



MindMap: Knowledge Graph Prompting Sparks Graph of Thoughts in Large Language Models

Yilin Wen¹

yilinwen510@gmail.com

Zifeng Wang¹

University of Illinois Urbana-Champaign, Champaign, IL

zifengw2@illinois.edu

Jimeng Sun²

University of Illinois Urbana-Champaign, Champaign, IL

jimeng@illinois.edu

ACL 2024

Introduction

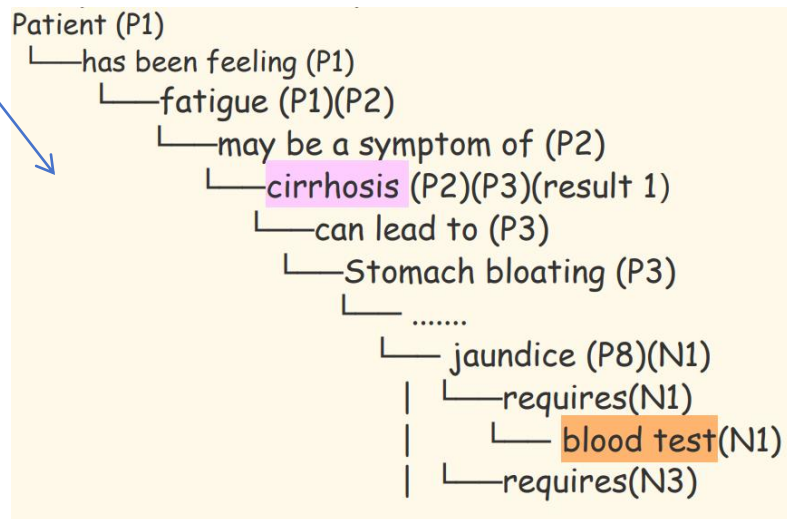


LLMs have achieved remarkable performance in **natural language understanding** and **generation tasks**.

HOWEVER, several challenges lie in the front of steering LLMs in production:

- **Inflexibility.** Pre-trained LLMs possess **outdated knowledge** and are inflexible to **parameter updating**.
Fine-tuning LLMs might **cost highly** to build training pipeline, or even cause **catastrophic forgetting**.
- **Hallucination.** LLMs are **notoriously** known to produce **hallucinations** with plausible-sounding but wrong outputs.
- **Transparency.** The knowledge is implicitly stored in LLM's parameters, thus **infeasible to be validated**.
The **inference process** in deep neural networks remains elusive to be interpretable.

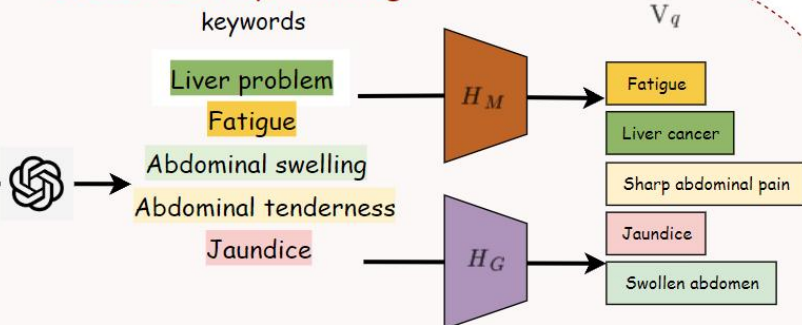
-



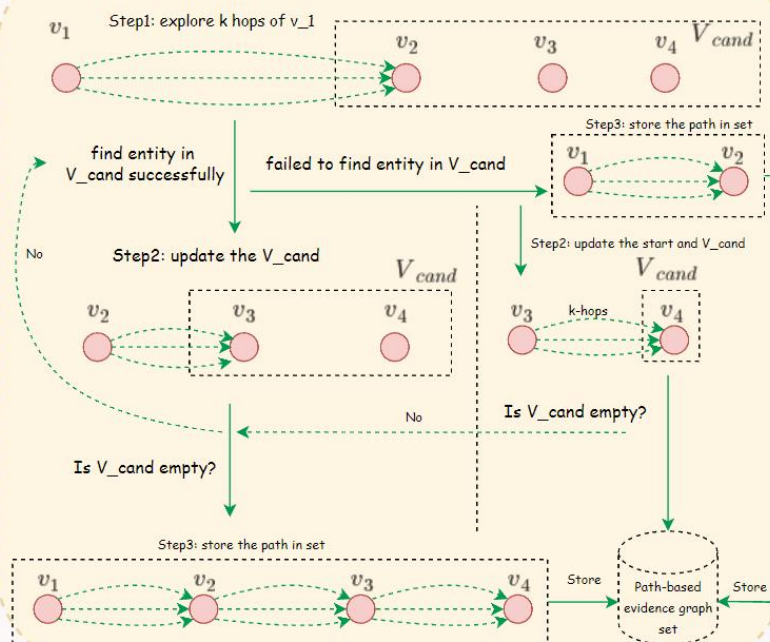
Framework

Evidence Graph Mining

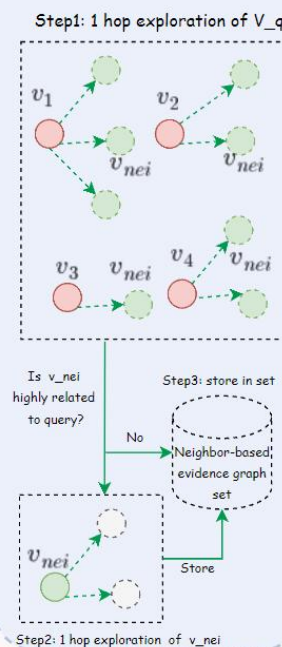
Question
Doctor, I have been feeling very fatigued lately and my abdomen feels swollen and tender. I am also experiencing jaundice in my eyes. Could it be a liver problem?



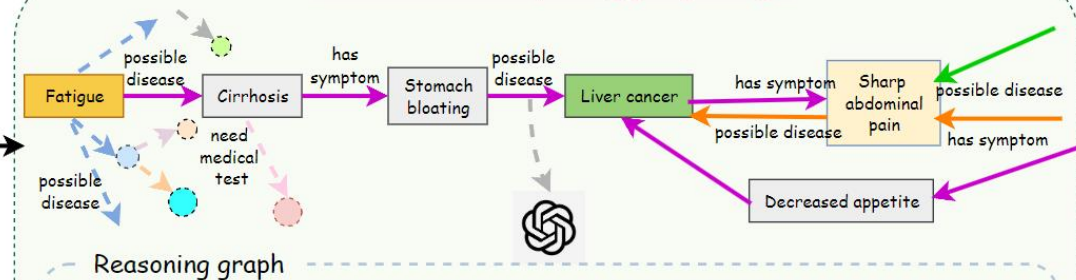
Path-based Evidence Graph Exploration



Neighbor-based Evidence Graph Exploration



Evidence Graph Aggregating



Reasoning graph

P1: Fatigue may be a symptom of Cirrhosis, which in turn may have Stomach bloating as a symptom. Stomach bloating may be a symptom of Liver cancer, which may have Swollen abdomen as a symptom. Swollen abdomen may also be a symptom of Liver cancer, which may have Sharp abdominal pain ...
N1: ...

LangChain

LLM Reasoning with mind map

Output1: (Summary)

Based on the symptom described, the patient may have cirrhosis, need blood test.... Recommended medications include antiviral medications....

Output2: (Inference process)

P1('Patient'-'>'has been feeling'-'>'fatigue')
->P2('fatigue'-'>'may be a symptom of'-'>'Cirrhosis')
->P3('Cirrhosis'-'>'can lead to'-'>'Stomach bloating'))
->...->result1('liver problem')
->N1('liver problem'-'>'requires'-'>'blood tests')->
...->N5(recommended medications)
->'depend on'-'>'...)->...

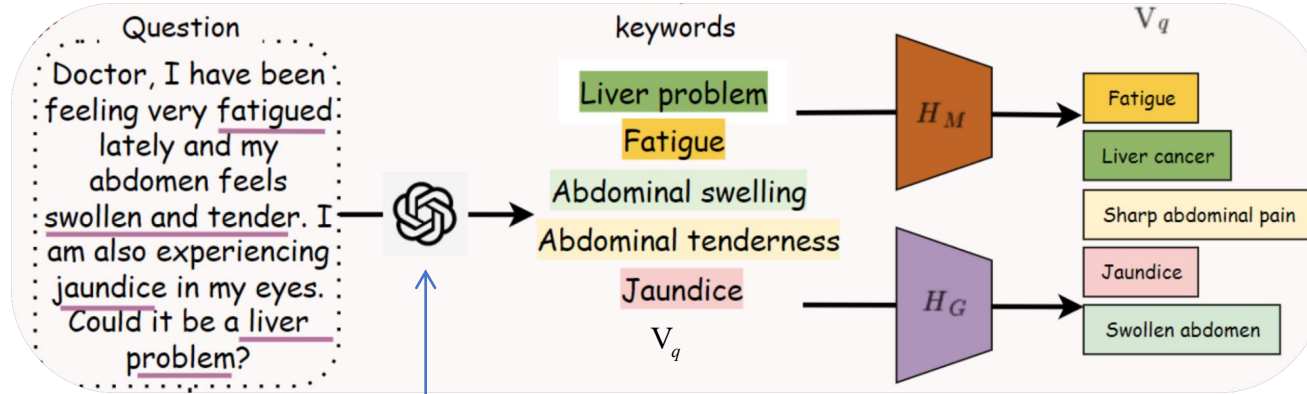
Output3: (Decision tree)

Patient (P1)
├─has been feeling (P1)
├─fatigue (P1)(P2)
├─may be a symptom of (P2)
├─cirrhosis (P2)(P3)(result 1)
├─can lead to (P3)
├─Stomach bloating (P3)
├─.....
├─jaundice (P8)(N1)
├─requires(N1)
├─blood test(N1)
├─requires(N3)
├─.....

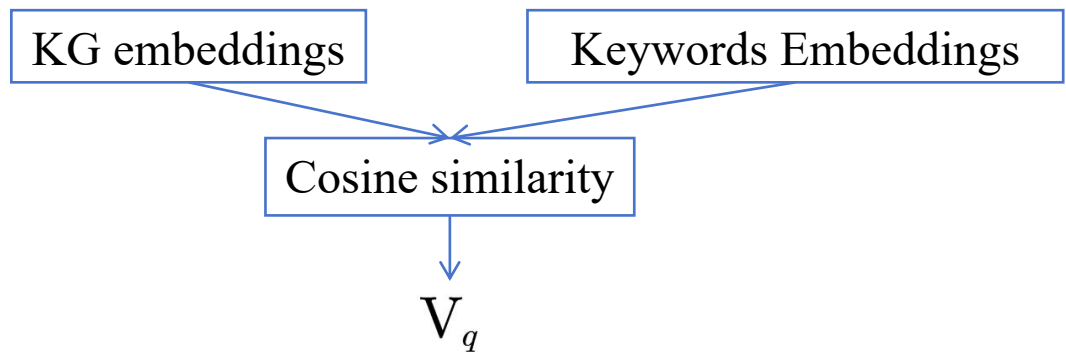
Step I: Evidence Graph Mining

Stage 1: Entity Recognition

Stage 2: Evidence Sub-graphs Exploration



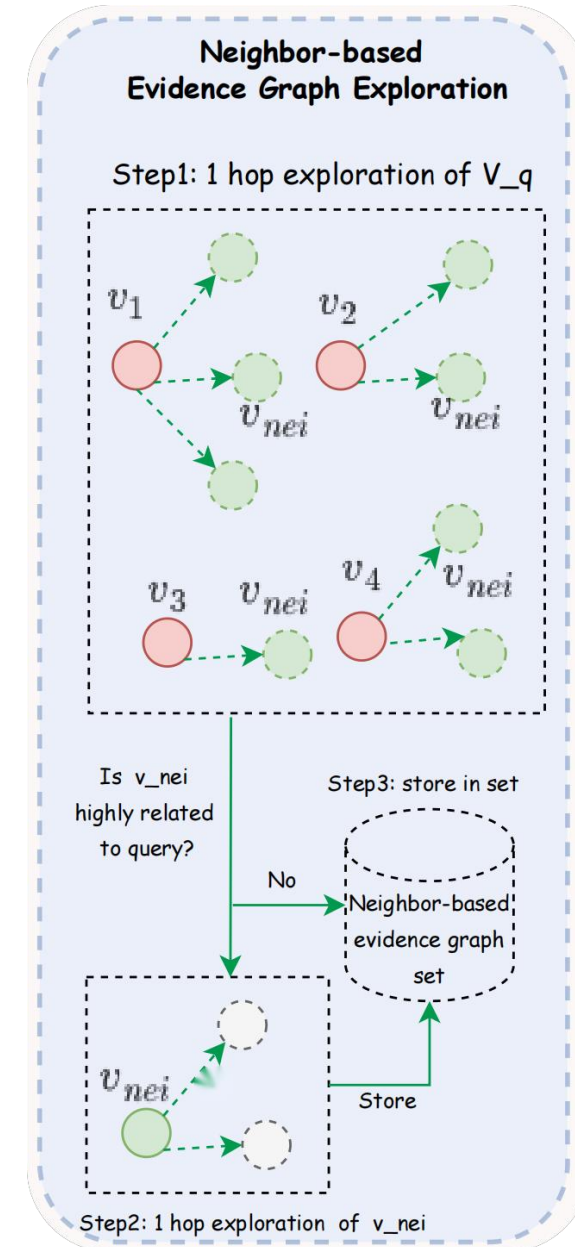
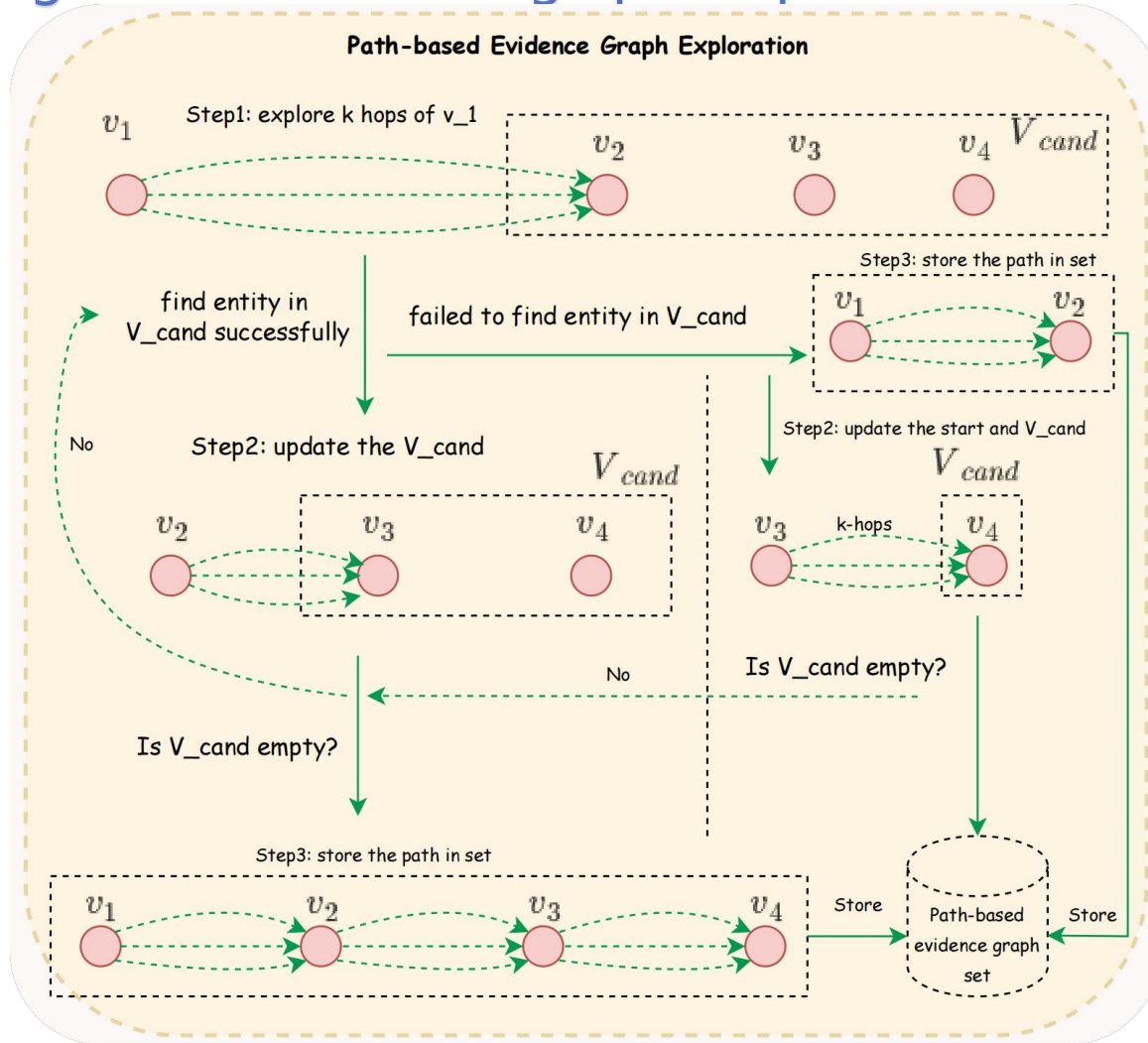
Instruction:\n'Learn to extract entities from the following medical answers.'\n\n### Input:\n<CLS>Okay, based on your symptoms, we need to perform some diagnostic procedures to confirm the diagnosis. We may need to do a CAT scan of your head and an Influenzavirus antibody assay to rule out any other conditions. Additionally, we may need to evaluate you further and consider other respiratory therapy or physical therapy exercises to help you feel better.<SEP>The extracted entities are \n\n ### Output:\n<CLS>Okay, based on your symptoms, we need to perform some diagnostic procedures to confirm the diagnosis. We may need to do a CAT scan of your head and an Influenzavirus antibody assay to rule out any other conditions. Additionally, we may need to evaluate you further and consider other respiratory therapy or physical therapy exercises to help you feel better.<SEP>The extracted entities are CAT scan of head (Head ct), Influenzavirus antibody assay, Physical therapy exercises; manipulation; and other procedures, Other respiratory therapy<EOS>



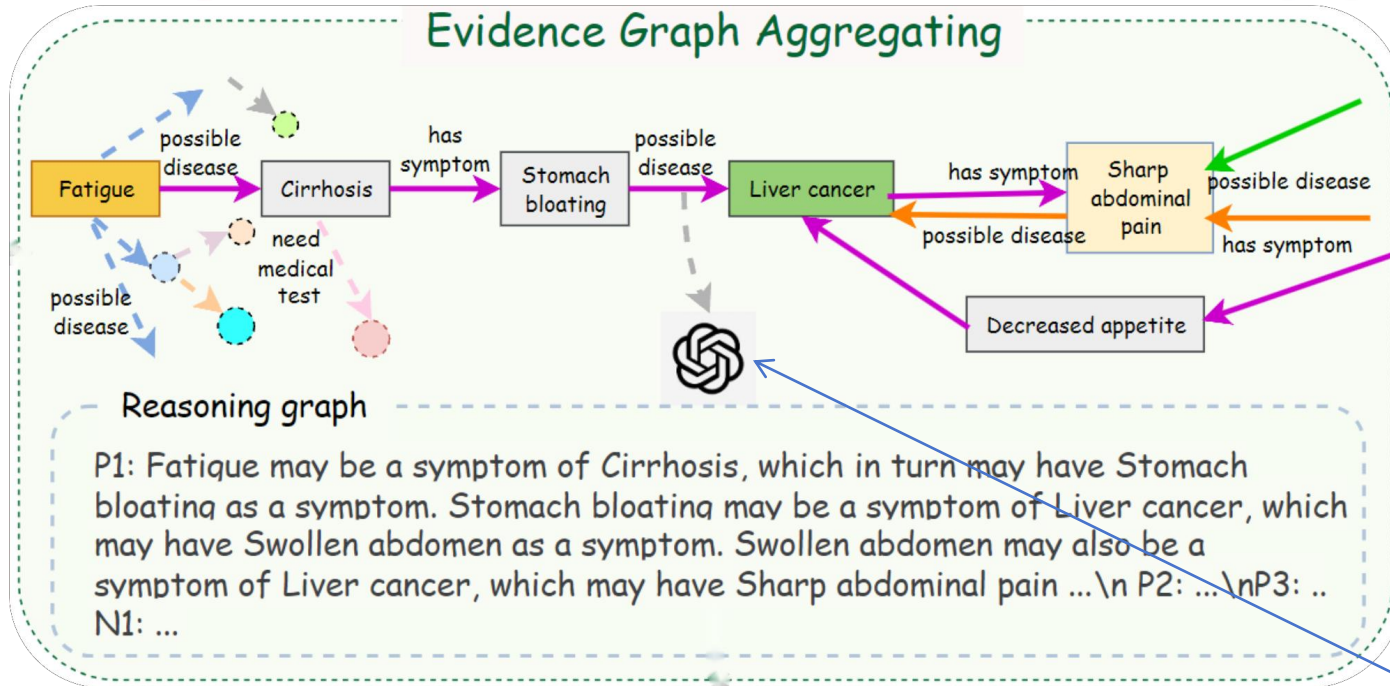
Step I: Evidence Graph Mining

Stage 1: Entity Recognition

Stage 2: Evidence Sub-graphs Exploration



Step II: Evidence Graph Aggregation



k path-based subgraphs
k neighbour-based subgraphs

Format into entity chain like:
(Fatigue, Nausea)-IsSymptomOf-(LiverProblem)

Assign sequence number like:
 $P_1:(Fatigue, Nausea)-IsSymptomOf-(LiverProblem)$
 $P_2, P_3, \dots; N_1, N_2, N_3, \dots$

- Simplify the subgraphs into a concise and consistent format that captured the essential information
- Leveraged LLM's natural language understanding and generation abilities to unify semantically similar entities and resolve potential ambiguities

```
template = """
There are some knowledge graph path. They follow entity->
relationship->entity format.
\n\n
{Path}
\n\n
Use the knowledge graph information. Try to convert them to
natural language, respectively. Use single quotation marks for
entity name and relation name. And name them as Path-based
Evidence 1, Path-based Evidence 2, ... \n\n

Output:
"""
```


Step III: LLM Reasoning with Mind Map

Part 1: Prompting for Graph Reasoning

Part 2: Synergistic Inference with LLM and KG Knowledge

```

SystemMessage(content= ""You are an excellent AI doctor, and you can diagnose diseases and recommend medications based on the symptoms in the conversation."" ),
HumanMessage(content= ""Patient input: "" + Question),
AIMessage(content= f ""Combine the knowledge you already have, you have some extra medical knowledge information in the following:\n\n ### "" + path_reasoning_graph + ""\n\n### "" + neighbor_reasoning_path),
HumanMessage(content= ""What disease does the patient have? What tests should patient take to confirm the diagnosis? What recommended medications can cure the disease? Think step by step.\n\n\n
Output1: The answer includes disease and tests and recommended medications.\n\n
Output2: Show me inference process as a string about extract what knowledge from which Path-based Evidence or Neighbor-based Evidence, and in the end infer what result. \n Transport the inference process into the
following format:\n Path-based Evidence number('entity name'-'>'relation name'-'>...)->Path-based Evidence number('entity name'-'>'relation name'-'>...)->Neighbor-based Evidence number('entity name'-'>'relation name'-'>...)-
>Neighbor-based Evidence number('entity name'-'>'relation name'-'>...)->result number('entity name')->Path-based Evidence number('entity name'-'>'relation name'-'>...)->Neighbor-based Evidence number('entity name'-'>'relation
name'-'>...). \n\n
Output3: Draw a decision tree. The entity or relation in single quotes in the inference process is added as a node with the source of evidence, which is followed by the entity in parentheses.\n\n
There is a sample:\n ... """)
  
```



Output1: (Summary)

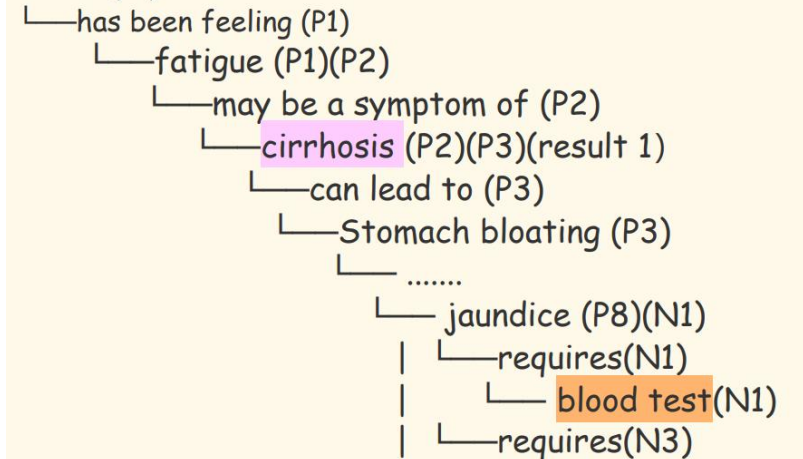
Based on the symptom described, the patient may have **cirrhosis**, need **blood test**.... Recommended medications include **antiviral medications**....

Output2: (Inference process)

```

P1('Patient'-'>'has been feeling'
->'fatigue')
->P2('fatigue'-'>'may be a
symptom of'-'>'Cirrhosis')
->P3('Cirrhosis'-'>'can lead to'-'>'Stomach bloating'))
->...->result1('liver problem')
->N1('liver problem'-'>'requires'-'>'blood tests')->
...->N5(recommended medications'
->'depend on'-'>...)->...
  
```

Patient (P1)



Step III: LLM Reasoning with Mind Map



Part 1: Prompting for Graph Reasoning

Part 2: Synergistic Inference with LLM and KG Knowledge

MindMap enables LLM to synergistically infer from both the **retrieved evidence graphs** and **its own knowledge**

- **Language Understanding.** LLM can **comprehend and extract** knowledge and the query in natural language
- **Knowledge Reasoning,** as LLM can perform **entity disambiguation** and **produce the final answer** based on the mind map
- **Knowledge Enhancement,** as LLM can leverage its **implicit knowledge** to **expand, connect, and improve** the information relevant to the query.

Experiments



Table 1: The statistics of the used datasets.

Dataset	GenMedGPT-5k	CMCQA	ExplainCPE
Domain	English Clinical Q&A	Chinese Long Dialogue	5-way Choice Question
Multi-task	Disease, Drug, Test	Disease, Drug, Test, Food	Option, Explanation
KG dataset	EMCKG	CMCKG	CMCKG
Question	714	468	400
Node	1122	62282	62282
Triple	5802	506490	506490
Relationship	6	12	12

Table 2: The BERTScore and GPT4 ranking of all methods for **GenMedGPT-5k**.

	BERT Score			GPT4 Ranking (Average)	Hallucination Quantify
	Precision	Recall	F1 Score		
MindMap	0.7936	0.7977	0.7954	1.8725	0.6070
GPT-3.5	0.7612	0.8003	0.7800	4.8571	0.5563
Tree-of-thought(TOT)	0.7202	0.7949	0.7554	-	0.5483
GPT4	0.7689	0.7893	0.7786	4.1764	0.5577
BM25 Retriever	0.7693	0.7981	0.7831	3.5546	0.5834
Embedding Retriever	0.7690	0.8038	0.7857	3.1232	0.5886
KG Retriever	0.7717	0.8030	0.7868	3.4159	0.5871

- *BERTScore* measures **semantic similarity** between the generated and reference answers.
- *GPT-4 Rating* ranks **answer quality** against ground truth and **compares pairs** of answers on four criteria:
 - Response diversity and integrity
 - Overall factual correctness,
 - Correctness of disease diagnosis,
 - Correctness of drug recommendation.
- *Hallucination Quantification* estimates the degree of deviation from the facts in the generated answers.

Experiments

MindMap v.s. Baselines on GenMedGPT-5k



Table 3: The pair-wise comparison by GPT-4 on the winning rate of MindMap v.s. baselines on diversity & integrity score (%), fact total match score (%), and disease diagnosis (%), on **GenMedGPT-5k**.

MindMap vs Baseline Metries	GPT-3.5			BM25 Retriever			Embedding Retrivaler			KG Reriever			GPT-4			TOT		
	Win	Tie	Lose	Win	Tie	Lose	Win	Tie	Lose	Win	Tie	Lose	Win	Tie	Lose	Win	Tie	Lose
Diversity & integrity	100	-	-	100	-	-	100	-	-	100	-	-	100	-	-	-	-	-
Total factualness	80.11	-	19.89	66.67	-	33.33	76.05	-	23.95	73.53	-	26.47	75.77	-	24.23	78.5	-	21.5
Disease diagnosis	84.73	0.14	15.13	75.91	1.26	22.83	77.03	1.96	21.01	66.67	2.94	30.39	73.11	1.40	25.49	75	24.6	0.3
Drug recommendation	88	5	7	87	8	5	72	13	15	74	19	7	83	8	9	87	5	8
Average	88.21	1.285	10.505	82.395	2.315	15.29	81.27	3.74	14.99	78.55	5.485	15.965	82.97	2.35	14.68	80.17	14.8	9.93

Experiments

MindMap v.s. Baselines on Chinese Long Dialogue Q&A



Table 4: The BERTScore and GPT-4 ranking of all methods for **CMCQA** dataset.

	BERT Score			GPT-4 Ranking
	Precision	Recall	F1 Score	(Average)
MindMap	0.9415	0.9321	0.9367	2.3
GPT-3.5	0.9385	0.9361	0.9372	3.4
GPT-4	0.9355	0.9358	0.9356	3.6
BM25 Retriever	0.9365	0.9348	0.9356	3.7
Embedding Retriever	0.9357	0.9334	0.9345	5.4
KG Retriever	0.9318	0.9348	0.9332	2.3

Table 5: The pair-wise comparison by GPT-4 on the winning rate of MindMap v.s. baselines on disease diagnosis and drug recommendation on CMCQA.

MindMap vs Baseline Metrics	GPT-3.5			BM25 Retriever			Embedding Retrivers			KG Reriever			GPT-4		
	Win	Tie	Lose	Win	Tie	Lose	Win	Tie	Lose	Win	Tie	Lose	Win	Tie	Lose
Disease diagnosis	35.68	39.96	24.36	30.98	50.21	18.80	37.18	42.74	20.08	34.40	45.51	20.09	27.99	47.22	24.79
Drug recommendation	47.32	30.62	22.06	47.11	29.34	23.55	44.97	32.12	22.91	44.33	31.26	24.41	44.11	29.76	26.12
Average	41.5	35.29	23.21	39.045	39.775	21.175	41.075	37.43	21.495	39.365	38.385	22.25	36.05	38.49	25.455

Experiments

MindMap v.s. Baselines on Multiple Choice Question

Table 6: The accuracy scores for **ExplainCPE**. We calculate the rates of correct, wrong, and failed responses.

Method	Accuracy Rate(%)		
	Correct	Wrong	Failed
GPT-3.5	52.2	47.2	0.5
BM25 Retriever	50	44.2	5.7
Embedding Retriever	54.2	45.2	0.5
KG Retriever	42	44	14
GPT-4	72	27.7	0.2
MindMap	61.7	37.7	0.5
w/o prompt template p ₁	53.5	46	0.5

Table 7: Quantitative comparison with BERTScore and GPT-4 preference ranking between MindMap and baselines in ExplainCPE dataset.

	BERT Score			GPT-4 Ranking (Average)
	Precision	Recall	F1 Score	
MindMap	0.9335	0.9376	0.9354	2.98
GPT-3.5	0.9449	0.9487	0.9467	3.0425
GPT-4	0.9487	0.9529	0.9507	3.0075
BM25 Retriever	0.9413	0.9411	0.9411	3.6675
Embedding Retriever	0.9440	0.9459	0.9449	4.3175
KG Retriever	0.9354	0.9373	0.9362	3.985

Experiments

Ablation Study



Table 8: The BERTScore and hallucination qualification of different component for GenMedGPT-5k.

	Tokens	BERT Score			Hallucination
	(Average)	Precision	Recall	F1 Score	Quantify
Path-only	1028	0.6310	0.7885	0.7002	0.3854
Neighbor-only	1236	0.6393	0.7930	0.7072	0.3894
MindMap	1431	0.7938	0.7987	0.7960	0.5890
Improved-path	+403	+0.1628	+0.0102	+0.0957	+0.2036
Improved-neigh	+195	+0.1545	+0.0057	+0.0888	+0.1996

Experiments

Case compared with other baselines

(a)
Question

Doctor, I have been feeling very fatigued lately and my abdomen feels swollen and tender. I am also experiencing jaundice in my eyes. Could it be a liver problem?

(b)
Reference

Based on your symptoms, it is possible that you may have **cirrhosis**, a liver disease. We will need to run a battery of medical tests to confirm the diagnosis, including hematologic tests, complete **blood count**, glucose measurement, an electrolyte panel, a lipid panel, liver function tests, and hemoglobin A1c measurement.

MindMap

Based on the symptoms described, the patient may have a liver problem, possibly **cirrhosis** or liver cancer, which can cause fatigue, swollen and tender abdomen, and jaundice. To confirm the diagnosis, the patient should undergo **blood tests** to check liver function and imaging tests such as an ultrasound or CT scan. Recommended medications depend on the underlying cause of the liver problem, and may include antiviral medications for viral hepatitis, immunosuppressants for autoimmune hepatitis, or diuretics for fluid buildup.

GPT-3.5

I'm not a doctor, but your symptoms could possibly be related to a liver problem. It's important to **consult** with a healthcare professional for an accurate diagnosis. They will be able to evaluate your symptoms, conduct any necessary tests, and provide you with proper guidance and treatment.

GPT-4

As an AI, I can suggest that these symptoms might indicate a liver problem, such as Hepatitis, Cirrhosis, or Liver disease. But, fatigue could also be linked to many other conditions. It's critical to speak directly to a healthcare provider who can accurately diagnose your condition based on a comprehensive medical evaluation. This could involve **answering detailed** questions about your symptoms and possibly undergoing lab tests or imaging studies.

(c)
Output

BM25 Retriever: Based on the symptoms described by the patient (fatigue, swollen and tender abdomen, and jaundice in the eyes), it is possible that the patient is suffering from a **liver problem**. However, the information provided about a disease characterized by **eye alignment disorder**, with symptoms such as eye deviation, diminished vision, double vision, cross-eyed, pain in the eye, abnormal movement of eyelid, foreign body sensation in eye, lack of growth, irregular appearing scalp, and swollen lymph nodes, is not relevant to the patient's current symptoms. Therefore, the information provided is not sufficient to make a diagnosis or recommend medications. In order to properly diagnose the patient, a physical examination and medical tests may be necessary, including blood tests to check liver function, ultrasound to examine the liver and other abdominal organs, and potentially a liver biopsy. Treatment for liver disease can vary depending on the specific diagnosis, but may include lifestyle changes, medications, or in severe cases, liver transplant. Therefore, I recommend that the patient schedule an appointment with a medical professional for further evaluation and diagnosis.

Embedding Retriever: Based on the patient's symptoms of fatigue, swollen and tender abdomen, and jaundice in the eyes, it is possible that the patient has a **liver problem**. However, the provided medical information is not related to liver problems, but rather to a medical condition called **Chalazion**, which affects the **eyes**. Therefore, it is recommended that the patient consults a medical doctor to undergo further tests and examinations to determine the cause of their symptoms and receive an accurate diagnosis. Some tests that a doctor may recommend include liver function tests, ultrasound, CT scan, or MRI of the abdomen. As for the medications to cure Chalazion, some of the options mentioned in the medical information include dexamethasone-tobramycin ophthalmic, erythromycin ophthalmic, and loteprednol-tobramycin ophthalmic. However, these medications are not related to the patient's current symptoms and are not recommended for the treatment of liver problems. The appropriate medications for liver problems depend on the specific diagnosis and should be prescribed by a medical doctor.

KG Retriever: Based on the symptoms described by the patient, it is possible that they have a liver problem. The combination of fatigue, swollen and tender abdomen, and jaundice in the eyes suggests that the patient may have **choledocholithiasis**, which is a condition where gallstones block the bile ducts in the liver. This can lead to an inflamed and swollen liver, causing the symptoms the patient is experiencing. To confirm the diagnosis, the patient should undergo a blood test to check for elevated levels of bilirubin and liver enzymes. An ultrasound or CT scan may also be necessary to visualize the bile ducts and liver. The recommended medication for choledocholithiasis is typically a procedure called an ERCP (Endoscopic Retrograde Cholangiopancreatography) to remove the gallstones from the bile ducts. Pain medication may be prescribed to help manage any discomfort during this procedure. It is important for the patient to seek medical attention as soon as possible to prevent further damage to the liver and to manage symptoms.

GPT-4
Rater



MindMap > KG Retriever > BM25 Retriever > KG Retriever > GPT-4 > GPT-3.5

Experiments

An Example of the MindMap

(a)
Question

我有一颗阻生齿埋在牙肉里,靠近磨牙的后面,磨牙前面有颗蛀牙拔掉了,所以是空的,现在唯一的一颗磨牙被后面的阻生齿顶歪了,请问我是该先拔智齿,再补牙,还是先补牙再拔后面的智齿? (女,23岁) 看到了吗? 对了,肉里的智齿老是长包发炎 对 就是那独立的一颗磨牙,后面牙肉里长了一颗智齿,前面蛀牙拔掉了前几年拍过,是倒着长的 那颗磨牙被顶歪了,所以我不知道该拔牙还是先补前面的 那拔多久才可以补 长在肉里的阻生齿不拔 会怎样呢 哦,谢谢",

Output1

根据患者的描述,您可能患有智齿阻生症。为了确认诊断,建议您进行口腔X光片检查。治疗阻生齿的推荐药物包括人工牛黄甲硝唑胶囊。建议您先拔除阻生齿,然后再进行补牙。忌吃辛辣食物和粘糕食物。

(b)
MindMap

Output2

路径证据1('阻生齿'-'可能导致'-'牙列拥挤')->路径证据2('牙列拥挤'-'并发症为'-'龋齿')->邻居证据1('龋齿'-'常用药品'-'丁硼乳膏和布洛芬片')->邻居证据2('阻生齿'-'常用药品'-'人工牛黄甲硝唑胶囊')->结果标号('智齿阻生症')->邻居证据3('阻生齿'-'诊断检查'-'口腔X光片检查')->结果标号('口腔X光片检查')->路径证据3('智齿阻生症'-'常用药品'-'人工牛黄甲硝唑胶囊')->结果标号('人工牛黄甲硝唑胶囊')->邻居证据4('智齿阻生症'-'常用药品'-'复方盐酸阿替卡因注射液')->结果标号('复方盐酸阿替卡因注射液')->邻居证据5('智齿阻生症'-'常用药品'-'丁硼乳膏和布洛芬片')->结果标号('丁硼乳膏和布洛芬片')

Output3

患者(路径证据 1)

- 阻生齿(路径证据 1)(路径证据 2)(邻居证据 2)(路径证据 3)
 - 可能导致(路径证据 1)
 - 牙列拥挤(路径证据 1)
 - 并发症为(路径证据 2)
 - 龋齿(路径证据 2)(邻居证据 1)
 - 常用药品(邻居证据 1)
 - 丁硼乳膏(邻居证据 1)(结果 2)
 - 布洛芬片(邻居证据 1)(结果 2)
 - 常用药品(邻居证据 2)
 - 人工牛黄甲硝唑胶囊(邻居证据 2)(结果 3)(路径证据 3)
 - 诊断检查(邻居证据 3)
 - 口腔 X 光片检查(邻居证据 3)(结果 4)
 - 常用药品(路径证据 3)
 - 人工牛黄甲硝唑胶囊(路径证据 3)(结果 3)
 - 复方盐酸阿替卡因注射液(结果 5)
 - 丁硼乳膏和布洛芬片(结果 6)