Analytics Vidhya

Chain of Natural Language Inferencing

Instructor

Bhaskarjit Sarmah

Vice President, Blackrock

Definition

- Chain of Natural Language Inferencing is a hierarchical framework designed to address and reduce hallucinations in text generated by large language models (LLMs).

  - Provides a structured approach to enhance the reliability of model-generated context without needing fine-tuning or specific prompts.

How Chain of Natural Language Inferencing Works

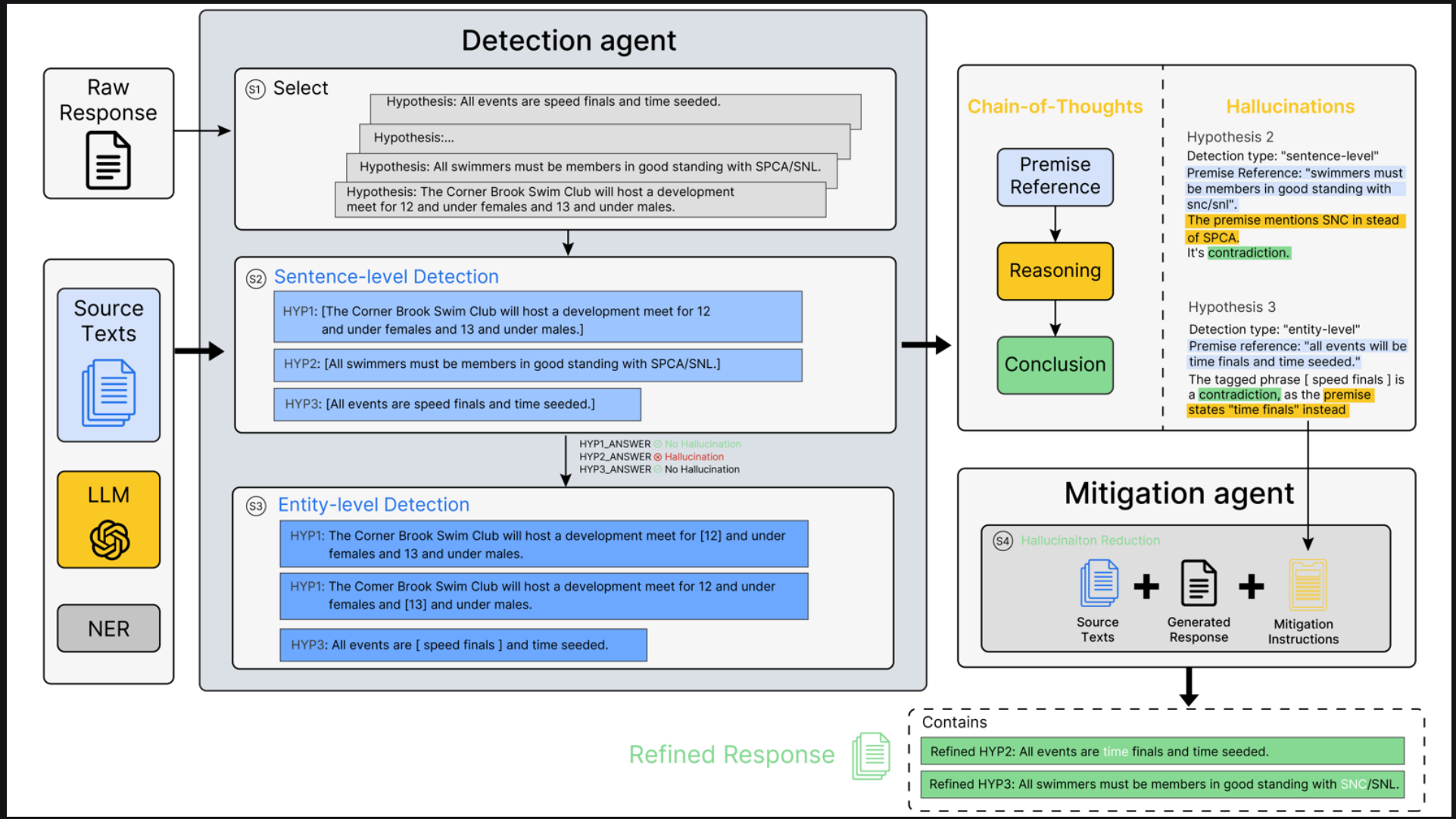
Detection Phase

Employs a detection agent to identify hallucinations in the text.

Mitigation Phase

Uses a mitigation agent to refine or remove detected inaccuracies, preserving text fluency and coherence.

Chain of Natural Language Inferencing: Example



Analytics Vidhya

Cons: Chain of Natural Language Inferencing

- May miss some hallucinations, mainly detecting those that lack grounding.
- Effectiveness is dependent on the LLM's accuracy, which is not foolproof.
- Post-processes rather than preventing hallucinations at the source.
- Less effective for brief or fact-sparse responses due to its segmentation reliance.

Thank You

Analytics Vidhya