ENTROPIC MEMORY: A THERMODYNAMICS-INSPIRED CONSOLIDATION MECHANISM FOR LIFELONG AGENT LEARNING

Jing Du
Khoury College of Computer Sciences
Northeastern University
du.jing2@northeastern.edu

Hang Zhao
Khoury College of Computer Sciences
Northeastern University
zhao.hang1@northeastern.edu

ABSTRACT

Large language model (LLM) agents often degrade over long interaction streams because memory accumulates noisy observations that reduce retrieval quality. We propose Entropic Memory, a two-tier memory consolidation mechanism that periodically transfers information from a hot working buffer into a cold long-term store. The method uses a free-energy objective to balance utility against embedding entropy, together with a temperature-controlled stochastic replacement rule. In the controlled Infinite Room environment under a fixed memory budget, Entropic Memory matches greedy importance sampling at 30% noise ($SR \approx 0.29$) and improves survival rate from 0.24 to 0.28 at 50% noise (+15% relative). Overall, these results indicate that entropy-aware consolidation improves robustness to distractors in this controlled continual-memory setting.

1 INTRODUCTION

LLM agents are limited by transient memory. Long-context modeling and RAG improve access to external information, but performance can still degrade as context grows and relevant evidence becomes harder to surface reliably (Lewis et al., 2020; Liu et al., 2024; Maharana et al., 2024). We propose Entropic Memory, inspired by thermodynamic annealing (Kirkpatrick et al., 1983; Sohl-Dickstein et al., 2015), which periodically consolidates a hot working store into a compact cold store by balancing utility against embedding entropy. The design is motivated by complementary learning systems in neuroscience (McClelland et al., 1995), but our goal is algorithmic memory management rather than biological realism.

Our contributions are: (1) we formalize memory consolidation as free energy minimization; (2) we introduce a stochastic acceptance rule that can escape local utility optima; and (3) we present controlled experiments in the Infinite Room environment showing that, under fixed capacity, Entropic Memory matches greedy importance sampling at 30% noise and improves over it by 15% at 50% noise.

2 METHODOLOGY

We model memory consolidation as minimizing Free Energy $F = E + \lambda S$. The internal energy $E(m) = -\text{Utility}(m)$ is derived from query relevance and recency, so that high utility corresponds to low energy. The entropy $S(m) = H(\mathbf{e}_m)$ measures the Shannon entropy of a memory's embedding distribution; noisy or ambiguous memories have high S and are penalized. A temperature parameter T regulates plasticity, with high T encouraging exploration and low T enforcing exploitation.

A candidate memory c replaces a victim v with probability $P = \min(1, \exp(-\Delta F/T))$, where $\Delta F = F(c) - F(v)$. This stochasticity prevents the agent from getting stuck in local utility maxima. See Figure 1.

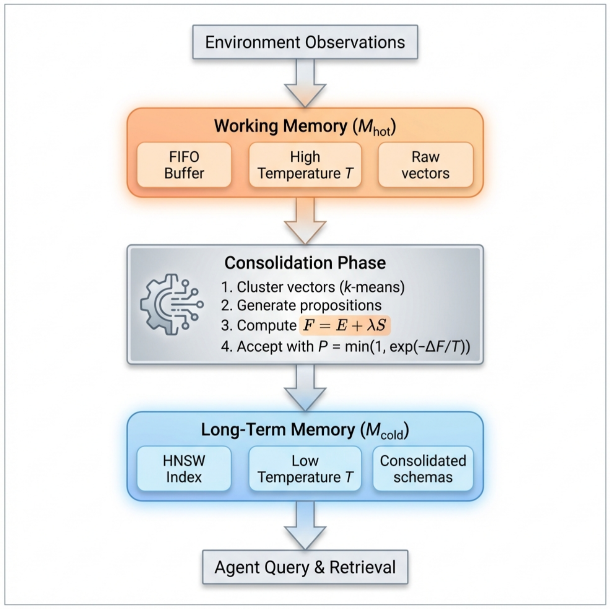

Figure 1: Entropic Memory architecture: observations enter M_{hot}, undergo consolidation via free energy minimization ($F = E + \lambda S$), and crystallize into M_{cold}.

3 Setup & Results

We use the "Infinite Room" environment with a 5,000-step horizon as a controlled testbed for memory consolidation. Concepts drift with rate $\delta = 0.005$, and retrieval uses cosine similarity threshold $\tau = 0.7$. At each step, the agent receives an observation (o_t), retrieves top-3 memories, and predicts an action a_t. Survival Rate (SR) is the fraction of steps where a_t matches the ground truth. Distractor observations are injected with probability p_{noise}. Info Density is defined as $SR \times (N/C)$, where $N = 50$ is the number of concepts and $C = 20$ is the memory capacity.

Table 1 shows performance at 30% noise. Entropic Memory matches the greedy Importance baseline ($SR \approx 0.29$).

Table 1: Performance at 30% Noise (Mean $\pm$ Std, 5 seeds). Entropic matches greedy Importance.

Method	**Survival**	**Hit@3**	**Info Density**	**Cap.**
Random	0.02 ± 0.00	0.02	0.05	20
FIFO	0.26 ± 0.04	0.31	0.65	20
LRU	0.27 ± 0.04	0.33	0.68	20
Importance	0.29 ± 0.04	0.35	0.73	20
Entropic (Ours)	**0.29** $\pm$ 0.04	**0.36**	**0.73**	20

At 50% noise, greedy Importance degrades to $SR = 0.24$, while Entropic Memory maintains $SR = 0.28$ (+15% relative improvement). This result suggests that the entropy term helps penalize noisy embeddings even when they are frequently accessed.

4 Discussion & Limitations

Entropic Memory offers an objective-driven alternative to heuristic pruning. Unlike MemGPT's OS-based paging, we use a continuous thermodynamic objective for replacement decisions. The main limitations are that our evaluation is currently restricted to a synthetic environment and that consolidation adds computational overhead (about 2s/cycle). Future work includes dynamic T scheduling, end-to-end evaluation with full LLM agents, and scaling to multi-agent systems.

REFERENCES

Karl Friston. The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11 (2):127–138, 2010. doi: 10.1038/nrn2787.

James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526, 2017. doi: 10.1073/pnas.1611835114.

Scott Kirkpatrick, C. Daniel Jr. Gelatt, and Mario P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–80, 1983. doi: 10.1126/science.220.4598.671.

Yann LeCun, Sumit Chopra, Raia Hadsell, Marc'Aurelio Ranzato, and Fu-Jie Huang. A tutorial on energy-based learning. In *Predicting Structured Data*. MIT Press, 2006.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive nlp tasks. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 9459–9474. Curran Associates, Inc., 2020. URL `https://proceedings.neurips.cc/paper_files/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf`.

Nelson F. Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. Lost in the middle: How language models use long contexts. *Transactions of the Association for Computational Linguistics*, 12:157–173, 2024. doi: 10.1162/tacl_a_00638.

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. doi: 10.18653/v1/2024.acl-long.747.

Yu. A. Malkov and D. A. Yashunin. Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(4):824–836, 2020. doi: 10.1109/TPAMI.2018.2889473.

James L. McClelland, Bruce L. McNaughton, and Randall C. O'Reilly. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3):419–457, 1995. doi: 10.1037/0033-295X.102.3.419.

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. doi: 10.48550/arXiv.2310.08560.

Joon Sung Park, Joseph C. O'Brien, Carrie J. 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*, 2023. doi: 10.1145/3586183.3606763.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning*, pp. 2256–2265. PMLR, 2015.

A APPENDIX

A.1 ALGORITHM DETAILS

The agent maintains two vector stores: M_{hot} (FIFO buffer for recent observations) and M_{cold} (HNSW index for consolidated knowledge; Malkov & Yashunin 2020). The consolidation process operates as follows:

At every K steps, the system enters a Sleep Phase. Vectors in M_{hot} are clustered using k-means. For each centroid C_j, the system generates a proposition P_j via LLM summarization. The utility of P_j is computed as cosine similarity to recent queries, weighted by recency. The storage cost $S(P_j)$ is computed as the entropy of the cluster's embedding distribution (high variance $\Rightarrow$ high S). The candidate competes with the lowest-utility memory in M_{cold}, and the transition is accepted with probability $P = \min(1, \exp(-\Delta F/T))$, where $\Delta F = F(P_j) - F(v)$ and v denotes the victim memory.

The Free Energy is $F = E + \lambda S$, where $E = -$Utility (high utility $\Rightarrow$ low E), S = embedding entropy (noisy memories have high S), and λ balances the two terms. High T encourages diverse memories (exploration); low T favors high-utility, low-entropy content (exploitation).

A.2 Theoretical Foundation

Our approach is motivated by two lines of prior work. First, the Free Energy Principle (Friston, 2010) provides a useful analogy for viewing memory as an internal model that should remain informative under changing observations. We use this connection only as high-level motivation; the method itself is an algorithmic consolidation rule rather than a neuroscientific model.

Second, the acceptance rule is directly connected to simulated annealing (Kirkpatrick et al., 1983). The temperature parameter controls the extent to which the system accepts worse local replacements early on, allowing the memory store to avoid purely greedy updates and to trade off exploration against exploitation.

A.3 Related Work

Memory in LLM Agents Park et al. (2023) introduced Generative Agents with reflection mechanisms that periodically synthesize observations into higher-level insights. MemGPT (Packer et al., 2023) proposes an OS-inspired hierarchical memory with explicit paging between main memory and external storage. In contrast, our method uses an objective-driven consolidation criterion (free energy minimization) with a tunable plasticity parameter.

Energy-Based Models We extend the intuition of Energy-Based Models (LeCun et al., 2006) to the semantic space of agent memory. Traditional EBMs operate on continuous representations for perception; we instead apply energy-based reasoning to discrete memory management decisions.

Continual Learning The catastrophic forgetting problem in neural networks (Kirkpatrick et al., 2017) motivates our approach. Rather than modifying network weights as in EWC, we address forgetting at the memory retrieval level, complementing frozen LLM backbones.

A.4 Experimental Setup

Experiments were conducted using a simulated "Infinite Room" environment with the following characteristics: agents must correctly identify and respond to $N = 50$ distinct concepts, each associated with specific survival actions. Concepts drift over time with rate $\delta = 0.005$, meaning approximately 25 concept-action associations change over the 5,000-step horizon. Additionally, 30% of observations are distractor noise unrelated to any survival-relevant concept.

The FIFO baseline uses a standard sliding window of size $C = 20$, replacing oldest memories when capacity is exceeded. The Entropic Agent uses the same capacity constraint but applies the Free Energy minimization criterion during the Sleep Phase (every $K = 100$ steps). All retrieval-based methods use the same embedding model for fair comparison. Figure 2 shows the full 5,000-step performance trajectory.

A.5 Limitations and Future Work

The current study has three main limitations. First, evaluation is restricted to a synthetic environment designed to isolate memory consolidation behavior. This makes it easier to interpret the effect of the objective, but it does not establish benefits on broader LLM-agent benchmarks or real deployment

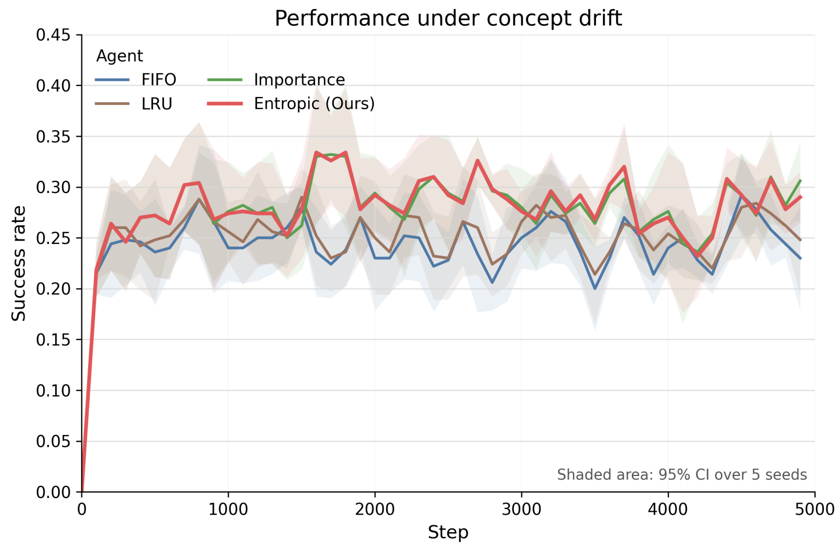

Figure 2: Survival Rate over 5,000 steps (95% CI, 5 seeds). Entropic and Importance outperform FIFO and LRU, illustrating the benefit of utility-based consolidation in this setup.

settings. Second, the Sleep Phase introduces nontrivial overhead: in our implementation, clustering and LLM summarization together take roughly 2 seconds per consolidation cycle. Third, the current system uses a fixed temperature $T = 1.0$ and fixed consolidation hyperparameters, which may not transfer across tasks without retuning.

These limitations suggest several concrete follow-up directions. The most immediate one is end-to-end evaluation with full LLM agents on longer-horizon tasks. A second direction is adaptive scheduling, where consolidation frequency and temperature are adjusted online rather than fixed in advance. A third direction is to study whether the same free-energy objective remains useful when memory is shared across agents or organized hierarchically across multiple storage tiers.

A.6 Ablation Study: Noise Resilience

We analyzed the robustness of Entropic Memory against increasing noise levels (10%, 30%, 50%). As shown in Figure 3, at 50% noise, Entropic Memory maintains $SR \approx 0.28$ while greedy Importance degrades to $SR \approx 0.24$ (+15%). This pattern is consistent with the entropy term filtering high-frequency noise that fools utility-only heuristics.

Table 2: Summary at 50% Noise. Info Density is computed as $SR \times (N/C)$ with $N = 50$ and $C = 20$.

Method	Survival	Info Density
Importance	0.24	0.60
Entropic (Ours)	0.28	0.70

Hit@3 was not logged in this ablation sweep.

A.7 Implementation Details

Embedding Model Observations are encoded using a 384-dimensional sentence embedding model. Similarity is computed via cosine distance with threshold $\tau = 0.7$ for retrieval.

Clustering K-means clustering with $k = 5$ is applied during each Sleep Phase. Centroids with fewer than 3 supporting vectors are discarded as noise.

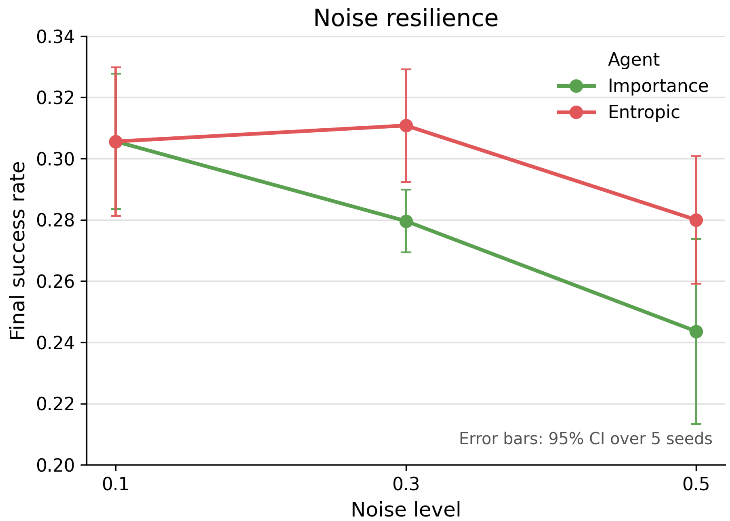

Figure 3: Impact of Noise Probability on Survival Rate. Entropic Memory shows stronger resilience than greedy Importance at high noise levels in this setup.

LLM Summarization Cluster summaries are produced using a 7B instruction-tuned model with temperature 0.3. Each summary is limited to 50 tokens.

Computational Cost Memory storage: M_{hot} (20 vectors $\times$ 384 dims $\times$ 4 bytes = 30KB), M_{cold} (20 propositions $\times$ 384 dims $\times$ 4 bytes = 30KB). Total: 60KB. Retrieval latency: 5ms average. Consolidation: 2s per cycle (offline).