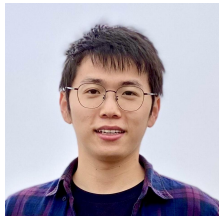
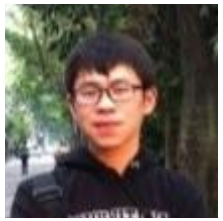
Microsoft

UCLA

Chameleon: Plug-and-Play Compositional Reasoning with Large Language Models


Pan Lu

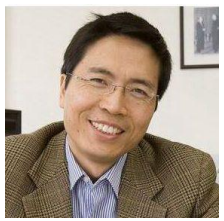

Baolin Peng

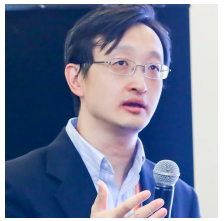

Hao Cheng


Michel Galley


Kai-Wei Chang


Ying Nian Wu


Song-Chun Zhu


Jianfeng Gao

NEURAL INFORMATION PROCESSING SYSTEMS

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.

Google python™ Bing

PromptPG: Dynamic Prompt Learning via Policy Gradient for Semi-structured Mathematical Reasoning, ICLR 2023

Tool Use

Hugging Face

Bing

OpenAI

GitHub

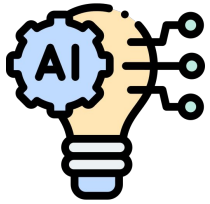
Bard AI

Google

python™

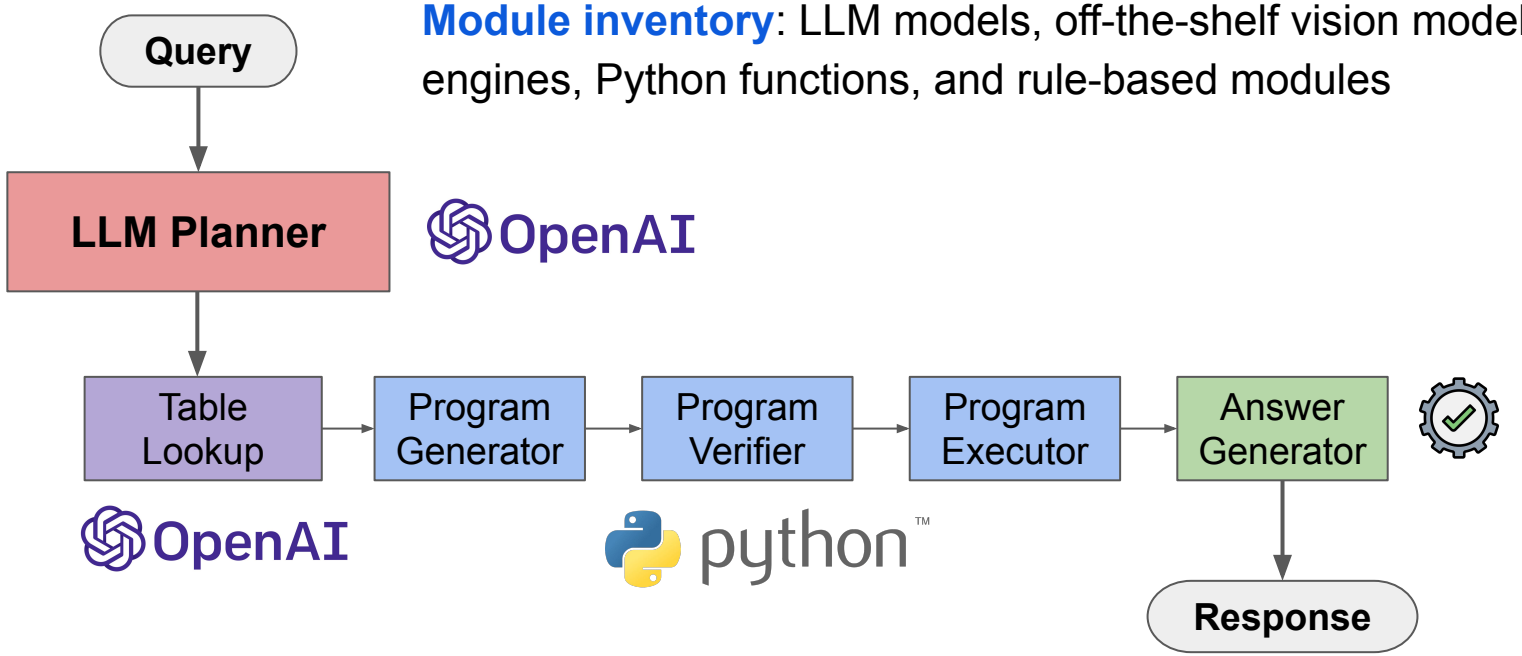
OpenAI
DALLE-2

How to **compose** these numerous tools to tackle complex tasks?



Tool-Augmented LLMs (LLM Agents)!

Chameleon: Module Inventory and LLM Planner



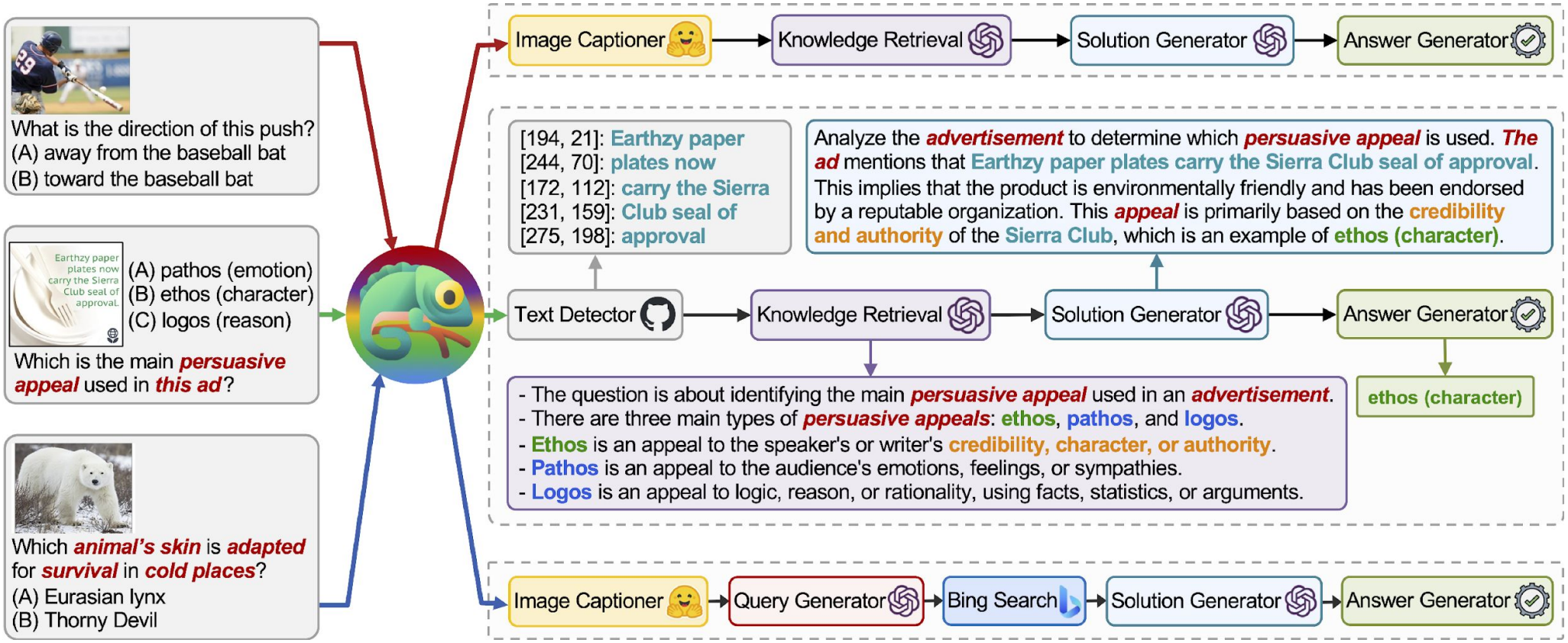
Module inventory: LLM models, off-the-shelf vision models, web search engines, Python functions, and rule-based modules

LLM planner: generate a program to compose tools by natural language instructions

Advantages:

- Efficiently extendable to using new modules
- Natural-language-like programs are less error-prone, easy to debug, and user-friendly
- Flexible to replace the underlying LLM for the planner as well as each module

Chameleon: Plug-and-Play Compositional Reasoning



Chameleon: LLM Planner

OpenAI

You need to act as a policy model, that given a question and a modular set, determines the sequence of modules that can be executed sequentially to solve the query.

The modules are defined as follows:

Query_Generator: This module generates a search engine query for the given question. Normally, we consider using "Query_Generator" when the question involves domain-specific knowledge.

Bing_Search: This module searches the web for relevant information to the question. Normally, we consider using "Bing_Search" when the question involves domain-specific knowledge.

Image_Captioner: This module generates a caption for the given image. Normally, we consider using "Image_Captioner" when the question involves the semantic understanding of the image, and the "has_image" field in the metadata is True.

Text_Detector: This module detects the text in the given image. Normally, we consider using "Text_Detector" when the question involves the unfolding of the text in the image, e.g., diagram, chart, table, map, etc., and the "has_image" field in the metadata is True.

Knowledge_Retrieval: This module retrieves background knowledge as the hint for the given question. Normally, we consider using "Knowledge_Retrieval" when the background knowledge is helpful to guide the solution.

Solution_Generator: This module generates a detailed solution to the question based on the information provided. Normally, "Solution_Generator" will incorporate the information from "Query_Generator", "Bing_Search", "Image_Captioner", "Text_Detector", and "Knowledge_Retrieval".

Answer_Generator: This module extracts the final answer in a short form from the solution or execution result.

Chameleon: LLM Planner

OpenAI

Below are some examples that map the problem to the modules.

Question: Compare the average kinetic energies of the particles in each sample. Which sample has the higher temperature?

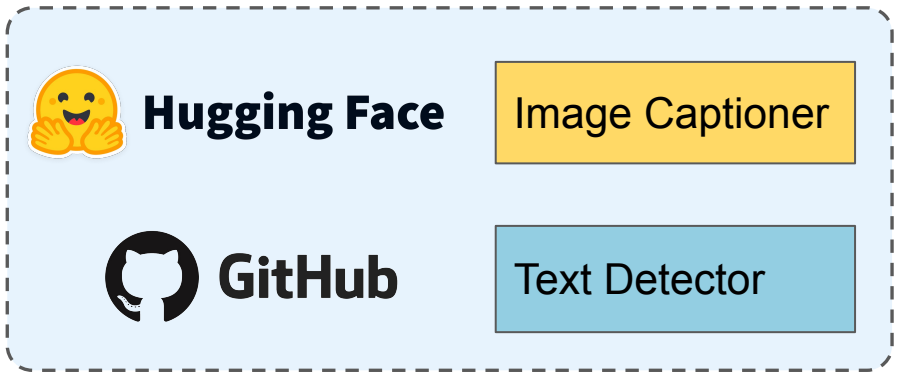
Context: The diagrams below show two pure samples of gas in identical closed, rigid containers. Each colored ball represents one gas particle. Both samples have the same number of particles.

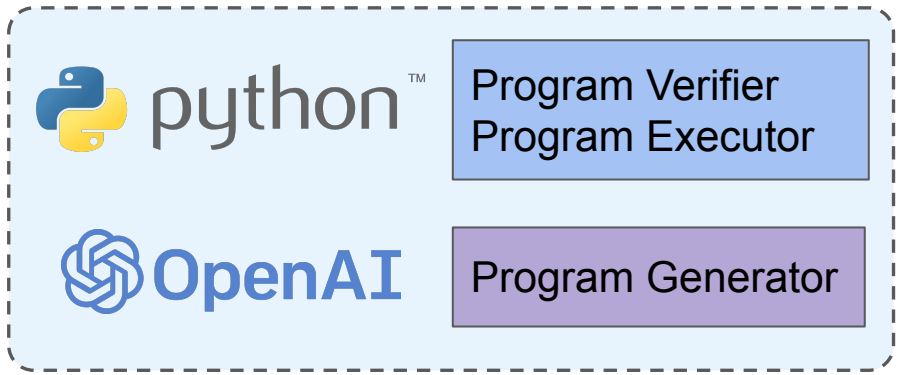
Options: (A) neither; the samples have the same temperature (B) sample A (C) sample B

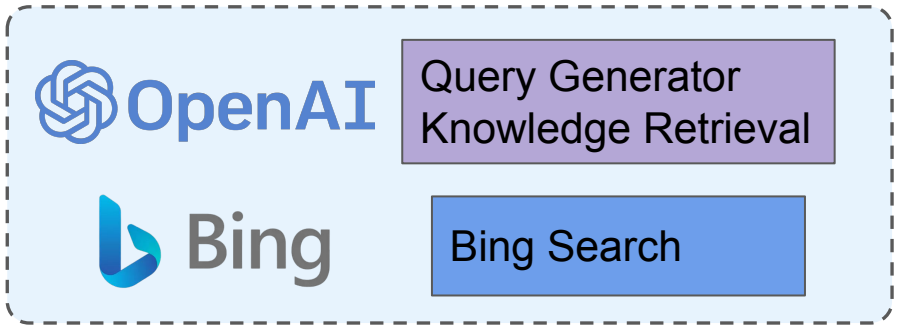
Metadata: ‘pid’: 19, ‘has_image’: True, ‘grade’: 8, ‘subject’: ‘natural science’, ‘topic’: ‘physics’, ‘category’: ‘Particle motion and energy’, ‘skill’: ‘Identify how particle motion affects temperature and pressure’

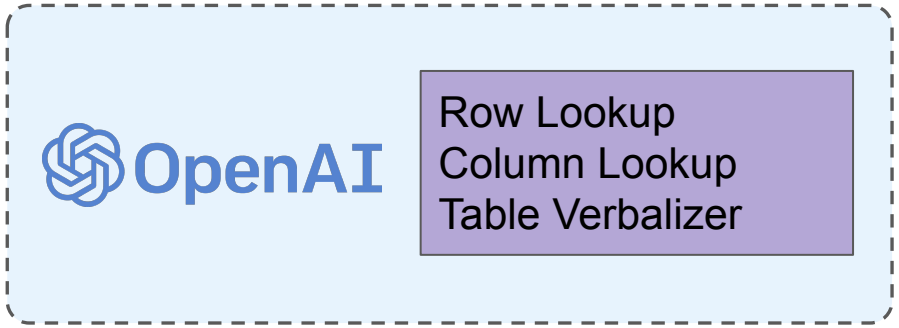
Modules: ["Text_Detector", "Knowledge_Retrieval", "Solution_Generator", "Answer_Generator"]

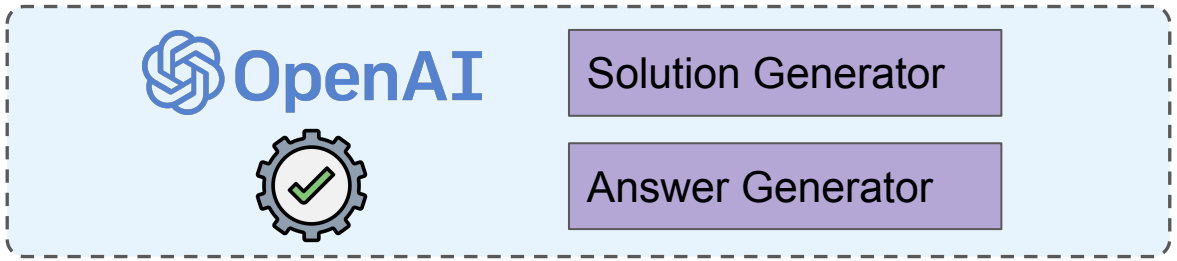
Chameleon: Module Inventory



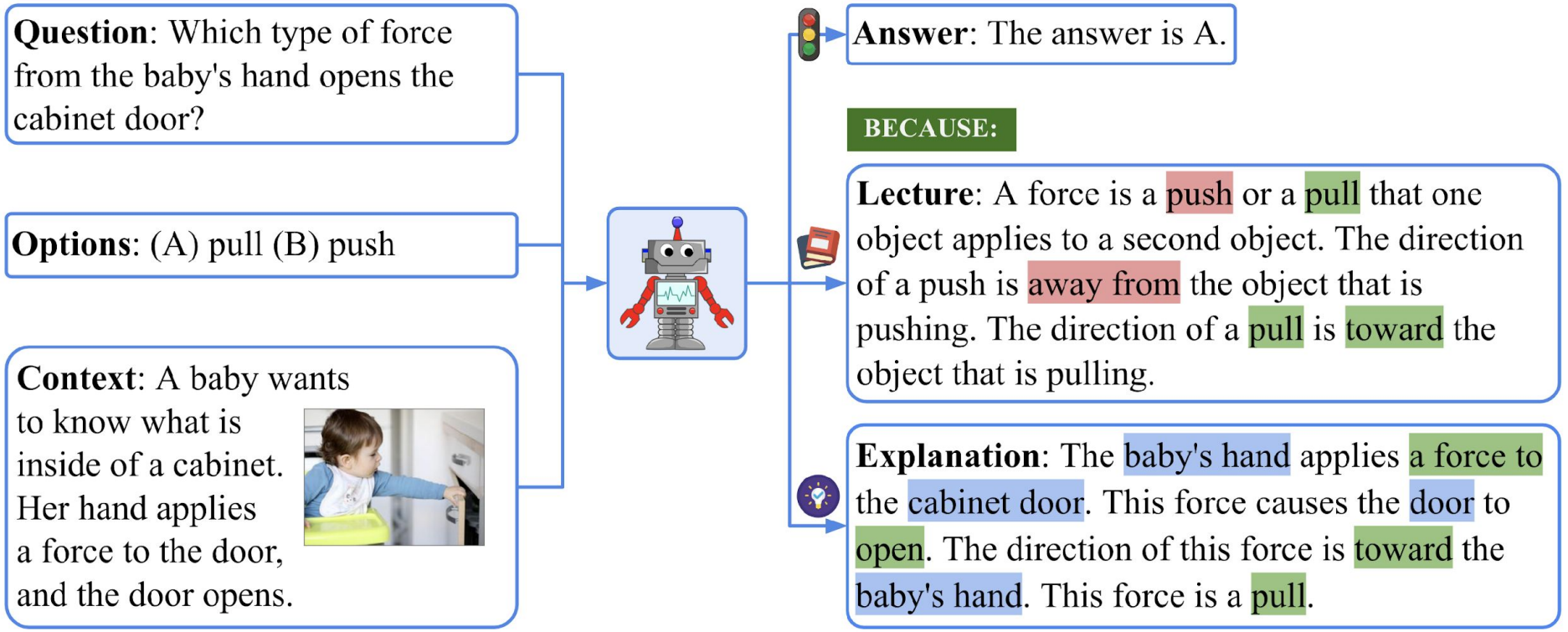








ScienceQA: Science Question Answering



Learn to Explain: Multimodal Reasoning via Thought Chains for Science Question Answering, NeurIPS 2022

https://scienceqa.github.io/explore.html

Domain Diversity in ScienceQA

Nature Science | Social Science | Language Science

Biology
Genes to traits
Classification
Adaptations
Traits and heredity
Ecosystems
Classification
Scientific names
Heredity
Ecological interactions
Cells
Plants
Animals
Plant reproduction

Earth Science
Weather and climate
Rocks and minerals
Astronomy
Fossils
Earth events
Plate tectonics

Physics
Materials
Magnets
Velocity and forces
Force and motion
Particle motion and energy
Heat and thermal energy
States of matter
Kinetic and potential energy
Mixture

Chemistry
Solutions
Physical and chemical change
Atoms and molecules
Chemical reactions

Engineering
Designing experiments
Engineering practices

Units and Measurement
Weather and climate

Geography
State capitals
Geography
Maps
Oceania: geography
Physical Geography
The Americas: geography
Oceans and continents
Cities
States

History
Colonial America
English colonies in North America
The American Revolution

World History
Greece
Ancient Mesopotamia
World religions
American history
Medieval Asia

Civics
Social skills
Government
The Constitution

Economics
Basic economic principles
Supply and demand
Banking and finance

Global Studies
Society and environment

Writing Strategies
Supporting arguments
Sentences, fragments, and run-ons
Word usage and nuance
Creative techniques
Audience, purpose, and tone
Pronouns and antecedents
Persuasive strategies
Editing and revising
Visual elements
Opinion writing

Vocabulary
Categories
Shades of meaning
Comprehension strategies
Context clues

Grammar
Sentences and fragments
Phrases and clauses

Figurative Language
Literary devices

Verbs
Verb tense

Capitalization
Formatting

Punctuation
Fragments

Phonology
Rhyming

Reference
Research skills

3 subjects

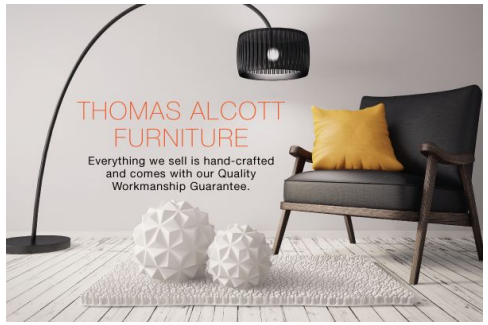
26 topics

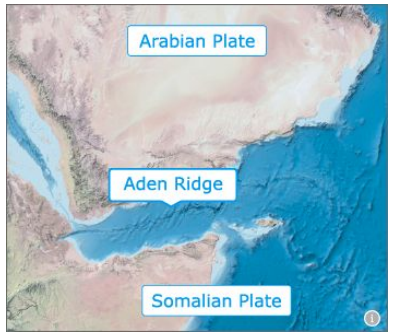
127 categories

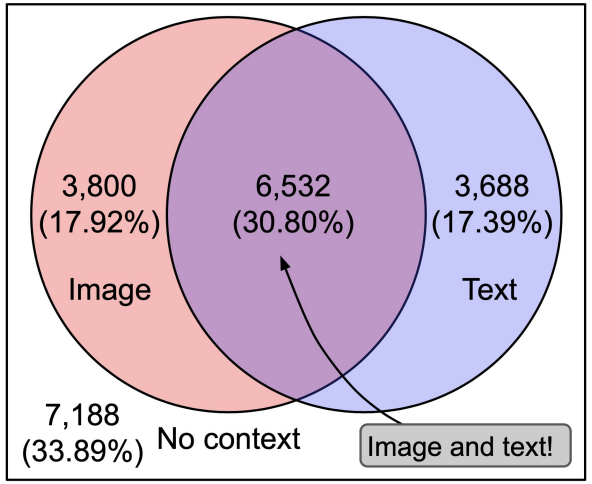
379 skills

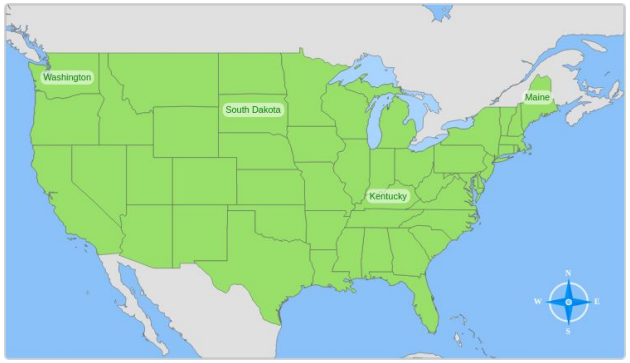
https://scienceqa.github.io/explore.html

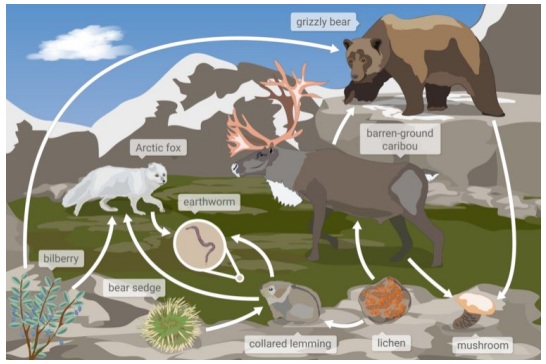
Context Diversity in ScienceQA

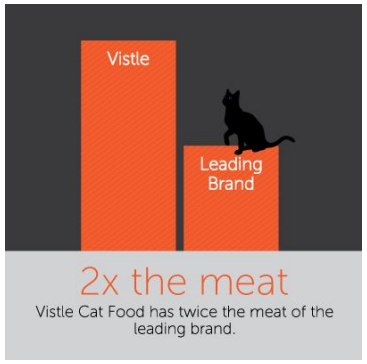


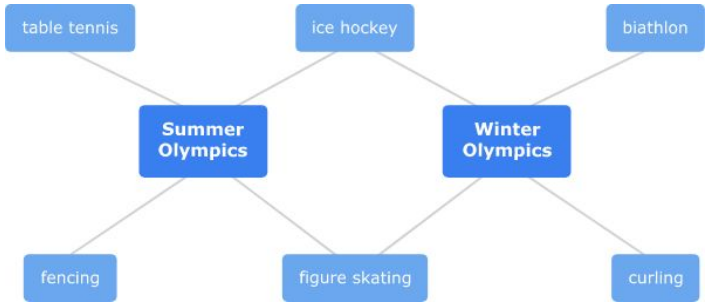




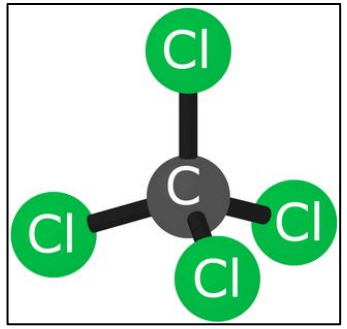


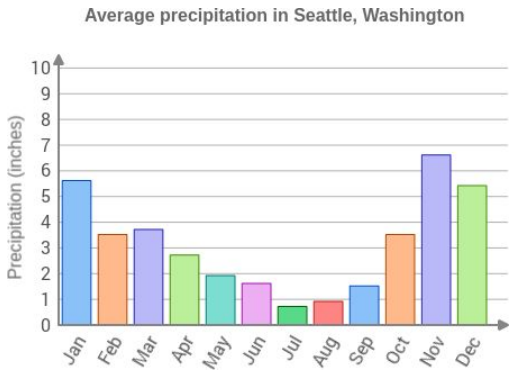




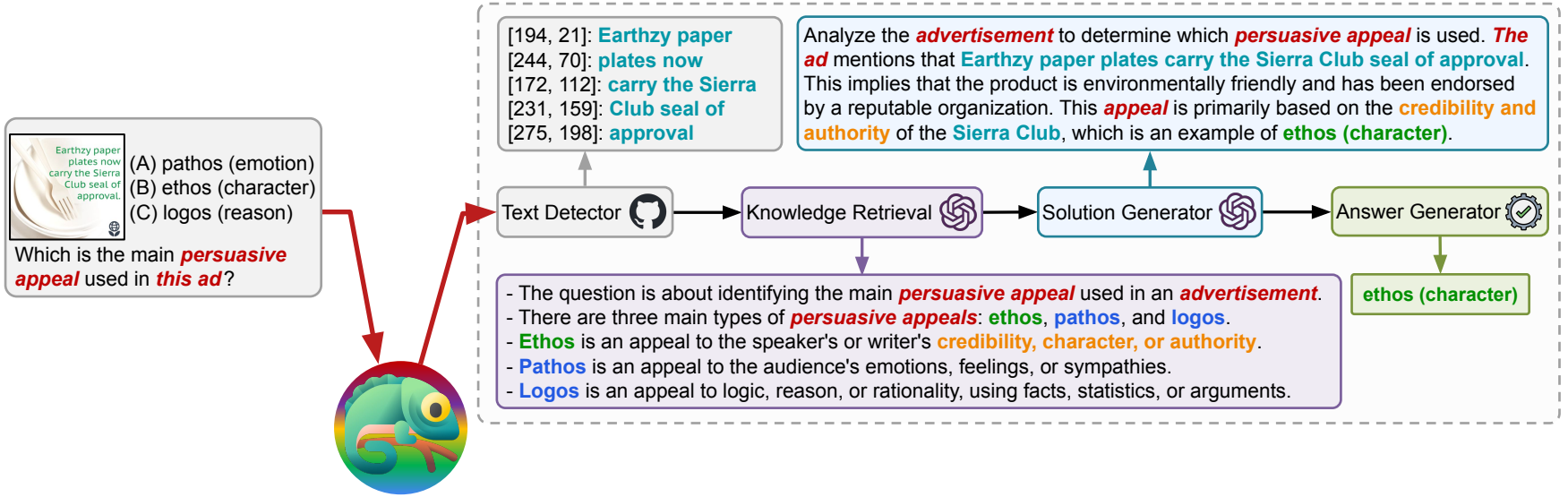


Planet	Volume (billions of km^3)	Primary composition
Mercury	60	rock
Venus	930	rock
Earth	1,090	rock
Mars	160	rock
Jupiter	1,431,280	gas
Saturn	827,130	gas
Uranus	68,330	ice
Neptune	62,530	ice

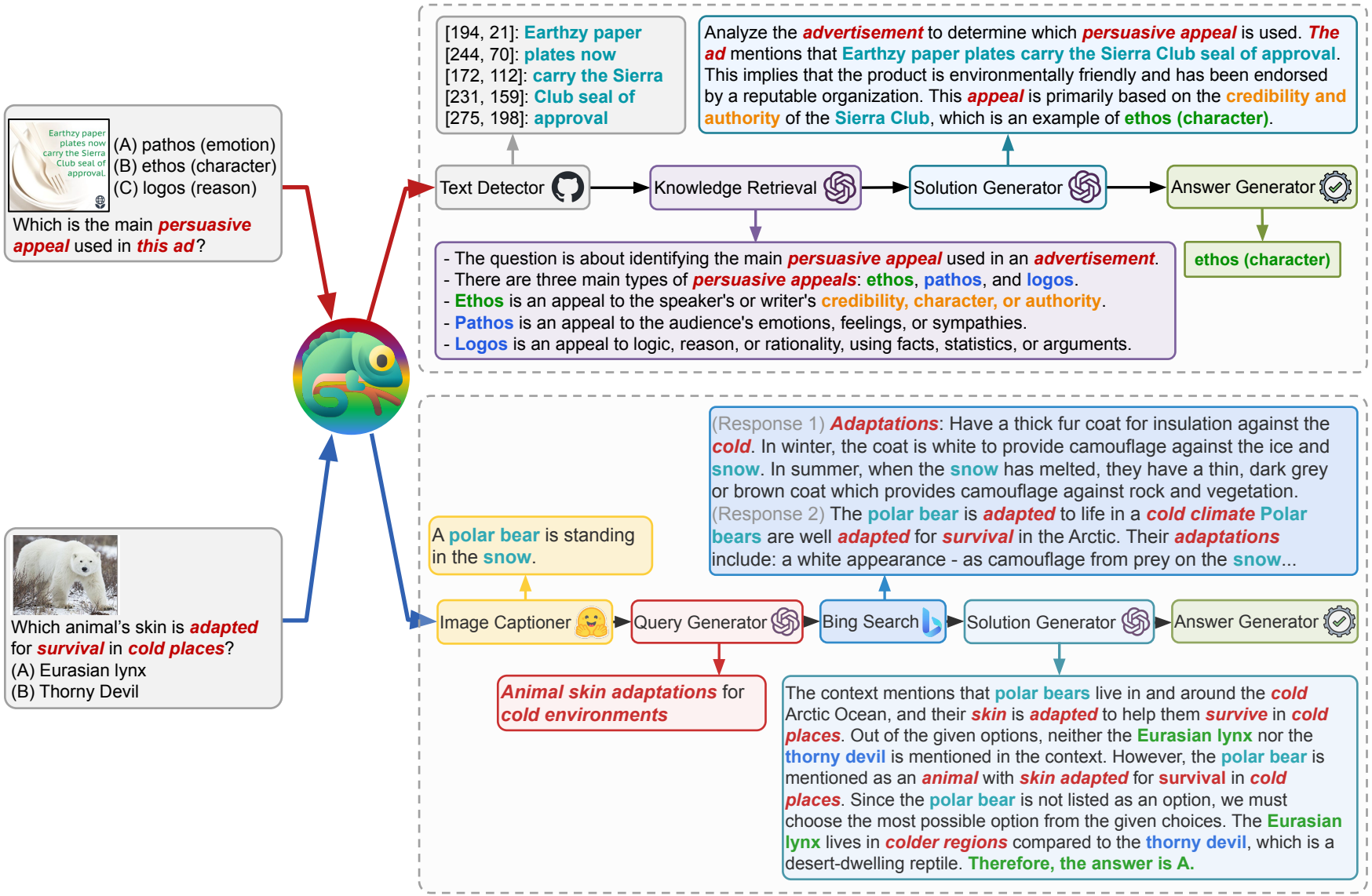




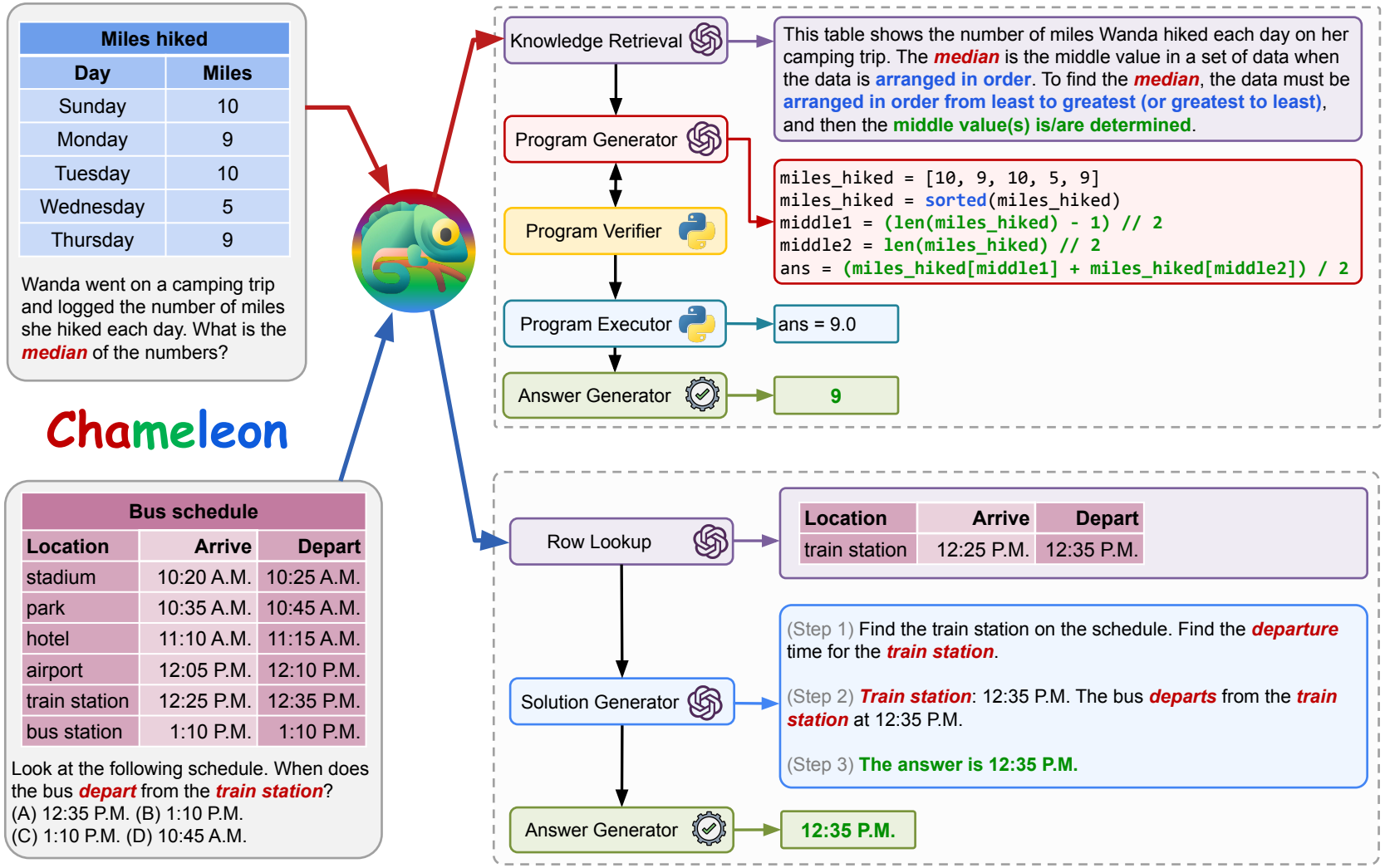
Chameleon for ScienceQA



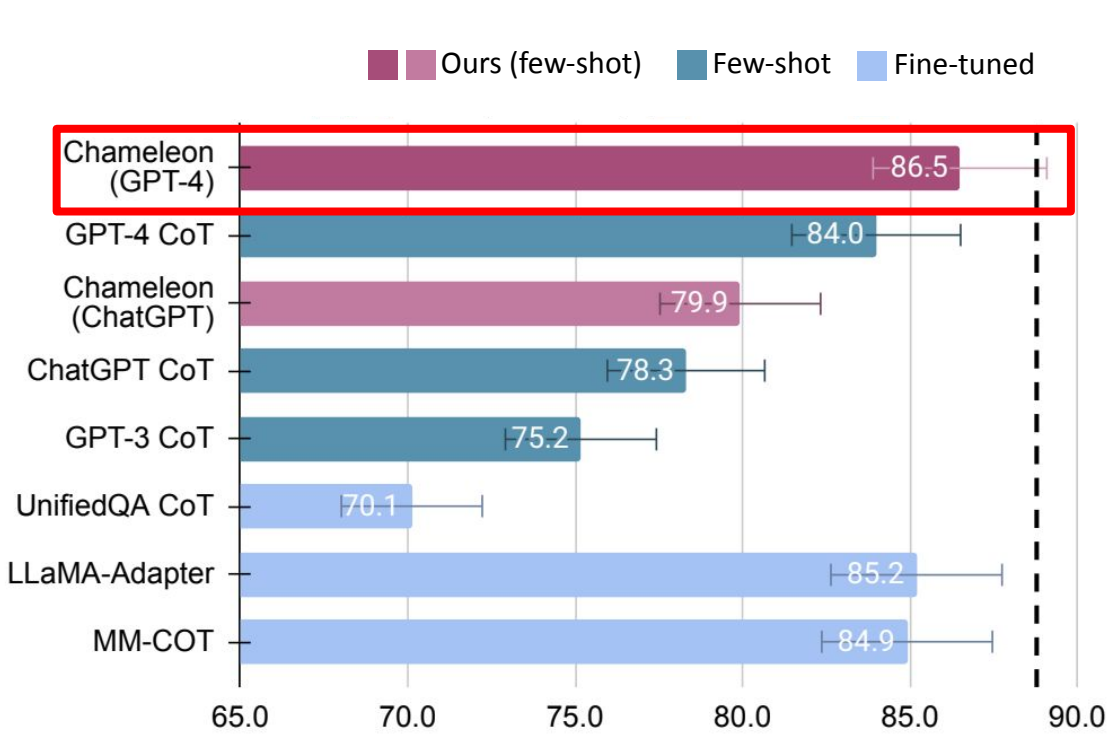
Chameleon for ScienceQA



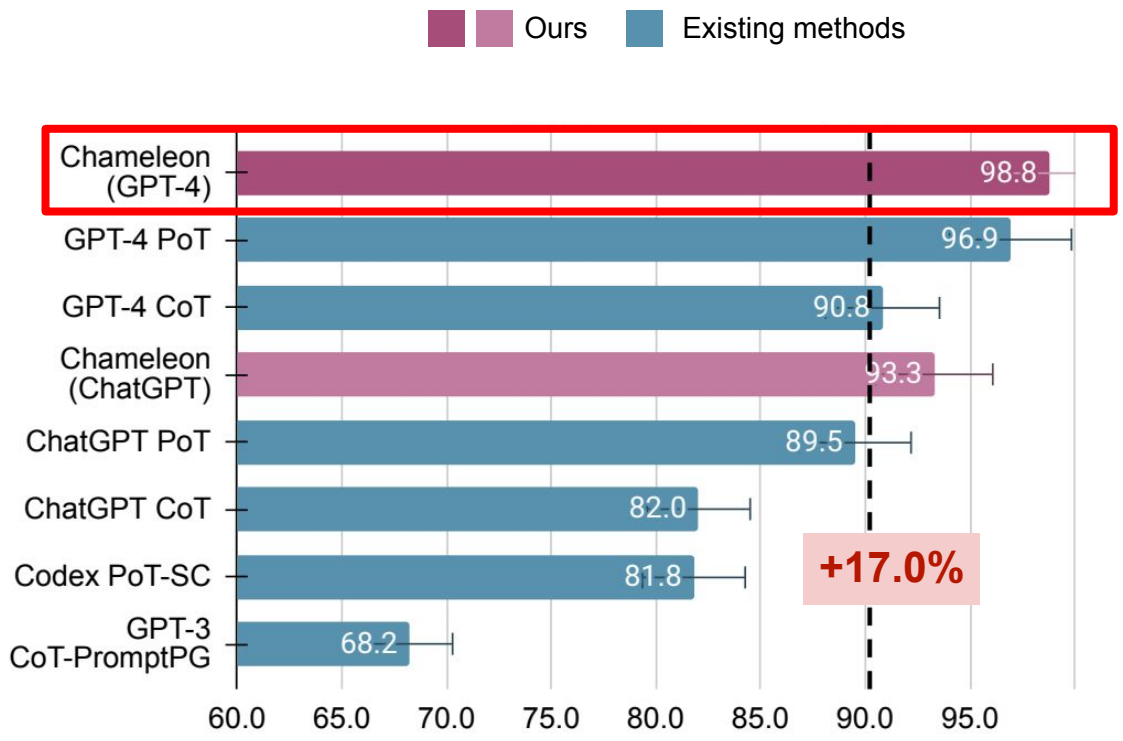
Chameleon for TabMWP (Tabular Math Word Problems)



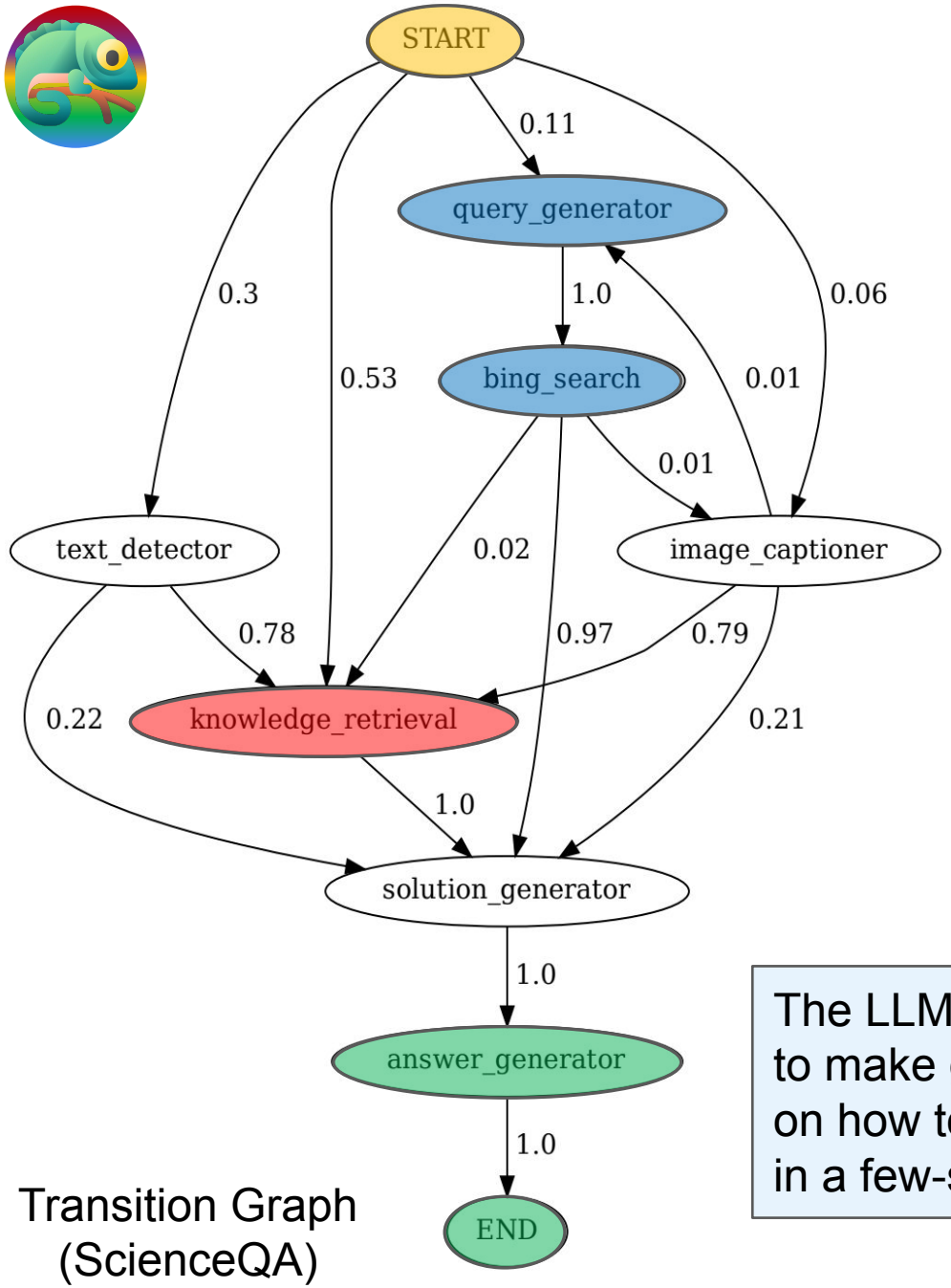
Results on ScienceQA and TabMWP



ScienceQA

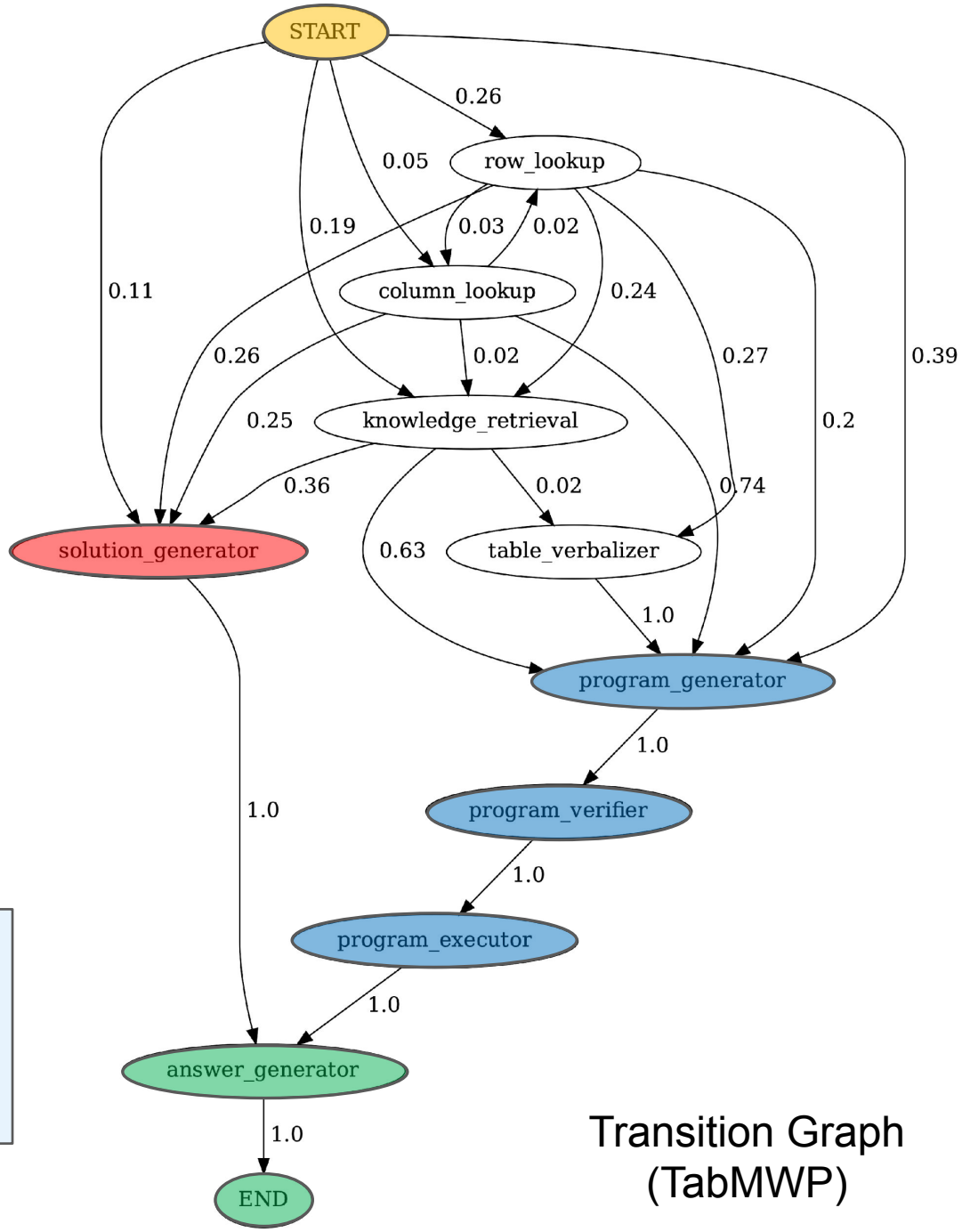


TabMWP



Transition Graph (ScienceQA)

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



Transition Graph (TabMWP)

NEURAL INFORMATION PROCESSING SYSTEMS

Microsoft

UCLA

Chameleon: Plug-and-Play Compositional Reasoning with Large Language Models

Project

https://chameleon-llm.github.io/

Code

https://github.com/lupantech/chameleon-llm

Thanks!