

Perception of Knowledge Boundary for Large Language Models through Semi-open-ended Question Answering

**Zhihua Wen^{1,†}, Zhiliang Tian^{1,*}, Zexin Jian¹, Zhen Huang¹,
Pei Ke², Yifu Gao¹, Minlie Huang², Dongsheng Li^{1,*}**

¹ College of Computer, National University of Defense Technology, Hunan, China

² Tsinghua University, Beijing, China

{zhwen, tianzhiliang, jianzexin21, gaoyifu, huangzhen, dsli}@nudt.edu.cn
kepei1106@outlook.com, aihuang@tsinghua.edu.cn

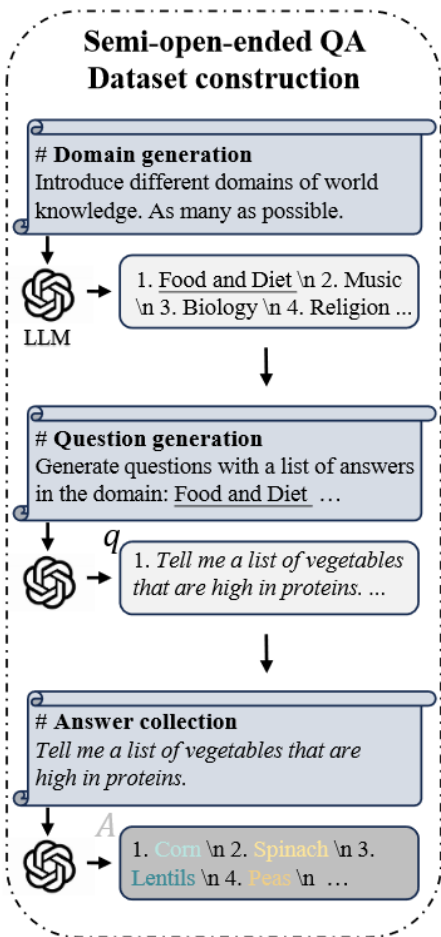
Motivation

- Large Language Models (LLMs) are susceptible to generating **non-factual responses**, i.e. hallucinations.
- The knowledge boundary of an LLM is vital as it describes **the limit of its factual understanding**.
- Existing studies neglect the importance of **semi-open-ended questions that can have multiple potential answers**.

Contribution

- Exploring Semi-Open-Ended Questions for LLM knowledge boundary perception
- Novel method to discover ambiguous answers
- Our dataset and method are effective in revealing GPT-4's unfamiliar knowledge

Dataset



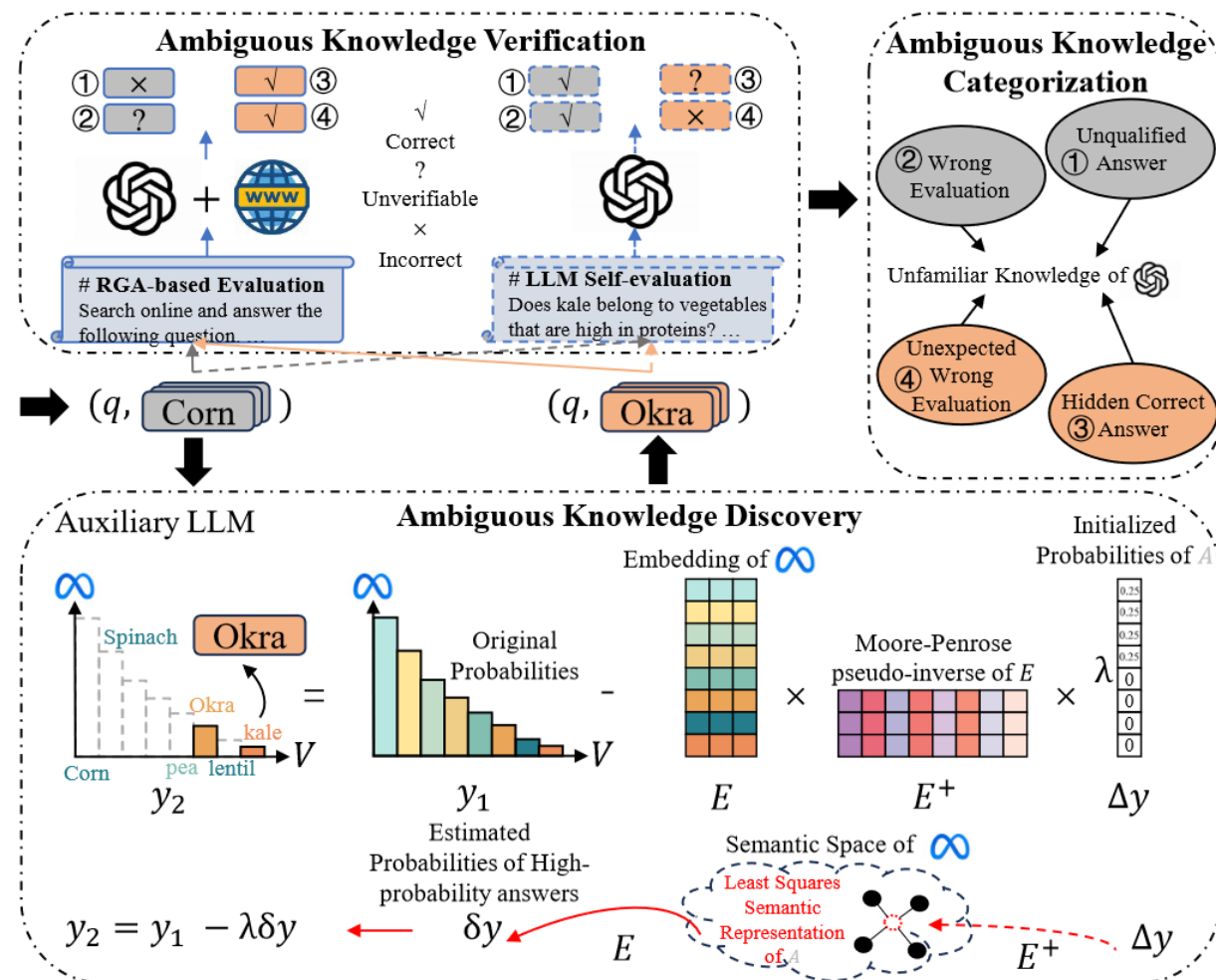
Dataset construction pipeline

Domain	Percentage	Question	Ambiguous Answer
Biology	4.93%	Tell me a list of trees that produce fruit with hard shells and are native to tropical regions.	Malabar Chestnut, Miracle Fruit Tree, Guapinol Tree, Peanut Tree, Sapodilla Tree, Desert Date, Castor Oil Plant, Chiclé Tree, Oriental Persimmon, Dika Tree, Langsat Tree, Pequi Tree, Karite Tree, Imbu Tree, Caraipa Tree
Music	4.20%	Tell me a list of films whose plot is centered on a historically significant event, but the narrative is from the perspective of an imagined protagonist.	The Hurt Locker, Zulu, Letters from Iwo Jima, Doctor Zhivago, Black Hawk Down, Les Miserables, Lawrence of Arabia, The Wind that Shakes the Barley, The Longest Day, Anthony Adverse, The Guns of Navarone, Lion of the Desert
Geology	4.09%	Tell me a list of current deserts across the world that were previously covered by an ancient sea.	Qaidam Basin, Dasht-e Kavir, Dasht-e Lut, The Registan Desert, The Franklin Basin, The Makran Desert, The Sonoran Desert, The Jornada Del Muerto, Great Karoo, Tanami Desert
Food and Diet	5.67%	Tell me a list of fruits that are sources of healthy fats, not including avocados and coconuts.	Sacha Inchi Nuts, Hemp Seeds, Chokeberries, Elderberries, Pine Nuts, Passion Fruit Seeds, Tibetan Goji Berries
Literature	1.68%	Tell me a list of female writers from the Victorian era whose work focuses on social reform.	Helen Taylor, Catherine Helen Spence, Octavia Hill, Rhoda Broughton, Elizabeth Missing Sewell, Emmeline Pankhurst, Alice Meynell, Elizabeth Robins, Mary Augusta Ward, Constance Garnett

Data samples

Method

- We reduce the probabilities of high-probability answers to induce more ambiguous answers.
- We compare the results of LLM self-evaluation and RAG-based evaluation.
- We categorize four types of ambiguous answers for the target LLM.



Findings

GPT-4

- performs poorly on the semi-open-ended questions and generates many unqualified answers.
- lacks understanding of the relevant knowledge.
- has limited ability to recognize its knowledge boundary.

The auxiliary model

- effectively supplements GPT-4 by identifying hidden correct answers.
- incur unexpected wrong evaluations during GPT-4 self-evaluation.
- discover situations where GPT-4 admits for its knowledge boundary.

Thanks!