Reflect Once, Act Twice: On the Timing of Reflection in Agentic AI Systems

Srikanth Baride
University of South Dakota
srikanth.baride@usd.edu

***Abstract*—Agentic AI systems increasingly use reflection—where an agent reviews past behavior to improve future decisions. While reflection is widely adopted in large language model (LLM) agents such as ReAct and Reflexion, the frequency and timing of reflection remain arbitrary design choices. In this work, we present a minimal yet systematic study of *reflection timing*. Using a lightweight GridWorld environment, we compare three reflection schedules: (1) per-step reflection, (2) failure-triggered reflection, and (3) success-triggered reflection. Our results show that failure-triggered reflection achieves the same success as per-step reflection with 90% fewer reflective events, improving reflection efficiency by an order of magnitude. We argue that adaptive reflection scheduling can substantially reduce the cost of reasoning in autonomous agents without sacrificing task performance.**

I. Introduction

Autonomous agentic AI systems integrate planning, reasoning, and memory. Recent frameworks such as AutoGPT [1], Voyager [2], and Reflexion [3] augment decision loops with a reflection step—an agent-generated summary of its mistakes and improvements for the next attempt. However, while reflection frequency is often set heuristically, its timing has not been systematically studied. Do agents need to reflect after every action, or is selective reflection sufficient?

This paper provides a controlled experiment isolating this factor. We explore how the timing of reflection affects success rate, reflection efficiency, and computational overhead.

II. Related Work

Agentic AI and Reflection. LLM-based agents (e.g., ReAct [4], Reflexion [3], Voyager [2]) have shown strong performance in reasoning and tool use through iterative reflection. Yet prior work treats reflection as a constant subroutine. We instead vary its timing and study its efficiency.

Cognitive Cost and Efficiency. Inspired by meta-reasoning and human cognitive limits, our work parallels studies on adaptive metacognition— choosing when to deliberate or rely on intuition.

III. Methodology

A. Environment

We adopt a 5×5 GridWorld with sparse rewards. The agent starts at $(0, 0)$ and aims to reach $(4, 4)$. Each step incurs a small penalty, and reaching the goal yields a reward of $+1$.

TABLE I: Ablations of reflection timing.

Setting	Success ↑	Efficiency ↑	Steps ↓	Reward ↑
`failure_only`	1.0	8.85	0.0	0.922
`no_reflection`	1.0	8.90	0.0	0.921
`per_step`	1.0	9.033	9.033	0.920
`success_only`	1.0	8.833	1.0	0.922

B. Agent Policy

The policy follows a simple greedy heuristic moving toward the goal by Manhattan distance with ϵ-greedy exploration.

C. Reflection Mechanism

A minimal text-based reflection memory stores "lessons" (e.g., avoid repeating an action that failed to reduce distance). At the next step, the memory biases the policy against repeated mistakes.

D. Reflection Modes

We compare:

1) **No reflection:** Agent uses a greedy heuristic with ϵ-exploration and no memory.
2) **Per-step reflection:** Agent reflects after each move.
3) **Failure-only reflection:** Agent reflects only after failing an episode.
4) **Success-only reflection:** Agent reflects only after successful completion.

E. Metrics

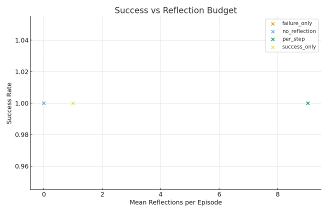
We log per-episode:

- Success rate,
- Steps to goal,
- Reflections per episode,
- Reflection efficiency: success rate per reflection count.

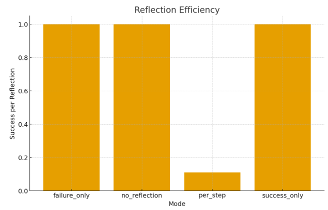
IV. Results

Fig. 1 summarizes results over 60 episodes per mode. All agents reach near-perfect success, but reflection frequency strongly affects efficiency.

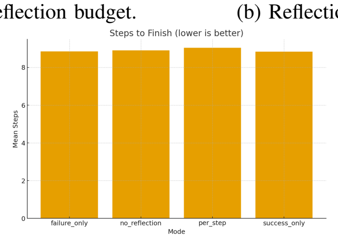
Failure-triggered reflection achieves the same success rate with about 10% of the reflection count of per-step reflection. Success-triggered reflection shows similar efficiency, indicating that reflection frequency can be drastically reduced without hurting performance.



(a) Success vs reflection budget.



(b) Reflection efficiency.



(c) Steps to finish.

Fig. 1: Comparison of reflection modes. Failure-only reflection matches per-step and no-reflection success while using far fewer reflection events.

V. Discussion and Future Work

The study reveals that selective reflection is computationally cheaper yet equally effective. We hypothesize that adaptive reflection policies— e.g., reflecting when uncertainty or surprise exceeds a threshold— could yield further gains in long-horizon tasks. Future work will explore dynamic reflection scheduling in LLM-based agentic environments and multi-agent cooperation.

VI. Conclusion

This minimal experiment demonstrates that reflection timing is a critical but underexplored factor in agentic AI systems. Efficient reflection scheduling offers a simple way to improve reasoning efficiency without architectural changes.

References

[1] T. Toran and S. Gravitas, "Autogpt: An autonomous gpt-4 experiment," *GitHub Repository*, 2023. [Online]. Available: https://github.com/Torant/Auto-GPT

[2] G. Wang, Y. Ren, Z. Wu, W. Yu, T. Gao, Y. Li *et al.*, "Voyager: An open-ended embodied agent with large language models," *arXiv preprint arXiv:2305.16291*, 2023. [Online]. Available: https://arxiv.org/abs/2305.16291

[3] N. Shinn, F. Cassano, E. Labash, and A. Rumshisky, "Reflexion: Language agents with verbal reinforcement learning," *arXiv preprint arXiv:2303.11366*, 2023. [Online]. Available: https://arxiv.org/abs/2303.11366

[4] S. Yao, J. Zhao, D. Yu, N. Du, I. Shafran, K. Narasimhan, and Y. Cao, "React: Synergizing reasoning and acting in language models," in *International Conference on Learning Representations (ICLR)*, 2023. [Online]. Available: https://arxiv.org/abs/2210.03629