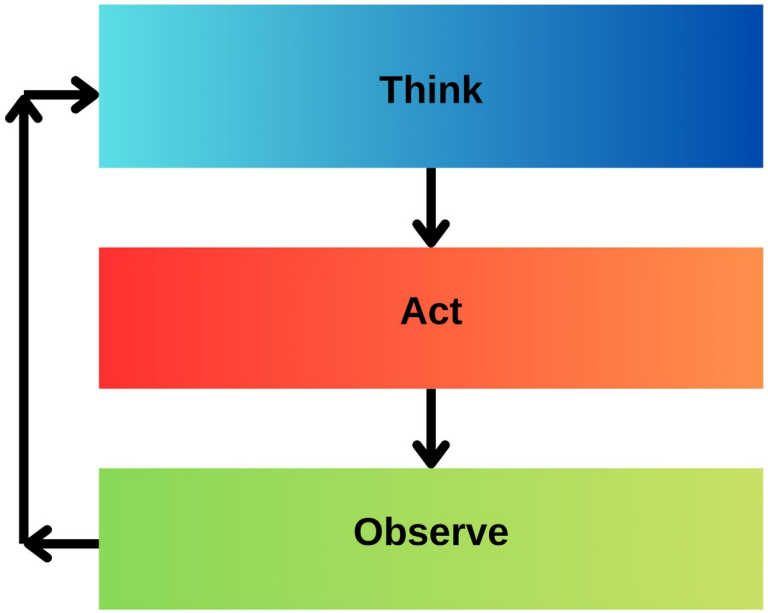
AI Agents

Thinking and Doing - Recent Work on Reasoning Models, Test-Time Scaling, AI Agents, and Agentic AI

ReAct [Reasoning + Acting]



Loop:
1. Send prompt to LLM
2. LLM replies with text output (may include Thought, Action, or Answer)
3. Display or record LLM output

4. If output contains "Action:" then:
   Extract the action command from output
   Execute the action (e.g., run search or query)
   Get the observation result from action execution
   Append the LLM output and the observation to the prompt

5. Else if output contains "Answer:" then:
   This is the final answer
   Stop the loop

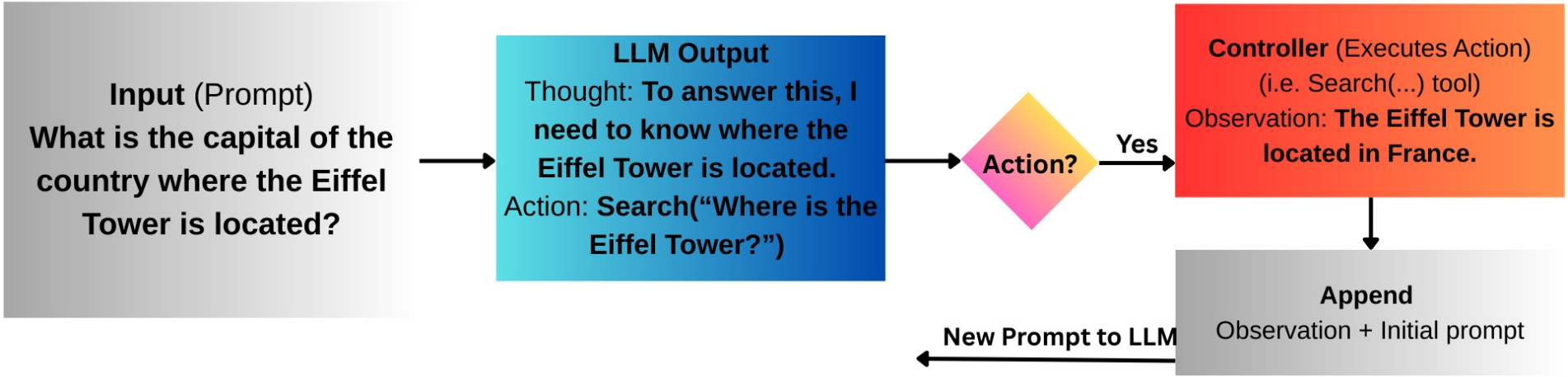
6. Else:
   Append the LLM output to the prompt
   Continue the loop

End loop

Yao, S. et al. (2023) 'ReAct: Synergizing Reasoning and Acting in Language Models'. arXiv. Available at: https://doi.org/10.48550/arXiv.2210.03629.

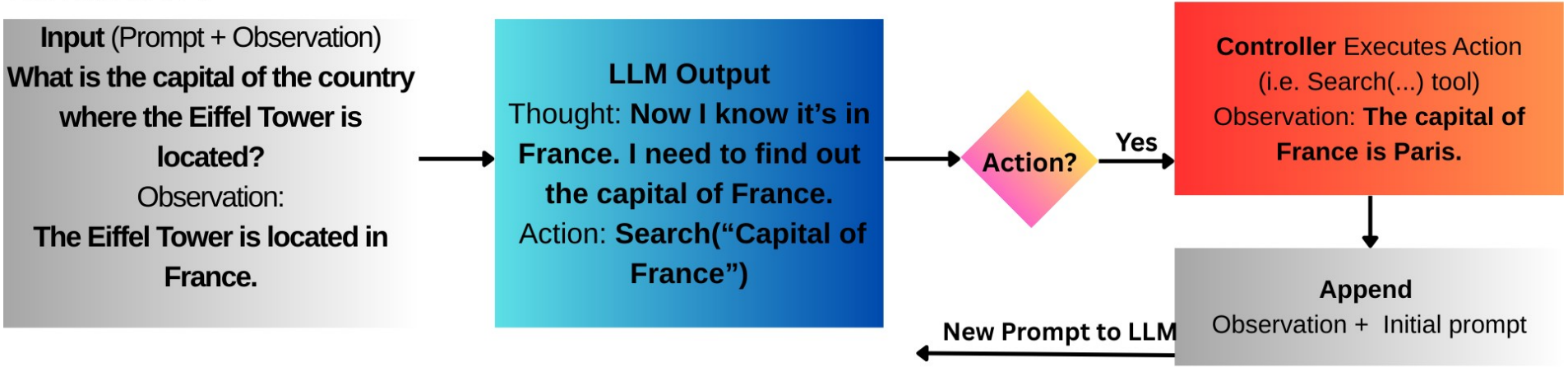
Prompt = "What is the capital of the country where the Eiffel is Located?"
How would ReAct tackle this?

Iteration 1:



The LLM has new information to work with. Initial prompt + the new fact.

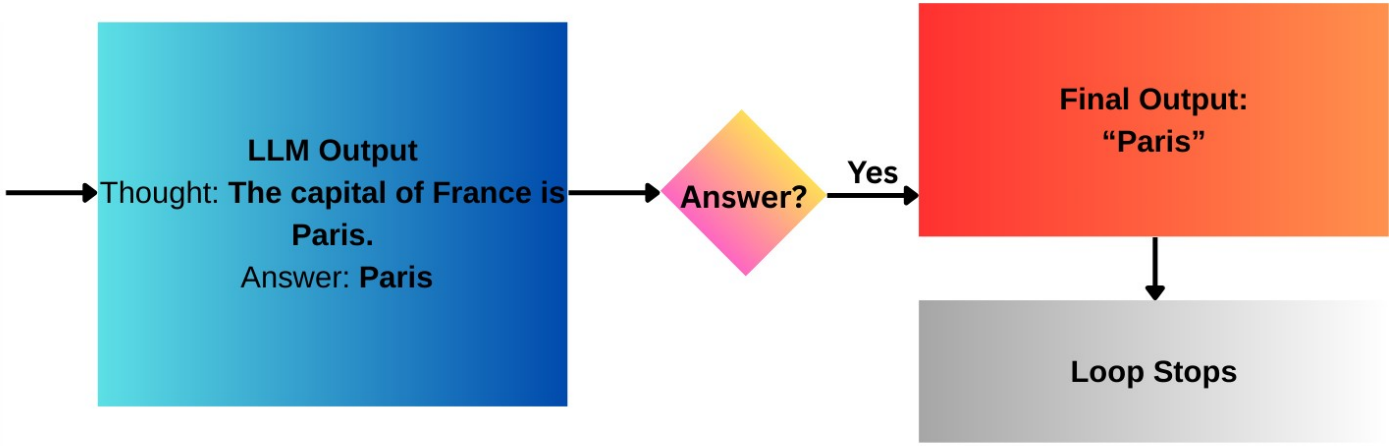
Iteration 2:



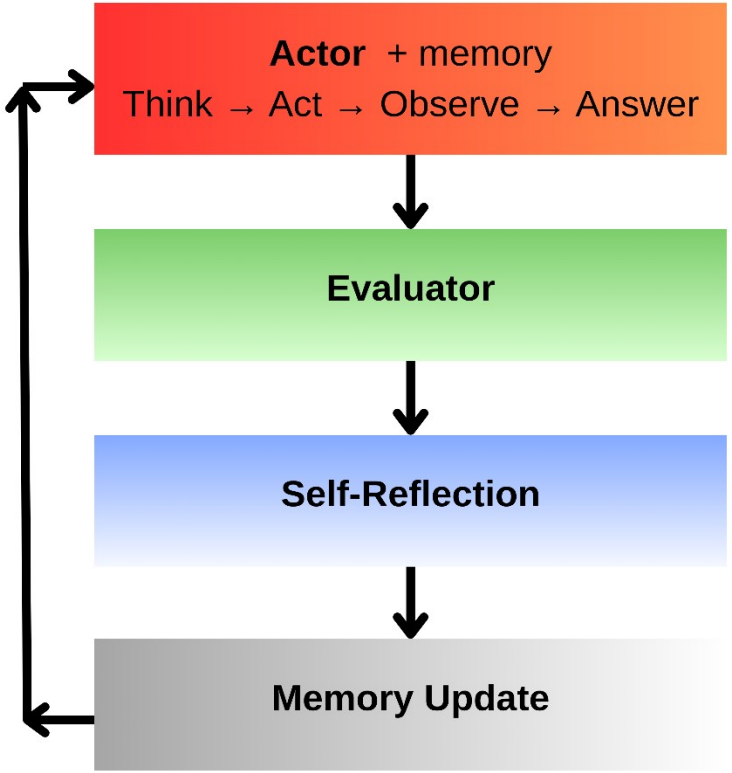
Then finally the LLM has all the pieces. It makes a final thought.

Iteration 3:

Input (Prompt + Observation)
What is the capital of the country where the Eiffel Tower is located?
Observation:
The Eiffel Tower is located in France.
Thought: **Now I know it's in France. I need to find out the capital of France.**
Action: **Search("Capital of France")**
Observation: **The capital of France is Paris.**



Reflexion



Input: current state(Prompt/ observation) + memory
Generates trajectory τt = [(s1,t1,a1),...,(sN,tN,aN)]

Input: trajectory τt
Output: Reward ex: (rt = 0 or 1)

Input: {current trajectory, reward, memory}
Output: verbal feedback srt

Append srt to long-term memory
Memory now contains: [srt-2, srt-1, srt]

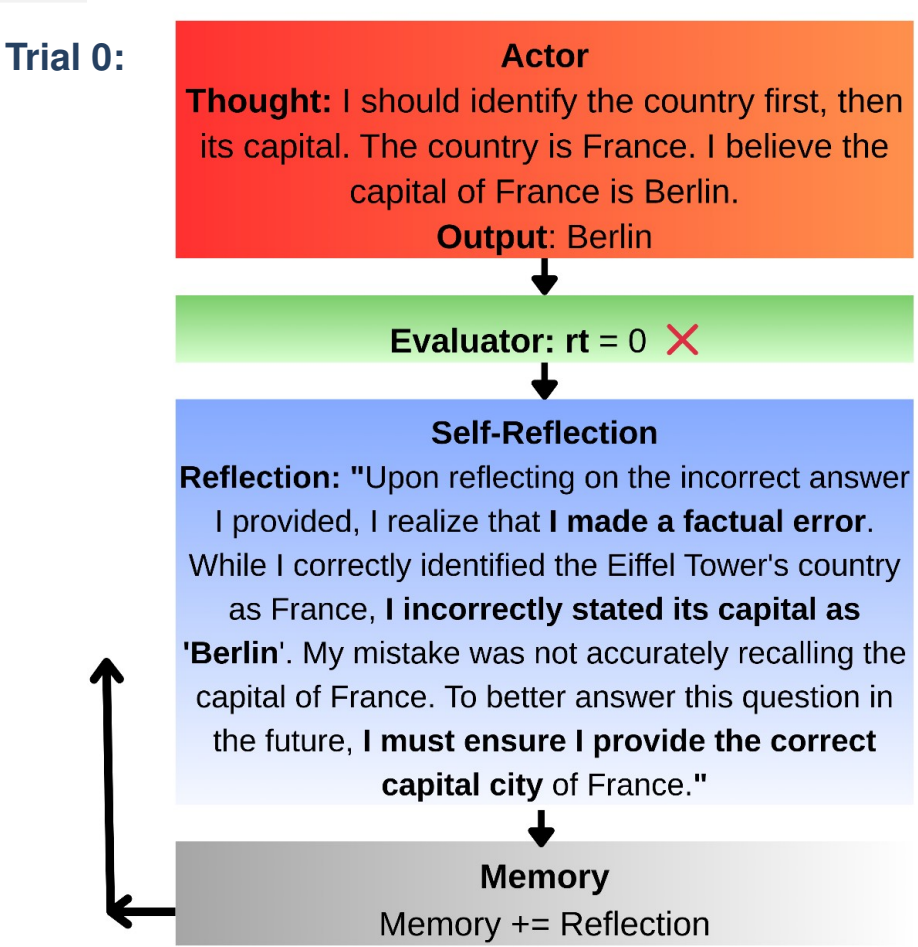
Shinn, N. et al. (2023) 'Reflexion: Language Agents with Verbal Reinforcement Learning'. arXiv. Available at: https://doi.org/10.48550/arXiv.2303.11366.

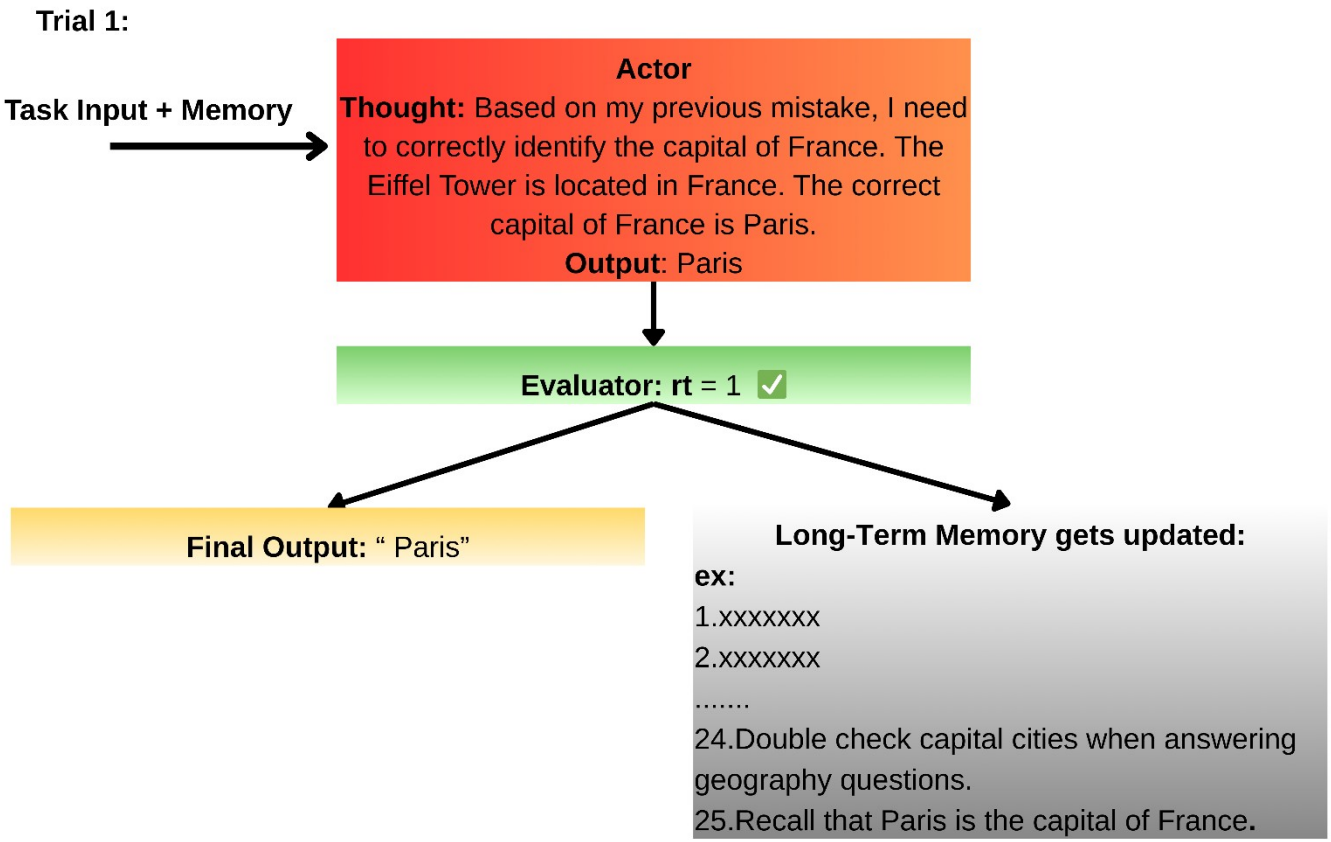
Reflexion

INPUT: task
LOOP:
  Actor receives prompt and memory
  Actor THINKS, ACTS, OBSERVES repeatedly
  Actor outputs a trajectory ending with a final answer

  Evaluator compares Actor's answer with correct answer

  IF Evaluator score (rt) is 1 (correct):
    SHOW final answer
    EXIT loop
  ELSE #score is 0 (Fail)
    Self-Reflection generates feedback
    Append feedback to memory
    CONTINUE loop




Trial 1:
Task Input + Memory
Actor
Thought: Based on my previous mistake, I need to correctly identify the capital of France. The Eiffel Tower is located in France. The correct capital of France is Paris.
Output: Paris
Evaluator: rt = 1
Final Output: “ Paris”
Long-Term Memory gets updated:
ex:
1.xxxxxxx
2.xxxxxxx
.....
24.Double check capital cities when answering geography questions.
25.Recall that Paris is the capital of France.

Evaluator

1. Reasoning Tasks: Exact Match (EM) Grading
- Binary reward: $r_t = 1$ if answer exactly matches ground truth, else 0

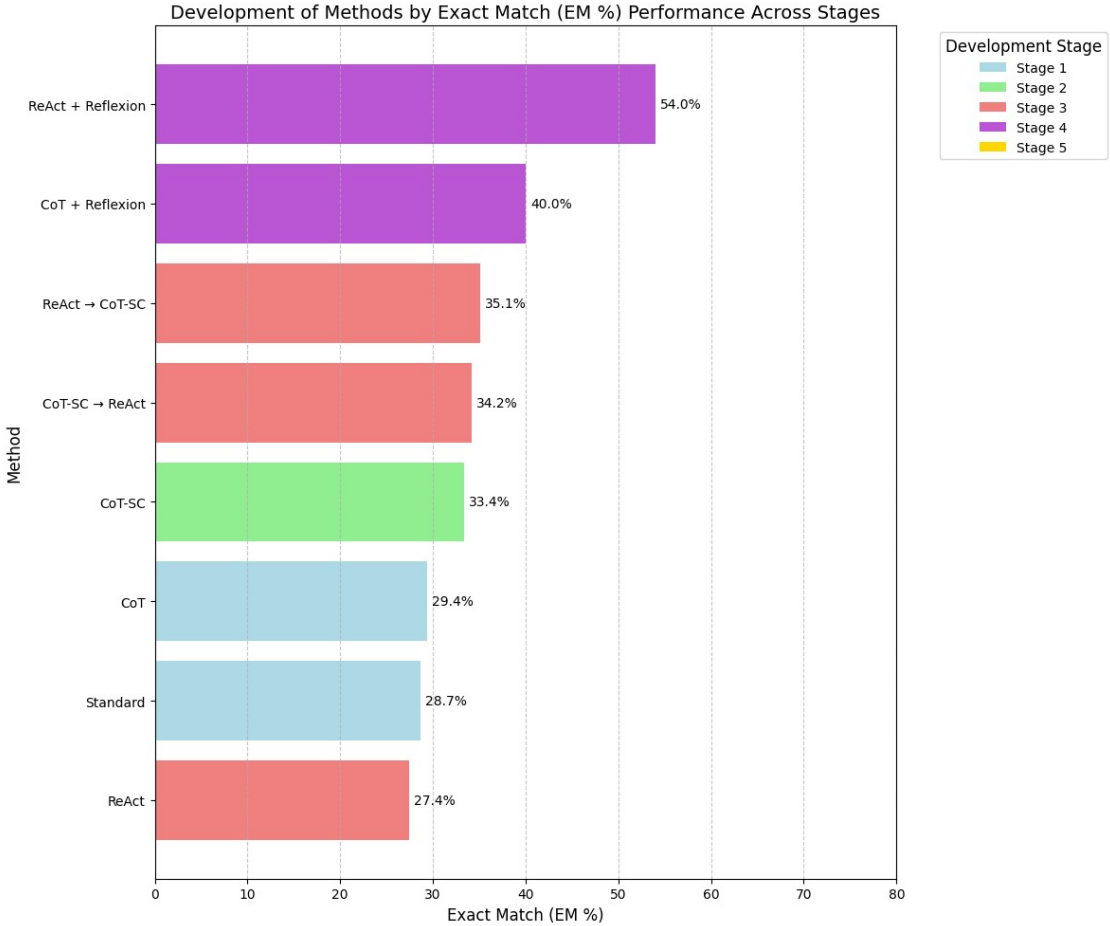
2.Decision making task: Predefined Heuristic Functions
-Task-specific logic to detect poor planning or action repetition

3. Programming Tasks: Unit Tests
-Agent creates unit tests, Tests must pass for $r_t = 1$

4. LLM as Evaluator :
- Another instance of an LLM used to interpret and critique outputs

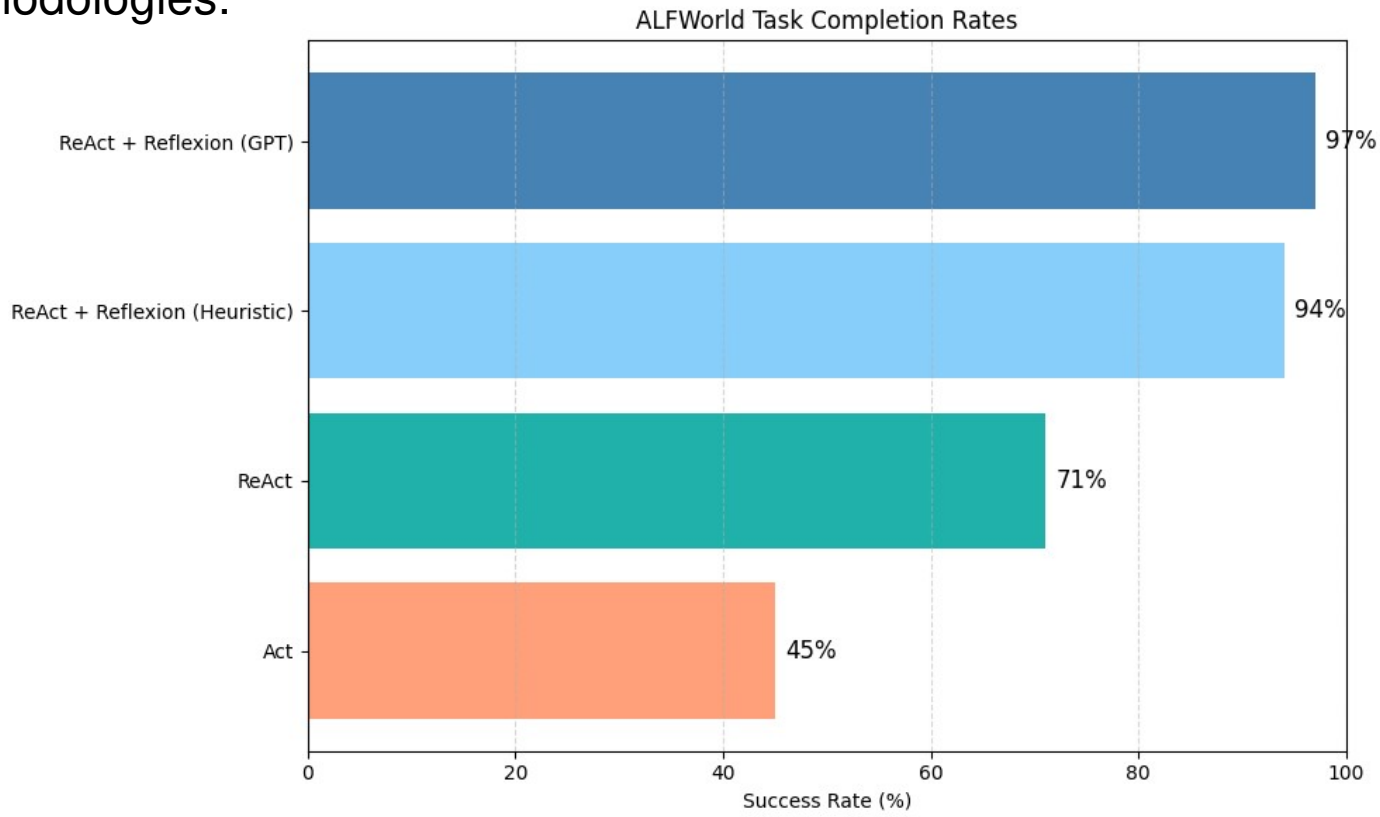
Development in Methodologies:

Dataset: HotpotQA
Model:PaLM-540B



ReAct (Yao, S. et al. (2023)), Reflexion (Shinn, N. et al. (2023))

Development in Methodologies:
Dataset: AlfWorld
Model:GPT-3



ReAct (Yao, S. et al. (2023)), Reflexion (Shinn, N. et al. (2023))