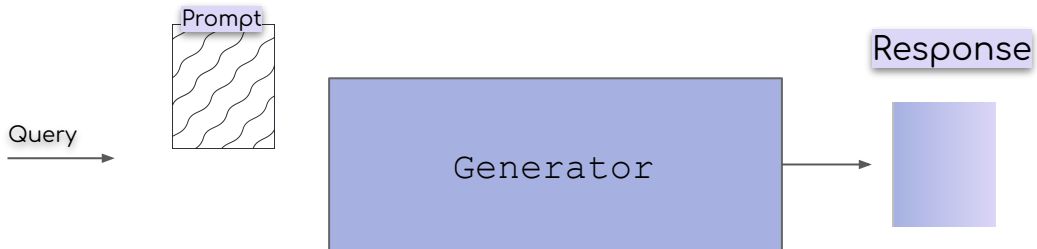


ReWOO: Decoupling Reasoning from Observations for Efficient Augmented Language Models

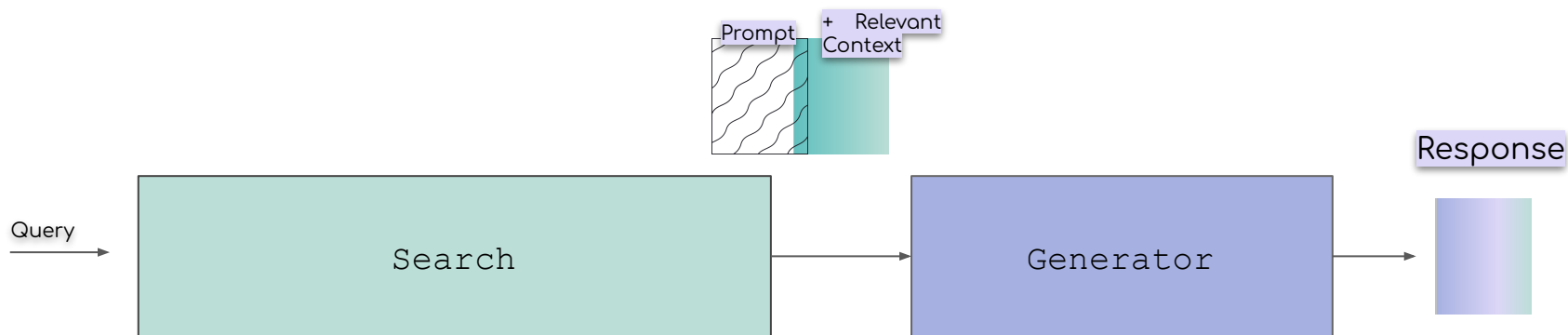


Introduction

Just the Generative Model



A Generative Model empowered by Search



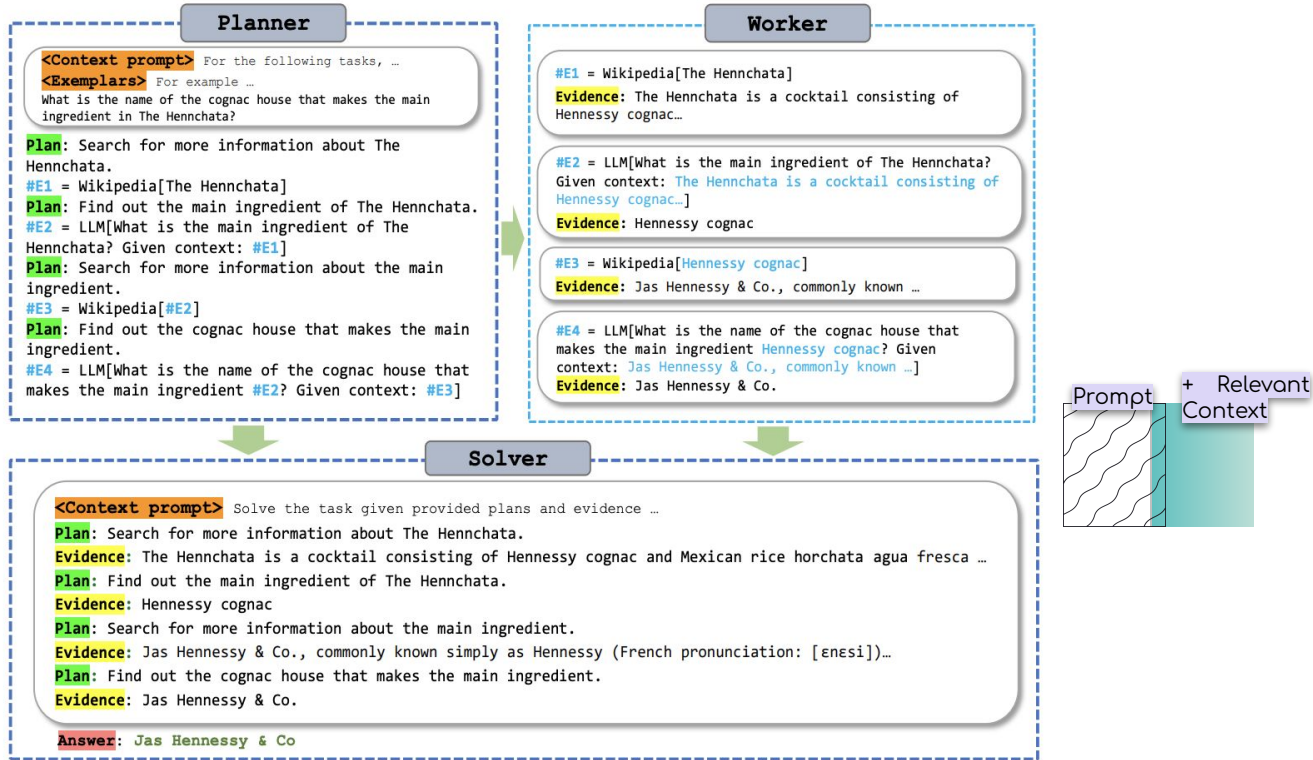


Figure 1: Workflow of ReWOO . Given a question, Planner composes a comprehensive blueprint of interlinked plans prior to tool response. The blueprint instructs Worker to use external tools and collect evidence. Finally, plans and evidence are paired and fed to Solver for the answer.

Introduction

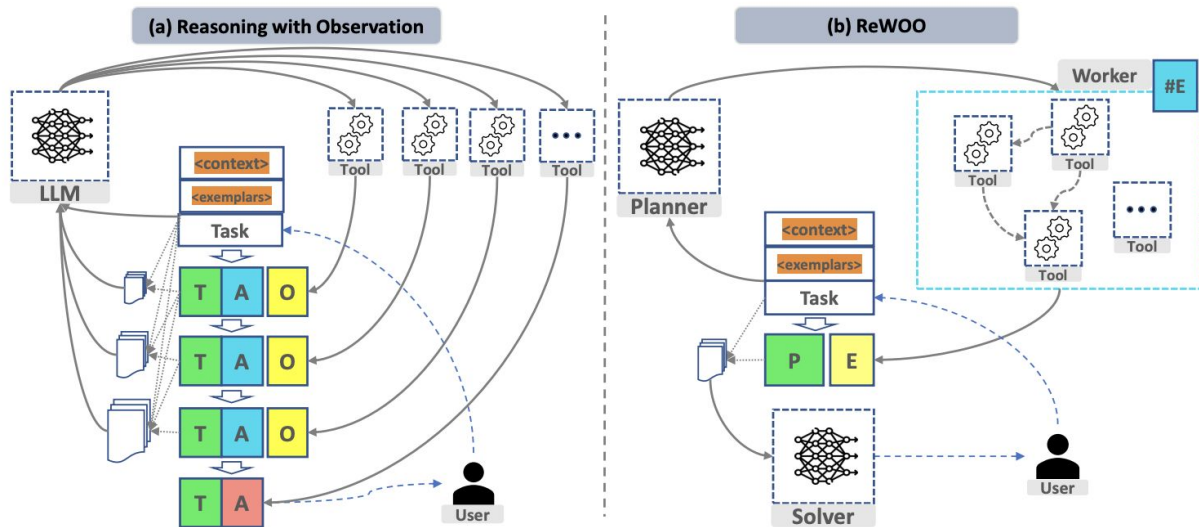


Figure 2: In **(a)** observation-dependent reasoning, the task requested from a user is first wrapped with context prompt and exemplars, then fed into an LLM to initiate a reasoning process. The LLM generates a thought(T) and an action(A), then waits for the observation(O) from tools. The observation is stacked into the prompt history to start the next LLM call. In ReWOO **(b)**, Planner produces at once a list of interdependent plans(P) and calls Worker to fetch evidence(E) from tools. The P and E are combined with the task, and then fed into Solver for the final answer. Note that in (a), the context and exemplars are repeatedly fed into the LLM, resulting in prompt redundancy.



References

Xu, Binfeng, et al. "[Rewoo: Decoupling reasoning from observations for efficient augmented language models.](#)" (2023).

Yao, Shunyu, et al. "[React: Synergizing reasoning and acting in language models.](#)" (2022).

ReWOO LangChain. <https://github.com/langchain-ai/langgraph/blob/main/examples/rewoo/rewoo.ipynb> . Github. 2024.

ReWOO Repository. Decoupling Reasoning from Observations for Efficient Augmented Language Models. Github.2024