



UCLA NLP Group

Chameleon: Compositional Reasoning with Large Language Models

Pan Lu
01/25/2024



Chain-of-Thought Prompting (CoT)

Question: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Standard Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27. ❌

Chain-of-Thought Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9. ✅

CoT
reasoning
processes

However, when it comes to complex problems...

Table:

Tour boat schedule					
Ocean City	8:15 A.M.	9:00 A.M.	9:15 A.M.	9:30 A.M.	10:00 A.M.
Whale Watch Harbor	9:30 A.M.	10:15 A.M.	10:30 A.M.	10:45 A.M.	11:15 A.M.
Oyster Lighthouse	10:15 A.M.	11:00 A.M.	11:15 A.M.	11:30 A.M.	12:00 P.M.
Fisherman's Cove	11:15 A.M.	12:00 P.M.	12:15 P.M.	12:30 P.M.	1:00 P.M.
Surfing Beach	12:00 P.M.	12:45 P.M.	1:00 P.M.	1:15 P.M.	1:45 P.M.

Question: Look at the following schedule. Haley is at Ocean City at 9:45 A.M. How soon can she get to Surfing Beach?

Options: (A) 1:45 P.M. (B) 12:00 P.M. (C) 10:30 A.M. (D) 12:15 P.M.



GPT-3
(text-davinci-002)

Solution:

(Step 1) Find 9:45 A.M. on the schedule.

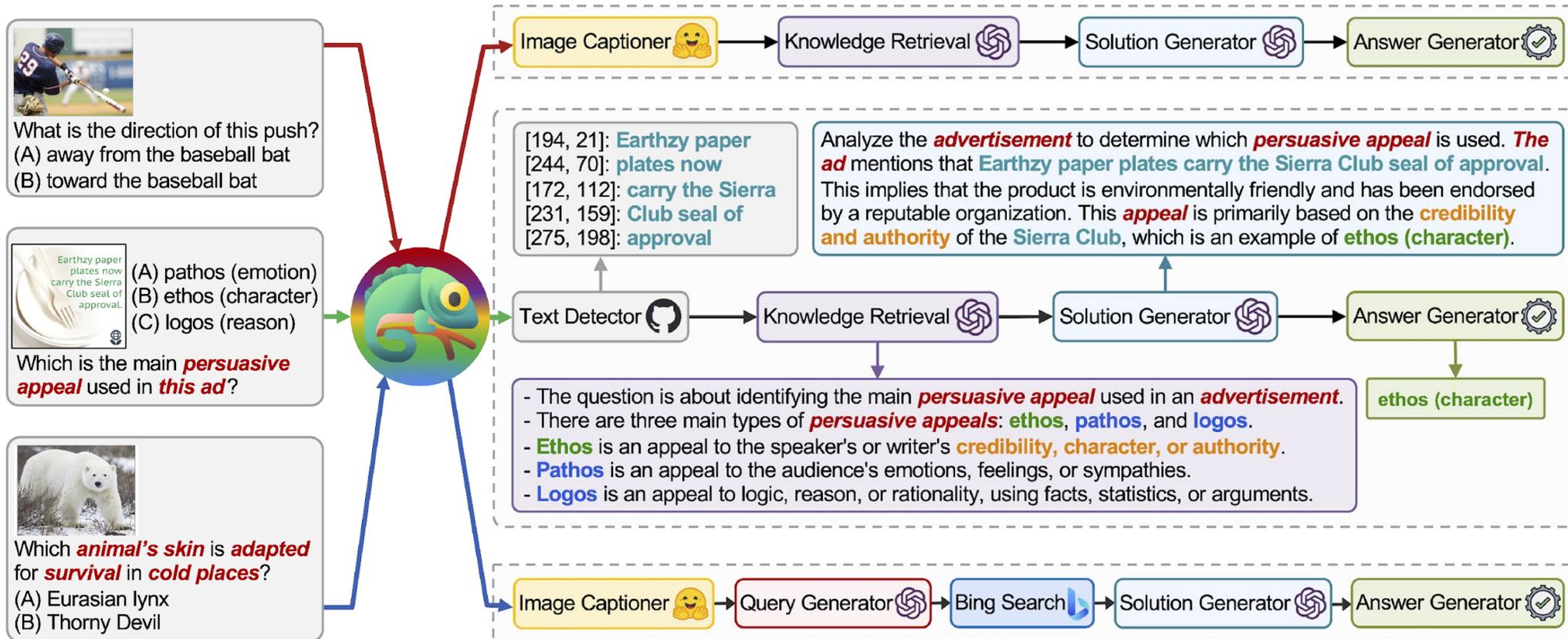
(Step 2) **The next stop is Whale Watch Harbor at 10:15 A.M.**

(Step 3) The earliest she can get to Surfing Beach is **12:00 P.M.** The **answer is 12:00 P.M.**

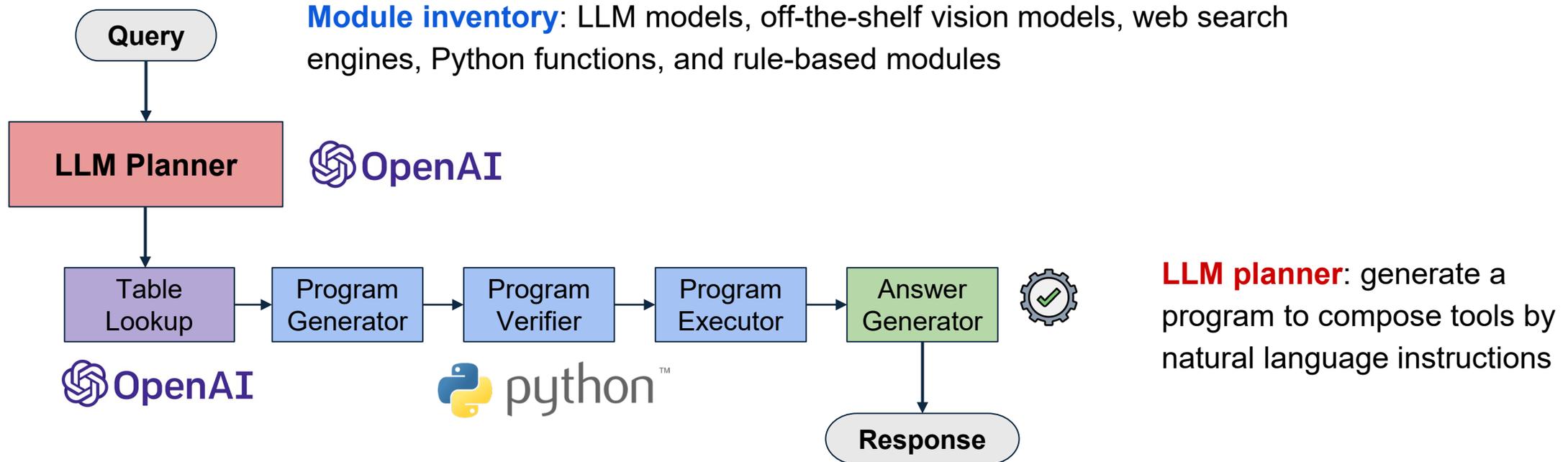
Tour boat schedule					
Ocean City	8:15 A.M.	9:00 A.M.	9:15 A.M.	9:30 A.M.	10:00 A.M.
Whale Watch Harbor	9:30 A.M.	10:15 A.M.	10:30 A.M.	10:45 A.M.	11:15 A.M.
Oyster Lighthouse	10:15 A.M.	11:00 A.M.	11:15 A.M.	11:30 A.M.	12:00 P.M.
Fisherman's Cove	11:15 A.M.	12:00 P.M.	12:15 P.M.	12:30 P.M.	1:00 P.M.
Surfing Beach	12:00 P.M.	12:45 P.M.	1:00 P.M.	1:15 P.M.	1:45 P.M.



Chameleon: Plug-and-Play Compositional Reasoning



Chameleon: Module Inventory and LLM Planner





Chameleon for Science Question Answering



Earthzy paper plates now carry the Sierra Club seal of approval.

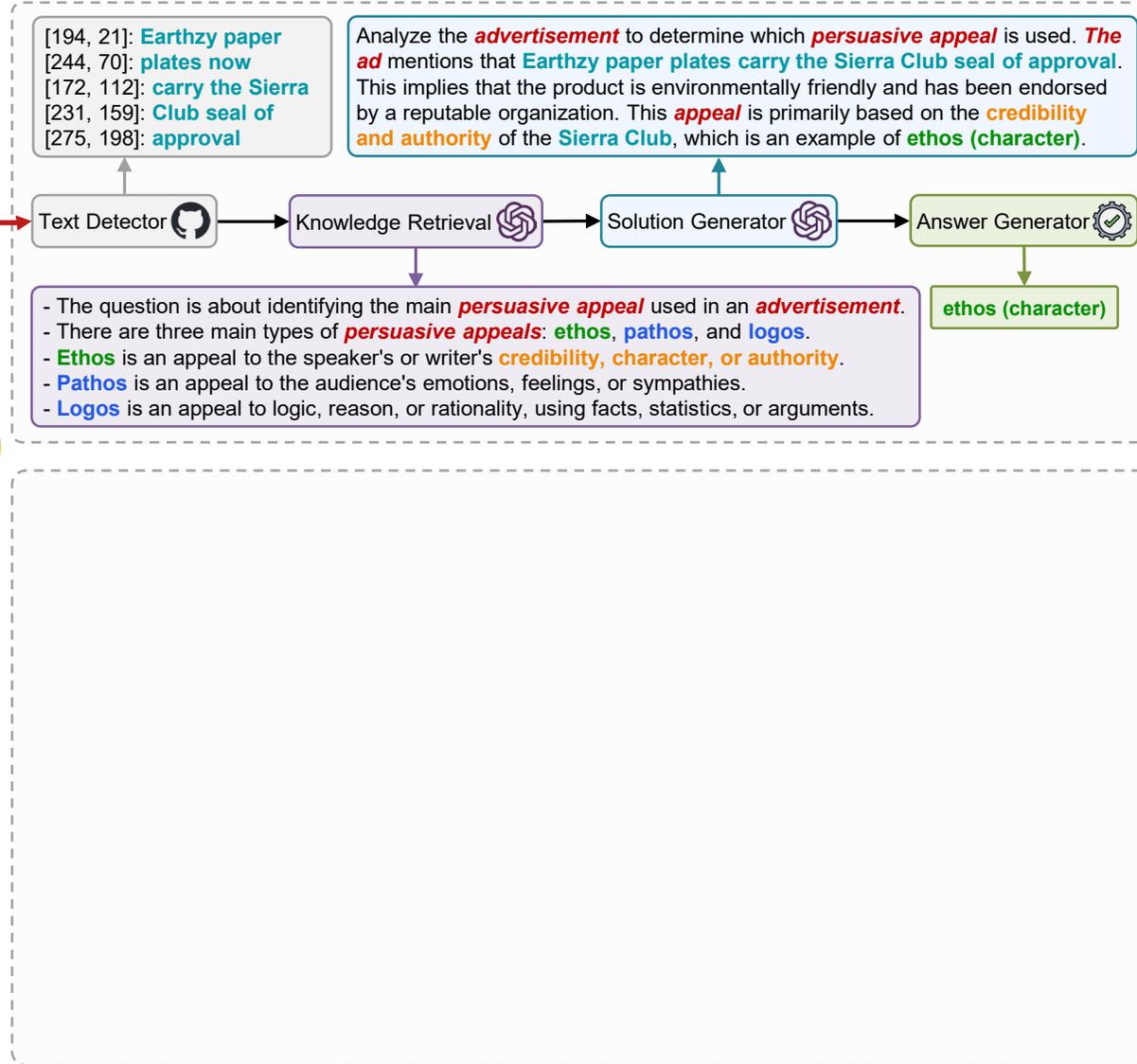
(A) pathos (emotion)
(B) ethos (character)
(C) logos (reason)

Which is the main **persuasive appeal** used in **this ad**?



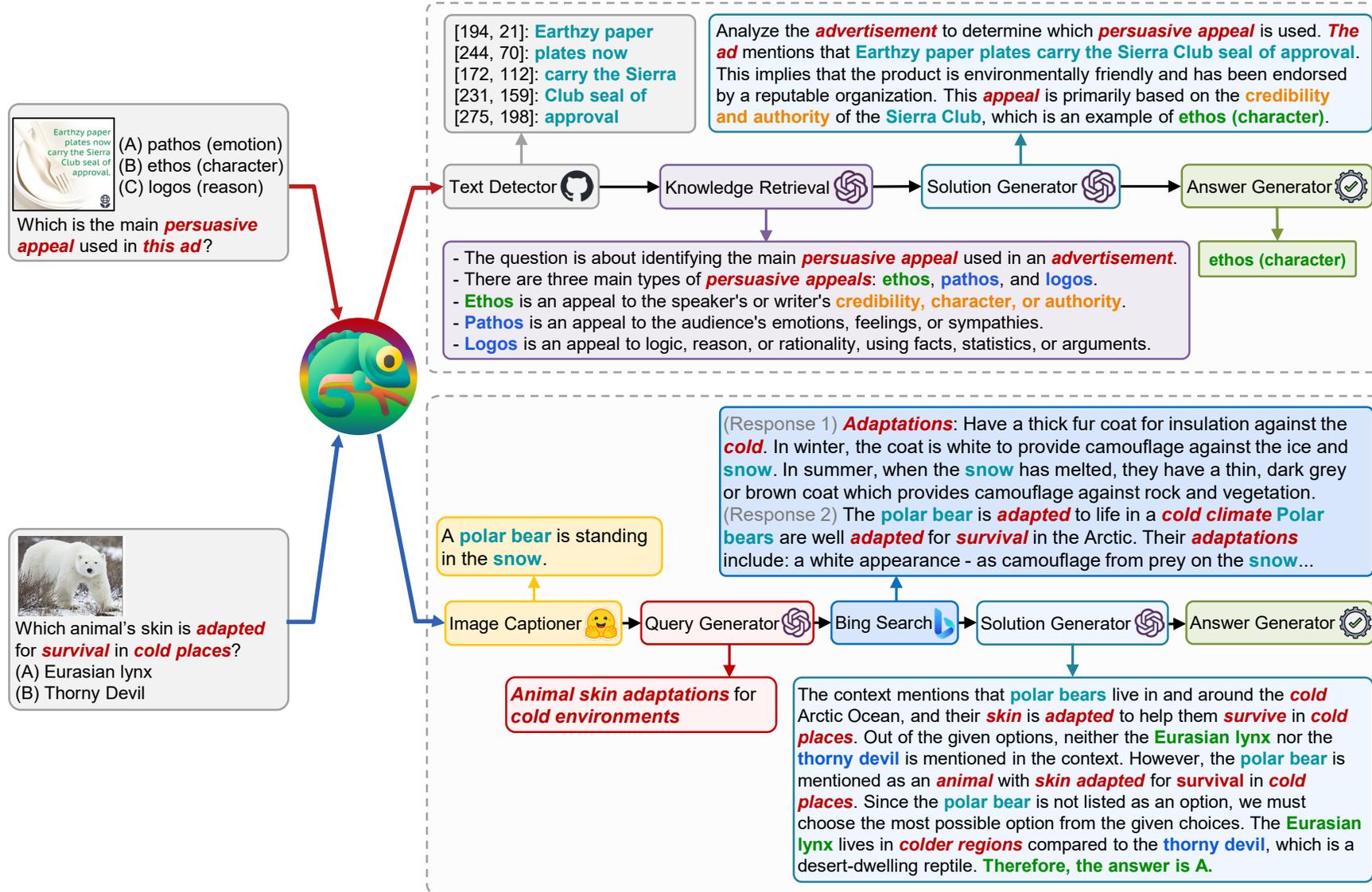

Which animal's skin is **adapted** for **survival** in **cold places**?

(A) Eurasian lynx
(B) Thorny Devil



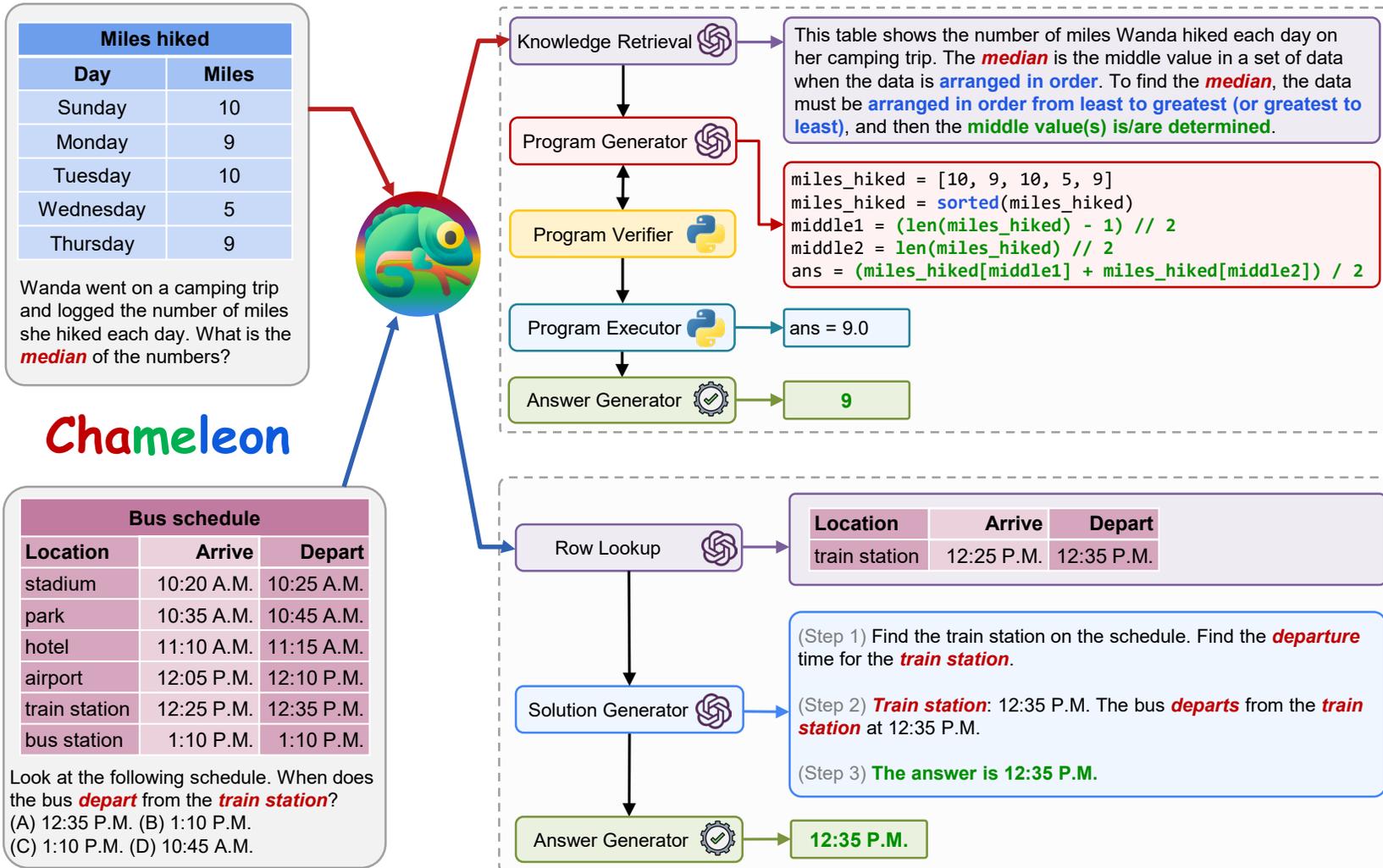


Chameleon for Science Question Answering



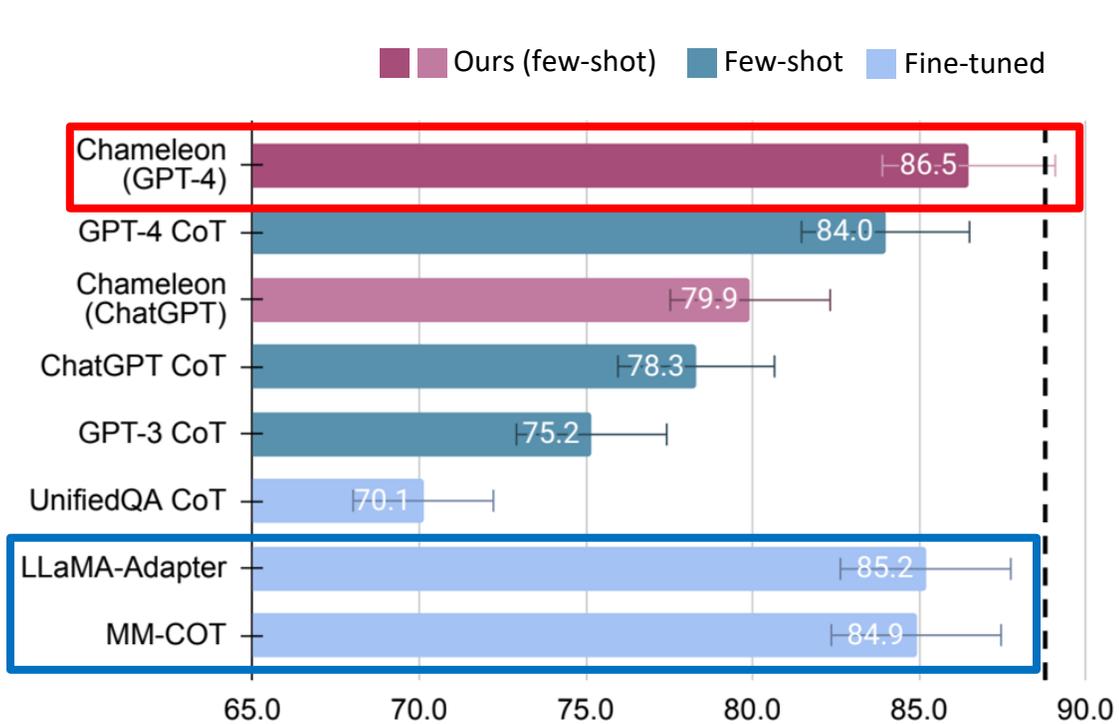


Chameleon for Mathematical Reasoning

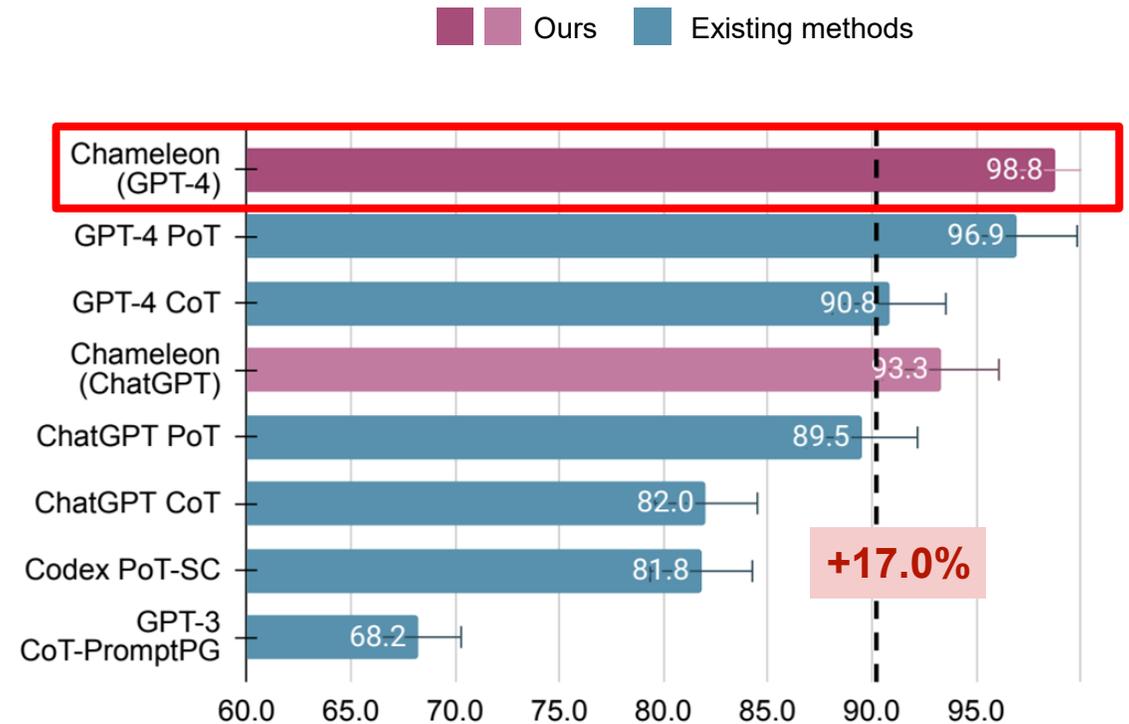




Experimental Results from Chameleon



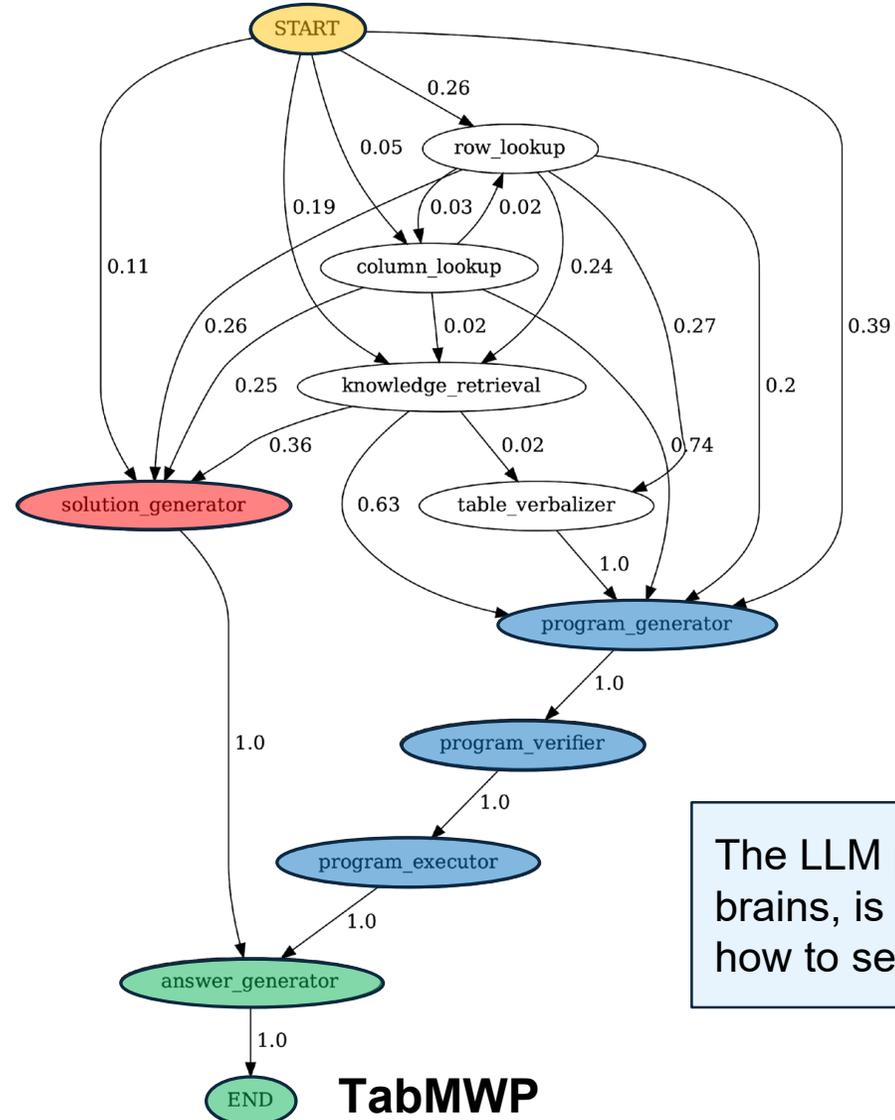
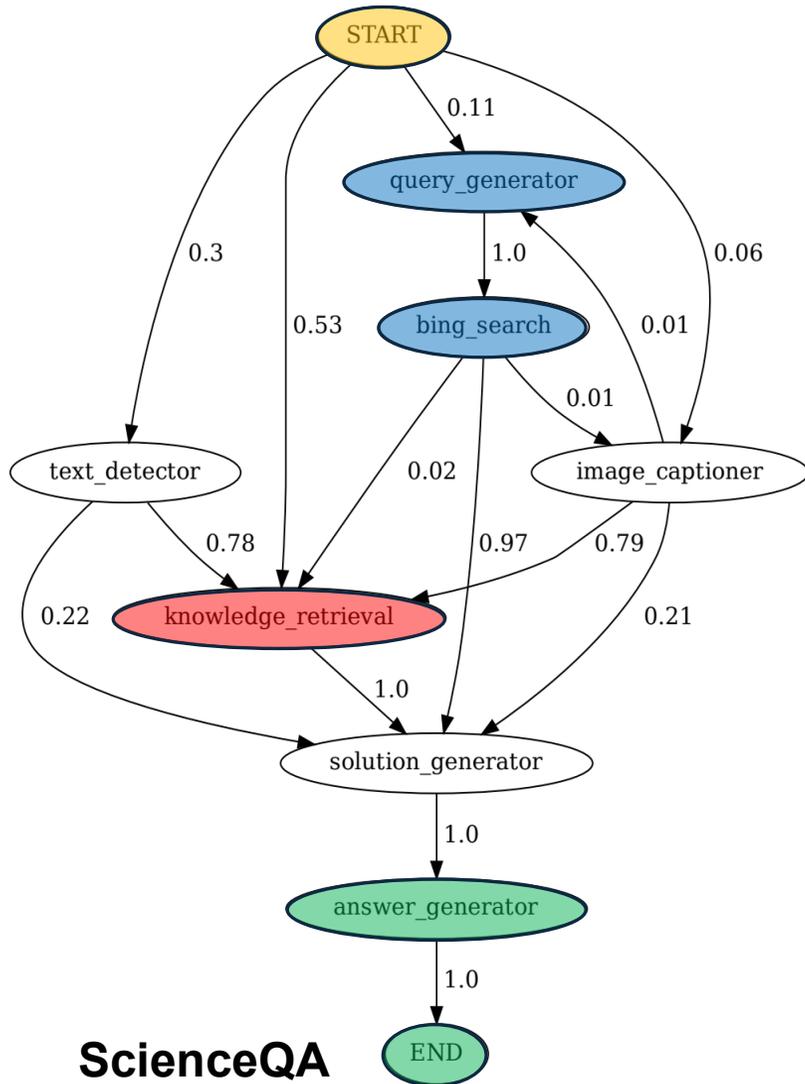
ScienceQA



TabMWP

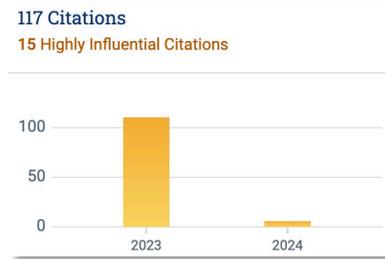


Transition Graphs from Chameleon

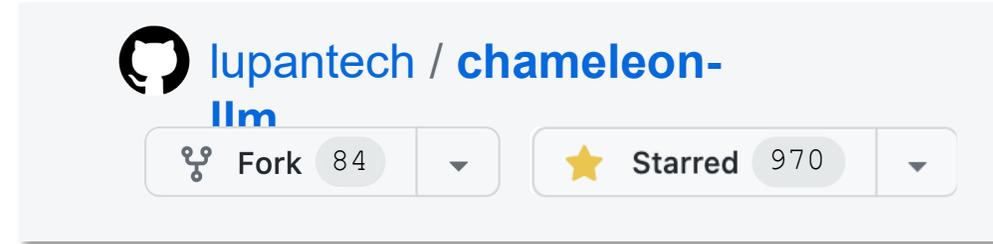


The LLM planner, mimicking our human brains, is able to make good decisions on how to sequence tools in a few-shot setup.

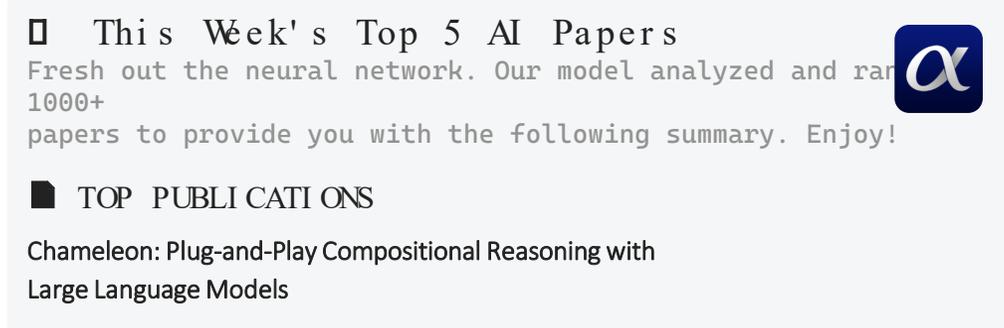
High Impacts of Chameleon



High Research Impact
(Cited 100+ after May 2023)

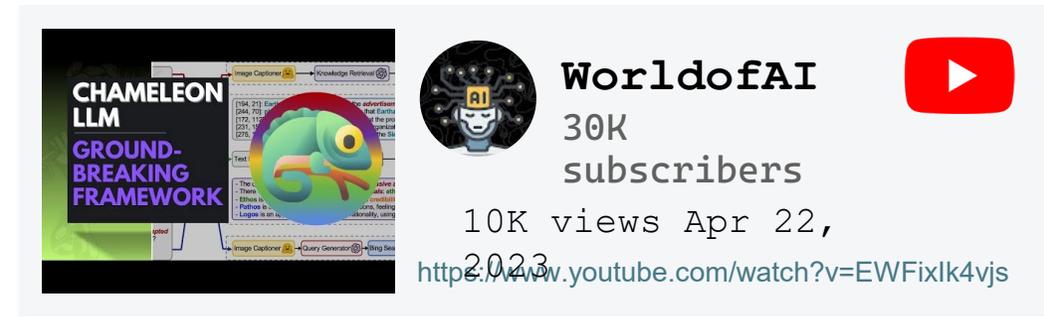


Open-Source Contribution
(Top trending program in May 2023)



Score: 9.9

Best Weekly AI Paper
(by AlphaSignal, 1st in 1682 = 0.06%)



High Feedback in the Community
(“GROUND-BREAKING FRAMEWORK”)

10+ Invited Talks



Thank you!

