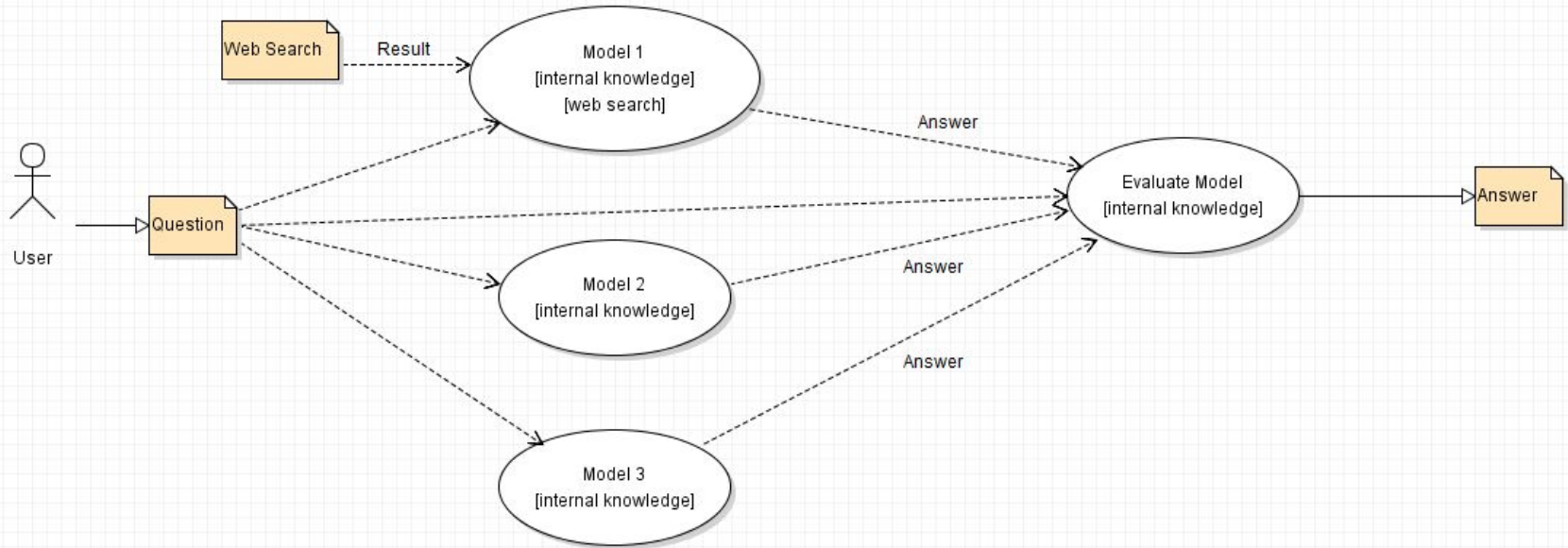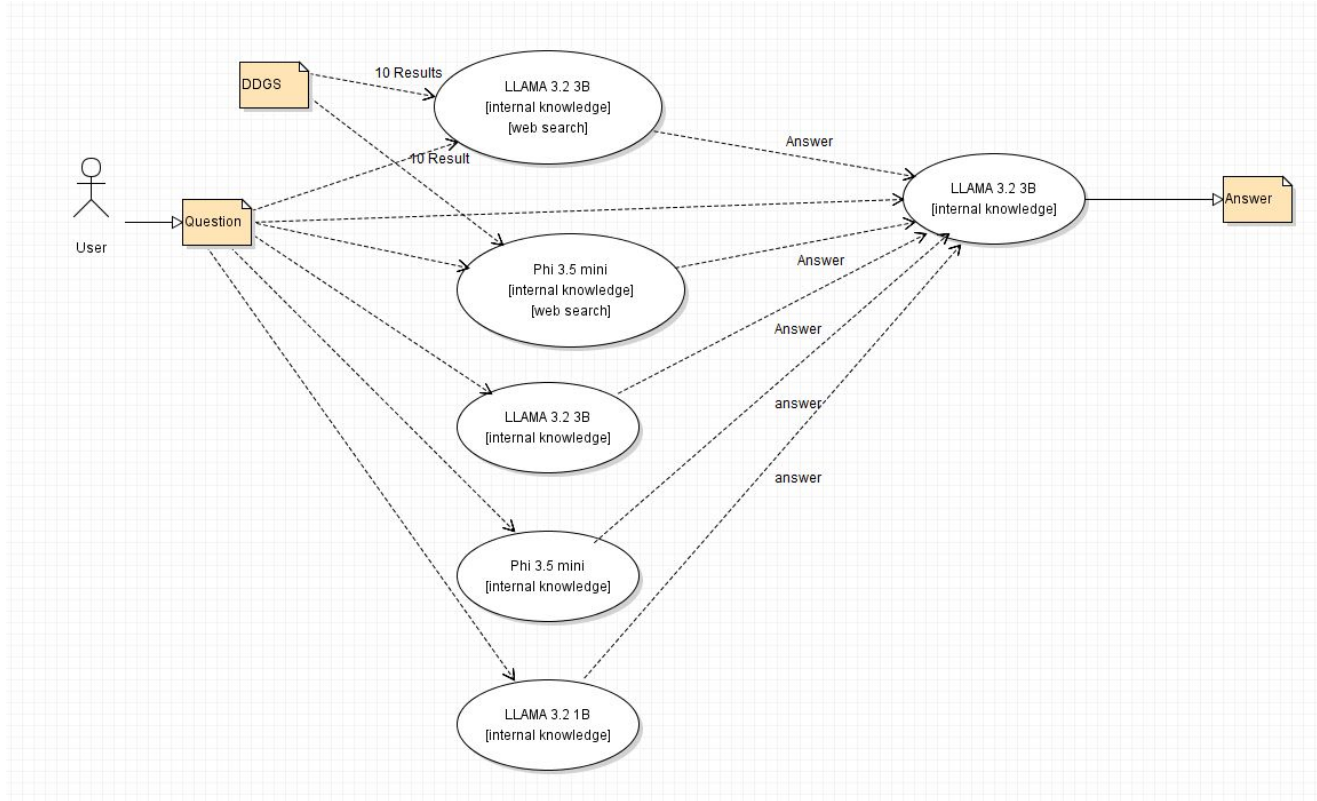# LLM Agents with Web Search

Yunrui Huang
11/13/2024

Framework Overview

Framework Overview (Implementation)

Agents Model

- Planning
  - Check if the web search is needed
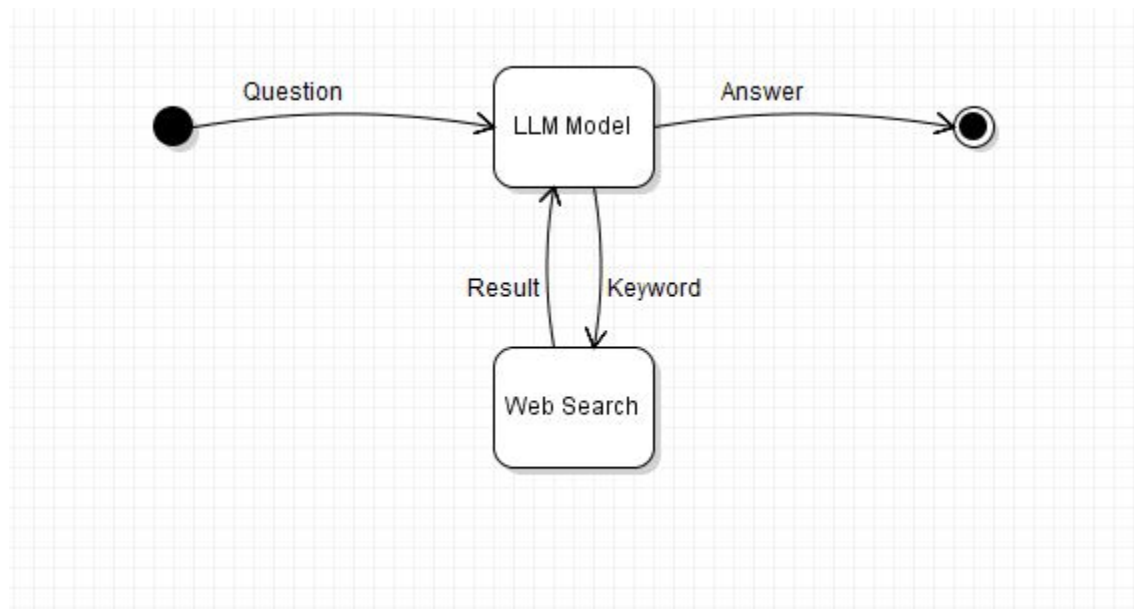 - Design the searching keyword

- Action
  - Using "SEARCH REQUEST" keyword to trigger tools
  - Get the top 10 results from DDGS

- Memory
  - Save the result and provide it to next prompts

Agents Model

LLM With Web Search

```python
def llm_with_web(user_question, model, tokenizer):
    chat = [
        {"role": "system", "content": "You are an assistant with access to a 'web search' tool that helps you find recent or sp
        {"role": "system", "content": "SEARCH REQUEST: When you need additional information to answer a question accurately, o
        {"role": "system", "content": "ANSWER: When you have the information to answer the question, output ANSWER: [your resp
        {"role": "system", "content": "Do not provide both a `SEARCH REQUEST` and `ANSWER` at the same time."},
        {"role": "user", "content": user_question}
    ]

    formatted_chat, response = pipe_line(chat, model, tokenizer)

    if response.startswith("SEARCH REQUEST"):
        print(response[16:])
        if "\n" in response:
            response, second_part = response.split("\n", 1)
        chat.extend(web_search(response[16:],10))

        formatted_chat, response = pipe_line(chat, model, tokenizer)

    return formatted_chat, response
```

Running Case

can you give me some information about SpaceX Starship Fifth Flight Test?

Starship flight test 5

文A 8 languages ∨

Article Talk

Read Edit View history Tools ∨

From Wikipedia, the free encyclopedia

This article is about the fifth integrated Starship/Super Heavy flight test. For the 2020 fifth flight test of the Starship segment, see List of Starship upper stage flight tests.

Starship flight test 5 was the fifth flight test of a SpaceX Starship launch vehicle. The prototype vehicles flown were the Starship Ship 30 upper-stage and Super Heavy Booster 12. This launch is notable for being the first time an orbital-class rocket has been caught out of mid air.

After launching and delivering the Starship upper stage into a suborbital trajectory heading toward a splashdown in the Indian Ocean, the Super Heavy booster turned around and fired its Raptor engines to return back to the launch site. As the booster approached the launch pad it slowed to a near hover and did a horizontal slide maneuver to line itself up with two massive "chopstick" arms on the launch tower, dubbed "Mechazilla." The arms then closed around the booster before the engines shut down.

The rocket launched on the morning of 13 October, one day after the Federal Aviation Administration (FAA) issued a launch permit that had been delayed since early August and after weeks of increasingly public feuding between SpaceX and the FAA.

Starship flight test 5

Booster 12 on final approach to the launch tower

Names Integrated Flight Test-5

Web Search Keywords

Llama 3.2 3B:

```
chat_with_web_Llama323B, response_with_web_Llama323B = llm_with_web(user_input,Llama323B_model, Llama323B_tokenizer)
chat_without_web_Llama323B, response_without_web_Llama323B = llm_without_web(user_input,Llama323B_model, Llama323B_tokenizer)
✓ 7.1s
Starting from v4.46, the `logits` model output will have the same type as the model (except at train time, where it will always be FP32)
SpaceX Starship SN5 launch
```

Phi 3.5 mini:

```
chat_with_web_Phi35mini, response_with_web_Phi35mini = llm_with_web(user_input,Phi35mini_model, Phi35mini_tokenizer)
chat_without_web_Phi35mini, response_without_web_Phi35mini = llm_without_web(user_input,Phi35mini_model, Phi35mini_tokenizer)
✓ 16.0s
Starting from v4.46, the `logits` model output will have the same type as the model (except at train time, where it will always be FP32)
SpaceX Starship 5 lunch menu details
```

Prompts with Searching Results

[{"role": system, "content": "You are an assistant with access to a 'web search' tool that helps you find recent or specific information online. Use the followin
    {"role": system, "content": "SEARCH REQUEST: When you need additional information to answer a question accurately, output SEARCH REQUEST: [query]. For exampl
    {"role": system, "content": "ANSWER: When you have the information to answer the question, output ANSWER: [your response]. For example: ANSWER: The recent Ca
    {"role": system, "content": "Do not provide both a `SEARCH REQUEST` and `ANSWER` at the same time."},
    {"role": user, "content": "can you give me some information about SpaceX Starship Fifth Flight Test?"},
    {'role': system, 'content': 'Information about SpaceX Starship Fifth Flight Test'},
    {'role': system, 'content': 'The flight test concluded at splashdown 1 hour, 5 minutes and 40 seconds after launch. The entire SpaceX team should take pride
    {'role': system, 'content': 'SpaceX launched its enormous Starship rocket for the fifth time ever Oct. 13, on a test flight that featured a mid-air catch of
    {'role': system, 'content': 'What we covered here. • SpaceX launched Starship, the most powerful rocket ever built, on its fifth test flight. • Liftoff of th
    {'role': system, 'content': 'Starship flight test 5 was the fifth flight test of a SpaceX Starship launch vehicle. The prototype vehicles flown were the Star
    {'role': system, 'content': 'SpaceX will try on Sunday during its fifth test flight of Starship, the giant next-generation rocket that is the most powerful e
    {'role': system, 'content': "SpaceX launched its fifth test flight of its Starship rocket on Sunday and made a dramatic first catch of the rocket's more than
    {'role': system, 'content': 'CNN —. Federal regulators granted SpaceX its long-awaited license to move forward with a fifth uncrewed test launch of Starship,
    {'role': system, 'content': 'SpaceX is hoping that this fifth test flight of its Starship megarocket marks a world first. Follow its progress by hitting watc
    {'role': system, 'content': "00:00. 00:00. Oct 13 (Reuters) - SpaceX in its fifth Starship test flight on Sunday returned the rocket's towering first stage b
    {'role': system, 'content': "Today's flight test, which was delayed 25 minutes while SpaceX waited for its launch range to be cleared of boats, marks the sec

Evaluation

```python
def evaluate_answer(user_question, answers, model, tokenizer):
    chat = [
        {"role": "system", "content": "You are an intelligent assistant with expertise in evaluating answers for q
        {"role": "system", "content": "Consider the following criteria to select the best answer:"},
        {"role": "system", "content": "Accuracy: How correct and factual the information is."},
        {"role": "system", "content": "Clarity: How clearly the answer is presented."},
        {"role": "system", "content": "After reviewing, state which answer is the best."},
        {"role": "system", "content": "For example: ANSWER: The answer [the best answer number] is the best."},
        {"role": "user", "content": user_question}
    ]
    chat.extend(answers)

    formatted_chat, response = pipe_line(chat, model, tokenizer)

    return formatted_chat, response
```

Combine Answers in Pormpts

```
answer = [{"role": "answer1", "content": response_with_web_Llama323B},
          {"role": "answer2", "content": response_without_web_Llama323B},
          {"role": "answer3", "content": response_with_web_Phi35mini},
          {"role": "answer4", "content": response_without_web_Phi35mini},
          {"role": "answer5", "content": response_without_web_Llama321B}]

chat_evaluate, response_evaluate = evaluate_answer(user_input,answer,Llama323B_model, Llama323B_tokenizer)
```
✓ 4.8s

Evaluation output

ANSWER: The answer 1 is the best. It provides accurate information about the SpaceX Starship Fifth Flight Test, which took place on November 10, 2021, at SpaceX's Starbase facility in Boca Chica, Texas. The test reached an altitude of approximately 33.5 kilometers (21 miles) and flew for around 2 minutes and 40 seconds, but unfortunately, the vehicle did not land vertically and disintegrated upon re-entry, resulting in a failed test.

Issue

- Lazy to search
  - When the LLM Found the relate result in their internal knowledge, even it is out date, the model will still using that information
- Wrong information
  - As the running case show, the Llama 3.2 treat the starship Fifth Flight Test as starship SN5
  - The SN5 is the internal knowledge it has but fifth flight test isn't
- Evaluation
  - Can't figure out the information correct or not

Summary

- Web search tool do help model get up to date information
- Model more tend to using internal knowledge
- Web search data would be ambiguous or bias
- Mult search rounds may helpful for clear information
- Evaluate model limitations in verifying the factual accuracy

References

- Guo, T., Chen, X., Wang, Y., Chang, R., Pei, S., Chawla, N. V., Wiest, O., & Zhang, X. (2024). *Large Language Model based Multi-Agents: A Survey of Progress and Challenges* (No. arXiv:2402.01680). arXiv. https://doi.org/10.48550/arXiv.2402.01680
- Starship flight test 5. (2024). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Starship_flight_test_5&oldid=1257221851
- *duckduckgo-search: Search for words, documents, images, news, maps and text translation using the DuckDuckGo.com search engine.* (Version 6.3.4). (n.d.). [Python; OS Independent]. Retrieved November 13, 2024, from https://github.com/deedy5/duckduckgo_search