How to Auto-optimize Prompts for Domain Tasks? Adaptive Prompting and Reasoning through Evolutionary Domain Knowledge Adaptation

Yang Zhao, Pu Wang, and Hao Frank Yang*

Johns Hopkins University | Whiting School of Engineering | Baltimore, MD

* Corresponding author, email: haofrankyang@jhu.edu

Motivation

[Task] How to write prompt for a domain specific task?

❑ LLMs does not perform well in some domain

- Restricted Data Access
- Real-Time and Rapidly Evolving

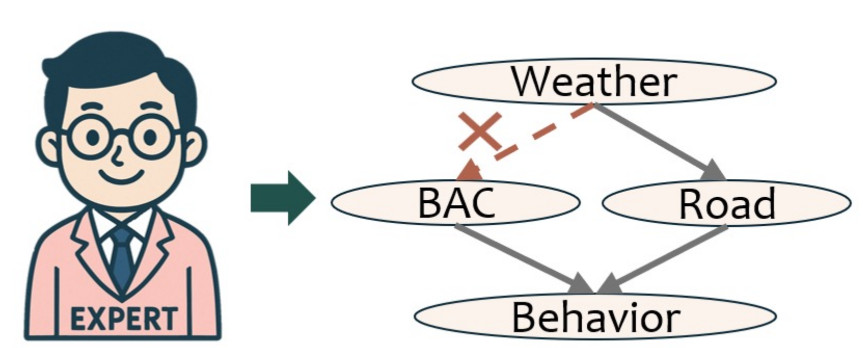
❑ Limitations of existing methods

	Fine-tuning	Retrieval-Augmented Generation (RAG)	Prompt Engineering
Performance	☆ ☆ ☆	☆ ☆	☆
Cost	$ $ $	$ $	$
Issues	Heavy; Expensive; Large-scale Ground Truth	External Databases	Limited Improvements; No New Knowledge

➢ Can we utilize expert knowledge for better prompt?

Challenges

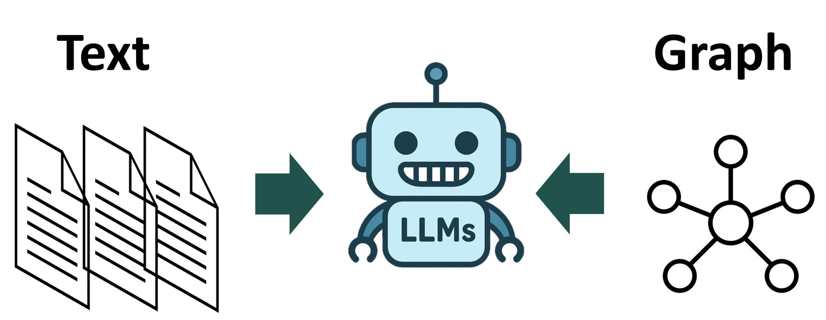
Flawed Knowledge



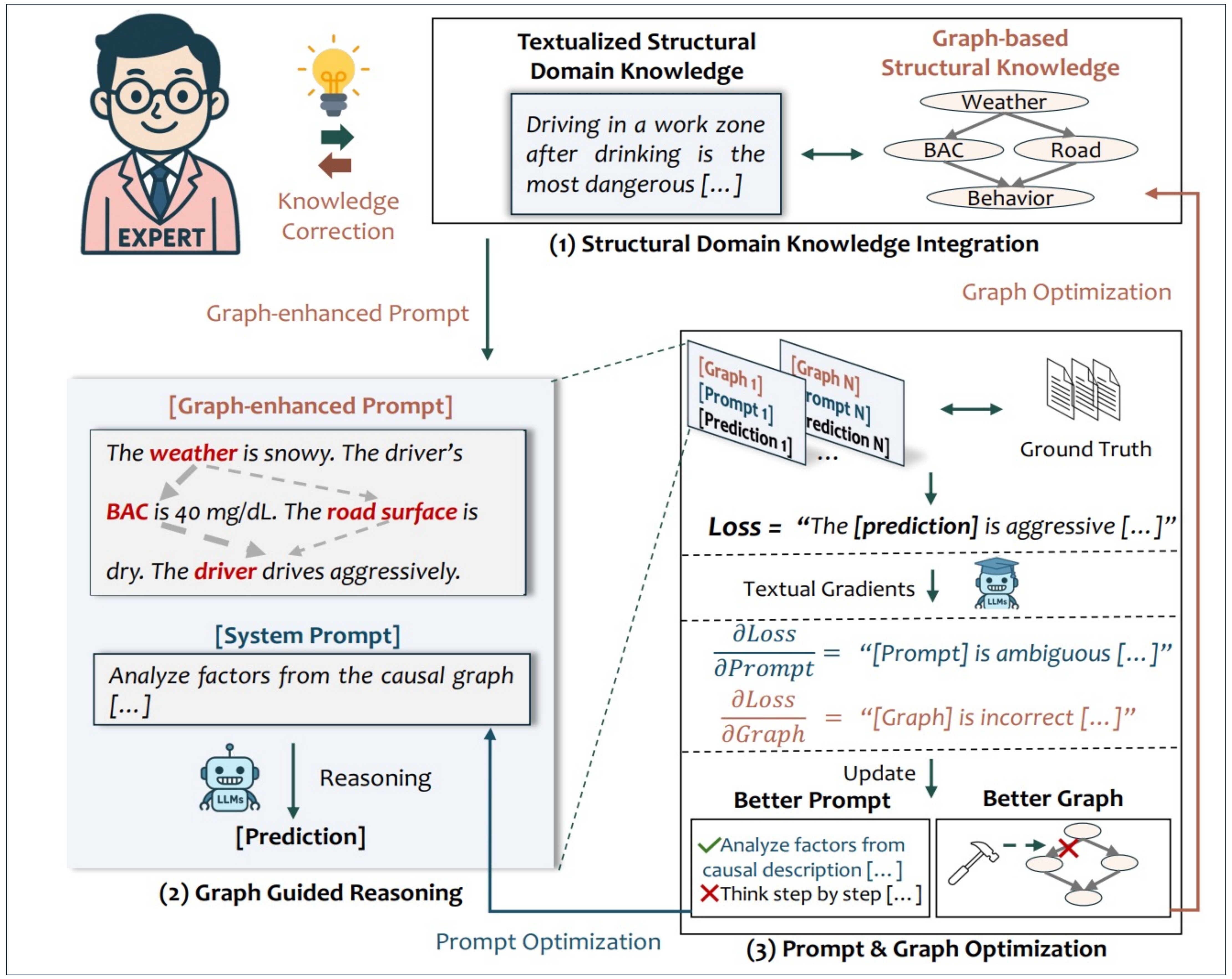
Dynamic Knowledge

❑Difference in transmissibility, risk, …

Text-Graph Alignment



Methodology



Lossless Domain-Specific Reasoning with Expert Knowledge Guidance

[Time] The crash occurred on April 29, 2022 at hour 16 (…)

Input

CAUSAL STATEMENT 1: [Person Status] affects [Severity]. The driver's BAC increases the probability of fatal crashes (…)

Semantic Causal Graph (SCG)

Instance-Specific Reasoning Guidance

$$p(y \mid x, \mathcal{G}) = p(y \mid x, z*(x, \mathcal{G}))$$

This transformation is lossless when:

(A1) SCG is independent from the reasoning process;

(A2) The reasoning guidance is uniquely determined.

Experiments

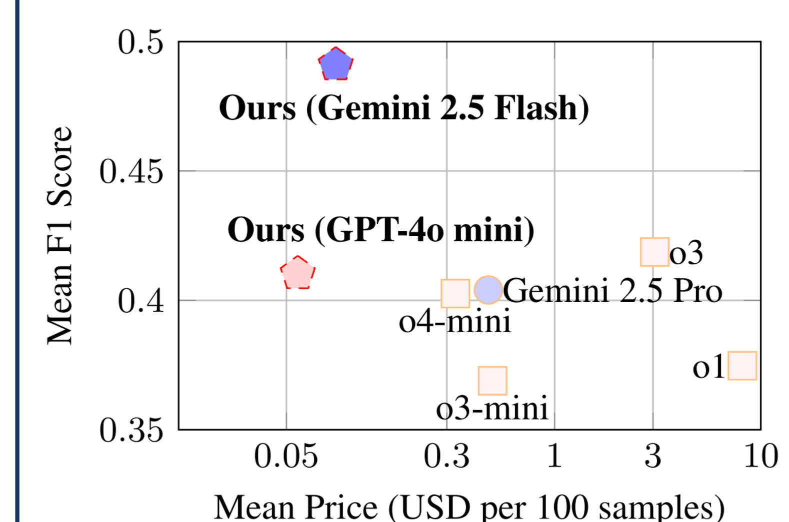
Tasks

❑ Crash Prediction ❑ Pandemic Trend ❑ Human Behavior

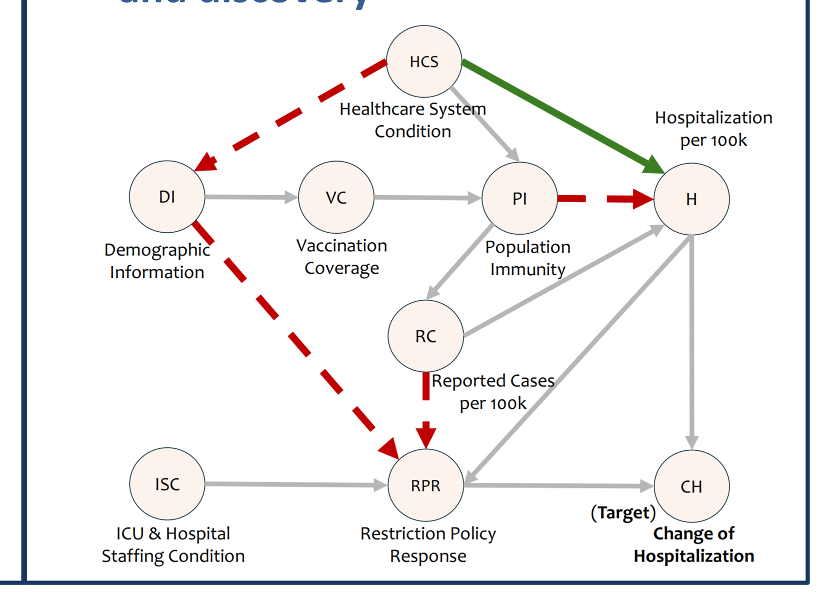
Performance

Forward Engine	Method	Pandemic		TrafficSafe		Mode Choice		Mean F1	
		Acc	F1	Acc	F1	Acc	F1	Value	Imp.
GPT-4o-mini (2024-07-18)	Organized Prompt	0.360	0.347	0.300	0.232	0.459	0.406	0.328	—
	Zero-Shot-CoT	0.370	0.361	0.280	0.221	0.494	0.435	0.339	3.2%
	ProTeGi	0.370	0.361	0.370	0.304	**0.529**	0.481	0.382	16.3%
	Auto-CoT	0.380	0.352	0.320	0.220	0.447	0.462	0.345	5.0%
	PHP	0.330	0.327	0.320	0.268	0.376	0.370	0.322	-2.0%
	TextGrad	0.380	0.359	0.300	0.243	0.506	0.432	0.345	5.0%
	EGO-Prompt	**0.410**	**0.399**	**0.380**	**0.333**	0.506	**0.498**	**0.410**	**24.9%**
Gemini-2.5-flash (preview-04-17)	Organized Prompt	0.470	0.470	0.340	0.319	0.459	0.392	0.394	—
	Zero-Shot-CoT	0.490	0.490	0.400	0.390	0.482	0.398	0.426	8.1%
	ProTeGi	0.470	0.482	0.420	0.400	0.494	0.425	0.435	10.6%
	Auto-CoT	0.380	0.352	0.340	0.230	0.447	0.462	0.348	6.0%
	PHP	0.520	0.515	0.360	0.334	0.494	0.436	0.428	8.6%
	TextGrad	0.470	0.483	0.380	0.374	0.494	0.428	0.428	8.6%
	EGO-Prompt	**0.540**	**0.546**	**0.430**	**0.428**	**0.518**	**0.499**	**0.491**	**24.6%**
GPT-5-mini (minimal reasoning effort)	Organized Prompt	0.420	0.387	0.330	0.265	0.435	0.435	0.362	—
	Zero-Shot-CoT	0.430	0.415	0.330	0.281	0.424	0.428	0.375	3.4%
	ProTeGi	0.420	0.413	0.320	0.267	0.471	0.454	0.378	4.4%
	Auto-CoT	0.410	0.408	0.310	0.280	0.506	0.470	0.386	6.6%
	PHP	0.430	0.397	0.320	0.240	0.435	0.431	0.356	-1.8%
	TextGrad	0.430	0.415	0.250	0.230	0.494	0.485	0.377	4.0%
	EGO-Prompt	**0.460**	**0.448**	**0.340**	**0.305**	**0.529**	**0.511**	**0.421**	**16.3%**

❑Comparable to reasoning models while costing less than 20%



❑ Automatic knowledge correction and discovery



Future Work

❑ Variability of LLMs in both the forward and backward processes

❑ Learn from input distributions rather than several data instances

Project Page: https://miemieyanga.github.io/EGOPrompt/