

Appendix for “Unbiased Reasoning for Knowledge-Intensive Tasks in Large Language Models via Conditional Front-Door Adjustment”

1 Case Study

1.1 Case on HotpotQA

Intermediate output details of a HotpotQA example.

External Knowledge: Tropical house: Tropical house, also known as trop house, is a subgenre of deep house, with elements of dancehall and Balearic house. Artists of the genre are often featured at various summer festivals such as Tomorrowland. The genre was popularized by artists including Thomas Jack, Kygo, Matoma, Lost Frequencies, Seeb and Klingande. Snoopstar: Snoop Star was a metasearch engine for filesharing networks, with which one could find music and video files in different file sharing networks and download simultaneously.

Question: At what summer festival might you find music by Bakermat?

(The correct answer: Tomorrowland)

Outputs from CoT-SC:

Answer voting results: Weekend Festival

Cluster Results:

CoT 1

1. Bakermat is a Dutch DJ and music producer known for his music that includes electronic, deep house, and tropical house influences.
2. Tropical house is a subgenre of deep house and is often featured at summer festivals.
3. The context mentions that artists of the tropical house genre are commonly featured at various summer festivals, including Tomorrowland.
4. Since Bakermat’s music fits within the tropical house genre, it is likely that he would perform at festivals that celebrate this genre.
5. Tomorrowland is specifically noted as a festival where tropical house artists are featured.

Final answer: [Tomorrowland]

CoT 2

1. Bakermat is a Dutch DJ and music producer known for his music that blends electronic, deep and tropical house, and techno with jazz and soul influences.
2. Tropical house, a subgenre of deep house, is often featured at various summer festivals.

3. The context mentions that artists from the tropical house genre, including Bakermat, perform at summer festivals like Tomorrowland.
 4. Therefore, Bakermat's music would likely be found at Tomorrowland, a prominent summer festival.
- Final answer: [**Tomorrowland**]

CoT 3

To determine at which summer festival you might find music by Bakermat, we can analyze the provided context step by step:

1. Identify Bakermat's Genre: Bakermat is known for his music that consists of a mix of electronic, deep and tropical house, and techno music, influenced by jazz and soul.
 2. Look for Relevant Festivals: We should find festivals that feature electronic or tropical house music.
 3. Review Festival Information: Tokyo Summer Festival: Primarily a music festival featuring various styles, but not specifically electronic or tropical house. Weekend Festival: This festival is specifically mentioned to have an artist lineup that mainly consists of EDM artists, which aligns with Bakermat's genre. All That! Music and More Festival: It was primarily for kids and featured musical guests from the show, not specifically aligned with Bakermat's genre. Dubrovnik Summer Festival: Focuses on classical performances and might not fit Bakermat's electronic style. Van Buren Popcorn Festival and Saskatchewan Festival of Words: These are not music festivals focused on electronic music.
 4. Conclusion: Based on the information provided, the **Weekend Festival** is the most likely place where Bakermat would perform, given its focus on electronic music and large attendance of music fans.
- Final answer: [**Weekend Festival**]

CoT 4

To determine at which summer festival you might find music by Bakermat, we can follow these steps:

1. Identify Bakermat's genre: Bakermat is known for his music that consists of a mix of electronic, deep and tropical house, and techno music.
2. Consider the festivals mentioned: We need to look for summer festivals that feature electronic or tropical house music.
3. Evaluate the festivals: The Van Buren Popcorn Festival does not focus on music genres but rather celebrates popcorn. -The Saskatchewan Festival of Words promotes literacy and does not focus on music. The Tokyo Summer Festival features various styles of music but is not specifically focused on Bakermat's genres. The Dubrovnik Summer Festival is more oriented towards classical and cultural performances. The All That! Music and More Festival was specifically for kids and featured various musical guests, but it's not a prominent festival for electronic music. The Weekend Festival in Helsinki is known for its lineup of EDM artists and has included various genres, making it a suitable venue for Bakermat's music.
4. Conclusion: The most likely summer festival where you might find music by Bakermat is the Weekend Festival.

Final answer: [**Weekend Festival**]

CoT 5

1. Identify the genre of music Bakermat produces: Bakermat is known for electronic, deep house, tropical house, and techno music with jazz and soul influences.
2. Determine which festivals are associated with this genre: Tropical house artists, including Bakermat, are often featured at summer festivals like Tomorrowland, which is known for electronic dance music.
3. Review the provided context for relevant festivals: The context mentions several summer festivals, including the Weekend Festival, which features a lineup mainly consisting of EDM artists and has hosted

various internationally acclaimed artists.

4. Conclude which festival aligns with Bakermat's music style: Since Bakermat's music is closely related to the genres featured at Weekend Festival, it is reasonable to conclude that he might perform there.

Final answer: [**Weekend Festival**]

Constructing Counterfactual External Knowledge:

The list of the top five entities most relevant to the question:

[Tropical house, Weekend Festival, Bakermat, EDM artists, Helsinki]

The list of counterfactual entities generated by the LLM:

[Tropical music, Summer Festival, Kygo, Electronic music artists, Finland]

Based on this, we can construct five counterfactual external knowledge according to the method in the text and obtain their corresponding probabilities.

$$P(e_1^*) = 0.258$$

$$P(e_2^*) = 0.221$$

$$P(e_3^*) = 0.194$$

$$P(e_4^*) = 0.172$$

$$P(e_5^*) = 0.155$$

Intervention Results:

CoT 1

$$P(c_1 | e_1^*, Q) = 5/10 = 0.5$$

$$P(c_1 | e_2^*, Q) = 6/10 = 0.6$$

$$P(c_1 | e_3^*, Q) = 8/10 = 0.8$$

$$P(c_1 | e_4^*, Q) = 7/10 = 0.7$$

$$P(c_1 | e_5^*, Q) = 6/10 = 0.6$$

$$P(A = \text{Tomorrowland} | c_1, e_1^*, Q) = 3/5 = 0.6$$

$$P(A = \text{Tomorrowland} | c_1, e_2^*, Q) = 3/6 = 0.5$$

$$P(A = \text{Tomorrowland} | c_1, e_3^*, Q) = 5/8 = 0.625$$

$$P(A = \text{Tomorrowland} | c_1, e_4^*, Q) = 4/7 = 0.57$$

$$P(A = \text{Tomorrowland} | c_1, e_5^*, Q) = 2/6 = 0.33$$

CoT 2

$$P(c_2 | e_1^*, Q) = 7/10 = 0.7$$

$$P(c_2 | e_2^*, Q) = 3/10 = 0.3$$

$$P(c_2 | e_3^*, Q) = 4/10 = 0.4$$

$$P(c_2 | e_4^*, Q) = 4/10 = 0.4$$

$$P(c_2 | e_5^*, Q) = 8/10 = 0.8$$

$$\begin{aligned}
P(A = \text{Tomorrowland} \mid c_2, e_1^*, Q) &= 5/7 = 0.71 \\
P(A = \text{Tomorrowland} \mid c_2, e_2^*, Q) &= 2/3 = 0.67 \\
P(A = \text{Tomorrowland} \mid c_2, e_3^*, Q) &= 3/4 = 0.75 \\
P(A = \text{Tomorrowland} \mid c_2, e_4^*, Q) &= 0/4 = 0 \\
P(A = \text{Tomorrowland} \mid c_2, e_5^*, Q) &= 6/8 = 0.75
\end{aligned}$$

CoT 3

$$\begin{aligned}
P(c_3 \mid e_1^*, Q) &= 7/10 = 0.7 \\
P(c_3 \mid e_2^*, Q) &= 3/10 = 0.3 \\
P(c_3 \mid e_3^*, Q) &= 8/10 = 0.8 \\
P(c_3 \mid e_4^*, Q) &= 6/10 = 0.6 \\
P(c_3 \mid e_5^*, Q) &= 4/10 = 0.4
\end{aligned}$$

$$\begin{aligned}
P(A = \text{Weekend Festival} \mid c_3, e_1^*, Q) &= 3/7 = 0.43 \\
P(A = \text{Weekend Festival} \mid c_3, e_2^*, Q) &= 2/3 = 0.67 \\
P(A = \text{Weekend Festival} \mid c_3, e_3^*, Q) &= 1/8 = 0.125 \\
P(A = \text{Weekend Festival} \mid c_3, e_4^*, Q) &= 2/6 = 0.33 \\
P(A = \text{Weekend Festival} \mid c_3, e_5^*, Q) &= 1/4 = 0.25
\end{aligned}$$

CoT 4

$$\begin{aligned}
P(c_4 \mid e_1^*, Q) &= 3/10 = 0.3 \\
P(c_4 \mid e_2^*, Q) &= 5/10 = 0.5 \\
P(c_4 \mid e_3^*, Q) &= 1/10 = 0.1 \\
P(c_4 \mid e_4^*, Q) &= 7/10 = 0.7 \\
P(c_4 \mid e_5^*, Q) &= 4/10 = 0.4
\end{aligned}$$

$$\begin{aligned}
P(A = \text{Weekend Festival} \mid c_4, e_1^*, Q) &= 1/3 = 0.33 \\
P(A = \text{Weekend Festival} \mid c_4, e_2^*, Q) &= 1/5 = 0.2 \\
P(A = \text{Weekend Festival} \mid c_4, e_3^*, Q) &= 0/1 = 0 \\
P(A = \text{Weekend Festival} \mid c_4, e_4^*, Q) &= 2/7 = 0.286 \\
P(A = \text{Weekend Festival} \mid c_4, e_5^*, Q) &= 2/4 = 0.5
\end{aligned}$$

CoT 5

$$\begin{aligned}
P(c_5 \mid e_1^*, Q) &= 4/10 = 0.4 \\
P(c_5 \mid e_2^*, Q) &= 2/10 = 0.2 \\
P(c_5 \mid e_3^*, Q) &= 5/10 = 0.5 \\
P(c_5 \mid e_4^*, Q) &= 9/10 = 0.9 \\
P(c_5 \mid e_5^*, Q) &= 3/10 = 0.3
\end{aligned}$$

$$\begin{aligned}
P(A = \text{Weekend Festival} \mid c_5, e_1^*, Q) &= 3/4 = 0.75 \\
P(A = \text{Weekend Festival} \mid c_5, e_2^*, Q) &= 0/2 = 0
\end{aligned}$$

$$P(A = \text{Weekend Festival} \mid c_5, e_3^*, Q) = 2/5 = 0.4$$

$$P(A = \text{Weekend Festival} \mid c_5, e_4^*, Q) = 3/9 = 0.33$$

$$P(A = \text{Weekend Festival} \mid c_5, e_5^*, Q) = 2/3 = 0.67$$

$$P(A = \text{Tomorrowland} \mid c_1, Q) = 0.258 \times 0.5 \times 0.6 + 0.221 \times 0.6 \times 0.5 + 0.194 \times 0.8 \times 0.625 + 0.172 \times 0.7 \times 0.57 + 0.155 \times 0.6 \times 0.33 = 0.34$$

$$P(A = \text{Tomorrowland} \mid c_2, Q) = 0.258 \times 0.7 \times 0.71 + 0.221 \times 0.3 \times 0.67 + 0.194 \times 0.4 \times 0.75 + 0.172 \times 0.4 \times 0 + 0.155 \times 0.8 \times 0.75 = 0.324$$

$$P(A = \text{Weekend Festival} \mid c_3, Q) = 0.258 \times 0.7 \times 0.43 + 0.221 \times 0.3 \times 0.67 + 0.194 \times 0.8 \times 0.125 + 0.172 \times 0.6 \times 0.33 + 0.155 \times 0.4 \times 0.25 = 0.191$$

$$P(A = \text{Weekend Festival} \mid c_4, Q) = 0.258 \times 0.3 \times 0.33 + 0.221 \times 0.5 \times 0.2 + 0.194 \times 0.1 \times 0 + 0.172 \times 0.7 \times 0.286 + 0.155 \times 0.4 \times 0.5 = 0.113$$

$$P(A = \text{Weekend Festival} \mid c_5, Q) = 0.258 \times 0.4 \times 0.75 + 0.221 \times 0.2 \times 0 + 0.194 \times 0.5 \times 0.4 + 0.172 \times 0.9 \times 0.33 + 0.155 \times 0.3 \times 0.67 = 0.175$$

$$P(A = \text{Tomorrowland} \mid do(Q)) = 0.34 + 0.324 = 0.664$$

$$P(A = \text{Weekend Festival} \mid do(Q)) = 0.191 + 0.113 + 0.175 = 0.479$$

The final answer produced by CFD-Prompting is **Tomorrowland**.

1.2 Case on MuSiQue

Intermediate output details of a MuSiQue example.

External Knowledge: Title - List of National Football League attendance figures — Supporting - False — Text - The National Football League is the largest live spectator sporting league in the world (excluding auto racing) in terms of average attendance. As of 2015, the NFL averaged 68,400 live spectators per game, and 17,510,312 total for the season.

Question: Who has the lowest batting average in the league that Jim Wilson's team plays for?
(The correct answer: **Bill Bergen**)

Outputs from CoT-SC:

Answer voting results: **Rob Deer**

Cluster Results:

CoT 1

To determine who has the lowest batting average in the league that Jim Wilson's team plays for, let's follow these steps:

1. Identify Jim Wilson's team**: According to the context, Jim Wilson played for the Cleveland Indians, as mentioned in Para 19.
2. Determine the league the Cleveland Indians belong to**: The Cleveland Indians are part of Major League Baseball (MLB), specifically in the American League (AL) Central division, as noted in Para 18.
3. Find the player with the lowest batting average in MLB**: From the context provided in Para 0, the record for the lowest career batting average for a player with more than 2,500 at-bats belongs to Bill Bergen, who recorded a batting average of .170.
4. Conclude the lowest batting average**: Since Bill Bergen has the lowest career batting average noted in the text, he holds this record in the league that Jim Wilson's team (Cleveland Indians) plays for. Therefore, the final answer is **[Bill Bergen]**.

CoT 2

1. First, we need to identify the team Jim Wilson played for. According to Para 19, he played for the Cleveland Indians.
2. Next, we should determine which league the Cleveland Indians are a part of. The Cleveland Indians compete in Major League Baseball (MLB) as a member of the American League (AL) Central division.
3. Now we need to find out who has the lowest batting average in Major League Baseball. According to Para 0, the record for the lowest career batting average for a player with more than 2,500 at-bats belongs to Bill Bergen, who recorded a .170 average.
4. Based on this information, we can conclude that Bill Bergen holds the record for the lowest batting average in the league (MLB) that Jim Wilson's team (Cleveland Indians) competes in. Therefore, the final answer is **[Bill Bergen]**.

CoT 3

Based on the provided context, we need to determine who has the lowest batting average in the league that Jim Wilson's team, the Seattle Mariners, plays for. Here's how we can analyze this step by step:

1. Identify Jim Wilson's Team**: From the context, Jim Wilson played for the Seattle Mariners after being released by the Cleveland Indians. The Mariners compete in Major League Baseball (MLB).
2. Determine the Lowest Batting Average in MLB**: The context provides information regarding batting averages. It mentions that the lowest career batting average for a player with more than 2,500 at-bats belongs to Bill Bergen, with an average of .170. It also states that the modern-era record for the lowest batting average for a player that qualified for the batting title is held by Rob Deer, who hit .179 in 1991.
3. Confirm the League**: Since Jim Wilson played for the Seattle Mariners, we can confirm we are looking for the lowest batting average within Major League Baseball.
4. Final Check of Averages**: The provided data indicates that Rob Deer has the lowest batting average for a player qualifying for the batting title in modern-era MLB at .179. This would indicate he is the player with the lowest batting average currently recognized in the context provided. Thus, following the analysis, the final answer is **[Rob Deer]**.

CoT 4

To determine who has the lowest batting average in the league that Jim Wilson's team (the Seattle Mariners) plays for, we will follow these steps:

1. Identify Jim Wilson's Team**: Jim Wilson played for the Seattle Mariners after being released by the Indians.
2. League Information**: The Seattle Mariners compete in Major League Baseball (MLB).
3. Research the Lowest Batting Average in MLB**: From the context, we know that Rob Deer holds the

modern-era record for the lowest batting average for a player that qualified for the batting title, which is .179 in 1991.

4. Determine if This Player's Record Applies**: Since Rob Deer is a player mentioned in the context, we consider him as the record holder for the lowest batting average.

5. Conclusion: Rob Deer has the lowest batting average in MLB as noted in the context. Therefore, the final answer is [**Rob Deer**].

CoT 5

To determine who has the lowest batting average in the league that Jim Wilson's team (the Seattle Mariners) plays for, we need to follow these steps:

1. Identify Jim Wilson's team: Jim Wilson played for the Seattle Mariners. 2. Determine the league the Seattle Mariners compete in: The Mariners are a member of Major League Baseball (MLB).

3. Look for the player with the lowest batting average in MLB: According to the context provided in Para 0, the modern-era record for the lowest batting average for a player that qualified for the batting title is held by Rob Deer, who hit .179 in 1991. Adam Dunn also had a low average but did not qualify for the batting title as he was six plate appearances short.

4. Conclusion: Based on the context, Rob Deer has the lowest batting average in the league that Jim Wilson's team (the Seattle Mariners) plays for.

Therefore, the final answer is [**Rob Deer**].

Constructing Counterfactual External Knowledge:

The list of the top five entities most relevant to the question:

[Rogers Hornsby, Bill Bergen, Napoleon Lajoie, Rob Deer, Adam Dunn]

The list of counterfactual entities generated by the llm:

[Stan Musial, Johnny Bench, Ty Cobb, Matt Williams, Ryan Howard]

Based on this, we can construct five counterfactual external knowledge according to the method in the text and obtain their corresponding probabilities.

$$P(e_1^*) = 0.258$$

$$P(e_2^*) = 0.221$$

$$P(e_3^*) = 0.194$$

$$P(e_4^*) = 0.172$$

$$P(e_5^*) = 0.155$$

Intervention Results:

CoT 1

$$P(c_1 \mid e_1^*, Q) = 7/10 = 0.7$$

$$P(c_1 \mid e_2^*, Q) = 6/10 = 0.6$$

$$P(c_1 \mid e_3^*, Q) = 7/10 = 0.7$$

$$P(c_1 \mid e_4^*, Q) = 8/10 = 0.8$$

$$P(c_1 \mid e_5^*, Q) = 5/10 = 0.5$$

$$P(A = \text{Bill Bergen} \mid c_1, e_1^*, Q) = 3/7 = 0.43$$

$$P(A = \text{Bill Bergen} \mid c_1, e_2^*, Q) = 5/6 = 0.83$$

$$P(A = \text{Bill Bergen} \mid c_1, e_3^*, Q) = 4/7 = 0.57$$

$$P(A = \text{Bill Bergen} \mid c_1, e_4^*, Q) = 5/8 = 0.625$$

$$P(A = \text{Bill Bergen} \mid c_1, e_5^*, Q) = 3/5 = 0.6$$

CoT 2

$$P(c_2 \mid e_1^*, Q) = 5/10 = 0.5$$

$$P(c_2 \mid e_2^*, Q) = 4/10 = 0.4$$

$$P(c_2 \mid e_3^*, Q) = 6/10 = 0.6$$

$$P(c_2 \mid e_4^*, Q) = 5/10 = 0.5$$

$$P(c_2 \mid e_5^*, Q) = 7/10 = 0.7$$

$$P(A = \text{Bill Bergen} \mid c_2, e_1^*, Q) = 4/5 = 0.8$$

$$P(A = \text{Bill Bergen} \mid c_2, e_2^*, Q) = 1/4 = 0.25$$

$$P(A = \text{Bill Bergen} \mid c_2, e_3^*, Q) = 5/6 = 0.83$$

$$P(A = \text{Bill Bergen} \mid c_2, e_4^*, Q) = 2/5 = 0.4$$

$$P(A = \text{Bill Bergen} \mid c_2, e_5^*, Q) = 4/7 = 0.57$$

CoT 3

$$P(c_3 \mid e_1^*, Q) = 5/10 = 0.5$$

$$P(c_3 \mid e_2^*, Q) = 7/10 = 0.7$$

$$P(c_3 \mid e_3^*, Q) = 6/10 = 0.6$$

$$P(c_3 \mid e_4^*, Q) = 3/10 = 0.3$$

$$P(c_3 \mid e_5^*, Q) = 5/10 = 0.5$$

$$P(A = \text{Rob Deer} \mid c_3, e_1^*, Q) = 3/5 = 0.6$$

$$P(A = \text{Rob Deer} \mid c_3, e_2^*, Q) = 2/7 = 0.29$$

$$P(A = \text{Rob Deer} \mid c_3, e_3^*, Q) = 3/6 = 0.5$$

$$P(A = \text{Rob Deer} \mid c_3, e_4^*, Q) = 1/3 = 0.33$$

$$P(A = \text{Rob Deer} \mid c_3, e_5^*, Q) = 1/5 = 0.2$$

CoT 4

$$P(c_4 \mid e_1^*, Q) = 6/10 = 0.6$$

$$P(c_4 \mid e_2^*, Q) = 3/10 = 0.3$$

$$P(c_4 \mid e_3^*, Q) = 5/10 = 0.5$$

$$P(c_4 \mid e_4^*, Q) = 4/10 = 0.4$$

$$P(c_4 \mid e_5^*, Q) = 6/10 = 0.6$$

$$P(A = \text{Rob Deer} \mid c_4, e_1^*, Q) = 3/6 = 0.5$$

$$P(A = \text{Rob Deer} \mid c_4, e_2^*, Q) = 1/3 = 0.33$$

$$P(A = \text{Rob Deer} \mid c_4, e_3^*, Q) = 2/5 = 0.4$$

$$P(A = \text{Rob Deer} \mid c_4, e_4^*, Q) = 2/4 = 0.5$$

$$P(A = \text{Rob Deer} \mid c_4, e_5^*, Q) = 2/6 = 0.33$$

CoT 5

$$P(c_5 \mid e_1^*, Q) = 5/10 = 0.5$$

$$P(c_5 \mid e_2^*, Q) = 3/10 = 0.3$$

$$P(c_5 \mid e_3^*, Q) = 4/10 = 0.4$$

$$P(c_5 \mid e_4^*, Q) = 6/10 = 0.6$$

$$P(c_5 \mid e_5^*, Q) = 7/10 = 0.7$$

$$P(A = \text{Rob Deer} \mid c_5, e_1^*, Q) = 2/5 = 0.4$$

$$P(A = \text{Rob Deer} \mid c_5, e_2^*, Q) = 2/3 = 0.67$$

$$P(A = \text{Rob Deer} \mid c_5, e_3^*, Q) = 1/4 = 0.25$$

$$P(A = \text{Rob Deer} \mid c_5, e_4^*, Q) = 2/6 = 0.33$$

$$P(A = \text{Rob Deer} \mid c_5, e_5^*, Q) = 3/7 = 0.43$$

$$P(A = \text{Bill Bergen} \mid c_1, Q) = 0.258 \times 0.7 \times 0.43 + 0.221 \times 0.6 \times 0.83 + 0.194 \times 0.7 \times 0.57 + 0.172 \times 0.8 \times 0.625 + 0.155 \times 0.5 \times 0.6 = 0.398$$

$$P(A = \text{Bill Bergen} \mid c_2, Q) = 0.258 \times 0.5 \times 0.8 + 0.221 \times 0.4 \times 0.25 + 0.194 \times 0.6 \times 0.83 + 0.172 \times 0.5 \times 0.4 + 0.155 \times 0.7 \times 0.57 = 0.318$$

$$P(A = \text{Rob Deer} \mid c_3, Q) = 0.258 \times 0.5 \times 0.6 + 0.221 \times 0.7 \times 0.29 + 0.194 \times 0.6 \times 0.5 + 0.172 \times 0.3 \times 0.33 + 0.155 \times 0.5 \times 0.2 = 0.213$$

$$P(A = \text{Rob Deer} \mid c_4, Q) = 0.258 \times 0.6 \times 0.5 + 0.221 \times 0.3 \times 0.33 + 0.194 \times 0.5 \times 0.4 + 0.172 \times 0.4 \times 0.5 + 0.155 \times 0.6 \times 0.33 = 0.203$$

$$P(A = \text{Rob Deer} \mid c_5, Q) = 0.258 \times 0.5 \times 0.4 + 0.221 \times 0.3 \times 0.67 + 0.194 \times 0.4 \times 0.25 + 0.172 \times 0.6 \times 0.33 + 0.155 \times 0.7 \times 0.43 = 0.196$$

$$P(A = \text{Bill Bergen} \mid do(Q)) = 0.398 + 0.318 = 0.716$$

$$P(A = \text{Rob Deer} \mid do(Q)) = 0.213 + 0.203 + 0.196 = 0.612$$

The final answer produced by CFD-Prompting is **Bill Bergen**.