Erase to Improve: Erasable Reinforcement Learning for Search-Augmented LLMs

Kang An, Ziliang Wang, Xuhui Zheng, Faqiang Qian, Weikun Zhang, Yuhang Wang, Yichao Wu

SenseTime, Shenzhen University, Nanjing University

Outline

Introduction & Challenges

Identify key challenges in current search-augmented LLMs for multi-hop reasoning.

Erasable Reinforcement Learning (ERL)

Introduce the novel ERL framework designed to address these challenges.

Experimental Results

Present experimental results showing significant improvements over the state of the art.

Conclusion & Future Work

Summarize contributions and discuss future research directions.

Introduction & Challenges

Background: Potential vs. Unreliability

Search-augmented LLMs show great potential but remain unreliable for complex multi-hop reasoning tasks.

01. Decomposition Errors

Incorrectly breaking down the main question into sub-queries, leading to a flawed starting point.

02. Retrieval Missing

Failing to retrieve key evidence even with correct sub-queries, leaving critical gaps in reasoning.

03. Reasoning Errors

Making logical mistakes when integrating retrieved information, propagating errors through the chain.

Catastrophic Fragility

A single error in any stage can derail the entire reasoning process, leading to an incorrect final answer.

Erasable Reinforcement Learning (ERL)

Core Idea: Human-like Self-Correction

To address catastrophic fragility, ERL enables the agent to emulate human-like self-correction, transforming brittle chains into robust ones.

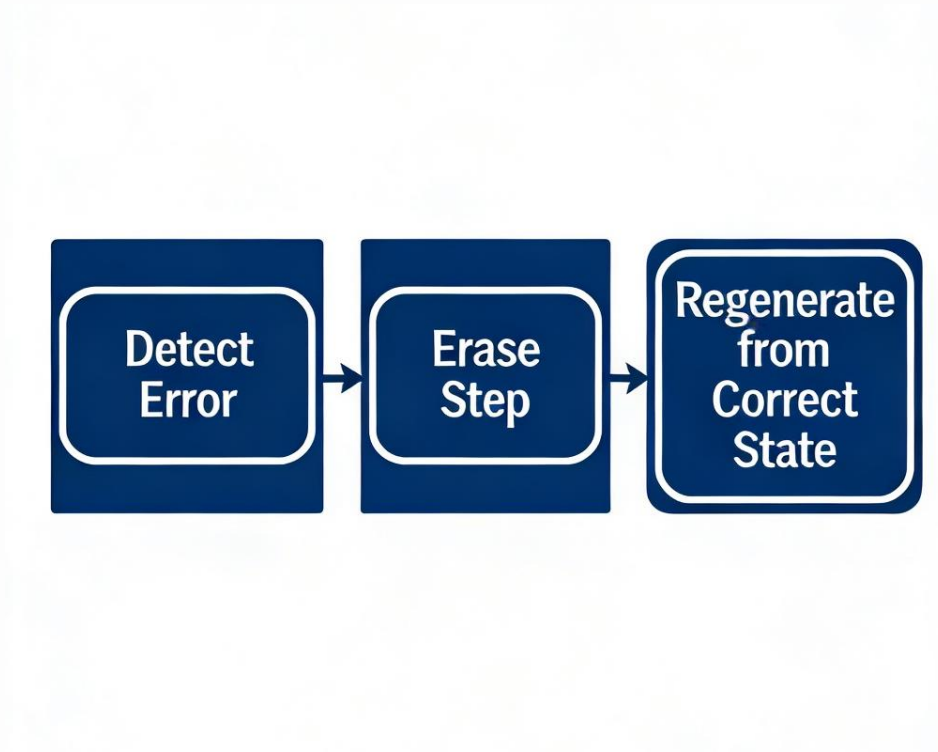
Key Mechanism

- **Detect:** Identify errors in reasoning steps.
- **Erase:** Remove the faulty step from history.
- **Regenerate:** Restart reasoning from the last correct state.

Result: Robust Recovery

The agent can recover from mistakes without restarting the entire process, significantly improving robustness.

ERL Framework Overview



Experimental Setup

Datasets

Evaluated on four multi-hop QA benchmarks:

- HotpotQA
- MuSiQue
- 2WikiMultiHopQA
- Bamboogle

Models (ESearch)

Implemented ERL framework on:

- 3 Billion parameters
- 7 Billion parameters

Baselines

Comparison against other methods:

- Search-R1
- ZeroSearch
- R-Search
- SSRL
- Research
- StepSearch

Key Results: New SOTA Achieved

ESearch-3B Model

Exact Match (EM)	F1 Score
+8.48%	**+11.56%**

Shows the most significant improvement over the previous SOTA.

ESearch-7B Model

Exact Match (EM)	F1 Score
+5.38%	**+7.22%**

Maintains consistent and significant gains across all benchmarks.

These results demonstrate consistent and significant improvements across all evaluated benchmarks, establishing new state-of-the-art performance for our ERL framework.

Conclusion & Future Work

Key Conclusions

- Identified three critical failure modes limiting search-augmented LLMs in multi-hop reasoning.
- Proposed ERL framework enabling fine-grained error detection and regeneration, boosting robustness.
- Established new SOTA results on four major benchmarks, validating ERL's effectiveness.

Future Directions

- Explore ERL application to even longer and more complex reasoning chains beyond current benchmarks.
- Investigate cross-domain generalization, such as mathematical reasoning and code generation.

Thank You!