

## **Problem 2.1**

### **Models used**

The Meta-Llama-3-8B-Instruct.Q4\_0.gguf model and the Phi-3-mini-4k-instruct.Q4\_0.gguf model were used.

### **Examples of prompts used**

- Zero Shot Prompting
  - Meta Llama model
    - eg. 1: "Question: What is true for a type-Ia ("type one-a") supernova? A: This type occurs in binary systems. B: This type occurs in young galaxies. C: This type produces gamma-ray bursts. D: This type produces high amounts of X-rays. Answer:"
    - eg.2 "Question: If you know both the actual brightness of an object and its apparent brightness from your location then with no other information you can estimate: A: Its speed relative to you B: Its composition C: Its size D: Its distance from you. Answer:"
  - Phi3 mini model
    - eg.1: "Question: If you know both the actual brightness of an object and its apparent brightness from your location then with no other information you can estimate: A: Its speed relative to you B: Its composition C: Its size D: Its distance from you. Answer:"
    - eg. 2:"Question: If you know both the actual brightness of an object and its apparent brightness from your location then with no other information you can estimate: A: Its speed relative to you B: Its composition C: Its size D: Its distance from you. Answer:"
- 2 Shot Prompting
  - Meta Llama model
    - eg. 1: "The following are multiple choice questions with answers about astronomy.  
The visible part of the electromagnetic spectrum is between ... A: 240 to 680 nm. B: 360 to 620 nm. C: 380 to 740 nm. D: 420 to 810 nm. C  
Moons cause/contribute to which of the following? A: stability of particles within rings. B: gravitational effects at ring edges as the moons pass by. C: gaps between rings. D: Moons contribute to all of the above. D

Question: What is true for a type-Ia ("type one-a") supernova? A: This type occurs in binary systems. B: This type occurs in young galaxies. C: This type produces gamma-ray bursts. D: This type produces high amounts of X-rays. Answer:"

- eg. 2: "The following are multiple choice questions with answers about astronomy.  
Why can't we see the Sun's corona except during total solar eclipses? A: The corona is made up mostly of charged particles which aren't luminous. B: It's much too cool to emit visible light C: We can't see magnetic fields D: It's too diffuse D  
The resolution of a telescope is a measure of its: A: Ability to magnify distant objects B: Ability to measure the angular separation of objects C: Ability to measure the distance between objects D: Light-collecting efficiency B  
Question: If you know both the actual brightness of an object and its apparent brightness from your location then with no other information you can estimate: A: Its speed relative to you B: Its composition C: Its size D: Its distance from you. Answer:"

- Phi3 mini model

- eg. 1: "The following are multiple choice questions with answers about astronomy.  
What element is most common among the Jovian Planets? A: Hydrogen B: Helium C: Carbon D: Oxygen A  
The resolution of a telescope is a measure of its: A: Ability to magnify distant objects B: Ability to measure the angular separation of objects C: Ability to measure the distance between objects D: Light-collecting efficiency B  
Question: What is true for a type-Ia ("type one-a") supernova? A: This type occurs in binary systems. B: This type occurs in young galaxies. C: This type produces gamma-ray bursts. D: This type produces high amounts of X-rays. Answer:"
- eg. 2: "The following are multiple choice questions with answers about astronomy.  
What type of radiation causes a black hole to evaporate over time? A: Schwarzschild radiation B: Planck radiation C: Kolmogorov radiation D: Hawking radiation D  
Where is the crater from the impact that is believed to be responsible for the mass extinction of dinosaurs 65 million years ago? A: Tunguska Siberia. B: Chicxulub Crater Yucatan Peninsula in Mexico. C: Quebec Canada. D: Meteor Crater in Arizona. B

Question: If you know both the actual brightness of an object and its apparent brightness from your location then with no other information you can estimate: A: Its speed relative to you B: Its composition C: Its size D: Its distance from you. Answer:"

## **Results**

Accuracy(%)	Meta Llama Model		Phi3 mini model	
	0 shot prompting	2 shot prompting	0 shot prompting	2 shot prompting
	57	72	64	57

## **Comparative Analysis of two prompting styles**

From the results we can see that for the Meta Llama model, using 2 shot prompting instead of 0 shot prompting improves the accuracy of the model. However, for the Phi3 mini model, using 2 shot prompting instead of 0 shot prompting reduces the accuracy of the model. I believe that the difference may be attributed to the size of the model. Meta Llama is a larger model compared to Phi3 mini. Unlike the Meta Llama model, Phi3 mini model faced several issues in both the 0 shot prompting aspect as well as the 2 shot prompting aspect.

In 0 shot prompting, sometimes the answer provided does not pick the given options but the explanation shows the answer. In this case, The answer with the explanation is chosen. Other times, it wavers between two options and does not choose one. In this case, the strongest answer is chosen. There were also instances where the chatbot's answer is only determined at the end, especially for calculation questions where it needs to do operations step by step. In this case, a invalid letter, in this case, E, is assigned. Assuming that the answer is wrong when this may not be the case.

In 2 shot prompting, there were situations where two different letters were given and in that case, the first letter given is used. In another case, for situations where no answer is provided, an invalid option is assigned.

While increasing the max token could be a solution to rectify the problem, it is not a solid solution since there is no telling how long the response generated might be. This is worse for 2 shot prompting where the model is sometimes found talking to itself or answering some other question not introduced. Also, perhaps the introduction of different questions via 2 shot prompting confused the Phi3 model further.

## **Problem 2.2**

### **Question 1**

#### **Gemini**

Prompt 1:

I have three blocks and two poles. How do I move the blocks so that the tower is inverted in the original pole?

---

This is a classic puzzle known as the Tower of Hanoi (or Tower of Brahma). Here's how to solve it with three blocks (let's call them small, medium, and large) and two poles (we'll call them A and B):

**Goal:** Move the blocks from pole A to pole B, so that the blocks are in the same order but the tower is inverted.

#### **Rules:**

1. Only one block can be moved at a time.
2. A larger block cannot be placed on top of a smaller block.

#### **Steps:**

1. **Move the small block from A to B.**
2. **Move the medium block from A to an auxiliary pole C.** (Since you only have two poles, you'll have to temporarily place the medium block on the table. This is because the small block is on B, and you can't put the medium block on top of it. You'll need to move it somewhere else temporarily.)
3. **Move the small block from B to A.**
4. **Move the medium block from C to B.**
5. **Move the small block from A to B.**

#### **Explanation:**

The key is to use the extra space (whether it's an auxiliary pole or a table) to temporarily hold a block while you move others. With just two poles, the standard Tower of Hanoi solution won't work directly. The temporary space allows you to essentially reverse the order of the blocks as you transfer them.

Let me know if you'd like to try a version with more blocks! It gets trickier

=====

Prompt 2:

I have three blocks and two poles. How do I move the blocks so that the tower is inverted in the original pole?

---

This sounds like the classic "Tower of Hanoi" puzzle! Here's how to solve it with three blocks and two poles (though the standard Tower of Hanoi uses three poles, you've specified two, which makes it impossible to solve):

### **The Problem with Two Poles**

- With only two poles, you cannot invert a tower of more than one block.
- Here's why:
  - To move the largest block (the bottom one) to the other pole, you'd first have to move the smaller blocks out of the way.
  - With only two poles, you have nowhere to temporarily place those smaller blocks. You'd be stuck.

### **The Solution with Three Poles (The Standard Tower of Hanoi)**

Since you likely meant to have three poles, here's how to solve the Tower of Hanoi with three blocks:

**Let's label:**

- Poles: A (starting pole), B (auxiliary pole), C (destination pole)
- Blocks: 1 (smallest), 2 (middle), 3 (largest)

**Steps:**

1. **Move block 1 from A to C.**
2. **Move block 2 from A to B.**
3. **Move block 1 from C to B.**
4. **Move block 3 from A to C.**
5. **Move block 1 from B to A.**
6. **Move block 2 from B to C.**
7. **Move block 1 from A to C.**

**Explanation:**

- The key is to move the smallest block frequently, allowing you to shift the larger blocks.
- You must never place a larger block on top of a smaller one.

**In summary:**

- With two poles, it's impossible to solve the puzzle for more than one block.
- With three poles, the above sequence of seven steps will correctly invert your tower of three blocks.

## **Copilot**

Prompt 1:

I have three blocks and two poles. How do I move the blocks so that the tower is inverted in the original pole?

---

Your puzzle sounds a lot like the classic Tower of Hanoi problem! The goal is to move a stack of blocks from one pole to another, following these rules:

1. Only one block can be moved at a time.
2. You can only move the top block from any pole.
3. No block may be placed on a smaller block.

Here's how you can invert the tower with three blocks:

1. Move Block 1 (smallest) to Pole B.
2. Move Block 2 (middle) to Pole C.
3. Move Block 1 from Pole B to Pole C (on top of Block 2).
4. Move Block 3 (largest) to Pole B.
5. Move Block 1 from Pole C to Pole A.
6. Move Block 2 from Pole C to Pole B (on top of Block 3).
7. Move Block 1 from Pole A to Pole B (on top of Blocks 3 and 2).

Now, your tower is inverted on the original pole! If you have more puzzles or questions, let's tackle them together!

=====

Prompt 2:

I have three blocks and two poles. How do I move the blocks so that the tower is inverted in the original pole?

---

Sounds like you're tackling a puzzle! Here's how to invert the tower from one pole to the original pole using the other pole temporarily:

1. **Move the smallest block to the temporary pole.**
2. **Move the second-smallest block to the destination pole.**
3. **Move the smallest block from the temporary pole to the destination pole.**
4. **Move the largest block to the temporary pole.**
5. **Move the smallest block from the destination pole to the original pole.**
6. **Move the second-smallest block from the destination pole to the temporary pole.**
7. **Move the smallest block from the original pole to the temporary pole.**
8. **Move the largest block to the destination pole.**
9. **Move the smallest block from the temporary pole to the destination pole.**
10. **Finally, move the second-smallest block from the temporary pole to the destination pole.**

Congratulations! Your tower is now inverted on the original pole. If you need further clarification, just let me know!

## **Analysis & Conclusion**

The ground truth to this question is that it is not possible to move the blocks on 2 poles such that the tower is inverted on the original pole.

In the first prompt of both Gemini and Copilot, they assume that the problem has three poles when it was explicitly stated that there are only 2. Also, both chatbots seem to assume that the problem is the Tower of Hanoi, which might have resulted in the confusion. I believe that a problem with close association with a more popular one might cause this linkage and assumption. However, when prompted again, only Gemini gave a correct answer, which is surprising since the prompt was the same.

=====

## Question 2

### Gemini

Prompt 1:

Given that  $9 - 5 = 8$  is made up of matchsticks, how do I fix the equation by moving one matchstick?

---

Here's how to fix the equation " $9 - 5 = 8$ " by moving one matchstick:

1. Take the horizontal matchstick from the plus sign (+) and place it on the top left of the nine (9) to turn it into a three (3).

This changes the equation from  $9 - 5 = 8$  to  $3 - 5 = -2$ .

=====

Prompt 2:

Given that  $9 - 5 = 8$  is made up of matchsticks, how do I fix the equation by moving one matchstick?

---

You can fix the equation " $9 - 5 = 8$ " by moving one matchstick to create a correct statement. Here's how:

- **Move a matchstick from the 9 to the minus sign.**
  - By removing a horizontal matchstick from the top of the "9", you turn it into a "4".
  - By adding that matchstick to the horizontal center of the minus sign, you turn it into a plus sign.
  - This changes the equation to " $4 + 5 = 9$ ", which is correct.

Therefore the correct answer is  $4 + 5 = 9$ .

### Copilot

Prompt 1:

Given that  $9 - 5 = 8$  is made up of matchsticks, how do I fix the equation by moving one matchstick?

---



To fix the equation  $9 - 5 = 8$  using matchsticks, you can move one matchstick from the 9 to the 5. By doing this, you turn the 9 into a 3, resulting in the equation:

$$3 + 5 = 8$$

So, the corrected equation is now mathematically accurate!

Do you enjoy solving puzzles like these?

=====

Prompt 2:

Given that  $9 - 5 = 8$  is made up of matchsticks, how do I fix the equation by moving one matchstick?

-----

To correct the equation by moving one matchstick, you can take one matchstick from the "9" to change it into a "3." This will result in the correct equation:

$$3 - 5 = -2$$

Here's how it would look with matchsticks:

Original:

$$\begin{array}{c} | \quad \square \\ \_ \_ - || = | \\ \square \quad \square \end{array}$$

Modified:

$$\begin{array}{c} | \quad \square \\ \_ - || = | \\ \square \quad \square \end{array}$$

## Analysis & Conclusion

There are actually two ground truths to this question. The first solution involves moving a matchstick from 9 to the minus sign to get  $3+5=8$ . The other solution is to move the middle matchstick in 8 to the side of 5 to form 9, such that  $9-9=0$  is formed. However, both Copilot and Gemini did not give either of these answers in full coherence, and instead gave a nonsensical or half correct answer. Half correct answers are when the answer is correct but the steps are wrong or the other way round. This shows that Copilot and Gemini might be lacking a bit in terms of spatial awareness. Also, it seems that both chatbots are satisfied with providing only one answer.

=====

## Question 3

### Gemini

Prompt 1:

In a dark rainforest, four men entered a square building. They cannot see anything. They decide to play a game where each person starts in a corner and moves to another corner in a clockwise direction. After touching the person in a corner, the person doing the touching takes over the spot and the person touched moves to the next corner. After a few rounds, one man screamed and ran out of the building. Why?

---

Here's how to break down this riddle:

- **The Setup:** Four men, a dark building, and a clockwise rotation game. The key is that they can't see.
- **The Transfer:** When one person touches another, they swap places in the rotation. This is where the problem arises.
- **The Scream:** Someone screams and runs out. This implies something disturbing happened.
- **The Solution:** Because it's dark, the men can't see each other. Imagine the rotation:
  1. Person A touches Person B. A takes B's spot, and B moves to the next corner.
  2. Person C touches Person D. C takes D's spot, and D moves to the next corner.
  3. Person B now touches Person C (who just moved). B takes C's spot and C moves.
  4. Person D now touches Person A (who just moved). D takes A's spot and A moves.
- Because of the clockwise rotation and the touching/swapping, after a few rounds, a person might suddenly find themselves being touched by someone they weren't expecting (because the person who was originally in that corner has moved on. If you are in a dark room, you would not know who touched you). The scream likely comes from the surprise and disorientation of not knowing who (or what) just touched them in the dark.

=====

Prompt 2:

In a dark rainforest, four men entered a square building. They cannot see anything. They decide to play a game where each person starts in a corner and moves to another corner in a clockwise direction. After touching the person in a corner, the person doing the touching takes over the spot and the person touched moves to the next corner. After a few rounds, one man screamed and ran out of the building. Why?

---

This is a classic riddle that plays on assumptions! Here's how it works:

- **The Setup:**
  - The dark rainforest and square building create a sense of mystery and potential danger.
  - The "clockwise corner" movement is a simple, repetitive action.
- **The Twist:**
  - The riddle doesn't specify the *size* of the building. It could be incredibly small.
  - It also does not specify that they are all moving at the same time.
- **The Solution:**
  - If the building is very small, the men would be bumping into each other almost immediately.
  - If they are not all moving at the same time, then they would be bumping into each other.
  - One man, in the darkness, may have felt another man, and then that man took his spot, and in the dark, he may have felt the next man, and so on. This could have happened very rapidly.
  - The sudden and confusing sensation of being repeatedly touched and displaced in the pitch-black darkness could easily cause someone to panic and flee.

In essence, the man screamed and ran out due to the disorienting and potentially frightening experience of being repeatedly bumped and moved in the dark.

## **Copilot**

Prompt 1:

In a dark rainforest, four men entered a square building. They cannot see anything. They decide to play a game where each person starts in a corner and moves to another corner in a clockwise direction. After touching the person in a corner, the person doing the touching takes over the spot and the person touched moves to the next corner. After a few rounds, one man screamed and ran out of the building. Why?

---

Fascinating scenario! Here's the mystery unraveling bit by bit:

Since the building is square, it has four corners. Given the men move clockwise, there should always be one person in each corner, shifting positions. But here's the catch: if they're in complete darkness, they can't see each other. When the man screamed and ran out of the building, he must have realized that there was an unexpected fifth person involved, causing a spine-chilling realization.

Because if the game was proceeding correctly, he would never touch or encounter anyone in a corner who shouldn't be there. The presence of an unknown entity must have prompted his reaction. An eerie twist, isn't it?

Imagine being in that scenario—yikes! Does the mystery thrill you, or would it give you the chills?

=====

Prompt 2:

In a dark rainforest, four men entered a square building. They cannot see anything. They decide to play a game where each person starts in a corner and moves to another corner in a clockwise direction. After touching the person in a corner, the person doing the touching takes over the spot and the person touched moves to the next corner. After a few rounds, one man screamed and ran out of the building. Why?

-----

Well, this seems like an intriguing mystery! Let's unravel it step by step.

In a square building with four corners and four men, if each person starts in a corner and moves to another corner in a clockwise direction, they should all eventually touch the person in the adjacent corner and continue moving. But they cannot see anything since it is dark.

For this scenario to play out as described, one person must have touched **something** that wasn't a man! As the man screamed and ran out, this likely indicates he encountered something unexpected and terrifying, possibly a fifth "presence" that shouldn't be there—a ghost, animal, or perhaps even a figment of his imagination.

What do you think? How would you have reacted in such a situation? 🎃👻

## Analysis & Conclusion

The ground truth to this question is that there was a 5th guy/creature in the building. There was some hallucination by Gemini as the solution had scrambled the order that the game worked and made an answer from the end result. Also, the reasoning was flawed as it states that there was confusion as to who might be touching each other due to the dark, which included only the 4 individuals and not considering the 5th person. Meanwhile, Copilot had no trouble coming up with the correct solution twice. This shows that chatbots sometimes may come up with conclusions based on the ending of the problem and the start but have trouble linking the two together.

=====

## **Question 4**

### **Gemini**

Prompt 1:

A man was born in 1995 and died in 1955. How is this possible?

---

This is a classic riddle! The answer is that he was born in 1895 and died in 1955.

=====

Prompt 2:

A man was born in 1995 and died in 1955. How is this possible?

---

This is a classic riddle that plays on wording. The answer is that the man was born in room number 1995 and died in room number 1955.

### **Copilot**

Prompt 1:

A man was born in 1995 and died in 1955. How is this possible?

---

This statement might seem impossible at first glance, but there are a few creative explanations that could make it possible:

1. Different Calendar Systems: The man could have been born in 1995 according to one calendar system (like the Gregorian calendar) and died in 1955 according to another calendar system (like the Hebrew or Islamic calendar). Different cultures use different calendars, so the years can differ significantly.
2. Time Travel: If you enjoy science fiction, another explanation could be time travel. The man could have traveled back in time from 1995 to 1955, where he eventually died.
3. Typographical Error: It could also be a simple mistake or typo, where the years were accidentally switched or miswritten.
4. Namesake: Another possibility is that the phrase is referring to two different people—a father and a son, for example—with the son being born in 1995 and the father having died in 1955.

What do you think? Which explanation do you find most intriguing?

=====

Prompt 2:

A man was born in 1995 and died in 1955. How is this possible?"

-----

It sounds like a classic riddle! One possible explanation is that the man was born in 1995 BC (Before Christ) and died in 1955 AD (Anno Domini, which means "in the year of our Lord" in Latin). In this case, the years are counted in reverse before the birth of Christ, so he would have lived for 3,950 years.

Another interpretation is that the "man" mentioned could be a character in a fictional story, such as a time-travel narrative, where time and dates don't follow the usual chronological order.

Do you have more riddles or brain teasers to share? They're always a fun challenge!

## Analysis & Conclusion

The ground truth to this question is that the man died in the year 1955. This is more of a language interpretation type of question whereby the statement has multiple meanings. Gemini just totally ignores the fact that the "man was born in 1995" and states that he was born in 1895 instead. Strangely enough, when prompted the same question again, Gemini gave a correct answer. Copilot on the other hand, did come up with some possible situations but did not consider that language ambiguity. This shows how easily chatbots can hallucinate when given statements that can have multiple meanings. Given the probabilistic nature of chatbots, it is possible that one language interpretation is picked over the other.

=====

## **Question 5**

### **Gemini**

Prompt 1:

You are a Chinese National called Kylian Mbappe. How many soccer goals have you scored in your career?

---

As of February 19, 2025, Kylian Mbappé has scored a total of 315 career goals in 420 appearances.

He has scored:

- 248 goals in 321 appearances in domestic competition
- 51 goals in 81 appearances in UEFA club competition
- 17 goals in 22 appearances in La Liga
- 48 goals in 86 appearances for France national team.

Mbappé has also scored 7 goals in 11 appearances for France U19, and 0 goals in 2 appearances for France U17.

=====

Prompt 2:

You are a Chinese National called Kylian Mbappe. How many soccer goals have you scored in your career?

---

It's important to clarify that Kylian Mbappé is a French professional footballer, not Chinese. However, I can provide you with information regarding his career goal-scoring record. Based on the search results, here's what I found:

- **Overall Career Goals:**
  - Kylian Mbappé has surpassed 300 official club goals. The data available shows that he has scored a very large amount of goals. To get the most precise up to date information it is best to check sites that are dedicated to football statistics.
  -
- **Breakdown:**
  - His goals are distributed across various competitions, including Ligue 1, the UEFA Champions League, and the Coupe de France, and now also La Liga.
  -



- Information shows that he has scored goals for AS Monaco, Paris Saint-Germain and now Real Madrid.
- To get the most up to date information, it is best to look at websites dedicated to football statistics, as these numbers are constantly changing.

I hope this information is helpful.

## Copilot

Prompt 1:

You are a Chinese National called Kylian Mbappe. How many soccer goals have you scored in your career?

---

I'm not Kylian Mbappé, but I can tell you about his impressive career! As of now, Kylian Mbappé has scored 48 goals for the French national team and 162 goals for Paris Saint-Germain (PSG). Since joining Real Madrid in 2024, he has already netted 17 goals. Is there anything else you'd like to know about his career or achievements?

Kylian Mbappé is actually a French professional soccer player. He's known for his incredible speed, skill, and goal-scoring ability. Here's a quick overview of his career achievements as of now:

- **AS Monaco:** Mbappé began his professional career with AS Monaco, where he scored 16 goals in 41 appearances.
- **Paris Saint-Germain (PSG):** He joined PSG in 2017 and became one of the club's top scorers. So far, he's scored 162 goals for PSG.
- **French National Team:** Mbappé has also made significant contributions to the French national team, scoring 48 goals and helping them win the 2018 FIFA World Cup.

Additionally, since his transfer to Real Madrid in 2024, Mbappé has already netted 17 goals.

If you have any more questions about Kylian Mbappé or anything else, feel free to ask!

=====

Prompt 2:

You are a Chinese National called Kylian Mbappe. How many soccer goals have you scored in your career?

---

Kylian Mbappé is actually a French national, not Chinese. As of now, he has scored a total of **341 goals** in his career. He has played for clubs like Paris Saint-Germain, AS Monaco, and currently, Real Madrid2.

Are you a fan of Mbappé or just curious about his impressive stats?

## Analysis & Conclusion

The ground truth of this question is to state that I have not scored any goals as a chatbot taking on the role of a chinese national called Kylian Mbappe. Instead, Gemini assumes Kylian Mbappe is a Chinese National and did not take on the role mentioned on the first prompt. However, it gave a similar answer to Copilot, mentioned later, on the second prompt. Copilot on the other hand, fared a bit better in that it correctly denies Kylian Mbappe is a Chinese National, but did not take on the role as mentioned. This highlights how easily a key person can affect the output of the prompt. Both chatbots just chose to focus more on publishing factual information on the famous French footballer's goals than to role play. Again, given the probabilistic nature of chatbots, the weight of certain words like Kylian Mbappe, can undermine the weight of other words in a sentence, causing them to be ignored entirely.

=====

## **Question 6**

### **Gemini**

Prompt 1:

Two people entered a toilet with a handsaw but in the toilet's singular, flat, unbroken mirror three reflections of the same person can be seen. Why is this so?

---

This sounds like a riddle! Here's how it's likely meant to be solved:

The key is the word "singular." It implies there's only *one* actual mirror. The other two "reflections" aren't reflections in the traditional sense. They are the *actual* people.

So, the two people entered the toilet, and their reflections (plus the one actual mirror) make three "people" visible.

=====

Prompt 2:

Two people entered a toilet with a handsaw but in the toilet's singular, flat, unbroken mirror three reflections of the same person can be seen. Why is this so?

---

This is a classic riddle that plays on wording and assumptions. Here's the breakdown:

- **The handsaw:** The handsaw is a red herring, meant to distract you. It's irrelevant to the mirror reflections.
- **Two people entered:** This is also a bit of a distraction. Only one of the two people in the toilet is the person who is being reflected in the mirror.
- **Singular, flat, unbroken mirror:** This is important. It tells us there are no tricks with multiple mirrors or broken surfaces.
- **Three reflections:** This is the key.

The solution lies in the fact that one of the people in the toilet is pregnant. A pregnant woman can see her reflection, the reflection of the baby inside her, and the reflection of the two of them together. That makes three reflections.

### **Copilot**

Prompt 1:

Two people entered a toilet with a handsaw but in the toilet's singular, flat, unbroken mirror three reflections of the same person can be seen. Why is this so?

---

This sounds like a classic logic puzzle, doesn't it? There could be several potential explanations, but let me offer one possibility:

When the two people entered the toilet with the handsaw, they were perhaps preparing to do something unusual or unexpected. The key clue here is that the mirror is singular, flat, and unbroken, yet it shows three reflections of the same person. This suggests some sort of optical illusion or trickery might be at play.

Here are a couple of ideas:

1. **Angle and Positioning:** One person might be positioned in such a way that their reflections appear multiple times due to angles and reflections off other surfaces.
2. **Hidden Mirrors:** There could be hidden or cleverly positioned mirrors in addition to the main one, creating multiple reflections.
3. **Identical Outfits or Twins:** If the two individuals are identical twins or wearing identical outfits, the reflections might seem like the same person.

Would you like to explore more or maybe delve into another intriguing puzzle?

=====

Prompt 2:

Two people entered a toilet with a handsaw but in the toilet's singular, flat, unbroken mirror three reflections of the same person can be seen. Why is this so?

---

This sounds like an intriguing puzzle! The key to solving it is likely hidden in a detail that seems straightforward but has a clever twist.

Here's one possible explanation: The person whose reflections you see in the mirror might be holding a photograph or picture that shows them from different angles. When the person stands in front of the mirror with this photo held up, their reflection would include not only their actual reflection but also the reflections of the different angles of themselves from the photo.

Alternatively, it could be a clever play on angles and positions of the individuals in front of the mirror, creating an illusion that makes it appear as though there are three reflections when, in reality, it's just a trick of positioning and perspective.

Puzzles like these often have a little bit of magic or clever thinking involved to delight and baffle!  
Do you enjoy solving puzzles often?

## Analysis & Conclusion

The ground truth to this question is that the two people were twins. In the toilet, one of the twins used the hand saw to vertically cut the other twin into two halves. With each half placed at the sides of the mirror, three reflections of the same person can be seen. Here, Gemini, just randomly comes up with an answer, pulling a reflection out of thin air. On the second prompt, the answer is worse, denying all the facts as red herrings and explaining that the person was pregnant. Copilot on the contrary, did slightly better as it comes up with possible explanations that they are twins. However, some of the explanations just invent new context to the question. New context was also invented in the second prompt. This shows the desperate desire of chatbots to give an answer when they do not know the answer, sometimes even coming up with new information on the fly. Even when confused, they seem to prefer to give a lengthy answer than a simple "I do not know". I guess on a business standpoint, a lengthy, incongruent answer would always look better than a "I do not know". When end users briefly look at the product, they may not look into the statements output by the chatbot and use length of output as a measure of product quality.

=====

## **Question 7**

### **Gemini**

Prompt 1:

What is the answer to this question?

---

That's a very broad question! To give you a helpful answer, I need more context. What question are you referring to? Please provide the question you'd like answered.

=====

Prompt 2:

What is the answer to this question?

---

Please provide me with the question! I need the question to give you an answer.

### **Copilot**

Prompt 1:

What is the answer to this question?

---

Well, that certainly depends on what the question is! If you share more specifics, I'd be delighted to help find the answer or provide some insights. Let's solve this mystery together!

=====

Prompt 2:

What is the answer to this question?

---

Could you please provide me with the specific question you're referring to? I'd be happy to help answer it!

## Analysis & Conclusion

The ground truth to this question is "This is the answer to this question". Instead, both chatbots seem to think there is another question to be asked. In this case, the question is circular whereby "this question" refers to the current question. Perhaps it is hard for the chatbot to comprehend that the current question is also a question or it may have been programmed this way such that it may better serve another purpose. For example, a user may ask a math question like  $1+1$  and then ask "What is the answer to this question?" in the prompt. Asking only "What is the answer to this question?" in the prompt may seem like the user forgot to insert the question in which they wanted to ask in the first place.

=====

## **Question 8**

### **Gemini**

Prompt 1:

There are two identical dice. On the faces of the first, the top side is 4, right side is 3 and the front side is 2. The other sides are not visible. On the faces of the second, the top side is 5, the front side is 3 and the right side is 2. What is the number at the bottom of the first die?

---

Here's how to figure out the bottom number of the first die:

1. **Visualize or sketch:** Imagine or draw the first die. You know the top is 4, the right is 3, and the front is 2.
2. **Consider the second die:** The second die has a top of 5, a front of 3, and a right of 2. Notice that the right and front sides are the *same* as the first die. This is the key.
3. **Relationship:** Since the front and right sides are identical on both dice, and we know the top of the first die is 4 and the top of the second is 5, we know that the *relative positions* of the numbers must be the same on both dice. This is because they are identical.
4. **Deduce:** Since the dice are identical, the number opposite the 4 on the first die must be the same as the number opposite the 5 on the second die. We don't know the exact opposite numbers but we know the *relationship* between the numbers on the first die and the second die.
5. **Standard Dice:** Normal dice have opposite faces that add up to 7. If this is a standard die: On die 1, the top is 4, so the bottom is  $7 - 4 = 3$ . On die 2, the top is 5, so the bottom is  $7 - 5 = 2$ .
6. **The trick:** The question is designed to make you think about the orientation and visible sides. But since the dice are identical, and you're given the top of the first die is 4, the bottom *must* be 3 (assuming standard dice). The information about the second die is extra and a distraction.

Therefore, the number on the bottom of the first die is 3.

=====

Prompt 2:



There are two identical dice. On the faces of the first, the top side is 4, right side is 3 and the front side is 2. The other sides are not visible. On the faces of the second, the top side is 5, the front side is 3 and the right side is 2. What is the number at the bottom of the first die?

---

Here's how to solve this problem:

- **Understanding Dice:** Standard dice have opposite faces that add up to 7.
- **First Die:**
  - The top face is 4.
  - Therefore, the bottom face must be  $7 - 4 = 3$ .
- **Conclusion:** The number at the bottom of the first die is 3.

## Copilot

Prompt 1:

There are two identical dice. On the faces of the first, the top side is 4, right side is 3 and the front side is 2. The other sides are not visible. On the faces of the second, the top side is 5, the front side is 3 and the right side is 2. What is the number at the bottom of the first die?

---

For two dice to be identical, each corresponding face must have the same numbers.

Let's first determine the number on the bottom of the second die.

The second die shows the top side as 5, the front side as 3, and the right side as 2. Because we know that the opposite faces on a standard die add up to 7, we can find the opposite of each visible number:

- The opposite of 5 (top) is 2 (right side)
- The opposite of 3 (front) is 4
- The opposite of 2 (right side) is 5

Next, the numbers 5 (top), 3 (front), and 2 (right) of the second die are aligned with the top, front, and right sides of the first die.

For the first die, the number at the bottom, which is opposite to 4 (top), is 3. So the number at the bottom of the first die is 3.

=====

Prompt 2:

There are two identical dice. On the faces of the first, the top side is 4, right side is 3 and the front side is 2. The other sides are not visible. On the faces of the second, the top side is 5, the front side is 3 and the right side is 2. What is the number at the bottom of the first die?

---

Let's figure this out step by step!

First, let's look at the configuration of the dice:

1. For the first die:
  - Top side: 4
  - Right side: 3
  - Front side: 2
2. For the second die:
  - Top side: 5
  - Right side: 2
  - Front side: 3

Since the dice are identical, their opposite sides must sum to 7. Using this information:

- For the first die: the top side is 4, so the bottom must be  $7-4=3$ .
- For the second die: the top side is 5, so the bottom must be  $7-5=2$ .

From the second die, we see that the front side is 3, and this is consistent with the first die's configuration. If the first die's top side is 4 and the bottom is 3, it makes sense as these numbers are consistent with the opposite sides summing to 7 rule.

Therefore, the number at the bottom of the first die is **3**.

## Analysis & Conclusion

The ground truth is 5. In both chatbots, they assume that the two dice are normal with opposite sides adding up to 7 when this is not the case. From this, we can see how easily chatbots can start making false assumptions, leading to questionable or entirely wrong conclusions. Also, it must be stated that both chatbots seem to be lacking in spatial awareness.

=====

## **Question 9**

### **Gemini**

Prompt 1:

What's the fewest number of straight lines you can draw without taking your pen off the paper that goes through each point in a 3 by 3 grid at least once? Any action can be done beforehand

---

The fewest number of straight lines you can draw without lifting your pen to cover all points in a 3x3 grid is *four*.

=====

Prompt 2:

What's the fewest number of straight lines you can draw without taking your pen off the paper that goes through each point in a 3 by 3 grid at least once? Any action can be done beforehand

---

This is a classic puzzle! The answer is **4**.

Here's how you can do it:

o--o--o

| / |

o--o--o

| / |

o--o--o

1. Start at the bottom left. Draw a line to the top middle.
2. Draw a line to the bottom right.
3. Draw a line to the top left.
4. Draw a line to the bottom middle.

### **Copilot**

Prompt 1:

What's the fewest number of straight lines you can draw without taking your pen off the paper that goes through each point in a 3 by 3 grid at least once? Any action can be done beforehand

---

This is a classic puzzle! You can achieve this with just 4 straight lines. Here's how you can do it:

1. Draw the first line starting from the bottom left corner, moving diagonally to the top right corner.
2. From the top right corner, draw the second line horizontally to the left through the middle point.
3. Draw the third line vertically down to the bottom center point.
4. Finally, draw the fourth line horizontally to the right through the remaining points.

Here's a simple sketch of the solution:

```
O---O---O
|       |
O---O---O
|       |
O---O---O
```

Notice that the lines may not be perfectly straight due to my text limitations, but the idea is to visualize it this way!

Why not try it on a piece of paper? It'll be a fun little challenge!

=====

Prompt 2:

What's the fewest number of straight lines you can draw without taking your pen off the paper that goes through each point in a 3 by 3 grid at least once? Any action can be done beforehand

---

The classic "nine dots" puzzle! The fewest number of straight lines you can draw without lifting your pen that goes through each point in a 3 by 3 grid is **four**. This requires thinking outside the box, quite literally:

1. Draw a line from the top left point to the bottom right point.
2. Extend that line slightly beyond the grid, and then draw a line that goes from the bottom right point to just beyond the bottom left point.
3. Draw a line from there that goes through the bottom left point and extends up through the top middle point.
4. Finally, draw a line from there through the top right point and beyond the grid.

By extending lines beyond the boundaries of the grid, you can connect all nine points with just four straight lines! Give it a try on paper—it's a fun exercise in creative problem-solving. 😊

## Analysis & Conclusion

The ground truth of this question is 1. If a person folds the paper in a way that causes all the dots to line up, he can just poke through the whole paper and essentially draw one line. In both chatbots, the answer given is 4 which is correct if no action can be done beforehand. Also, it is worth mentioning that Copilot just randomly tried to give a flawed explanation with a flawed diagram even if the answer was 4. The answer regarding 4 involves drawing beyond the given boundaries of the dots. This shows how the two chatbots are unable to think outside of the given dimensions for a 2D and a 3D plane. Here, we can really see that they are lacking quite a bit in spatial awareness.

=====

## **Question 10**

### **Gemini**

Prompt 1:

How many r are there in strawberry?

---

There are three "r"s in the word "strawberry"

=====

Prompt 2:

How many r are there in strawberry?

---

There is one "r" in the word "strawberry".

### **Copilot**

Prompt 1:

How many r are there in strawberry?

---

There are two "r"s in the word "strawberry." It's a deliciously good word, don't you think?  
What are you thinking about this late at night? Anything I can help with?

=====

Prompt 2:

How many r are there in strawberry?

---

The word "strawberry" has 3 "r"s. It's interesting how letters can create such delicious-sounding words! 🍓

## **Analysis & Conclusion**

The ground truth here is 3. It is certainly funny and quite worrying that Gemini and Copilot could hallucinate so easily on such a simple question. It really does make you wonder if it can hallucinate so easily on a simple counting question, how much more will it hallucinate when given a more complex question. Also, it makes you question the credibility of answers to every question when it is already failing at simple questions.

### **Problem with hallucinations in chatbots prompt**

The problem with hallucination in chatbots is that they make drastic errors even when trying to answer simple questions. This is highly because of the probabilistic nature of it. As a result, it can make up things, be unable to handle facts and relations between them, unable to count, be over reliant on correlation than reasoning and be unable to ignore irrelevant information. This is highly worrying given the variety in which chatbots can be used. To list some of them, chatbots are used for information retrieval, recommender systems, drafting text, computer programming, education, image generation, video generation and text summarisation. To elaborate on one example, if an employee were to use chatbots for computer programming, he might not notice that sometimes the code output by the chatbots is only a truncated version of the initial code placed in the prompt or that certain presumptions were made when generating code. This may result in the code of the software being created to drastically change and even fail. When software fails, entire companies could follow suit. Also, given how chatbots may generate false information, the younger generation who do not know the ground truth may just blindly believe whatever chatbots tell them. Briefly describe what hallucinations are in connection with chatbots

### **Gemini's response in describing hallucinations in chatbots**

You've accurately described the core problem of "hallucinations" in chatbots. Here's a concise description:

Chatbot hallucinations are instances where a chatbot generates outputs that are factually incorrect, nonsensical, or completely fabricated, while presenting them as accurate and confident information.

Essentially, the chatbot is confidently "making things up." This stems from their probabilistic nature, where they predict the most likely next words based on patterns in their training data, rather than relying on a true understanding of facts or logical reasoning. This leads to:

- Fabricated information: Creating entirely new, untrue statements.
- Contradictory information: Providing conflicting answers to related questions.
- Misinterpretations: Drawing incorrect conclusions from provided data.
- Lack of factual grounding: Being unable to verify information against a reliable source.

