Project: ReSearch - Integrating Search and Reasoning with Reinforcement Learning for Large Language Models

Technical Summary Report

1. Paper Summary

Title: ReSearch: Integrating Search and Reasoning with Reinforcement Learning for Large Language Models

Link: https://arxiv.org/abs/2503.19470

ReSearch proposes a novel framework that integrates search and reasoning in large language models (LLMs) using reinforcement learning (RL). The core innovation is GRPO (Generalized Recurrent Policy Optimization), a new RL algorithm that enables better reasoning and memory-aware decision making across multiple steps.

2. Key Contributions

- RL-based integration of search into reasoning workflows.
- GRPO algorithm enabling optimization over reasoning sequences.
- Demonstrates significant improvements over SFT (Supervised Fine-Tuning) and RAG (Retrieval-Augmented Generation).

3. Reinforcement Learning Setup

- Environment: Agents perform sequences of actions including search, reasoning, and answering.
- Actions: [search, read, reason, answer]
- Rewards:
  - +1 for correct answer
  - +0.2 for helpful search result

- -0.1 for redundant or irrelevant search

4. GRPO: Generalized Recurrent Policy Optimization

- Handles recurrent policy structures with memory.

- Suitable for partially observable and non-differentiable environments.

- Outperforms PPO, A2C in long-horizon reasoning tasks.

5. Results

Quantitative:

- Accuracy improved by 15-20% on complex QA datasets.

Qualitative:

- LLMs showed better evidence use and reduced hallucinations.

Working Implementation Overview

Repository:

GitHub - Agent-RL/ReSearch: https://github.com/Agent-RL/ReSearch

Installation:

git clone https://github.com/Agent-RL/ReSearch.git

cd ReSearch

pip install -r requirements.txt

Running Experiments:

python train.py --config configs/multihop_qa.yaml

Modules:

- env/: Search + reasoning environment

- models/: GRPO model + value nets

- train.py: GRPO training loop

- configs/: Task configurations

Evaluation:

- Metrics: Accuracy, avg. search steps, reasoning chain length

- Ablation: GRPO vs PPO, with vs without search

Presentation Slides Outline

Slide 1: Title Slide

- Project Title

- Team Members

Slide 2: Motivation

- Why integrate search and reasoning?

- Limitations of RAG and SFT

Slide 3: ReSearch Framework Overview

- Diagram of RL setup

- Action flow

Slide 4: What is GRPO?

- Definition

- Advantages over PPO/A2C

Slide 5: Experiments & Results

- Accuracy improvements

- Sample reasoning trace

Slide 6: Technical Challenges

- RL instability

- Latency from external search

Slide 7: Real-World Use Cases

- AI copilots

- Knowledge assistants

- Research tools

Slide 8: ReSearch vs RAG vs SFT

- Comparison table

Slide 9: Discussion Points

- Feasibility in production systems

- Future improvements

Discussion Points

1. ReSearch vs RAG vs SFT

   - RAG retrieves but doesnt reason.

   - SFT has static knowledge.

   - ReSearch dynamically searches + reasons.

2. Challenges:

- Long episode credit assignment.
- RL convergence.
- Web API delays.

3. Future Potential:

- Can power reflective LLM agents.
- Fits well into tool-augmented pipelines.
- Enables adaptive, interactive AI systems.