s3: You Don't Need That Much Data to Train a Search Agent

Paper　Code

Pengcheng Jiang, Xueqiang Xu, Jiacheng Lin, Jinfeng Xiao*, Zifeng Wang, Jimeng Sun, Jiawei Han

University of Illinois Urbana-Champaign

30th Anniversary EMNLP 2025 Suzhou, China November 4-9

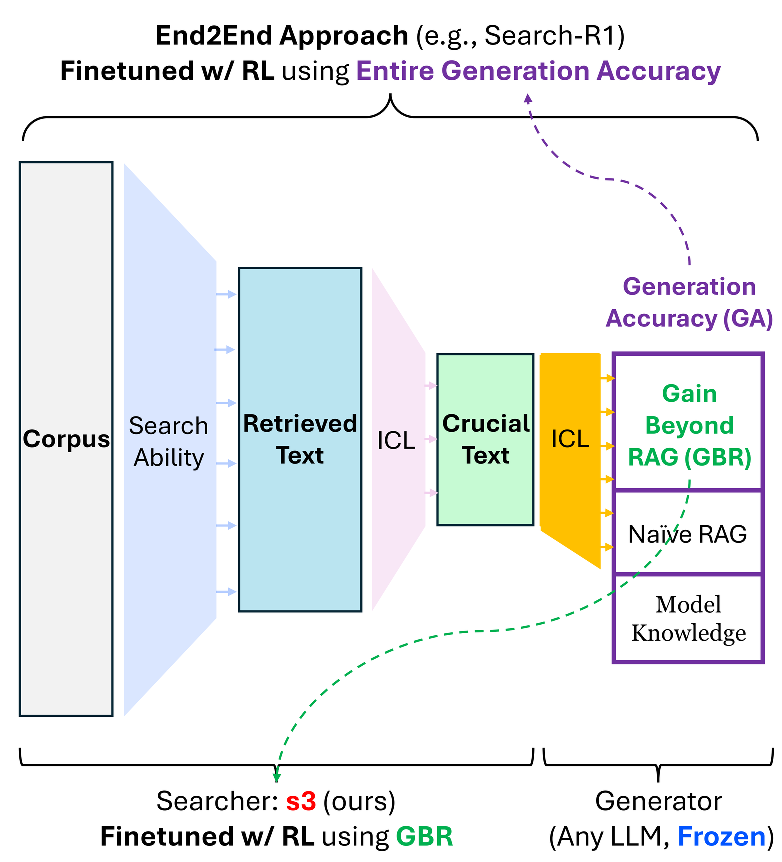
Background

Why search and generation should be decoupled?

End-to-end methods like Search-R1 **cannot** effectively improve search capability, as it fuses search and generation tuning, making the **contribution of each component ambiguous**.

For RAG, correct generation can be achieved by:

- LLM's own knowledge
- Naïve RAG
- **Gain Beyond**



Why exact match (EM) is a horrible metric/reward for open question-answering task with LLMs?

Very obvious. All the papers following Search-R1's evaluation setting should re-consider its feasibility, especially they compare to prompting-based (untuned) methods like IRCoT, with EM.

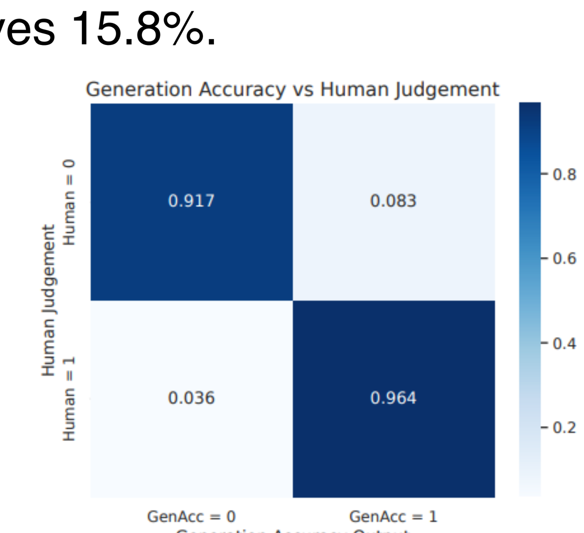
Why Exact Match Falls Short - An Example

Golden answer: "Barack Obama"
LLM response: "The 44th President of the United States was Barack Obama."
Exact match: 0 (response $\neq$ golden)
Generation Accuracy: 1 (span_check succeeds)

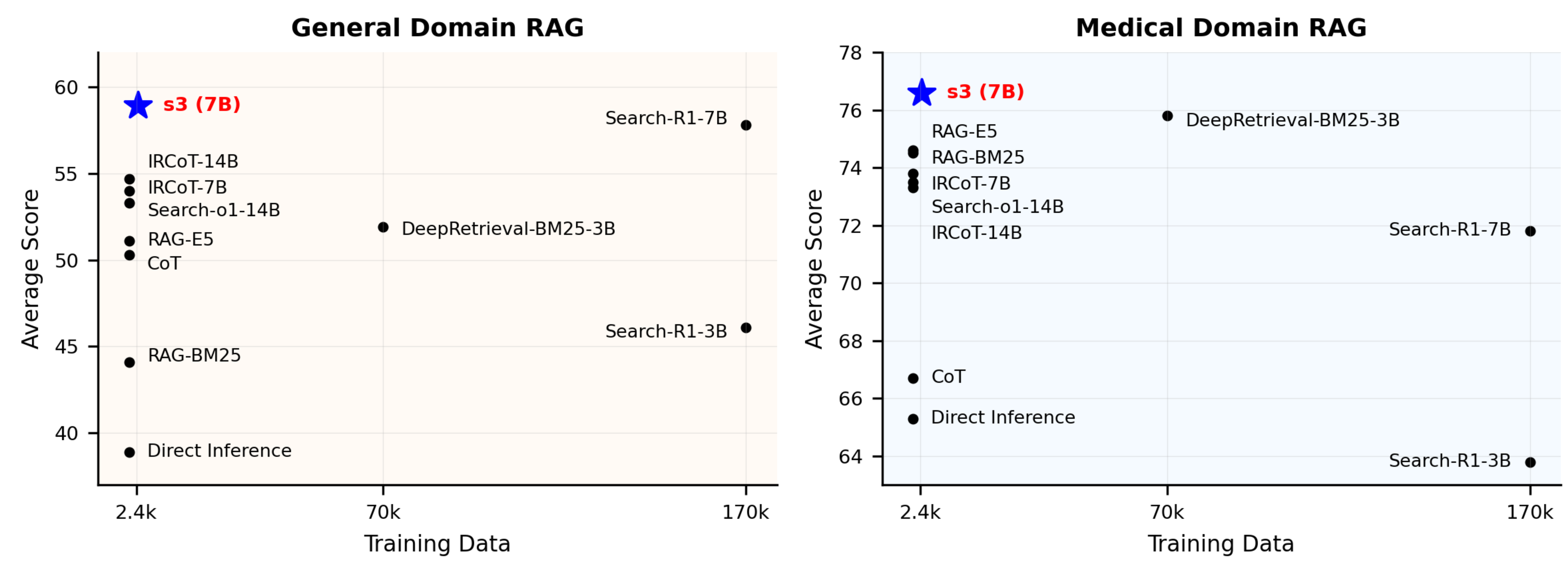
We introduce a new metric **GenAcc** – checking answer span in the response, with LLM-as-a-Judge, which achieves **96.4% alignment** with human judgement while EM achieves 15.8%.

Evaluation Flow of Generation Accuracy

Input: Prediction p, Gold Answers $\mathcal{A}$
Step 1: Normalize p and $\mathcal{A}$ (lowercase, remove punctuation and articles).
Step 2: span_check $\to$ If any $a \in \mathcal{A}$ is a token span in p, return GenAcc = 1.
Step 3: judge_check $\to$ Prompt LLM: *"Does p contain any of $\mathcal{A}$?"*
Step 4: Return GenAcc = 1 if LLM says yes; else 0.

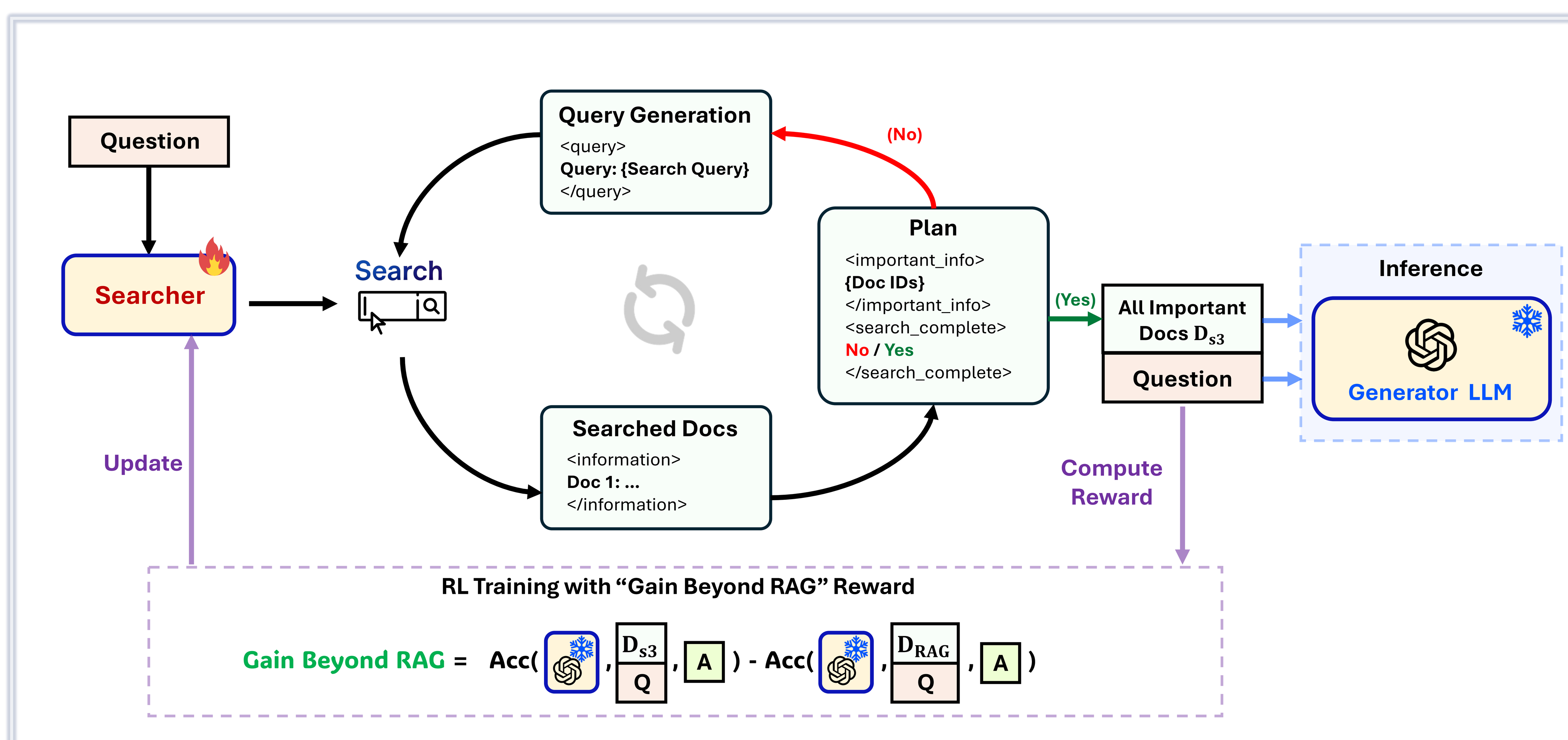


Main Result (see full version in the paper)



1. s3 outperforms all previous methods with **70x less training data** than Search-R1.
2. It also shows its **robustness to domain transfer** (from general to medical), without training on medical data.
3. **Searcher-only** is **much better** than end-to-end optimization for RAG.

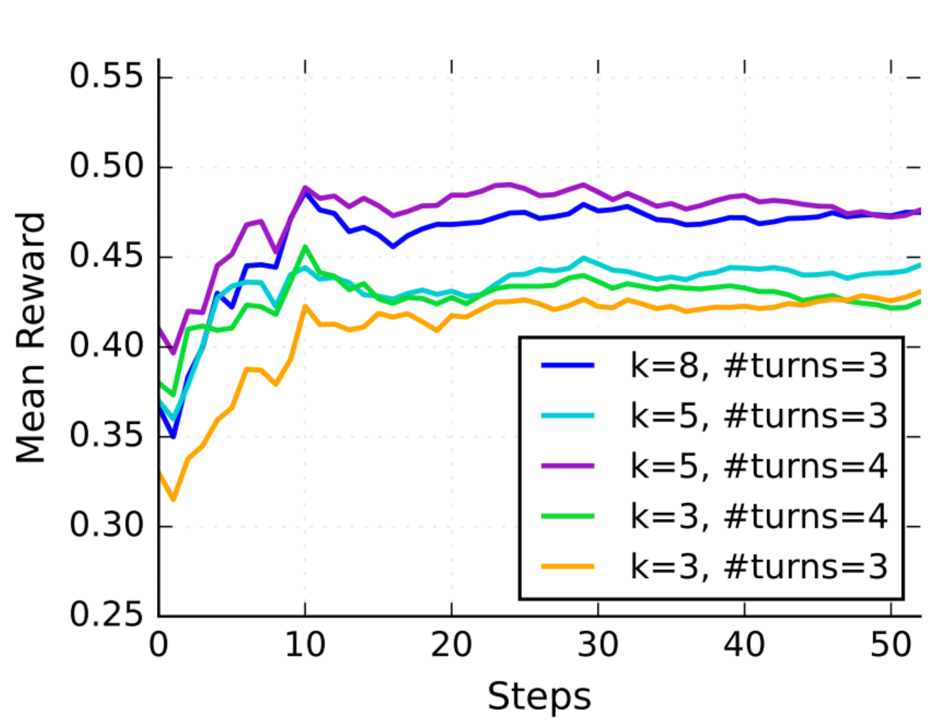
s3 Framework



: trained with RL　: frozen　D_{s3} D_{RAG} : searched docs by s3 / naïve RAG　Q : question　A : golden answer

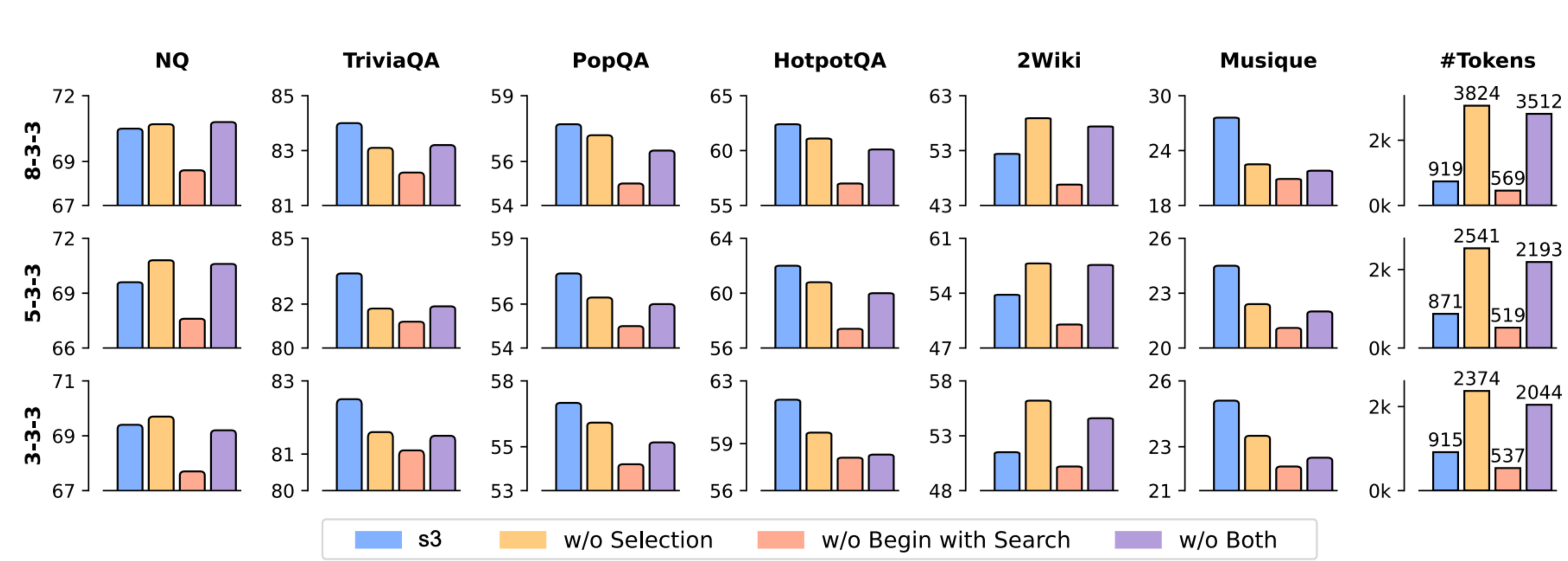
(Gain Beyond RAG: only reward the (1-0) case or penalize the (0-1) case to Searcher)

LLMs are already good searchers. GBR boosts them.



- With human-like evaluation metric, we can see LLM could search well at the beginning, which matches our results of prompting based methods
- Gain Beyond RAG as reward leads to faster and better convergence.

Ablation Study of Proposed Components



Findings:

- **Begin with search** by original question is important, as it can avoid the trajectory deviate from the original query intent.
- **Selection** process can effectively reduce the token consumption, without compromising the overall performance.

Metric Matters

	GenAcc	LLMJudge	Span	EM
General QA	58.9	59.6	57.1	50.5
Medical QA	76.6	77.3	74.3	70.3

Future Directions

- s3 shows that we can efficiently train a **task-specific auxiliary agent** while keeping the main reasoning model frozen. **This modular paradigm can scale to many other tasks**.
- Although search and answering are decoupled, the answering stage remains unoptimized. A natural extension is to train a lightweight answering-specific agent that reasons more effectively over the searched context.

Code & Contact

Code

Starred 766

(Give a star and) try it out!

https://github.com/pat-jj/s3
Patrick (Pengcheng) Jiang, pj20@illinois.edu