Analytics Vidhya

Chain-of-verification

Instructor

Bhaskarjit Sarmah

Vice President, Blackrock

Definition

- Chain-of-Verification enables large language models to self-check and refine their responses through a series of validation questions.

- It includes a 'factored' approach to improve response accuracy by preventing the propagation of initial errors.

How Chain-of-Verification Works

Initial Responses

The model first drafts a provisional answer to the problem at hand.

Verification Questions

It then poses and addresses verification questions related to the initial answer.

Final Responses

Using insights from the verification step, the model produces a more accurate and validated final response.

Chain-of-Verification: Example

Query

Name some politicians who were born in NY, New York.

Baseline Response

Here are some politicians who were born in NY, New York:
1. Hillary Clinton - former secretary of state and former Democratic presidential nominee
2. Donald Trump - former president of the United States
3. Michael Bloomberg - former Mayor of New York city and former Democratic presidential candidate
---<list continues...>

Plan Verifications

- Where was Hillary Clinton born?
- Where was Donald Trump born?
- Where was Michael Bloomberg born?

Execute Verifications

Hillary Clinton was born in Chicago, Illinois, USA on October 26, 1947.
Donald Trump was born on June 14, 1946, in Queens, NY, USA
Michael Bloomberg was born on Feb 14, 1942, in Boston, Massachusetts, USA

Final Verified Response

Here are some politicians who were born in NY, New York:
1. Donald Trump - former president of the USA
2. Alexandra Ocasio-Cortez -Democratic member of the U.S. House of Representatives.

Cons: Chain-of-Verifications

- It reduce but doesn't entirely prevent incorrect model-generated content.
- Tackles factual errors more than subtle reasoning or opinion inaccuracies.
- Adds clarity but at a higher computational cost due to increased output length.
- Its effectiveness is constrained by the model's inherent capabilities, such as identifying and knowing what it knows.

Thank You