

ナレッジグラフ推論チャレンジ 2024 応募シート

1. 応募者に関する情報

- 氏名またはチーム名：Zihao Yuan
- 所属：Zhejiang University
- メールアドレス（代表）：maskyuan@zju.edu.cn
- 応募者に学生が含まれる： はい
- 応募者の代表が学生である： はい

2. 応募部門：一般部門

3. 構築したナレッジグラフについて

- 構築対象としたナレッジグラフ

Mystery in True detective benchmark.

In the narrative reasoning process of mystery genres, exceptional detectives typically follow a three-stage cognitive paradigm of "clue collection - clue organization - logical deduction." Our study focuses on the performance differences of large language models (LLMs) in such knowledge-intensive reasoning tasks, with particular attention to the distinctions between their capabilities in traditional natural language processing tasks. While Transformer-based LLMs excel in basic tasks such as text generation and semantic understanding, they show significant limitations in mystery story-based reasoning scenarios that require clue extraction and combinatorial analysis.

As a structured data representation method, knowledge graphs enable humans to start from a particular node and gain a comprehensive understanding of all related information. For the Transformer architecture, each token at any given position is related to other tokens at different positions, which facilitates the construction of the graph. If the goal is to solve mystery-type problems, is there a significant difference between directly using LLMs for reasoning and first constructing a knowledge graph and then performing reasoning?

Currently, research in this area remains lacking. Using the True Detective benchmark proposed by Maksym Del et al., we explore the application of LLMs in knowledge graph construction and reasoning. Our work is innovative in several key areas: 1) it demonstrates that LLMs are not particularly adept at reasoning with mystery stories; 2) we construct a knowledge graph centered around options,

covering characters, locations, times, objects, etc., and use it for reasoning. When compared to Chain-of-Thought (CoT) reasoning, we find no significant improvement in LLMs' reasoning capabilities. Their accuracy is around 40%, and it is difficult to achieve further improvement.; 3) although our method makes the knowledge graph constructed by LLMs somewhat more accessible for human use, enabling LLMs to perform autonomous reasoning still presents challenges.



Figure 1: If we judge solely based on the clues about the character, it's easy to choose "a." But can LLMs detect in the text that "a" has never appeared?

- 構築したナレッジグラフの基本情報
 - データサイズ
 - Gemini-2.0-Flash CoT reasoning data: 328kb
 - Gemini-2.0-Flash CoT with graphs reasoning data: 957kb
 - Gemini-2.0-Flash-thinking CoT reasoning data: 263kb
 - Gemini-2.0-Flash-thinking CoT with graphs reasoning data: 1047kb
 - データ形式
 - The information is recorded in text form, with part of it represented as a knowledge graph in triples.
- 構築したナレッジグラフのデータの入手先

Mystery in True detective benchmark.

4. ナレッジグラフ構築に用いた「言語モデル」および「構築手法」について

- ナレッジグラフ構築に用いた「言語モデル」

Gemini-2.0-flash-exp(Google ai studio api) and

Gemini-2.0-flash-thinking-exp-01-21(Google ai studio api)

- ナレツジグラフ構築に用いた「データ」

The True Detective benchmark contains 191 stories. We designed prompts with the goal of selecting answers and constructed a knowledge graph centered around the provided options.
Basically, we construct input as: {Story content} and {Answer options}.
- ナレツジグラフの構築手法の説明

The prompt used will be provided in the attachment. Here, we introduce several modules included in our prompt, which are the tasks that LLMs need to complete.

 - **Story background and global information extraction.**

After reading a mystery story, we ensure that the model has a basic factual understanding of the story as a whole, so that LLMs do not overlook elements associated with the story. For example, if the story takes place in Arizona, a major backdrop, the model will also pay attention to other locations of the same level, such as other states and regions mentioned in the text.
 - **Knowledge graph construction and information extraction.**

The options usually include categories such as characters, locations, time, and objects. For each category, we ensure that the model collects relevant clues from the text without omission, constructing the knowledge graph in the form of triples, as shown in Figure 2.
 - **Analyze Each Option**

For each option and clue, we conduct Factual Analysis, Contradiction Analysis, and Relevance Analysis, cross-checking whether the clue contradicts any specific object.
 - **Final Conclusion**

Based on the previous analysis, we provide the model with several common types of mystery stories for identification, while also informing the model about how clues are typically abstracted and used in each category. This serves as a tool for the model to apply the prior content and the tool to complete the reasoning process. The main categories include: Whodunit (Classic Detective Mysteries), Howdunit (Method-

Based Mysteries), Why-dunit (Psychological or Motivational Mysteries), Cozy Mystery, Hardboiled/Noir Mystery, Police Procedural, Legal or Courtroom Mystery, Historical Mystery, and Supernatural Mystery.

- Person: Record all information related to individuals, including:
 - Quotes: (Speaker, said, Quote) Example: (John, said, "I went to the park yesterday")
 - Actions: (Actor, did, Action) Example: (Mary, ran, across the street)
 - Personality Traits: (Person, is, Trait) Example: (David, is, kind)
 - Hobbies: (Person, enjoys, Hobby) Example: (Sarah, enjoys, reading)
 - Relationships: (Person1, is related to, Person2) Example: (Tom, is friend of, Peter)
 - Any other information related to the person
- Location: Record all information related to places, including:
 - Objects Present: (Location, has, Object) Example: (Bedroom, has, desk) / (Living Room, does not have, television)
 - People Present: (Location, has, Person) Example: (Park, has, John)
 - Events Occurring: (Location, event occurred, Event) Example: (Restaurant, event occurred, argument)
- Time: Record all information related to time, including:
 - Events Happening: (Time, event happened, Event) Example: (9 PM, event happened, power outage)
 - Person's Actions: (Time, Person, Action) Example: (7 AM, Mary, woke up)
 - Temporal Relationships: (Event1, is before/after/at the same time as, Event2) Example: (Arrival, is before, Departure)
- Object: Record all information related to objects, including:
 - Usage: (Object, is used for, Purpose) Example: (Key, is used for, opening door)
 - Ownership: (Person, owns, Object) Example: (John, owns, phone)
 - Location: (Object, is located in, Location) Example: (Book, is located in, bookshelf)
 - State/Condition: (Object, is, Condition) Example: (Vase, is, broken)
- Object: Record all information related to objects, including:
 - Usage: (Object, is used for, Purpose) Example: (Key, is used for, opening door)
 - Ownership: (Person, owns, Object) Example: (John, owns, phone)
 - Location: (Object, is located in, Location) Example: (Book, is located in, bookshelf)
 - State/Condition: (Object, is, Condition) Example: (Vase, is, broken)

Figure 2: Content to be extracted for different categories.

● パフォーマンス情報

Access is made via API, and the environment runs on WSL with Ubuntu 20.04,

using Python version 3.11.

● 参考情報

Del, Maksym, and Mark Fishel. "True detective: a deep abductive reasoning benchmark undoable for GPT-3 and challenging for GPT-4." arXiv preprint arXiv:2212.10114 (2022).

Team, Gemini, et al. "Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context." arXiv preprint arXiv:2403.05530 (2024).

Takahiro Kawamura, Shusaku Egami, Kotaro Tamura, Yasutomo Sozono, Takanori Ukai, Yusuke Koyanagi, Fumito Nishino, Seiji Okajima, Katsuhiko Murakami, Kunihiro Takamatsu, Aoi Sugiura, Shun Shiramatsu, Xiang Yu, Koji Furusaki : 2018

5. 構築したナレッジグラフの評価

一般部門

As this is the first attempt to construct a knowledge graph on this benchmark, we try to measure the extent to which this graph aids reasoning for both LLMs and humans, from the perspective of "helpfulness for reasoning." We take one of the stories, "*Who Stole the Rock of Ages*," (mystery is shown in Figure) as an example and explain by observing the answers generated by LLMs using the knowledge graph. Additionally, we explore whether using this graph is helpful for humans.

● For LLMs:

Model Name	CoT w/o Graphs	CoT w/ Graphs
Gemini-2.0-Flash	35.08%	37.17%
Gemini-2.0-Flash-thinking	41.36%	36.65%

Table 1: The accuracy of two models when using CoT for reasoning and when assisted by the knowledge graph for reasoning, with a total of 191 questions.

By comparing the data in Table 1, we can easily see that even when the corresponding knowledge graph is provided, the overall performance in mystery reasoning tasks was not significantly affected. By specifically examining the answers in Figures 4 and 5, we found that with the help of the knowledge graph, the model correctly identified the places each character visited, especially the mention of Juan driving into the sunlight. This presents a key contradiction: Juan was driving from east to west, so he would not be affected by the sunlight, which is crucial for the reasoning. However, the LLMs were distracted by the mention of sand in the text, overlooked the

contradiction, and proceeded to reason based on another fact, leading to the incorrect answer.

Who Stole the Rock of Ages

The Deer Valley Rock Art Center is a repository of art and history, preserving ancient petroglyphs and pictographs for the modern world to study and enjoy. Maintained as a part of Arizona State University, it is ideally situated near caches of desert rock, upon which natives could leave their mark in the clay-coated surfaces. The meaning of this art will forever be uncertain, as the artists died long before any art critics had the chance to ask. Since the artists are long gone, their work is priceless, valuable beyond any dollar measure. Sadly, such value can lead to problems as well. Dr. Welsh could barely contain himself. First, someone had delivered, practically to his doorstep, the most astounding piece of art ever, and now, not a week later, it had been stolen. It was intolerable. How could anyone stoop so low as to steal this prize of humanity, this "Stona Lisa" of petroglyphs? Only money could drive a person to act so fiendishly, he reasoned. Disgusted, he shook his head and waited. Sure enough, his four research students showed up for their Monday morning assignments, right on time. Three of them made it on time, anyway. Denise showed up five minutes late, her leg in a cast. Dr. Welsh was too preoccupied to notice, so she didn't get the usual lecture on punctuality. Dr. Welsh knew that one of these four was the culprit. His other staff all had perfect alibis, and the evidence pointed clearly to an inside job. The alarm had been deactivated Sunday, at 11:03 a.m., and reactivated just three minutes later. The thief had known the codes and known exactly where to find their prize quickly – back in the preservation lab. So far, no one had informed the students of the theft. He decided to keep it that way, for the moment. Only the thief would know. Aside from the time of the theft, Dr. Welsh had only one other clue to work with. A trail of sand had fallen from the thief's pants or shoes as he or she walked through the building. In Arizona, such a thing wouldn't be unusual, but a quick lab test had shown that the sand didn't match the local sand – it had a different composition, and came from somewhere else. The students had just had a three-day weekend, so the culprit had probably gotten sandy while on vacation somewhere. It was time to begin. Dr. Welsh took a deep breath to compose himself, and went to greet his students. "Welcome back everyone. I trust you all had a relaxing weekend." Skye Smith piped up immediately. "It was good, but not as nice as being here." Juan Carde snorted at her. "Are you kidding? Anything is more relaxing than work." When Dr. Welsh raised an eyebrow, Juan hurried to adjust his wording. "But working here is still better than other places, I mean." "Oh, would you two stop sucking up so we can get our assignments?" Denise wasn't in the mood to be patient. "I want to go sit down already." Denise was usually such a sweet girl that she caught Dr. Welsh's attention. "My apologies Denise, I hadn't noticed your leg before now." He waved them all into the lobby, where chairs were plentiful. "There's no rush to get to work on a Monday, but we can certainly talk sitting down." Once they were seated, Dr. Welsh knew what he wanted to ask everyone, and had the perfect place to start. "However did you manage to break your leg this weekend?" Denise grimaced, either in pain or embarrassment; it was difficult to tell which. "I thought it would be a good idea to hit the slopes up in Aspen, Colorado. About the time I tumbled down the mountainside Saturday morning, I started to rethink that idea." Dr. Welsh needed a little more information. "How did you manage to get home with that leg?" "After they had me patched up, they contacted my parents for me. They came and drove me home after I was released from the hospital Sunday afternoon. They said that next year I should just take them along so that they could enjoy the skiing part too. That's sympathy for you." Dr. Welsh nodded thoughtfully. "What about you Mr. Gaigon? You've been awfully quiet this morning." Jim quickly looked up from the floor. (Had he been avoiding Dr. Welsh's eyes, or sneaking a nap?) "I'm sorry professor, I didn't – YAWN – sleep much last night. I spent the weekend in Roswell, New Mexico, visiting the International UFO Museum and Research Center. I came back late yesterday afternoon, but I tell you, that desert sun coming west was murder on my eyes. I kept dreaming about bright spots and aliens all night long. It was creepy." Juan scoffed at his story. "UFO's, ha! You think the sun was bad on your drive? Coming back from Long Beach, California yesterday afternoon was just as bright, and I hate driving into the sun. Oh, but a weekend in the surf was well worth it." Dr. Welsh looked surprised. "Why Juan, I didn't know you were a surfer." Now it was Juan's turn to look a bit embarrassed. "Well, I'm not much of one yet, I'm still learning. When I say I spent my time in the surf, I mean exactly that. I kept falling off my board. I think the people were calling me 'Can't-Stay-On Juan' by the time I left." That got a chuckle out of Dr. Welsh, despite his usually professional demeanor, and despite all that was on his mind. "We'll promise to let you be just plain Juan here, never fear. What about you Skye, how was your vacation? You look a bit tanned." "And well I should," answered Skye, flipping back her long brown hair. I flew down to Cozumel, Mexico for some scuba diving with the fishes. I'll bet you didn't know that I scuba either." "Well actually..." began Dr. Welsh, but Skye went on after inhaling. "It was beautiful, just like always, with the sparkling water and the beautiful fish, and so much to see, and the perfect sun for tanning. Oh, I could have stayed forever, except that I would never do that to you. Even so, I almost didn't make it back. My plane was delayed, and delayed again, and it didn't get back here until one this morning. But I'm still not as sleepy as Jim. I slept on the plane. I always say – get your sleep when there's nothing to do, or else you'll miss it when there is." She glared over at her snoring colleague. "Okay Jim, you can wake up now, I'm done." Jim yawned and stretched. "I feel much better now," he managed. "That's good Jim," began Dr. Welsh, "because I have some important news for you all." They all perked up, attentive to what he would say. "Yesterday morning, someone stole the 'Stona Lisa' we received last week." Four gasps caused him to pause. "Worse, the theft was by one of our own – one of you four, and I'm afraid that I now know who that person is. You all have alibis, but the guilty person is lying. Please don't do anything silly, a police officer is patiently waiting in the next room."

Figure 3: Story content. Answer options are: (a) Denise Hurst; (b) Jim Gaigon; (c) Juan Carde; (d) Skye Smith, answer is (c) Juan Carde.

Answer of Gemini-2.0-Flash

- Goal: Who stole the "Stona Lisa" petroglyph?
- Clue Graph:
 - (a) Denise Hurst
 - (Denise, showed up, five minutes late)
 - (Denise, has, leg in a cast)
 - (Denise, is usually, sweet)
 - (Denise, went to, Aspen, Colorado)
 - (Denise, fell down, mountainside Saturday morning)
 - (Denise, was patched up at, hospital)
 - (Denise's parents, drove her, home Sunday afternoon)
 - (Denise, wanted, to sit down)
 - (Denise, was not, patient)
 - (Denise, was asked about, her leg)
 - (b) Jim Gaigon
 - (Jim, was quiet, this morning)
 - (Jim, looked up from, floor)
 - (Jim, didn't sleep much, last night)
 - (Jim, spent weekend in, Roswell, New Mexico)
 - (Jim, visited, International UFO Museum and Research Center)
 - (Jim, came back, late yesterday afternoon)
 - (Jim, said the desert sun coming west, was murder on his eyes)
 - (Jim, kept dreaming about, bright spots and aliens all night)
 - (Jim, was sleepy)
 - (Jim, was snoozing)
 - (c) Juan Carde
 - (Juan, snorted at, Skye)
 - (Juan, adjusted his wording, after being called out)
 - (Juan, spent weekend in, Long Beach, California)
 - (Juan, went, surfing)
 - (Juan, kept falling off his board)
 - (Juan, was called 'Can't-Stay-On Juan')
 - (Juan, hates driving into the sun)
 - (Juan, scoffed at, Jim's UFO story)
 - (Juan, was a bit, embarrassed)
 - (d) Skye Smith
 - (Skye, said it was good, but not as nice as being here)
 - (Skye, was sucking up)

- (Skye, went to, Cozumel, Mexico)
 - (Skye, went, scuba diving)
 - (Skye, was tanned)
 - (Skye, said the water was sparkling)
 - (Skye, had perfect sun for tanning)
 - (Skye, plane was delayed)
 - (Skye, didn't get back until, one this morning)
 - (Skye, was not sleepy)
 - (Skye, glared at, Jim)
- Analysis:
 - **Type of Story:** This is a classic Whodunit mystery. The focus is on identifying the culprit from a limited pool of suspects. The detective (Dr. Welsh) is presented with clues and uses logic to deduce the guilty party.
 - **Reasoning:**
 1. **Elimination based on evidence:** The thief left a trail of "non-local sand". This clue eliminates Denise who was in Colorado (snow, not sand), and possibly Juan, who was surfing in Long Beach which is a well-known, heavily populated beach where sand composition is probably well-known. It should not have confused Dr. Welsh's expert. While there could have been some sort of unusual sand at the UFO center, Dr. Welsh did not identify that as a possible location with a different sand composition. Jim said that he was tired from the long drive. The clues point to Skye. Cozumel has distinctive coral sand. Dr. Welsh himself says "You look a bit tanned", indicating that he is observing closely and picking up on details.
 2. **Alibi contradictions:** Each student provides a plausible weekend trip. However, Skye was in Cozumel, Mexico, known for its coral sand. The sand in the lab didn't match the local sand, and the culprit had a three-day weekend. It is probable that she did not change her clothes/shoes on the flight home so that the sand would still be on her.
 3. **Logical deduction:** The thief knew the alarm codes and the location of the petroglyph. This means it was someone familiar with the facility and its security protocols. The students were the only people who could plausibly have that level of knowledge besides the staff, who have alibis. She was also the only student who does not appear to be under duress.
 4. **Motivation:** The text suggests that the motive was money. Any of the students could have this motive, making it a weaker clue than the physical evidence.
 - Answer: (d) Skye Smith

Figure 4: The answer generated by Gemini-2.0-Flash under the given prompt ultimately made the wrong choice.

Answer of Gemini-2.0-Flash-Thinking

- Goal: Who is the culprit?
- Clue Graph:
 - ***(a) Denise Hurst**
 - (Denise, showed up, 5 minutes late)
 - (Denise, has, leg in cast)
 - (Denise, went to, Aspen, Colorado)
 - (Denise, went for, skiing)
 - (Denise, tumbled down, mountainside Saturday morning)
 - (Denise, was, patched up)
 - (Hospital, contacted, Denise's parents)
 - (Denise's parents, drove, Denise home)
 - (Denise, released from hospital, Sunday afternoon)
 - (Denise, said, "I thought it would be a good idea to hit the slopes up in Aspen, Colorado.")
 - (Denise, said, "About the time I tumbled down the mountainside Saturday morning, I started to rethink that idea.")
 - (Denise, said, "After they had me patched up, they contacted my parents for me. They came and drove me home after I was released from the hospital Sunday afternoon.")
 - (Denise, said, "They said that next year I should just take them along so that they could enjoy the skiing part too. That's sympathy for you.")
 - (Denise, said, "Oh, would you two stop sucking up so we can get our assignments?")
 - (Denise, said, "I want to go sit down already.")
 - (Denise, is usually, sweet girl)
 - ***(b) Jim Gaigon**
 - (Jim, looked up from, floor)
 - (Jim, been, quiet this morning)
 - (Jim, yawned)
 - (Jim, stretched)
 - (Jim, didn't sleep, much last night)
 - (Jim, spent weekend, in Roswell, New Mexico)
 - (Jim, visiting, International UFO Museum and Research Center)
 - (Jim, came back, late yesterday afternoon)
 - (Desert sun, coming west, murder on Jim's eyes)
 - (Jim, kept dreaming about, bright spots and aliens)
 - (Jim, said, "I'm sorry professor; I didn't – YAWN – sleep much last night.")
 - (Jim, said, "I spent the weekend in Roswell, New Mexico, visiting the International UFO Museum and Research Center.")
 - (Jim, said, "I came back late yesterday afternoon, but I tell you, that desert sun coming west was murder on my eyes. I kept dreaming about bright spots and aliens all night long. It was creepy.")
 - (Jim, said, "I feel much better now.")

- ***(c) Juan Carde**
 - (Juan, snorted at, Skye)
 - (Juan, went to, Long Beach, California)
 - (Juan, went for, surf)
 - (Juan, is, not much of surfer yet)
 - (Juan, kept falling off, his board)
 - (People, were calling Juan, 'Can't-Stay-On Juan')
 - (Juan, drove back from, Long Beach, California yesterday afternoon)
 - (Sun, was bad on, Juan's drive)
 - (Juan, hates, driving into the sun)
 - (Juan, said, "Are you kidding? Anything is more relaxing than work.")
 - (Juan, said, "But working here is still better than other places, I mean.")
 - (Juan, said, "UFO's, ha! You think the sun was bad on your drive? Coming back from Long Beach, California yesterday afternoon was just as bright, and I hate driving into the sun.")
 - (Juan, said, "Oh, but a weekend in the surf was well worth it.")
 - (Juan, said, "Well, I'm not much of one yet, I'm still learning. When I say I spent my time in the surf, I mean exactly that. I kept falling off my board. I think the people were calling me 'Can't-Stay-On Juan' by the time I left.")
- ***(d) Skye Smith**
 - (Skye, piped up, immediately)
 - (Skye, flipped back, her long brown hair)
 - (Skye, flew down to, Cozumel, Mexico)
 - (Skye, went for, scuba diving)
 - (Skye, got, tanned)
 - (Water in Cozumel, is, sparkling)
 - (Fish in Cozumel, is, beautiful)
 - (Sun in Cozumel, is, perfect for tanning)
 - (Skye's plane, was, delayed)
 - (Skye's plane, delayed, again)
 - (Skye's plane, got back, one this morning) (Monday morning)
 - (Skye, slept, on the plane)
 - (Skye, said, "It was good, but not as nice as being here.")
 - (Skye, said, "And well I should," answered Skye, flipping back her long brown hair. I flew down to Cozumel, Mexico for some scuba diving with the fishes. I'll bet you didn't know that I scuba either.")
 - (Skye, said, "It was beautiful, just like always, with the sparkling water and the beautiful fish, and so much to see, and the perfect sun for tanning. Oh, I could have stayed forever, except that I would never do that to you. Even so, I almost didn't make it back. My plane was delayed, and delayed again, and it didn't get back here until one this morning.")
 - (Skye, said, "But I'm still not as sleepy as Jim. I slept on the plane. I always say – get your sleep when there's nothing to do, or else you'll miss it when there is.")
 - (Skye, glared over at, Jim)
 - (Skye, said, "Okay Jim, you can wake up now, I'm done.")

- Analysis:
This is a **Whodunit** (Classic Detective Mystery). We need to identify the culprit among the four students. The key clues are the non-local sand and the timeframe of the theft (Sunday 11:03-11:06 am).

- **Denise:** Alibi is Aspen, Colorado, with a broken leg injury on Saturday. Aspen is mountainous and unlikely to be the source of the non-local sand. Her injury and parent escort makes her less mobile and less likely to be the thief who needed to be quick and quiet.

- **Jim:** Alibi is Roswell, New Mexico. Roswell is in the desert and could be the source of non-local sand. His "late Sunday afternoon" return is vague and he could have easily returned earlier to commit the theft.

- **Juan:** Alibi is Long Beach, California. Long Beach has coastal sand, which is non-local and could match the clue. His "Sunday afternoon" return is also vague.

- **Skye:** Alibi is Cozumel, Mexico. Cozumel is a tropical island with coral sand, a very distinctive type of non-local sand. Her plane arriving at "one this morning" (Monday) is presented as a reason for potential tiredness, but it does **not** alibi her for Sunday morning when the theft occurred. In fact, it implies she **could** have been back on Sunday and committed the theft. Coral sand is highly distinctive and easily tracked if lab tested.

Reasoning: Skye Smith's alibi is the weakest when examined closely against the time of the theft. Her destination, Cozumel, is the most likely source of a truly "non-local" and distinctive sand (coral sand). The detail about the sand not matching local sand is a crucial clue, pointing to someone who visited a location with very different sand. While Jim (Roswell) and Juan (Long Beach) also visited places with non-local sand, coral sand from Cozumel is arguably more uniquely "non-local" to Arizona. Furthermore, Skye's alibi focuses on her late arrival on Monday, subtly drawing attention away from her whereabouts on Sunday morning.

- Answer: (d) Skye Smith

Figure 5: The answer generated by Gemini-2.0-Flash-Thinking under the given prompt ultimately made the wrong choice.

● **For Humans:**

Since the constructed knowledge graph is centered around the options, we will use the graph created by Gemini-2.0-Flash-Thinking as an example to explain how humans should use such a graph.

In Figure 6, We find the highlighted content quite strange. How is it possible that when driving from two different states to the same destination at the afternoon, both would simultaneously face the sun? So we further prompt the Gemini: ***“When Juan drove back to Arizona, he started from Long Beach, and when Jim drove back to Arizona, he started from New Mexico. Can both of them see the sun?”***

The model soon realize its mistake and the answer has been corrected, its analysis is shown in Figure 7.

Clue Graph	
<p>- Clue Graph:</p> <ul style="list-style-type: none"> - “(a) Denise Hurst” <ul style="list-style-type: none"> - (Denise, showed up, 5 minutes late) - (Denise, has, leg in cast) - (Denise, went to, Aspen, Colorado) - (Denise, went for, skiing) - (Denise, tumbled down, mountainside Saturday morning) - (Denise, was, patched up) - (Hospital, contacted, Denise’s parents) - (Denise’s parents, drove, Denise home) - (Denise, released from hospital, Sunday afternoon) - (Denise, said, “I thought it would be a good idea to hit the slopes up in Aspen, Colorado.”) - (Denise, said, “About the time I tumbled down the mountainside Saturday morning, I started to rethink that idea.”) - (Denise, said, “After they had me patched up, they contacted my parents for me. They came and drove me home after I was released from the hospital Sunday afternoon.”) - (Denise, said, “They said that next year I should just take them along so that they could enjoy the skiing part too. That’s sympathy for you.”) - (Denise, said, “Oh, would you two stop sucking up so we can get our assignments?”) - (Denise, said, “I want to go sit down already.”) - (Denise, is usually, sweet girl) - “(b) Jim Gaigon” <ul style="list-style-type: none"> - (Jim, looked up from, floor) - (Jim, been, quiet this morning) - (Jim, yawned) - (Jim, stretched) - (Jim, didn’t sleep, much last night) - (Jim, spent weekend, in Roswell, New Mexico) - (Jim, visiting, International UFO Museum and Research Center) - (Jim, came back, late yesterday afternoon) - (Desert sun, coming west, murder on Jim’s eyes) - (Jim, kept dreaming about, bright spots and aliens) - (Jim, said, “I’m sorry professor, I didn’t – YAWN – sleep much last night.”) - (Jim, said, “I spent the weekend in Roswell, New Mexico, visiting the International UFO Museum and Research Center.”) - (Jim, said, “I came back late yesterday afternoon, but I tell you, that desert sun coming west was murder on my eyes. I kept dreaming about bright spots and aliens all night long. It was creepy.”) - (Jim, said, “I feel much better now.”) 	<ul style="list-style-type: none"> - “(c) Juan Carde” <ul style="list-style-type: none"> - (Juan, snorted at, Skye) - (Juan, went to, Long Beach, California) - (Juan, went for, surf) - (Juan, is, not much of surfer yet) - (Juan, kept falling off, his board) - (People, were calling Juan, “Can’t-Stay-On Juan”) - (Juan, drove back from, Long Beach, California yesterday afternoon) - (Sun, was bad on, Juan’s drive) - (Juan, hates, driving into the sun) - (Juan, said, “Are you kidding? Anything is more relaxing than work.”) - (Juan, said, “But working here is still better than other places, I mean.”) - (Juan, said, “UFO’s, ha! You think the sun was bad on your drive? Coming back from Long Beach, California yesterday afternoon was just as bright, and I hate driving into the sun.”) - (Juan, said, “Oh, but a weekend in the surf was well worth it.”) - (Juan, said, “Well, I’m not much of one yet, I’m still learning. When I say I spent my time in the surf, I mean exactly that. I kept falling off my board. I think the people were calling me ‘Can’t-Stay-On Juan’ by the time I left.”) - “(d) Skye Smith” <ul style="list-style-type: none"> - (Skye, piped up, immediately) - (Skye, flipped back, her long brown hair) - (Skye, flew down to, Cozumel, Mexico) - (Skye, went for, scuba diving) - (Skye, got, tanned) - (Water in Cozumel, is, sparkling) - (Fish in Cozumel, is, beautiful) - (Sun in Cozumel, is, perfect for tanning) - (Skye’s plane, was, delayed) - (Skye’s plane, delayed, again) - (Skye’s plane, got back, one this morning) (Monday morning) - (Skye, slept, on the plane) - (Skye, said, “It was good, but not as nice as being here.”) - (Skye, said, “And well I should,” answered Skye, flipping back her long brown hair. I

Figure 6: Clue graph generated by Gemini-2.0-flash-Thinking. The highlighted parts in red are the ones that humans find contradictory after reading.

Corrected Answer	
<p>## Analysis of Sun Visibility and Answer Change</p> <p>Yes, the analysis of sun visibility does suggest a need to reconsider the answer.</p> <p>**Recap of Sun Visibility Analysis:**</p> <p>* **Jim:** Driving West from Roswell in the afternoon - **Consistent** with seeing the sun in his eyes. * **Juan:** Driving East from Long Beach in the afternoon - **Inconsistent** with seeing the sun in his eyes.</p> <p>**Impact on Suspect Credibility:**</p> <p>* **Jim:** His story gains credibility due to the accurate sun detail. * **Juan:** His story loses credibility due to the inaccurate sun detail. This raises a red flag about the truthfulness of his alibi.</p> <p>**Re-evaluating Suspects:**</p> <p>* **Denise:** Alibi remains strong (Aspen, broken leg). Sand type still less likely. * **Jim:** Alibi strengthened by consistent sun detail. Less likely now. * **Juan:** Alibi weakened by inconsistent sun detail. Suspicion increases significantly. Beach location still relevant for sand clue. * **Skye:** Alibi remains suspicious due to being overly elaborate. Beach location relevant for sand clue. Sun detail doesn't directly impact her.</p> <p>**Shifting Suspicion:**</p> <p>The inconsistent sun detail for Juan is a significant factual error in his alibi. In a detective scenario, such errors are strong indicators of deception. While Skye's alibi was initially suspicious due to its "too perfect" nature, Juan's alibi now has a concrete, verifiable flaw.</p>	<p>**Revised Conclusion and Answer:**</p> <p>Considering the new evidence from the sun visibility analysis, suspicion shifts more strongly towards **Juan Carde**. His beach location aligns with the sand clue, and the geographically incorrect statement about driving into the sun significantly undermines his alibi's credibility. This suggests he might be lying about his weekend activities or at least fabricating details to sound relatable to Jim, which is suspicious behavior in this context.</p> <p>While Skye's elaborate alibi is still questionable, Juan's alibi now has a clear factual flaw, making him a more compelling suspect.</p> <p>**Therefore, based on the comprehensive re-evaluation incorporating the sun visibility analysis, the answer should be changed to:**</p> <p>**Answer: (c) Juan Carde**</p>

Figure 7: Corrected answer after prompted "When Juan drove back to Arizona, he started from Long Beach, and when Jim drove back to Arizona, he started from New Mexico. Can both of them see the sun?" and "So do you need to change the answer?"

As shown in Figure 7, when doubts regarding contradictions in the knowledge graph are expressed, the model can reflect these doubts in its corrected reasoning, ultimately arriving at the correct answer. From this perspective, the knowledge graph enhances the interaction efficiency between humans and large models.

6. ナレッジグラフの構築に利用したプログラム（オプション）

We are developing a more complex clue analysis system to help the model better perceive the overall environment, further improving the quality of the knowledge graph and the model's analytical capabilities, with the goal of fully automating the reasoning process.

7. 資料の共有について

応募フォーム

- 公開の可否：
 - (○) 公開してよい
 - () 非公開とする

応募したナレッジグラフ

- 公開の可否：
 - ☐ 公開してよい
 - ☐ 非公開とする
- 公開形式：
 - ☐ ナレッジグラフ推論チャレンジのサイトで公開
 - ☐ 独自のサイトで公開してリンクを希望

→公開先 URL (※) :

応募したプログラム等

- 公開の可否：
 - ☐ 公開してよい
 - ☐ 非公開とする
- 公開形式：
 - ☐ ナレッジグラフ推論チャレンジのサイトで公開
 - ☐ 独自のサイトで公開してリンクを希望

→公開先 URL (※) :