# Retrieve, Summarize, Plan: Advancing Multi-hop Question Answering with an Iterative Approach


Zhouyu Jiang
Ant Group
Shanghai, China
jiangzhouyu.jzy@antgroup.com

Mengshu Sun
Ant Group
Beijing, China
mengshu.sms@antgroup.com

Lei Liang
Ant Group
Hangzhou, Zhejiang, China
leywar.liang@antgroup.com

Zhiqiang Zhang
Ant Group
Hangzhou, Zhejiang, China
lingyao.zzq@antgroup.com


**WWW2025**

# Background

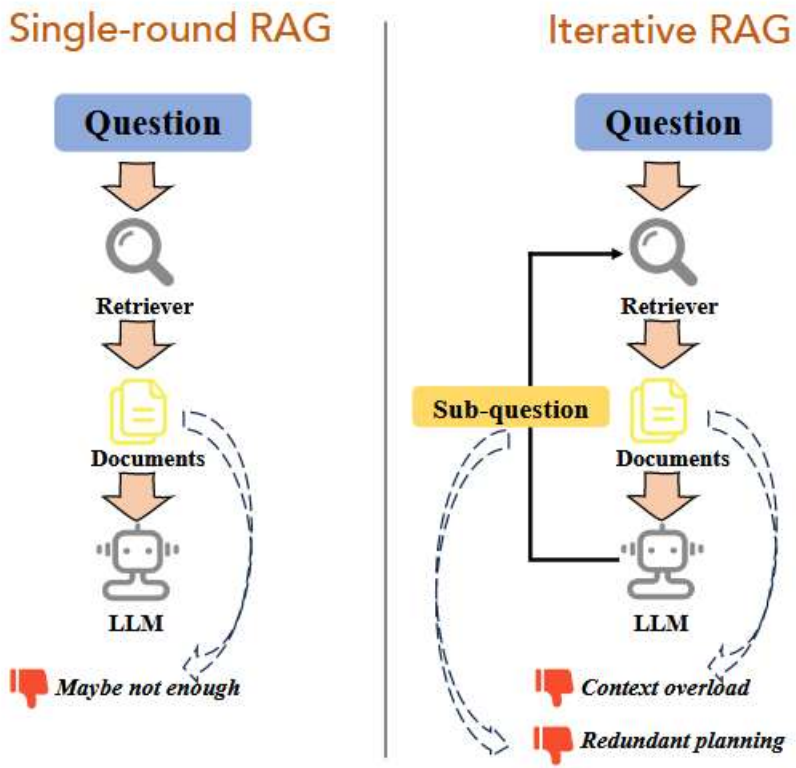



Figure 1: RAG pipelines illustration and challenges faced.

**Context overload：**

每轮检索到的内容越来越多，LLM 的上下文窗口会爆掉。

**Redundant planning：**

LLM 在每轮推理中可能重复制定计划、重复思考已经解决的子问题。

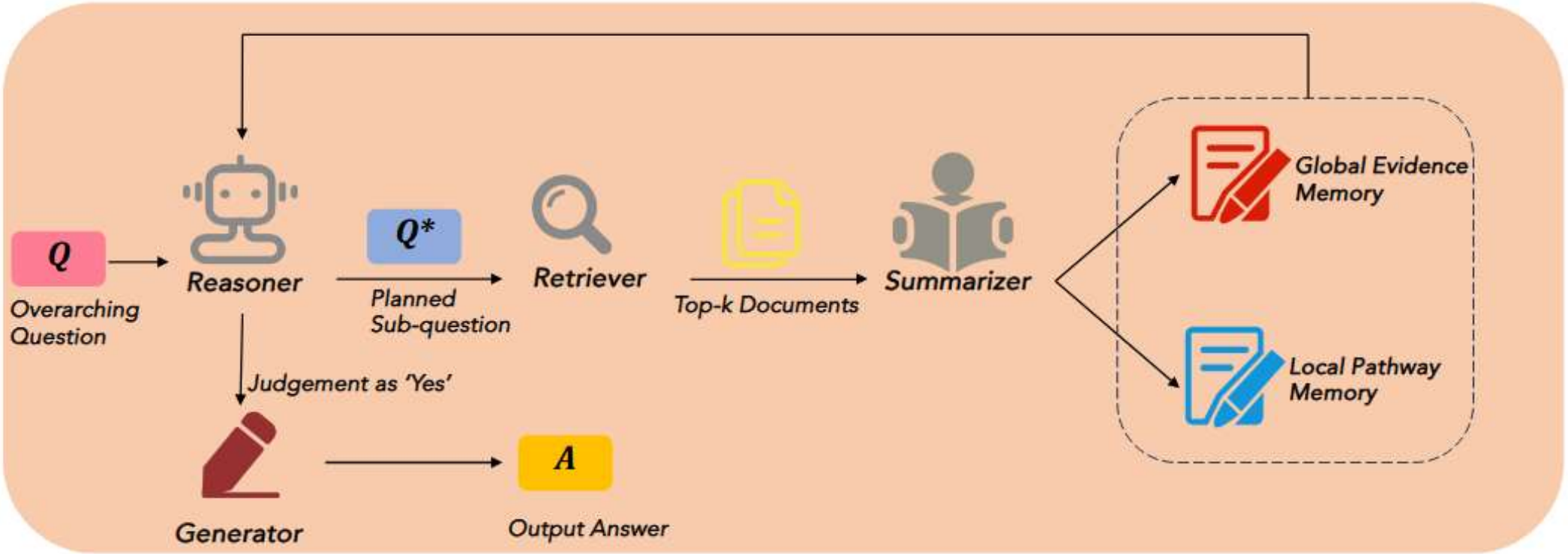
# ReSP



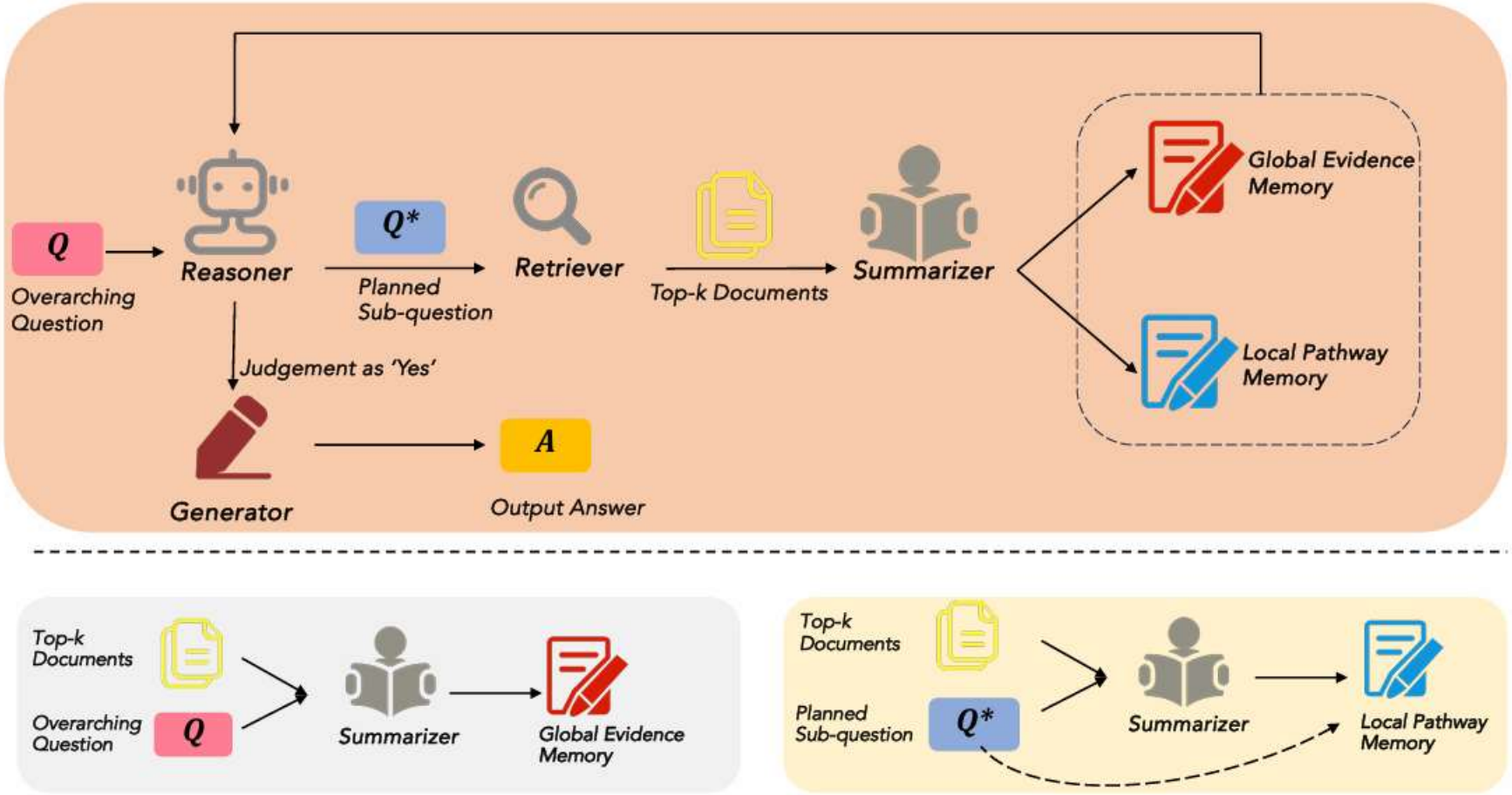
> **Q: Which university is attended by the author of the book *The Old Man and the Sea*?**

这是一个典型 **2Wiki 多跳问题**:

1. *The Old Man and the Sea* → 作者是谁
2. 作者 → 上过哪所大学



# Dual-Function Summarizer




## Global Evidence Memory：

从检索到的文档中提炼出 对总问题有帮助的信息摘要。告诉系统“我们已经有足够的证据”，从而避免过度规划

## Local Pathway Memory：

让模型知道自己已经探索过哪些子问题，从而避免重复规划

# Experiments



**Table 2: Performance comparison on HotpotQA and 2WikiMultihopQA. All methods except Self-RAG utilize fine-tuning-free Llama3-8B-instruct for generation. Self-RAG uses trained selfrag-llama2-7B released by authors. We run experiments five times and report the average token-level F1 score of answer strings. The result of ReSP outperforms baseline models in t-test at $p < 0.05$ level. The best results are in bold and the second best results are underlined.**

| Method | Pipeline type | HotpotQA | 2WikiMultihopQA |
|---|---|---|---|
| Direct Prompting | – | 28.4 | 33.9 |
| Standard RAG | Single-round RAG | 38.6 | 20.1 |
| SuRe [16] | Single-round RAG | 33.4 | 20.6 |
| RECOMP[32] | Single-round RAG | 37.5 | 32.4 |
| REPLUG[25] | Single-round RAG | 31.2 | 21.1 |
| Iter-RetGen[24] | Iterative RAG | 38.3 | 21.6 |
| Self-RAG[2] | Iterative RAG | 29.6 | 25.1 |
| FLARE [12] | Iterative RAG | 28.0 | <u>33.9</u> |
| IRCoT[28] | Iterative RAG | <u>43.1</u> | 32.4 |
| ReSP(ours) | Iterative RAG | **47.2** | **38.3** |

# Experiments



**Table 3: Comparison of computation time.**

| Method | HotpotQA | 2WikiMultihopQA |
|---|---|---|
| Standard RAG | 1.1min | 1.2min |
| IRCoT | 41.8min | 48.2min |
| ReSP | 9.5min | 13.7min |

Additionally, as shown in Table 3, we provide a comparison of computation time. We record the runtime of standard RAG, IRCoT, and ReSP on 200 instances, maintaining the same experimental environment described before. The results indicate that the iterative ReSP takes longer compared to the single-round standard RAG, which may make it less suitable for time-sensitive scenarios unless specialized engineering optimizations are performed. However, compared to IRCoT, which is also an iterative method, ReSP significantly reduces runtime by compressing context and minimizing redundant planning while achieving better performance.

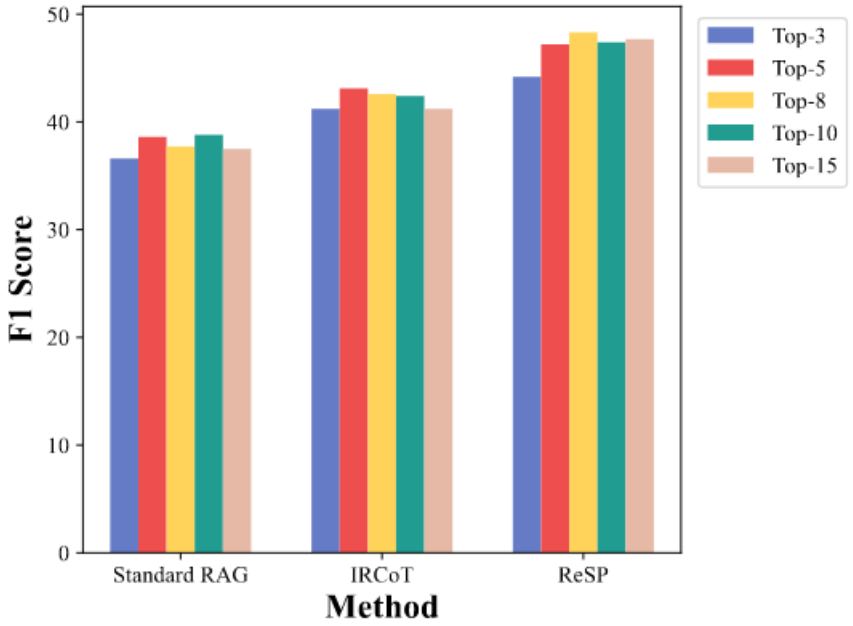



**Figure 4: Bar chart of the performance variations of different RAG methods with varying numbers of retrieved documents per iteration. We report the token-level F1 score obtained from testing on the HotpotQA dataset.**

# Experiments



| Issue | Question | IRCoT | ReSP |
|---|---|---|---|
| Over-planning | Victor Varnada directed Twisted Fortune, which starred which brother of Eddie Murphy? | **Round 0 Thought:** "The text does not mention Twisted Fortune, it talks about Eddie Murphy's career and his brother Charlie Murphy. What's the next step?" | **Round 0 Thought:** [Global Evidence Memory]: "Victor Varnado directed the black comedy Twisted Fortune, which starred Charlie Murphy, a brother of Eddie Murphy." –>[Judgment]: "Yes" –>[Response]: "Charlie Murphy" |
| Repetitive planning | What is Rachelle Amy Beinart's role in the film that follows a group of rebels on a mission? | **Round 0 Thought:** "Rachelle Amy Beinart is not mentioned in any of the provided Wikipedia articles." **Round 1 Thought:** "Rachelle Amy Beinart is not mentioned." ... | **Round 0 Thought:** [Sub-question]: "Who is Rachelle Amy Beinart?" **Round 1 Thought:** [Sub-question]: "What film mentioned in the provided passages features a group of rebels on a mission?" ... |

# Thanks