and Sercan O. Arik. Chain of agents: Large language models collaborating on long-context tasks. In *Advances in Neural Information Processing Systems*, 2024.

Lianmin Zheng et al. Sglang: Efficient execution of structured language model programs. *arXiv preprint arXiv:2312.07104*, 2024.

Zijian Zhou, Ao Qu, Zhaoxuan Wu, Sunghwan Kim, Alok Prakash, Daniela Rus, Jinhua Zhao, Bryan Kian Hsiang Low, and Paul Pu Liang. Mem1: Learning to synergize memory and reasoning for efficient long-horizon agents. *arXiv preprint arXiv:2506.15841*, 2025.

A Summary of Notations

For convenience, Table 3 summarizes the main notations used throughout the paper.

Table 3: Summary of key notations used throughout the paper.

Notation	Meaning
	System Variables
$t \in [0, T]$	Discrete execution step.
$\mathcal{N}$	Set of collaborative LLM agents.
N	Number of agents.
$X_t = (\mathcal{S}_t, m_t)$	Global system state.
$\mathcal{S}_t = \{x_t^i\}_{i \in \mathcal{N}}$	Collection of local agent states.
x_t^i	Local runtime state of agent i.
m_t	Shared memory–token state.
s_t	Shared memory size.
$\bar{\eta}_t$	Average memory noise ratio.
H_t	Cumulative token consumption.
$o_t^i = (x_t^i, m_t)$	Local observation of agent i.
u_t^i	Task progress of agent i.
$\bar{u}_t$	Average collaborative progress.
	Memory Governance
$a_t^i = (k_t^i, w_t^i, d_t^i)$	Memory operation of agent i.
k_t^i	Retrieval intensity.
w_t^i	Memory write action.
d_t^i	Memory deletion intensity.
τ_t^w	Adaptive write threshold.
	Population Coordination
μ_t	Empirical population measure.
Ω_t	Population operation load.
ζ_t	Coordination signal generated by CRC.
	System Metrics
$E_t(a)$	Memory–token efficiency.
P_t	Runtime memory–token pressure.
Γ_t	Adaptive governance intensity.
$\mathbf{c} = (c_k, c_w, c_d)$	Token cost coefficients for retrieval, writing, and deletion.
	Resource Constraints
$S_{\max}$	Maximum shared-memory capacity.
$H_{\max}$	Maximum token budget.

B Detailed Experimental Setup

Tasks and datasets. We evaluate shared-memory reasoning across single- and multi-objective settings. The multi-objective workloads cover fact aggregation, sequential reasoning, collaborative planning, and noisy-memory retrieval, instantiated using *HotpotQA* Yang et al. (2018), *2WikiMultihopQA* Ho et al. (2020), *Bamboogle* Press et al. (2023), and *MuSiQue* Trivedi et al. (2022). We additionally use synthetic collaborative environments to isolate memory growth and memory–token efficiency during prolonged execution. The single-objective setting evaluates the four multi-hop QA datasets individually.

Reasoning horizons and system scale. The multi-objective experiments consider 2, 4, 8, and 16 concurrent reasoning objectives, while the single-objective experiments represent shorter task-specific reasoning workloads. We evaluate different execution horizons by varying the number of collaborative rounds and the frequency of memory interaction. Unless otherwise stated, diagnostic analyses use the 8-objective setting with Qwen3-8B over 150 interaction rounds. Scalability experiments further vary the collaborating population from 10 to 500 agents and test multiple model backbones.

Models and shared-memory protocol. We implement all methods using *Qwen3-1.7B* and *Qwen3-8B* Yang et al. (2025). Backbone generalization is evaluated using *Llama-3.1-8B* Grattafiori et al. (2024) and *Mistral-7B* Jiang et al. (2023) without model-specific tuning. All agents interact through a unified shared-memory repository. MANGO operates entirely at test time and updates only memory actions and runtime states, without modifying model parameters. For controlled comparison, all methods use identical prompts, memory interfaces, memory initialization, and inference budgets, together with common retrieval and memory settings where applicable.

Evaluation metrics. Reasoning quality is measured using Exact Match (EM) and token-level F1 between normalized predictions and ground-truth answers. Total token consumption measures cumulative usage over execution, while peak prompt length captures the largest per-step context. We also report average shared-memory size, memory noise ratio, and inference latency. The memory noise ratio measures the proportion of low-value or redundant records accumulated during execution. Memory–token efficiency $E_t(a)$, defined in Definition 2, is the primary system-level metric and relates collaborative progress to memory occupancy, token expenditure, and memory noise.

Baselines. The baselines cover three complementary memory-management regimes. *FullMem* retains all shared-memory records without governance, providing an unrestricted-memory reference. *ReAct* Yao et al. (2022) interleaves reasoning and action generation. Memory-augmented methods include *Mem1* Zhou et al. (2025), which introduces memory-aware retrieval for long-context agents; *MemPO* Li et al. (2026), which organizes memory through preference-aware selection; and *A-Mem* Xu et al. (2026), which dynamically regulates memory usage. We additionally evaluate two controlled memory policies: *Retrieval-only* Xu et al. (2024) retrieves existing records without adaptive writing or deletion, whereas *Sliding-window* Xiao et al. (2024) retains only recent records within a fixed-size window.

Reproducibility Configuration and Statistical Protocol. We distinguish two evaluation regimes. The *broad* experiments (multi-objective, single-objective, ablation, scalability) are run with seed 42 and reported descriptively. The *paired long-horizon study* uses three independent seeds $\{42, 43, 44\}$: within each seed, all compared methods share the same objective set, visitation order, and initial memory state, ensuring a controlled comparison.

The study uses `Qwen3-8B` with greedy decoding (temperature 0, Qwen3), applied uniformly across all methods. The initial shared memory contains 16 records with a controlled mixture of relevant and distractor entries, kept identical across methods within each seed. MANGO's hyperparameters follow the default configuration released with the code. For single-objective experiments, each dataset provides 50 questions evaluated over at most 16 interaction steps.

We treat the independently seeded trajectory as the primary statistical unit. Confidence intervals for trajectory curves are obtained via moving-block bootstrap to account for temporal correlation. Experiments ran on NVIDIA GeForce RTX 3090 GPUs. Complete configurations and analysis scripts are provided in the released artifact.

B.1 Memory Quality and Noise Evaluation

This subsection details the measurement protocol for memory noise and retrieval quality used across all experiments.

Record structure. Each memory record is a tuple

$$m = \big(x_m,\, \ell_m,\, t_m,\, a_m,\, Q_m,\, z_m\big), \tag{14}$$

where x_m is the stored text, ℓ_m is its task topic, $t_m \in \mathbb{N}$ is the creation step, $a_m \geq 0$ is the cumulative access count, $Q_m \in [0,1]$ is stored quality metadata, and $z_m \in \{0,1\}$ is a noise indicator. We set $z_m = 1$ for any record produced by the predefined distractor or noise-generation branch, and $z_m = 0$ for a relevant or verified record. The indicator is assigned deterministically at write time and is never modified thereafter, making the noise measurement independent of model predictions.

Memory-noise ratio. The noise ratio of the shared memory at step t is

$$\rho_t^{\mathrm{eval}} = \frac{\big|\{m \in \mathcal{M}_t : z_m = 1\}\big|}{\max\big(1,\, s_t\big)}, \tag{15}$$

where $s_t = |\mathcal{M}_t|$. Because z_m is a deterministic binary label fixed at write time, ρ_t^{eval} is threshold-free and exactly reproducible across methods, seeds, and execution horizons. We distinguish ρ_t^{eval} from the runtime noise ratio $\bar{\eta}_t$ in Definition 1: the two coincide under perfect noise labelling, but ρ_t^{eval} serves purely as a diagnostic and does not influence any governance decision.

Retrieval quality score. To assess the usefulness of retrieved context at step t, we evaluate:

$$\widehat{q}_t = \mathrm{clip}_{[0,1]}\Big(\lambda_r\, r_t + \lambda_c\, c_t + \lambda_n\, (1 - n_t)\Big), \tag{16}$$

where $\lambda_r, \lambda_c, \lambda_n \geq 0$ with $\lambda_r + \lambda_c + \lambda_n = 1$. The terms r_t, c_t, and n_t measure on-topic coverage, query–record similarity, and retrieved-context noise, respectively. We set $\widehat{q}_t = 0$ when $\mathcal{R}_t = \varnothing$. All coefficients are fixed by default and remain unchanged across settings; their exact values are included in the released experimental configuration.

Summary statistics. We report the following aggregate metrics over a horizon of T steps:

$$\overline{s} = \frac{1}{T}\sum_{t=1}^{T} s_t, \qquad \overline{\rho} = \frac{1}{T}\sum_{t=1}^{T} \rho_t^{\mathrm{eval}}, \qquad \rho_{\mathrm{final}} = \rho_T^{\mathrm{eval}}, \tag{17}$$

where $\overline{s}$ is the average repository size, $\overline{\rho}$ is the average memory-noise ratio, and ρ_{final} is the noise ratio at the final interaction step. The record-labelling rule, similarity computation, and metric aggregation in equation 17 are applied identically across all compared methods; only the writing, retention, and deletion policies differ. Memory statistics are evaluated independently of task correctness, which is measured by Exact Match and token-level F1.

C Runtime Overhead Analysis

To quantify the deployment overhead of MANGO, Table 4 reports the average per-step runtime and additional working memory of each governance module under the Qwen3-8B setting with eight collaborative agents. CRC, MTSP, and AGC incur only 0.0261 ms, 0.0138 ms, and 0.0722 ms per interaction step, respectively, resulting in a total governance overhead of 0.1121 ms.

The additional working memory is similarly small, requiring 2.35 KB, 1.97 KB, and 11.45 KB for the three modules, respectively, for a total overhead of only 15.77 KB. Compared with the runtime and memory footprint of a single LLM inference step, this overhead is negligible. These measurements complement the complexity analysis in Section 4, demonstrating that the lightweight theoretical complexity of MANGO translates into minimal runtime overhead in practice.

Algorithm 1 Collective Resource Coordinator (CRC)

Require: Local states $\{x_t^i\}_{i\in\mathcal{N}}$; previous memory actions $\{a_{t-1}^i\}_{i\in\mathcal{N}}$; agent progress $\{u_t^i\}_{i\in\mathcal{N}}$; global memory-token state m_t
Ensure: Population-level coordination signal ζ_t
1: Construct the empirical population state $\mu_t \leftarrow \frac{1}{|\mathcal{N}|}\sum_{i\in\mathcal{N}} \delta_{x_t^i}$
2: Compute the average population state $\bar{x}_t \leftarrow \frac{1}{|\mathcal{N}|}\sum_{i\in\mathcal{N}} x_t^i$
3: Compute average progress $\bar{u}_t \leftarrow \frac{1}{|\mathcal{N}|}\sum_{i\in\mathcal{N}} u_t^i$
4: Compute the remaining collaborative demand $\Delta_t \leftarrow 1 - \bar{u}_t$
5: Estimate the memory-operation load $\Omega_t \leftarrow \frac{1}{|\mathcal{N}|}\sum_{i\in\mathcal{N}} \left(\omega_r r_{t-1}^i + \omega_w w_{t-1}^i + \omega_d d_{t-1}^i\right)$
6: Construct the coordination signal $\zeta_t \leftarrow (\bar{x}_t, \bar{u}_t, \Delta_t, \Omega_t, m_t)$
7: **return** ζ_t

Algorithm 2 Memory-Token State Profiler (MTSP)

Require: Coordination signal ζ_t; shared memory $\mathcal{M}_t$; cumulative token consumption H_t; budgets $S_{\max}$ and $H_{\max}$; quality threshold θ_q
Ensure: Runtime pressure P_t and memory-noise ratio ρ_t
1: Measure the current memory size $s_t \leftarrow |\mathcal{M}_t|$
2: Compute normalized memory pressure $p_t^m \leftarrow s_t / S_{\max}$
3: Compute normalized token pressure $p_t^h \leftarrow H_t / H_{\max}$
4: Estimate the memory-noise ratio $\rho_t \leftarrow |\{m \in \mathcal{M}_t : q(m) < \theta_q\}| / \max(1, s_t)$
5: Compute the runtime pressure P_t according to Equation (7)
6: **return** (P_t, ρ_t)

D MANGO GOVERNANCE PROCEDURES

This appendix presents the complete execution procedures and pseudo-code of MANGO. At each reasoning step, the *Collective Resource Coordinator (CRC)* first constructs a population-level coordination signal, as detailed in Algorithm 1. The *Memory-Token State Profiler (MTSP)* then estimates the current runtime pressure and memory noise through Algorithm 2. Finally, the *Adaptive Governance Controller (AGC)* converts these signals into agent-specific retrieval, writing, and deletion actions following Algorithm 3. These procedures are executed sequentially throughout the reasoning horizon.

The CRC aggregates local agent states, progress, and previous memory operations. It exposes their collective effect through a compact coordination signal without replacing the reasoning policy of any individual agent.

The MTSP converts raw runtime measurements into a compact state that jointly reflects memory growth, token consumption, and memory noise. The resulting state allows MANGO to distinguish useful memory accumulation from increasingly costly or unreliable memory.

The AGC closes the feedback loop. It combines collective demand with the profiled memory-token state and produces agent-specific controls over the complete memory lifecycle.

Table 4: **Governance overhead of MANGO.** Average per-step runtime and additional working memory of each governance module under the Qwen3-8B setting with eight collaborative agents.

Method	Avg. Runtime (ms)	Extra Memory (KB)
CRC	0.0261	2.35
MTSP	0.0138	1.97
AGC	0.0722	11.45
MANGO (Total)	**0.1121**	**15.77**

Algorithm 3 Adaptive Governance Controller (AGC)

Require: Coordination signal ζ_t; runtime pressure P_t; memory-noise ratio ρ_t; shared memory $\mathcal{M}_t$
Ensure: Agent-specific memory actions $\{a_t^i\}_{i\in\mathcal{N}}$
for each agent $i \in \mathcal{N}$ **do**
  Compute the governance intensity Γ_t^i according to Equation (8)
  Determine the retrieval budget r_t^i according to Equation (9)
  Evaluate each candidate record and determine $w_t^i(m)$
  Determine the deletion intensity d_t^i according to Equation (10)
  Rank retained records by importance, redundancy, and recency
  Remove the lowest-ranked records until the deletion target d_t^i is reached
  Set the memory action $a_t^i \leftarrow (r_t^i, w_t^i, d_t^i)$
end for
return $\{a_t^i\}_{i\in\mathcal{N}}$

E System Insights

The empirical results suggest that MANGO does not merely reduce memory usage, but fundamentally changes the evolution of the shared-memory system. To better understand this behavior, we analyze the coupled dynamics induced by adaptive memory governance.

Recall that the shared-memory evolution follows Eq. (1), while the governance intensity is determined by Eq. (8). Combining the two equations yields the expected memory increment:

$$\mathbb{E}[\Delta s_t] = \mathbb{E}\left[\sum_i w_t^i\right] - \mathbb{E}\left[\sum_i d_t^i\right]. \tag{18}$$

Since both writing and deletion are governed by the adaptive governance intensity Γ_t, stronger governance simultaneously suppresses unnecessary writes and promotes memory cleanup. Consequently,

$$\frac{\partial \mathbb{E}[\Delta s_t]}{\partial \Gamma_t} < 0, \tag{19}$$

indicating that increasing governance pressure continuously reduces the expected memory growth rate.

Similarly, the expected increase in cumulative token consumption satisfies

$$\mathbb{E}[\Delta H_t] = c_k \mathbb{E}[k_t] + c_w \mathbb{E}[w_t] + c_d \mathbb{E}[d_t], \tag{20}$$

where retrieval, writing, and deletion contribute to the incremental token cost and are jointly regulated according to the runtime state. As collaborative progress advances, retrieval demand gradually decreases while deletion becomes increasingly active, shifting the system from exploration-oriented execution toward memory consolidation. To characterize this transition, we define the exploration–consolidation ratio:

$$R_t = \frac{\mathbb{E}[k_t] + \mathbb{E}[w_t]}{\mathbb{E}[d_t] + \epsilon}, \tag{21}$$

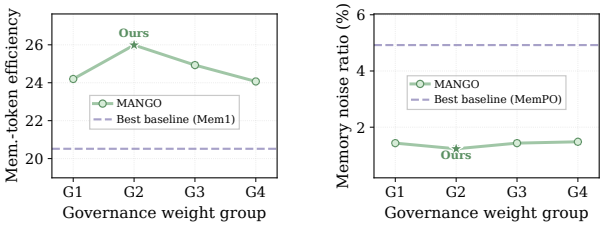
where larger values correspond to exploration-dominated execution, whereas smaller values indicate consolidation-oriented memory management. As illustrated in Figure 7, R_t decreases steadily throughout execution, showing that MANGO progressively shifts from aggressive memory expansion during early reasoning to selective memory maintenance as collaborative progress advances.

This transition also explains the stable memory–token dynamics observed in Figure 6. Rather than explicitly minimizing memory size, MANGO continuously reduces the marginal growth rate of shared memory, thereby limiting future retrieval overhead and preventing long-term amplification of token consumption. Consequently, the system gradually converges to a resource-efficient operating regime while maintaining collaborative reasoning performance.

Discussion and Insights. Together, the theoretical analysis and experimental observations reveal several system-level insights. First, memory quality is more critical than memory quantity: effective governance should regulate the *evolution* of shared memory rather than simply limiting its size, since low-value or redundant records continuously increase future retrieval overhead while contributing

Table 5: **Governance-intensity weight groups** used in the sensitivity analysis. Group G2 is the default configuration used in the main text.

Group	κ_1	κ_2	κ_3
G1	0.6	0.8	1.3
G2 (Ours)	**1.0**	**1.0**	**1.0**
G3	1.3	1.2	0.8
G4	1.5	1.3	0.6



(a) Memory–token efficiency (b) Memory noise ratio

Figure 11: **Sensitivity of MANGO to governance-intensity weight groups.** Memory–token efficiency and memory noise remain stable under moderate parameter perturbations, demonstrating the robustness of the proposed governance policy.

little to collaborative reasoning. Second, memory–token efficiency emerges from stable system dynamics rather than explicit token minimization. By adaptively coordinating retrieval, writing, and cleanup, MANGO progressively suppresses memory growth and prevents long-term resource amplification through closed-loop feedback. Finally, governance naturally evolves with collaborative progress, shifting from exploration-oriented retrieval and writing toward consolidation-oriented memory maintenance as execution proceeds. These observations consistently explain the theoretical properties and empirical gains of MANGO, and suggest a general design principle for resource-efficient long-horizon multi-agent LLM systems.

F Sensitivity Analysis

To evaluate the robustness of MANGO to governance parameter selection, we vary the governance-intensity weights $(\kappa_1, \kappa_2, \kappa_3)$ in Eq. equation 8:

$$\Gamma_t = \kappa_1 P_t + \kappa_2 \Omega_t + \kappa_3 (1 - \bar{u}_t). \tag{22}$$

Here, P_t, Ω_t, and $(1 - \bar{u}_t)$ denote the runtime memory–token pressure, population-level operation load, and remaining collaborative demand, respectively. All other hyperparameters are fixed to the main experimental protocol (8-objective HotpotQA+NQ, Qwen3-8B, 150 interaction steps). We compare four weight groups (Table 5), where G2 is the default configuration used throughout the paper and the remaining groups perturb the relative contributions of the three control factors.

Figure 11 shows that MANGO remains stable under moderate perturbations of the governance-intensity weights. Although G2 consistently achieves the best overall trade-off, the neighboring configurations exhibit only minor variations in both memory–token efficiency and memory noise ratio, indicating that the framework does not rely on finely tuned parameter values. These empirical observations are consistent with Proposition 1, suggesting that the theoretical stability guarantee also translates into robust practical behavior under parameter perturbations.

G Detailed Experimental Results

In this section, we provide the complete single-objective multi-hop QA results reported in Tables 6 and 7. Table 6 reports the detailed memory–token efficiency, total token consumption, and peak prompt length for all four datasets under both Qwen3-1.7B and Qwen3-8B. Across all settings, MANGO consistently achieves the highest memory–token efficiency while maintaining substantially lower token consumption and prompt lengths than competing memory-augmented methods. These results demonstrate that the proposed governance strategy effectively improves resource utilization even in shorter reasoning workloads, rather than only under long-horizon multi-objective execution.

Table 7 further reports the corresponding memory noise ratios. Compared with all baselines, MANGO consistently maintains the cleanest shared memory throughout execution, indicating that

adaptive retrieval regulation and quality-aware memory cleaning remain effective even when memory growth is relatively moderate. Together, these detailed results reinforce the observations in the main text, showing that MANGO simultaneously improves resource efficiency and memory quality across diverse multi-hop reasoning tasks and model backbones.

Table 6: **Results on Single-objective multi-hop QA datasets.**

Single-objective Multi-hop QA															
Dataset		**2WikiMultihopQA**			**Bamboogle**			**HotpotQA**			**Musique**			**Avg**	
Backbone	**Method**	**TT**	**PT**	**Eff**	**TT**	**PT**	**Eff**	**TT**	**PT**	**Eff**	**TT**	**PT**	**Eff**	**EM**	**F1**
Qwen3-1.7B	FullMem	13.51	0.27	1.70	15.57	0.28	1.72	15.06	0.62	1.77	15.21	0.31	1.64	25.78	33.50
	ReAct	24.74	0.44	1.89	26.87	0.43	1.97	26.82	1.03	1.54	25.82	0.49	1.88	25.04	30.50
	Mem1	13.67	0.27	1.38	15.78	0.30	1.47	14.59	0.62	1.53	15.88	0.32	1.36	25.34	32.50
	MemPO	13.59	0.29	1.47	15.74	0.31	1.49	13.83	0.55	1.60	15.71	0.29	1.54	28.72	**34.50**
	A-Mem	19.33	0.40	1.35	23.13	0.39	1.33	21.98	0.97	1.30	21.98	0.41	1.29	**29.30**	34.00
	Retrieval-only	17.55	0.36	1.24	20.86	0.34	1.33	17.02	0.83	1.46	19.41	0.38	1.38	26.17	33.00
	Sliding-window	17.60	0.37	1.21	20.92	0.37	1.30	20.01	0.86	1.40	20.34	0.40	1.26	27.89	32.50
	MANGO	**12.31**	**0.25**	**4.77**	**13.61**	**0.25**	**4.52**	**8.51**	**0.35**	**3.65**	**12.73**	**0.28**	**4.61**	26.68	25.00
Qwen3-8B	FullMem	14.79	0.28	1.81	15.01	0.30	1.77	13.21	0.56	1.68	14.88	0.34	1.66	30.65	32.50
	ReAct	23.72	0.44	1.88	26.33	0.44	2.03	20.52	1.02	1.32	24.48	0.49	1.89	33.59	35.50
	Mem1	14.25	0.30	1.45	15.56	0.30	1.59	11.42	0.55	1.51	15.52	0.33	1.35	33.02	36.50
	MemPO	14.58	0.29	1.52	15.48	0.30	1.58	9.87	0.53	1.41	14.80	0.32	1.52	32.18	37.00
	A-Mem	20.55	0.42	1.39	22.05	0.42	1.35	15.02	0.88	1.21	22.25	0.44	1.35	34.84	37.50
	Retrieval-only	12.54	0.27	1.57	12.79	0.29	1.73	10.64	0.61	1.81	12.48	0.30	1.74	32.20	34.00
	Sliding-window	18.03	0.38	1.24	20.17	0.36	1.33	13.87	0.76	1.32	18.88	0.41	1.24	**36.93**	**39.50**
	MANGO	**12.16**	**0.26**	**4.66**	**12.20**	**0.27**	**4.59**	**6.52**	**0.37**	**3.74**	**12.45**	**0.29**	**4.60**	35.15	31.00

Note: TT denotes total token consumption ($\times 10^4$), PT denotes peak prompt length ($\times 10^3$), and Eff. denotes memory–token efficiency η (higher is better). Results are averaged over 50 questions with 16 interaction steps. Avg. F1/EM are macro-averages across all datasets. Best results are highlighted.

Table 7: **Memory noise ratio (%) on single-objective multi-hop QA.**

Single-objective Multi-hop QA									
Method	**Qwen3-1.7B**				**Qwen3-8B**				**Avg.↓**
Dataset	**2Wiki**	**Bamb.**	**Hotpot**	**MuSiQue**	**2Wiki**	**Bamb.**	**Hotpot**	**MuSiQue**	
FullMem	28.43	30.53	27.31	33.13	28.69	30.23	27.41	33.12	29.91
ReAct	29.49	28.61	32.01	28.86	30.18	29.26	31.74	28.57	29.84
Mem1	36.12	35.82	33.04	38.27	36.13	35.37	32.14	38.43	35.79
MemPO	35.00	36.63	31.94	37.03	35.88	35.46	31.07	36.34	34.92
A-Mem	37.18	39.52	36.40	39.29	37.13	39.22	33.44	38.77	37.62
Retrieval	39.61	41.48	32.47	38.20	33.47	33.72	28.91	30.72	34.82
SlideWin	39.75	42.81	37.07	43.34	40.39	42.21	34.04	42.49	40.26
MANGO	**12.50**	**12.96**	**20.96**	**10.86**	**12.56**	**12.90**	**21.35**	**10.58**	**14.33**

Note: Values denote average memory noise ratios (%) over the interaction horizon. The final column reports the macro-average over all settings. Lower is better.

H Additional Theoretical Results and Proofs

This appendix formalizes the mechanisms summarized in Section 4. The analysis establishes sufficient conditions under which MANGO exhibits bounded memory and noise dynamics, permutation-invariant population coordination, local robustness, and improved memory–token efficiency. These results characterize the operating regime of the framework rather than asserting unconditional convergence or performance improvement for every workload.

H.1 Analytical Setup and Sufficient Conditions

Definition 3 (Closed-Loop Memory–Token Dynamics)**.** *Under the MANGO policy, the memory–token state evolves as*

$$m_{t+1} = \Phi(m_t, \mu_t, a_t, \xi_t), \tag{23}$$

where $m_t = (s_t, \bar{\eta}_t, H_t)$, $a_t = \{a_t^i\}_{i \in \mathcal{N}}$, and ξ_t represents stochastic retrieval and reasoning outcomes.

Let $\mathcal{F}_t$ denote the execution history, $W_t = \sum_{i \in \mathcal{N}} w_t^i$ the number of admitted writes, $D_t = \sum_{i \in \mathcal{N}} d_t^i$ the number of removed records, and $c_t = H_{t+1} - H_t$ the token increment. We consider the following sufficient conditions.

(C1) **Bounded feasible actions.** For finite $W_{\max}$, $D_{\max}$, and $c_{\max}$,

$$0 \le W_t \le W_{\max}, \qquad 0 \le D_t \le \min\{D_{\max}, s_t + W_t\}, \qquad 0 \le c_t \le c_{\max}. \tag{24}$$

This condition follows when retrieval, writing, and deletion actions are projected onto their feasible ranges.

(C2) **Pressure-responsive memory drift.** For some $\kappa_s \in (0, 1]$ and $b_s \ge 0$,

$$\mathbb{E}[D_t - W_t \mid \mathcal{F}_t] \ge \kappa_s s_t - b_s. \tag{25}$$

This condition describes a regime in which cleanup eventually offsets admitted writes as memory pressure increases.

(C3) **Controlled noise drift.** For some $\kappa_\eta \in (0, 1]$ and $b_\eta \ge 0$,

$$\mathbb{E}[\bar{\eta}_{t+1} \mid \mathcal{F}_t] \le (1 - \kappa_\eta)\bar{\eta}_t + b_\eta. \tag{26}$$

This condition describes a regime in which quality-aware admission and cleanup prevent persistent noise amplification.

(C4) **Exchangeable population dynamics.** Conditional on the current memory–token state and collective action, the memory transition and one-step resource cost are invariant to permutations of agent identities and depend on the agent configuration through its empirical state distribution.

Conditions (C2) and (C3) are drift certificates rather than universal properties of every execution. For example, if $\mathbb{E}[W_t \mid \mathcal{F}_t] \le \bar{W}$ and $\mathbb{E}[D_t \mid \mathcal{F}_t] \ge \kappa_s s_t - \bar{D}$, then (C2) holds with $b_s = \bar{W} + \bar{D}$. The ratio b_s/κ_s therefore characterizes the memory scale induced by the balance between admitted writes and cleanup.

H.2 Persistent-Memory Externality

Lemma 2 (Persistent-Memory Externality)**.** *Consider a record m admitted at step t. Let $p_{\tau,j}(m)$ denote the probability that m remains available and is retrieved by agent j at a future step $\tau > t$. Its expected downstream context-token cost is $\mathrm{Ext}_t(m)$ in Equation* (11)*. Hence, $\mathrm{Ext}_t(m) \ge 0$, and it is nondecreasing in the number of future retrieval opportunities.*

Proof. Let $I_{\tau,j}(m)$ indicate whether record m remains in shared memory and enters the retrieved context of agent j at step τ. Its downstream context-token contribution is

$$C_t(m) = \ell(m) \sum_{\tau=t+1}^{T} \sum_{j \in \mathcal{N}} I_{\tau,j}(m). \tag{27}$$

By linearity of expectation,

$$\mathbb{E}[C_t(m)] = \ell(m) \sum_{\tau=t+1}^{T} \sum_{j \in \mathcal{N}} \mathbb{E}[I_{\tau,j}(m)] \tag{28}$$

$$= \ell(m) \sum_{\tau=t+1}^{T} \sum_{j \in \mathcal{N}} p_{\tau,j}(m) = \mathrm{Ext}_t(m). \tag{29}$$

Every term is nonnegative because $\ell(m) \ge 0$ and $p_{\tau,j}(m) \ge 0$. Extending the horizon or adding a retrieval opportunity therefore adds only nonnegative terms. □

The lemma quantifies the direct context cost induced by persistent reuse. It neither requires monotonic memory growth nor includes secondary reasoning costs caused by processing the retrieved record; such effects may further increase the externality.

H.3 Population-Level Representation

The empirical population state is

$$\mu_t = \frac{1}{N} \sum_{i \in \mathcal{N}} \delta_{(x_t^i, a_{t-1}^i, u_t^i)}. \tag{30}$$

It preserves the population configuration while removing dependence on arbitrary agent identities.

Lemma 3 (Population-Level Sufficiency under Exchangeability)**.** *Under (C4), μ_t is sufficient for predicting the next memory–token state and any permutation-invariant one-step resource cost. In particular, for every measurable set $\mathcal{B}$, there exists a transition kernel K such that*

$$\Pr(m_{t+1} \in \mathcal{B} \mid \mathcal{F}_t, a_t) = K(\mathcal{B} \mid \mu_t, m_t, a_t)\,. \tag{31}$$

Any coordination signal ζ_t through which this kernel further factorizes inherits the same sufficiency property.

Proof. Under (C4), permuting agent identities does not change the conditional transition or one-step resource cost. Ordered configurations containing the same collection of local state–action–progress tuples therefore belong to the same equivalence class represented by μ_t. Consequently, their conditional transitions can be expressed through a common kernel depending on μ_t, m_t, and a_t, which gives Equation (31).

If there exists a kernel $\widetilde{K}$ such that

$$K(\mathcal{B} \mid \mu_t, m_t, a_t) = \widetilde{K}(\mathcal{B} \mid \zeta_t, a_t), \tag{32}$$

then the ordered collection of agent identities provides no additional transition information beyond ζ_t. □

The lemma does not imply that arbitrary population averages preserve every heterogeneous interaction. MANGO uses a structured low-dimensional coordination signal as a practical approximation; exact sufficiency requires the stated factorization.

H.4 Proof of Closed-Loop Resource Stability

Proof of Theorem 1. From Equation (1),

$$s_{t+1} = s_t + W_t - D_t. \tag{33}$$

Applying Equation (12) gives

$$\mathbb{E}[s_{t+1} \mid \mathcal{F}_t] = s_t - \mathbb{E}[D_t - W_t \mid \mathcal{F}_t] \tag{34}$$

$$\leq (1 - \kappa_s) s_t + b_s. \tag{35}$$

Taking total expectations and recursively expanding yield

$$\mathbb{E}[s_t] \leq (1 - \kappa_s)^t s_0 + \frac{b_s}{\kappa_s} \left(1 - (1 - \kappa_s)^t\right). \tag{36}$$

Therefore,

$$\sup_{t \geq 0} \mathbb{E}[s_t] \leq \max\left\{ s_0, \frac{b_s}{\kappa_s} \right\}. \tag{37}$$

Applying the same argument to Equation (12) gives

$$\mathbb{E}[\bar{\eta}_t] \leq (1 - \kappa_\eta)^t \bar{\eta}_0 + \frac{b_\eta}{\kappa_\eta} \left(1 - (1 - \kappa_\eta)^t\right), \tag{38}$$

and hence

$$\sup_{t \geq 0} \mathbb{E}[\bar{\eta}_t] \leq \max\left\{ \bar{\eta}_0, \frac{b_\eta}{\kappa_\eta} \right\}. \tag{39}$$

If the token-budget gate enforces $c_t \leq H_{\max} - H_t$, then

$$H_{t+1} = H_t + c_t \leq H_{\max} \tag{40}$$

whenever $H_t \leq H_{\max}$. Induction from $H_0 \leq H_{\max}$ establishes the token bound. □

The memory bound is informative when b_s/κ_s lies below the scale reached by unmanaged execution. Similarly, the noise bound is nontrivial when $b_\eta/\kappa_\eta < 1$; otherwise, only the natural feasibility bound $\bar{\eta}_t \in [0,1]$ is obtained. The theorem is therefore a sufficient stability certificate rather than an unconditional convergence result.

H.5 Feasible Region and Token Boundary

Corollary 1 (Feasible Invariant Region)**.** *Suppose memory and token actions are projected onto their feasible sets. If the system is initialized in*

$$\mathcal{R} = \{(s, \bar{\eta}, H) : 0 \le s \le S_{\max},\ 0 \le \bar{\eta} \le 1,\ 0 \le H \le H_{\max}\}, \tag{41}$$

then $\mathcal{R}$ is invariant under the resulting closed-loop dynamics.

Proof. Feasible projection prevents admitted writes from violating $S_{\max}$ and deletion from exceeding available memory. Hence, $0 \le s_{t+1} \le S_{\max}$ whenever $0 \le s_t \le S_{\max}$. Since $\bar{\eta}_{t+1}$ is a ratio, it remains in $[0,1]$. The token-budget gate similarly guarantees $0 \le H_{t+1} \le H_{\max}$. □

Corollary 2 (Boundary of Cumulative-Token Stability)**.** *Without a finite token budget or termination condition, if $\mathbb{E}[c_t \mid \mathcal{F}_t] \ge c_{\min} > 0$, then*

$$\mathbb{E}[H_t] \ge H_0 + t c_{\min}. \tag{42}$$

Consequently, cumulative token consumption cannot remain uniformly bounded over an infinite horizon.

Proof. The token dynamics give

$$H_t = H_0 + \sum_{\tau=0}^{t-1} c_\tau. \tag{43}$$

Taking expectations and applying the tower property yield

$$\mathbb{E}[H_t] = H_0 + \sum_{\tau=0}^{t-1} \mathbb{E}[\mathbb{E}[c_\tau \mid \mathcal{F}_\tau]] \tag{44}$$

$$\ge H_0 + t c_{\min}. \tag{45}$$

The lower bound diverges as $t \to \infty$. □

This corollary delineates the scope of Theorem 1: MANGO regulates the rate of token consumption within a finite execution budget, but does not assume that cumulative token consumption can decrease.

H.6 Robustness to Runtime-State Perturbations

The runtime state used by the controller is estimated online. We therefore characterize how bounded profiling errors affect the resulting actions.

Theorem 4 (Robustness under Profiling Perturbations)**.** *Let $z_t = (P_t, \Omega_t, \bar{u}_t, u_t^i, \rho_t)$ and suppose $\|\widehat{z}_t - z_t\|_\infty \le \varepsilon$ and $|\widehat{q}_t^i - q_t^i| \le \varepsilon$. If $\tau(P,\rho)$ is L_τ-Lipschitz and $0 \le \Gamma_t, \widehat{\Gamma}_t \le \Gamma_{\max}$, then*

$$|\widehat{r}_t^i - r_t^i| \le L_r \varepsilon, \qquad L_r = \lambda_P + \lambda_\Omega + \lambda_u, \tag{46}$$

$$|\widehat{d}_t^i - d_t^i| \le L_d \varepsilon, \qquad L_d = \lambda_d(\kappa_1 + \kappa_2 + \kappa_3 + \Gamma_{\max}). \tag{47}$$

The writing decision remains unchanged whenever

$$\left|q_t^i - \tau(P_t, \rho_t)\right| > (1 + L_\tau)\varepsilon. \tag{48}$$

Proof. Because the governance intensity is affine,

$$|\widehat{\Gamma}_t - \Gamma_t| \le \kappa_1 |\widehat{P}_t - P_t| + \kappa_2 |\widehat{\Omega}_t - \Omega_t| + \kappa_3 |\widehat{\bar{u}}_t - \bar{u}_t| \tag{49}$$

$$\le (\kappa_1 + \kappa_2 + \kappa_3)\varepsilon. \tag{50}$$

The clipping operator is non-expansive, so applying it to Equation (9) gives

$$|\widehat{r}_t^i - r_t^i| \le \lambda_P|\widehat{P}_t - P_t| + \lambda_\Omega|\widehat{\Omega}_t - \Omega_t| + \lambda_u|\widehat{u}_t^i - u_t^i| \tag{51}$$
$$\le (\lambda_P + \lambda_\Omega + \lambda_u)\varepsilon. \tag{52}$$

Similarly, because $0 \le \rho_t, \widehat{\rho}_t \le 1$,

$$|\widehat{d}_t^i - d_t^i| \le \lambda_d \left|\widehat{\Gamma}_t(1-\widehat{\rho}_t) - \Gamma_t(1-\rho_t)\right| \tag{53}$$
$$\le \lambda_d \left(|\widehat{\Gamma}_t - \Gamma_t| + \Gamma_{\max}|\widehat{\rho}_t - \rho_t|\right) \tag{54}$$
$$\le \lambda_d(\kappa_1 + \kappa_2 + \kappa_3 + \Gamma_{\max})\varepsilon. \tag{55}$$

Finally,

$$\left|\tau(\widehat{P}_t, \widehat{\rho}_t) - \tau(P_t, \rho_t)\right| \le L_\tau \varepsilon. \tag{56}$$

Together with $|\widehat{q}_t^i - q_t^i| \le \varepsilon$, the signed distance between candidate quality and the writing threshold changes by at most $(1 + L_\tau)\varepsilon$, proving Equation (48). □

Controller-parameter perturbations. Let $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \kappa_3)^\top$ and $\mathbf{z}_t = (P_t, \Omega_t, 1 - \bar{u}_t)^\top$, so that $\Gamma_t = \boldsymbol{\kappa}^\top \mathbf{z}_t$.

Proposition 1 (Stability under Controller-Parameter Perturbations)**.** *Suppose* $0 \le P_t \le P_{\max}$, $0 \le \Omega_t \le \Omega_{\max}$, *and* $0 \le \bar{u}_t \le 1$. *For* $\widehat{\boldsymbol{\kappa}} = \boldsymbol{\kappa} + \Delta\boldsymbol{\kappa}$,

$$|\widehat{\Gamma}_t - \Gamma_t| \le L_\kappa \|\Delta\boldsymbol{\kappa}\|_2, \qquad L_\kappa = \sqrt{P_{\max}^2 + \Omega_{\max}^2 + 1}. \tag{57}$$

Under the action mappings in Equations (9) *and* (10)*,*

$$|\widehat{r}_t^i - r_t^i| = 0, \tag{58}$$
$$|\widehat{d}_t^i - d_t^i| \le \lambda_d L_\kappa \|\Delta\boldsymbol{\kappa}\|_2. \tag{59}$$

If $\tau(P_t, \rho_t)$ *does not directly depend on* $\boldsymbol{\kappa}$*, the writing decision is also unaffected by these coefficient perturbations.*

Proof. The perturbed intensity satisfies

$$\widehat{\Gamma}_t - \Gamma_t = (\Delta\boldsymbol{\kappa})^\top \mathbf{z}_t. \tag{60}$$

By Cauchy–Schwarz,

$$|\widehat{\Gamma}_t - \Gamma_t| \le \|\Delta\boldsymbol{\kappa}\|_2 \|\mathbf{z}_t\|_2 \tag{61}$$
$$\le L_\kappa \|\Delta\boldsymbol{\kappa}\|_2. \tag{62}$$

The retrieval rule in Equation (9) does not directly depend on $\boldsymbol{\kappa}$. Applying non-expansiveness of clipping to Equation (10) and using $0 \le 1 - \rho_t \le 1$ give the deletion bound. The writing claim follows because $\tau(P_t, \rho_t)$ is unchanged. □

These are local sensitivity bounds rather than invariance guarantees under arbitrary profiling or parameter errors. Continuous retrieval and deletion controls vary smoothly within the stated region, while the discrete writing decision additionally requires a nonzero margin from its threshold.

H.7 Efficiency-Dominance Boundary

Theorem 5 (Efficiency-Dominance Boundary)**.** *Consider MANGO, denoted by* M*, and a reference policy* B *evaluated on the same task and horizon. Suppose*

$$U_t^{\mathrm{M}} \ge U_t^{\mathrm{B}} - \Delta_U, \qquad R_t^{\mathrm{M}} \le R_t^{\mathrm{B}} - \Delta_R, \tag{63}$$

where $0 \le \Delta_U < U_t^{\mathrm{B}}$ *and* $0 < \Delta_R < R_t^{\mathrm{B}}$. *A sufficient condition for* $E_t^{\mathrm{M}} \ge E_t^{\mathrm{B}}$ *is*

$$\frac{\Delta_R}{R_t^{\mathrm{B}}} \ge \frac{\Delta_U}{U_t^{\mathrm{B}}}. \tag{64}$$

The improvement is strict when the relative resource reduction strictly exceeds the worst-case relative progress reduction.

Proof. The assumed bounds imply

$$E_t^{\mathrm{M}} = \frac{U_t^{\mathrm{M}}}{R_t^{\mathrm{M}}} \geq \frac{U_t^{\mathrm{B}} - \Delta_U}{R_t^{\mathrm{B}} - \Delta_R}. \tag{65}$$

It is therefore sufficient that

$$\frac{U_t^{\mathrm{B}} - \Delta_U}{R_t^{\mathrm{B}} - \Delta_R} \geq \frac{U_t^{\mathrm{B}}}{R_t^{\mathrm{B}}}. \tag{66}$$

Because the denominators are positive, this is equivalent to $R_t^{\mathrm{B}}\Delta_U \leq U_t^{\mathrm{B}}\Delta_R$, which gives Equation (64). Strict inequality gives strict efficiency improvement. □

Corollary 3 (Strict Pareto Dominance). *If $U_t^{\mathrm{B}} > 0$, $U_t^{\mathrm{M}} \geq U_t^{\mathrm{B}}$, and $R_t^{\mathrm{M}} < R_t^{\mathrm{B}}$, then $E_t^{\mathrm{M}} > E_t^{\mathrm{B}}$.*

Proof. Since the resource burdens are positive,

$$\frac{U_t^{\mathrm{M}}}{R_t^{\mathrm{M}}} \geq \frac{U_t^{\mathrm{B}}}{R_t^{\mathrm{M}}} > \frac{U_t^{\mathrm{B}}}{R_t^{\mathrm{B}}}. \tag{67}$$

□

The theorem is a comparison certificate rather than an unconditional claim of per-step improvement. It identifies when the proportional reduction in resource burden compensates for a bounded reduction in collaborative progress.

H.8 Computational Complexity

Proposition 2 (Computational Complexity). *Let N denote the number of agents and $|\mathcal{M}_t|$ the number of shared-memory records. With incrementally maintained runtime statistics and linear-time record selection, MANGO requires*

$$O(N + |\mathcal{M}_t|) \tag{68}$$

governance time per execution step. It maintains $O(1)$ aggregate state and at most $O(|\mathcal{M}_t|)$ scalar record metadata. If deletion candidates are fully sorted, the worst-case time is

$$O(N + |\mathcal{M}_t| \log |\mathcal{M}_t|)\,. \tag{69}$$

Proof. CRC aggregates a constant-size summary from each agent in $O(N)$ time. With incrementally maintained statistics, MTSP updates memory size, token expenditure, and aggregate noise in $O(1)$ time. AGC computes retrieval and writing controls with a constant number of operations per agent, contributing another $O(N)$ term.

Updating one governance score per memory record requires $O(|\mathcal{M}_t|)$ time. Linear-time selection identifies deletion candidates without fully sorting them, yielding $O(N + |\mathcal{M}_t|)$ total governance time. If the candidates are fully sorted, the record-processing term becomes $O(|\mathcal{M}_t| \log |\mathcal{M}_t|)$.

CRC and MTSP retain only constant-size aggregate state. AGC may additionally maintain one scalar score per record, requiring at most $O(|\mathcal{M}_t|)$ auxiliary metadata. □

The proposition concerns only the governance layer in Section 3; it excludes the cost of LLM inference and any external retrieval backend.

I Limitations and Future Work

MANGO is designed as a lightweight online governance framework for long-horizon multi-agent LLM systems. While our evaluation demonstrates consistent performance across diverse reasoning tasks, multiple LLM backbones, and varying agent populations, many opportunities remain for extending memory governance to broader execution settings, according to more complex evolving deployment scenarios. Future work includes exploring more diverse collaboration patterns, richer memory organizations, and tighter integration between memory governance and system-level resource management to further improve the scalability and efficiency of large-scale agentic systems.

J AI USE STATEMENT.

This paper was edited for grammar and language clarity using GPT-5.6-Sol. We reviewed and verified all AI-assisted text, languages, and take full responsibility for the final content of this work.