GIVE: Structured Reasoning of LLM with Knowlegde Graph Inspired Veracity Extrapolation

Jiashu He [1]   **Mingyu Derek Ma** [2]   **Jinxuan Fan** [3]   **Dan Roth** [1]   **Wei Wang** [2]   **Alejandro Ribeiro** [1]

[1] University of Pennsylvania [2] University of California, Los Angeles [3] University of California, Berkeley

Reasoning requires knowledge

Scaling doesn't solve everything.

Question:
Traumatic aortic injury: does the anatomy of the aortic arch influence aortic trauma severity?

Chain-of-Thought ➡ **lack of internal knowledge**

Prompt : Let's think step by step.
Response: First, the anatomy of the aortic arch *may influence* aortic trauma severity, but it is *not a definitive* yes or no answer. Second, there are *various factors* to consider, such as the type of trauma. Therefore, the correct answer should be *maybe*. ✗
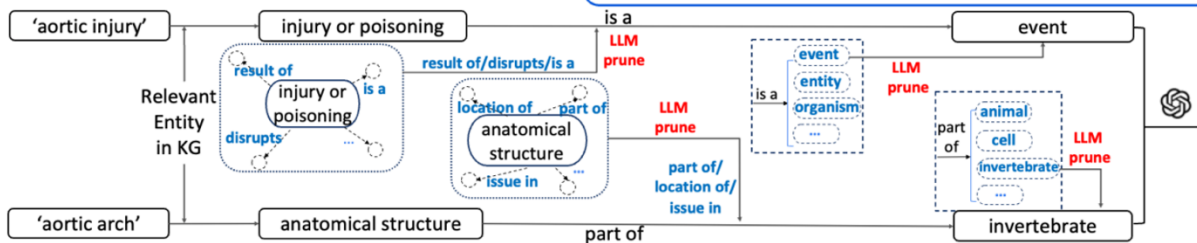
-If we only rely on **parametric knowledge** of LLM, we cannot generate logic reasoning on the corner domain questions, because the models are rarely trained on them.

-In fact, for any size LLM, there will **always** be unsolvable problems in a single pass (Dziri et al., 2024).

"Gold context" is not always feasible

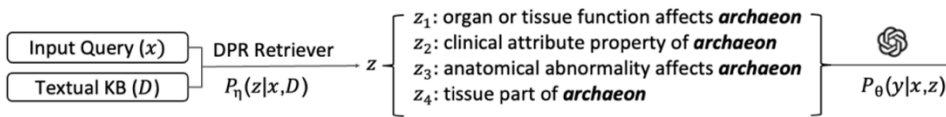**Assume accessibility to comprehensive knowledge base is not generally applicable.**

Entity groups to bridge query and knowledge

Question:
Traumatic aortic injury: does the anatomy of the aortic arch influence aortic trauma severity?
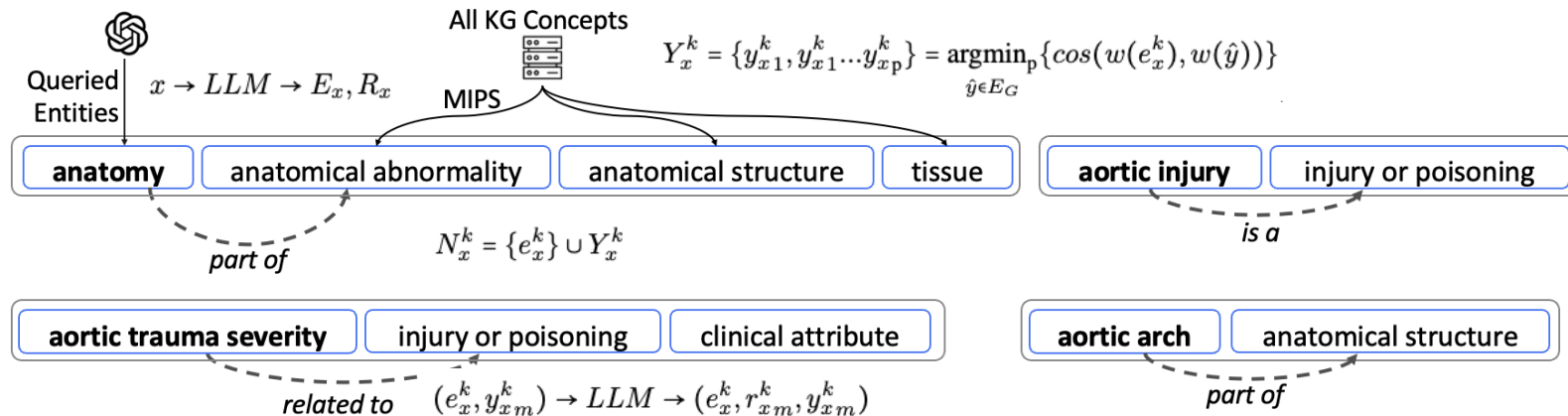
Expert guided extrapolation

Identify the possible relation sets between two groups:

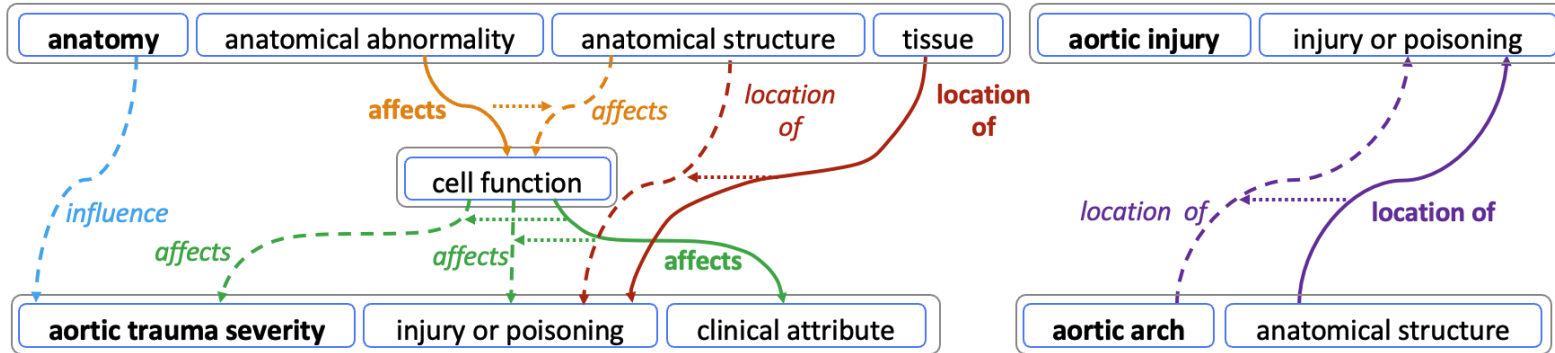$$R_G^{ij} = \{r, (u, r, v) \in \mathcal{E}_G, u \in N_i, v \in N_j\} \quad R^{ij} = R_x \cup R_G^{ij}$$

Question:
Traumatic aortic injury: does the anatomy of the aortic arch influence aortic trauma severity?

GIVE: Graph Inspired Veracity Extrapolation

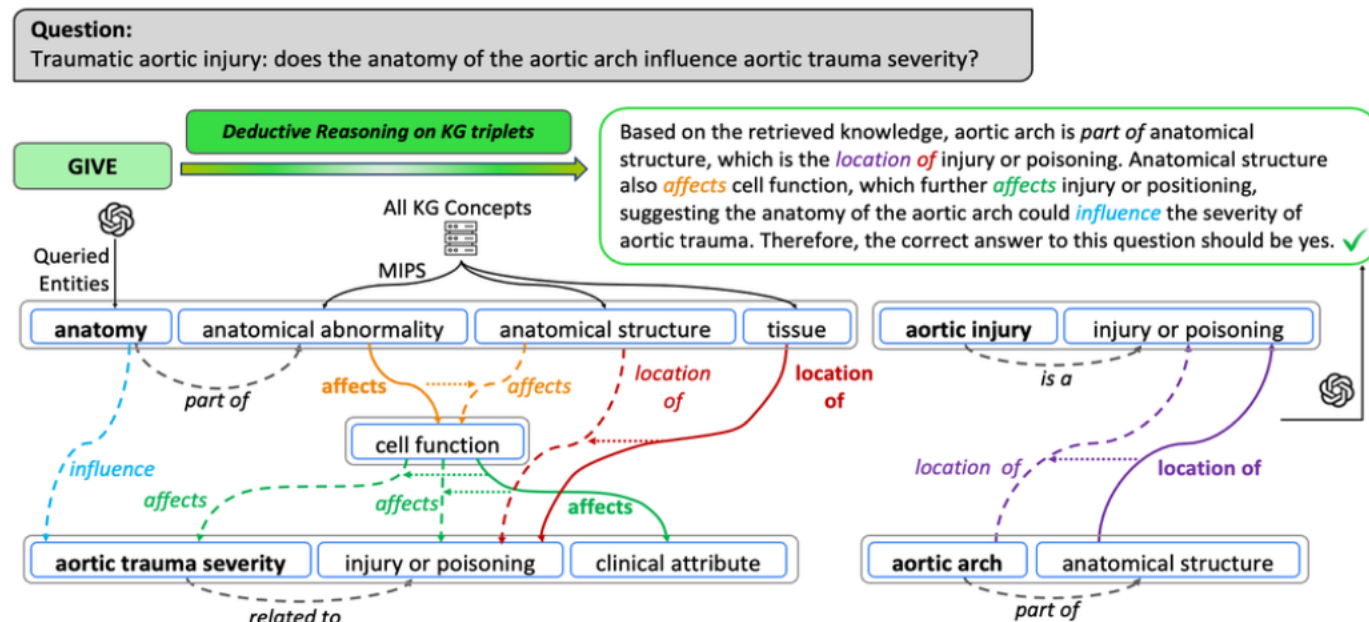**Figure 2:** Reasoning process of GIVE. Solid lines indicate expert information, while dashed lines depict results from the "veracity extrapolation" procedure. GIVE first constructs an entity group for each queried concept, then induce inner-group connections using its internal knowledge. The expert's cross-group connections serve as evidence, guiding the LLM to extrapolate the veracity of potential relationships among similar concepts.

Results on biomedical tasks

# Method/Dataset	GPT3.5-turbo			GPT4			GPT4o-mini			Meta-Llama-3.1-70B-Instruct		
	PubMedQA	BioASQ	ProcessBank	PubMedQA	BioASQ	ProcessBank	PubMedQA	BioASQ	ProcessBank	PubMedQA	BioASQ	ProcessBank
	Internal knowledge reasoning											
1 I/O prompt	46.2	43.5	67.3	42.2	88.2	64.8	23.4	88.7	79.4	48.0	91.0	85.4
2 CoT	48.6	63.5	70.9	37.8	80.4	59.3	23.8	79.3	81.4	50.4	91.3	84.3
	External knowledge (text) reasoning											
3 RAG	13.4	40.9	67.3	26.4	24.3	78.9	15.2	16.3	84.9	49.8	45.4	84.4
	External knowledge (KG) reasoning											
4 ToG	17.6	18.0	66.8	19.1	15.4	81.4	21.8	10.1	84.4	38.4	31.0	85.9
5 GraphRAG	23.4	10.3	71.3	26.5	11.2	80.9	22.6	10.1	84.9	/	/	/
	Structured reasoning with internal and external knowledge(Our method)											
5 GIVE$_a$	44.4	82.6	72.9	50.0	**90.0**	82.7	26.0	**89.5**	85.9	56.0	91.7	86.4
6 GIVE$_{a+c}$	49.8	86.1	**73.9**	**50.2**	80.6	**83.3**	**27.4**	81.9	**87.4**	**56.2**	91.7	**86.9**
7 GIVE$_{a+c+e}$	**53.6**	**88.2**	73.4	43.4	87.8	82.7	27.2	81.9	86.9	56.0	**92.6**	86.4
8 Best Gain(+%)	7.4/40.2/36.0	44.7/47.3/77.9	6.6/6.6/7.1	12.4/23.8/31.1	9.6/65.7/78.8	24.0/4.4/2.4	4.0/12.2/5.6	10.2/73.2/79.4	8.0/2.5/3.0	8.2/6.4/17.8	1.6/47.2/61.6	2.6/2.5/1.0

Older model VS later model: scaling training does **not** solve everything

Vanilla VS RAG: hallucination occurs when we directly equip the model with limited information

Takeaways

1. We cannot solely rely on parametric or non-parametric knowledge to build specialized LLM in corner domains or limited personalized data.

2. This issue can be resolved by providing the limited retrieved knowledge, along with some "bridging" information for the model to connect them with the query.

3. In the era of LRM, retrieval methods should not focus on gold answer (context) matching, but identifying generalizable patterns or clues for LLM to observe and Reflect, thus conduct deductive/inductive reasoning.