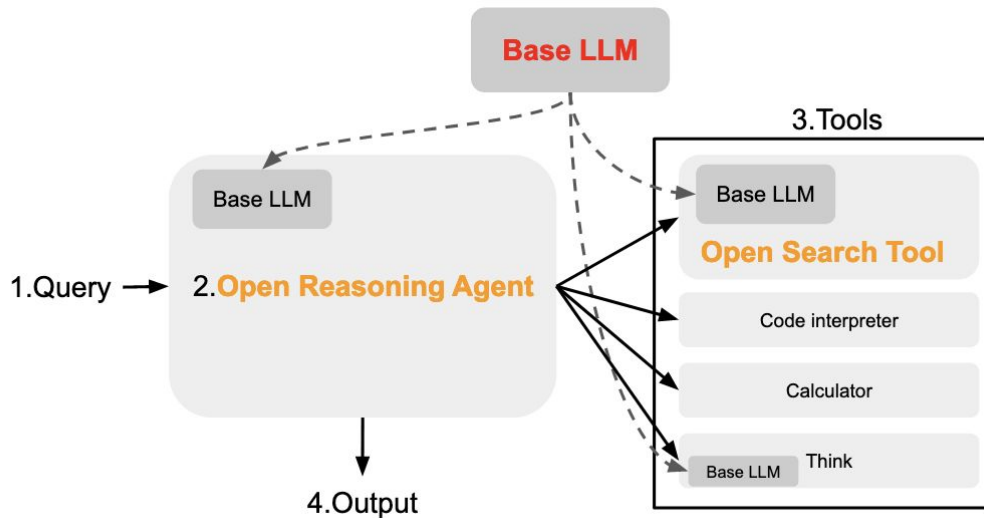


AI Agent Self-Consistency

Severin Bratus, David Dinucu-Jianu, Ilias Mc Auliffe

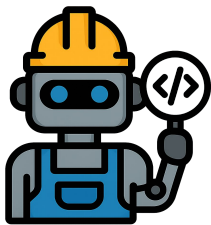
Motivation

- Main idea: open-source AI enabled search (e.g. alternative to Google/Bing/DuckDuckGo and Perplexity/ChatGPT Search)
- The LLM acts as the agent and combines reasoning with external tools to provide the best possible response to a user query

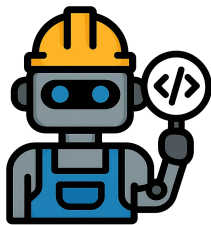


Query

As of August 1, 2024, which country were holders of the FIFA World Cup the last time the UEFA Champions League was won by a club from London?

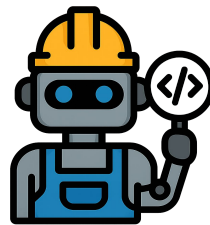


France

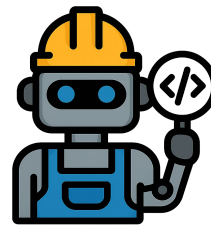


France

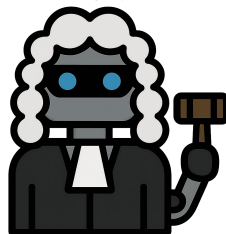
...



Spain



No answer



The country
was France

Dataset

In physics, when speaking of classical mechanics, there is an infamous problem that involves taking the initial positions and velocities of three point masses that orbit each other and attempting to calculate their trajectories. There is no general closed-form solution for this infamous problem. French mathematicians in the 18th century focused on solving this problem in regards to astronomical motion, specifically how the Moon rotates on its apsides. Their work led to a solution using Newton's laws of physics and the Discrete Fourier Transformation (DFT), which ultimately changed how sailors were able to determine longitude at sea. The inventor of the device that would revolutionize naval navigation using these new principles and proofs spent how many years testing and perfecting his work?

Pierre-Simon Laplace



History of longitude

Article [Talk](#)

From Wikipedia, the free encyclopedia

The **history of longitude** describes the centuries-long effort by astronomers, cartographers and navigators to discover a means of determining the [longitude](#) (the east-west position) of any given place on Earth. The measurement of longitude is important to both [cartography](#) and [navigation](#). In particular, for safe ocean navigation, knowledge of both [latitude](#) and longitude is required, however latitude can be determined with good accuracy with local astronomical observations.

A **marine chronometer** is a precision [timepiece](#) that is carried on a ship and employed in the determination of the ship's position by [celestial navigation](#). It is used to determine [longitude](#) by comparing [Greenwich Mean Time](#) (GMT), and the time at the current location found from observations of celestial bodies. When first developed in the 18th century, it was a major technical achievement, as accurate knowledge of the time over a long sea voyage was vital for effective [navigation](#), lacking electronic or communications aids. The first true chronometer was the life work of one man, [John Harrison](#), spanning **31 years** of persistent experimentation and testing that revolutionized naval (and later aerial) navigation.





Marine chronometer



Dataset

Split (1)
test · 824 rows

Search this dataset

Unnamed: 0 int64	Prompt string · lengths	Answer string · lengths	wikipedia_link_1 string · lengths	wikipedia_link_2 string · lengths
 0→82 10.1%	 195→271 16%	 1→137 95.4%	 59→79 21.1%	 33→47 46.2%
0	If my future wife has the same first name as the 15th first lady of the United States' mother and her surname is the same as the second assassinated president's mother's maiden name, what is my future wife's name?	Jane Ballou	https://en.wikipedia.org/wiki/President_of_the_United_States	https://en.wikipedia.org/wiki/President_of_the_United_States
1	Imagine there is a building called...	37th	https://en.wikipedia.org/wiki/Charlotte_Bront%C3%A9	https://en.wikipedia.org/wiki/Charlotte_Bront%C3%A9
2	How many years earlier would...	87	https://en.wikipedia.org/wiki/Punxsutawney_Phil	https://en.wikipedia.org/wiki/Punxsutawney_Phil
3	As of August 1, 2024, which...	France	https://en.wikipedia.org/wiki/FIFA_World_Cup	https://en.wikipedia.org/wiki/FIFA_World_Cup

Single Agent Chain-of-Thought + Tools

New run

If you add up the birth years of Emma Watson, Daniel Radcliffe and Rupert Grint, what is the sum?

LiteLLMModel - fireworks_ai/accounts/fireworks/models/qwen2p5-72b-instruct

Step 1

- Executing parsed code:

```
# Perform web searches to find the birth years
emma_birth_year = web_search(query="Emma Watson birth year")
print("Emma Watson birth year:", emma_birth_year)
daniel_birth_year = web_search(query="Daniel Radcliffe birth year")
print("Daniel Radcliffe birth year:", daniel_birth_year)
rupert_birth_year = web_search(query="Rupert Grint birth year")
print("Rupert Grint birth year:", rupert_birth_year)
```

Step 2

- Executing parsed code:

```
# Extracting the birth years from the search results
emma_birth_year = 1990
daniel_birth_year = 1989
rupert_birth_year = 1988

# Calculating the sum of the birth years
sum_birth_years = emma_birth_year + daniel_birth_year + rupert_birth_year
print(f"The sum of the birth years is {sum_birth_years}.")
final_answer(sum_birth_years)
```

Execution logs:

The sum of the birth years is 5967.

Out - Final answer: 5967

Results (easy)

Query: As of August 2024, who was president of the United States the last time The Beach Boys topped the chart on the Billboard Hot 100?

Ensemble Answers: ['Ronald Reagan', 'Ronald Reagan', 'Ronald Reagan', 'Ronald Reagan']

Final Result: Ronald Reagan

Results (hard)

Query: I can't recall who I'm trying to think of. This person was the partner of someone in the 1984 Olympic keelboat competition. Their partner's sister was the only American who placed in the 2003 Pan American Games for sailing in the men's or women's competition (not including the open events).

Answers: ['Richard Coxon', 'Stephen Erickson', 'Richard Coxon', 'Richard Coxon', 'Diane Crocker', 'Colin Beashel', 'Stevie Erickson', 'Unknown, but possibly the sister of William E. Buchan', 'Richard Coxon']

Final Result: Richard Coxon

Results (hard)

Query: I can't recall who I'm trying to think of. This person was the partner of someone in the 1984 Olympic keelboat competition. Their partner's sister was the only American who placed in the 2003 Pan American Games for sailing in the men's or women's competition (not including the open events).

Answers: ['Richard Coxon', 'Stephen Erickson', 'Richard Coxon', 'Richard Coxon', 'Diane Crocker', 'Colin Beashel', 'Stevie Erickson', 'Unknown, but possibly the sister of William E. Buchan', 'Richard Coxon']

Final Result: Richard Coxon

Thank you