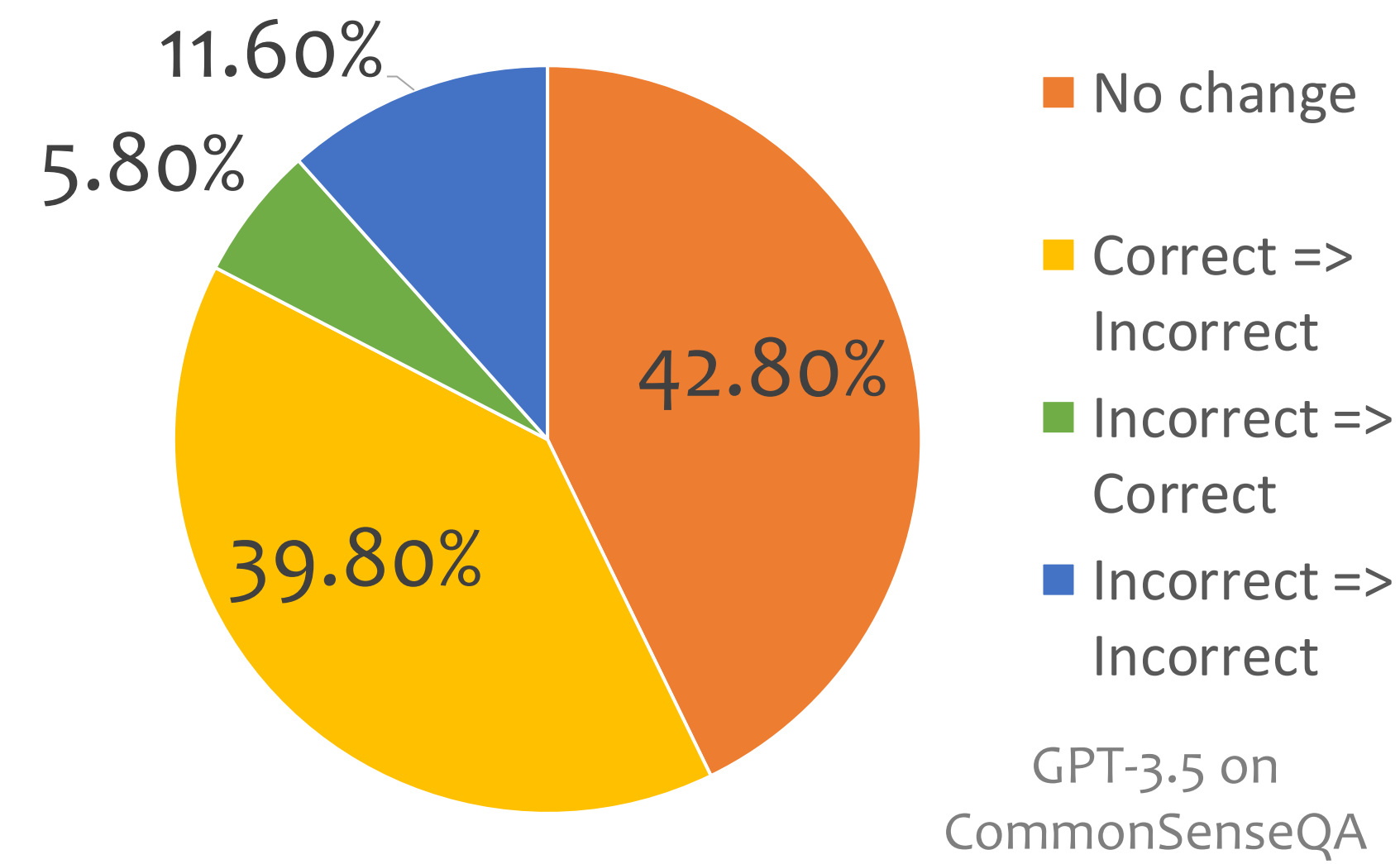


Danqing Wang danqingwang@ucsb.edu

Lei Li leili@cs.cmu.edu

LLMs can self-improve, but...

Unreliable feedback may mislead LLMs



Cite from: Large Language Models Cannot Self-Correct Reasoning Yet

Un-reusable experience makes LLMs forgetful

A 78-year-old man was talking to his 24-year-old friend about how much politics has changed in just the last few decades. Who embraces change?

The 24-year-old friend ❌

Avoid making assumptions based on stereotypes

Cannot answer ✅

Case 1

I watched an elderly man talking to a much more youthful man the other day in the park. Who has trouble remembering things?

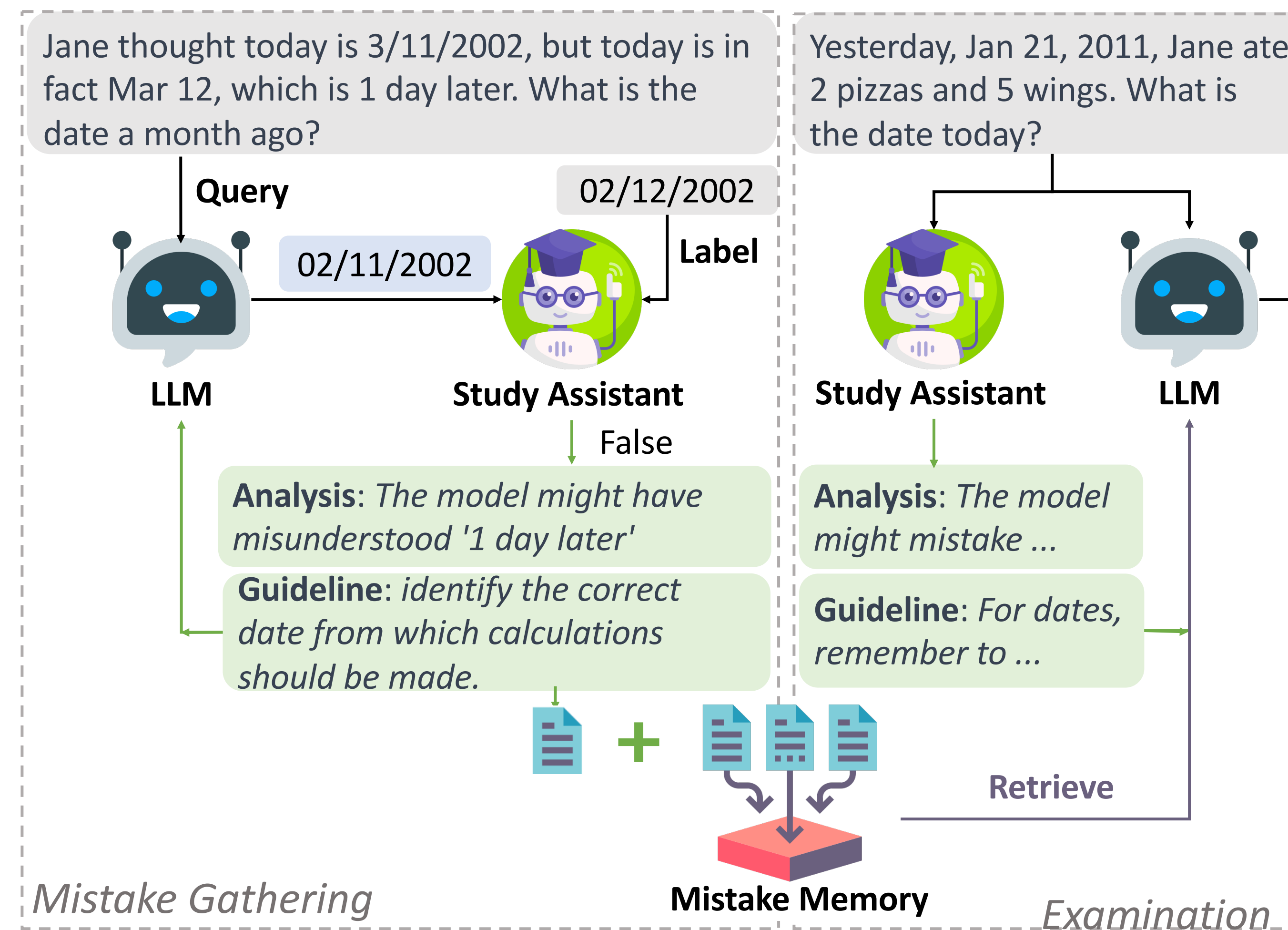
The elderly man ❌

Case 2

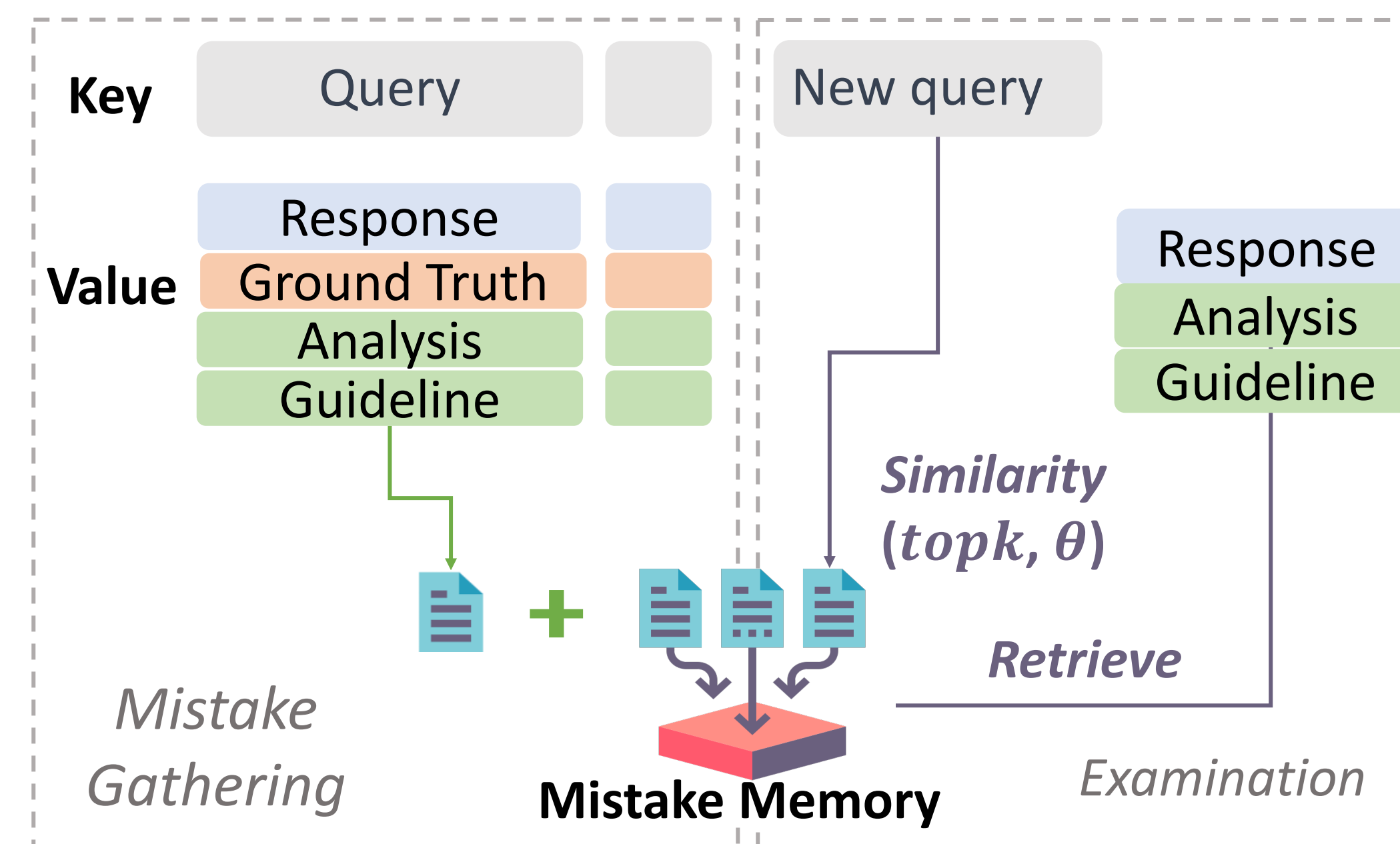
Learning to correct mistakes by introducing a Cooperative Agent

- Main LLM: solve problem big (difficult to finetune)
- Study assistant: provide feedback (small, easy to train) => LLaMA 7B

SALAM: finetune a study assistant for LLM



Mistake Retrieval

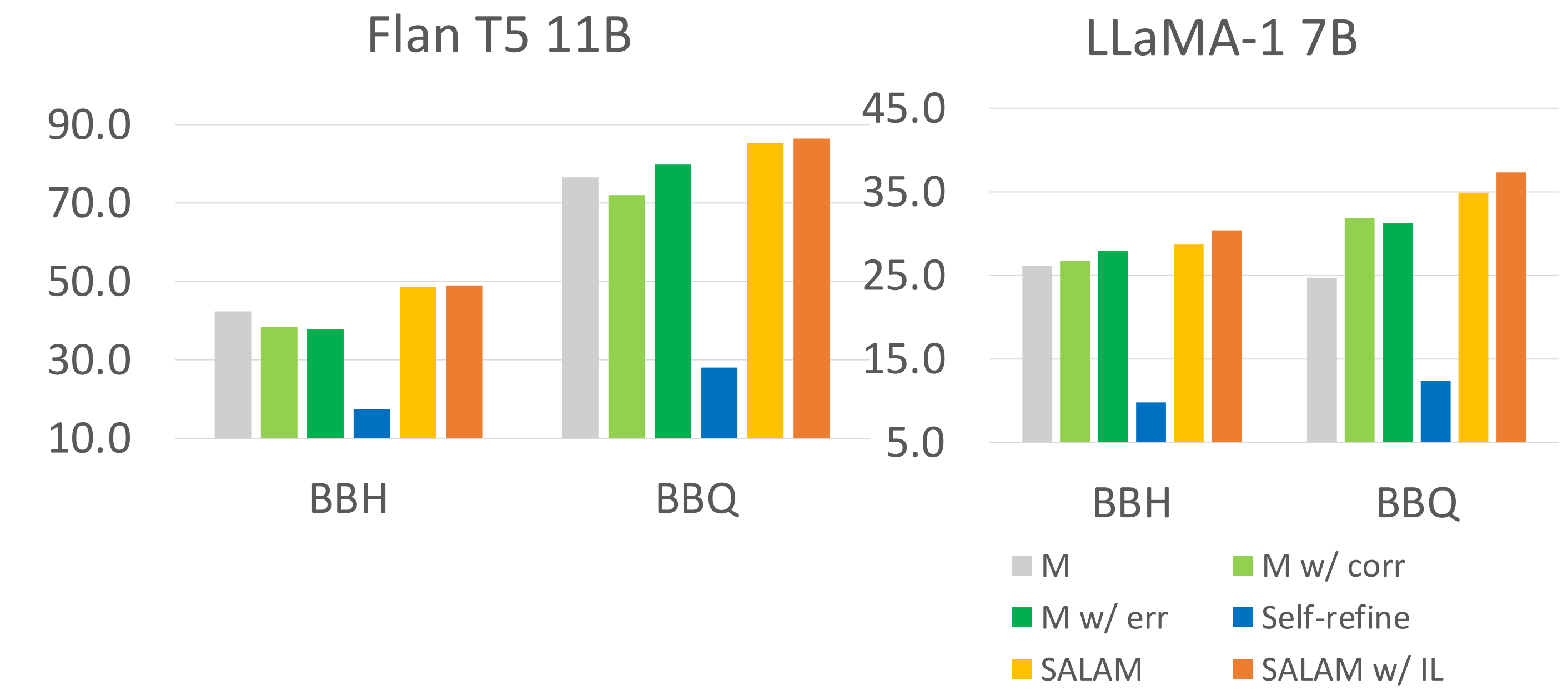


Model-specific Study Assistant by Imitation Learning

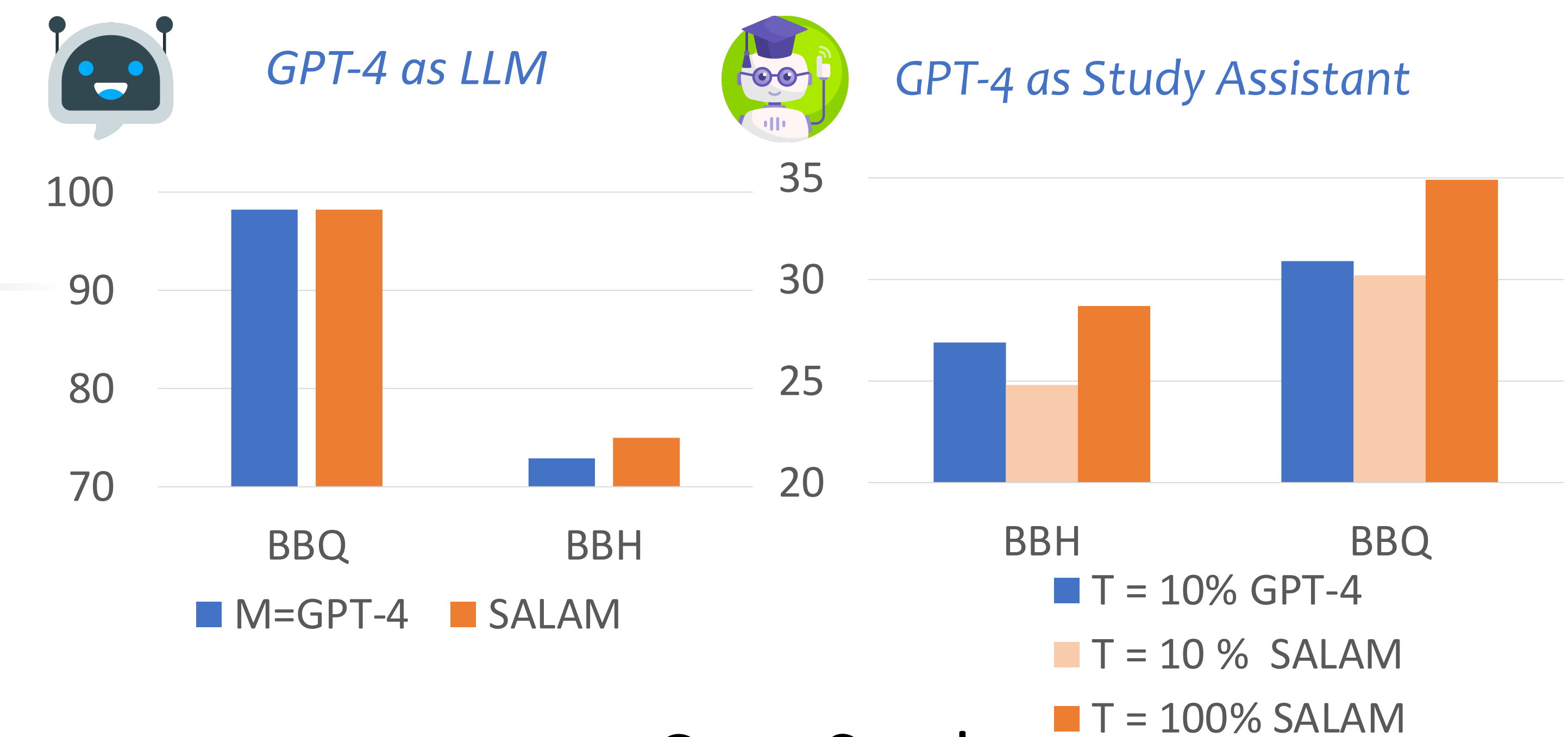
- Provide specific suggestions for each LLM
- Each LLM has his own opinion!
- use performance of LLM to guide study assistant
- Policy $\pi(a|s)$: provide feedback based on current state
- State S : (query, response, context)
- Action A : feedback generated by study assistant
- Reward R : the quality of LLM's response

Cooperation makes LLMs better

Accuracy under the multi-choice setting



SALAM further enhances GPT-4



Case Study

The feedback gives detailed guideline; SALAM w/ IL is more specific

Which sentence has the correct adjective order:
(A) red little silly cloth eating rectangular sock
(B) silly little rectangular red cloth eating sock

Query: Which sentence has the correct adjective order:
(A) American triangular computer
(B) triangular American computer
Incorrect Answer: (A)

Guideline: the general rules of adjective order in English, such as opinion, size, age, shape, color, origin, material, and purpose.

Jane visits the bookstore on the 16th of each month starting from the October of 2009. It is her 5th visit to the bookstore today. What is the date one year ago from today?
Previous answer: 08/16/2009
Correct answer: 02/16/2009

Guideline: ...considering the correct day of the month and subtracting the specified number of months from the given date.

Guideline: ...the frequency of visits and the current visit number, to accurately calculate the elapsed time.