

5-shot Prompt Exemplars

SST-2

- Standard Prompting:

Q: For the sentence: "will find little of interest in this film , which is often preachy and poorly acted", is the sentiment in this sentence positive or negative?

A: The sentiment is negative.

Q: For the sentence: "gorgeous and deceptively minimalist", is the sentiment in this sentence positive or negative?

A: The sentiment is positive.

Q: For the sentence: "that would make lesser men run for cover .", is the sentiment in this sentence positive or negative?

A: The sentiment is positive.

Q: For the sentence: "of marivaux 's rhythms , and mira sorvino 's limitations as a classical actress", is the sentiment in this sentence positive or negative?

A: The sentiment is negative.

Q: For the sentence: "viewers of barney 's crushingly self-indulgent spectacle", is the sentiment in this sentence positive or negative?

A: The sentiment is negative.

- Chain-of-thought Prompting:

Overall procedure: First, understand the sentence and identify key concepts. Second, consider the context of the sentence. Determine the sentiment based on the meaning of key words and sentence context.

Q: For the sentence: "will find little of interest in this film , which is often preachy and poorly acted", is the sentiment in this sentence positive or negative?

A: First, the sentence is a critique of a film. Key phrases include: "little of interest", "preachy", and "poorly acted". Second, the sentence is reviewing a film. All key phrases carry negative connotations. The sentiment is negative.

Q: For the sentence: "gorgeous and deceptively minimalist", is the sentiment in this sentence positive or negative?

A: First, the key phrases in the sentence are "gorgeous" and "deceptively minimalist". Second, the sentence is describing something's appearance or style. "Gorgeous" is positive, and "deceptively minimalist" suggests a surprising or unexpected simplicity, which can be seen as positive. The sentiment is positive.

Q: For the sentence: "that would make lesser men run for cover .", is the sentiment in this sentence positive or negative?

A: First, the key phrases in the sentence are "lesser men" and "run for cover". Second, the context implies a situation or task that is challenging or intimidating. The overarching sentiment celebrates the bravery, resilience, or capability of the person being described. The sentiment is positive.

Q: For the sentence: "of marivaux 's rhythms , and mira sorvino 's limitations as a classical actress", is the sentiment in this sentence positive or negative?

A: First, the key phrases in the sentence are "marivaux's rhythms" and "mira sorvino's limitations as a classical actress". Second, the context seems to comment on the style of Marivaux and the abilities of Mira Sorvino. The term "limitations" indicates a shortfall or inadequacy in Mira Sorvino's acting in classical roles. The sentiment is negative.

Q: For the sentence: "viewers of barney 's crushingly self-indulgent spectacle", is the sentiment in this sentence positive or negative?

A: First, the key phrases in the sentence are "barney's" and "crushingly self-indulgent spectacle". Second, the context appears to be a commentary on a spectacle associated with Barney. The term "crushingly self-indulgent" implies excessive self-interest and a lack of consideration for the audience. The sentiment is negative.

- **Metacognitive Prompting:**

Overall procedure: 1. Sentence understanding. 2. Preliminary identification of the sentiment. 3. Critically assess the preliminary analysis. If not sure, try to reassess. 4. Confirm the final answer and explain the reasoning. 5. Confidence evaluation and explanation.

Q: For the sentence: "will find little of interest in this film , which is often preachy and poorly acted", is the sentiment in this sentence positive or negative?

A: 1. The sentence is providing an opinion about a film, suggesting that viewers might not find much to appreciate in it and criticizing it for being overly moralistic and having subpar acting. 2. The preliminary sentiment appears to be negative. 3. The terms identified clearly lean toward a negative connotation. "Little of interest" implies a lack of engaging content. "Preachy" often means trying too hard to convey a moral lesson, and "poorly acted" straightforwardly critiques the performances. It's hard to find a positive spin on any of these phrases. 4. The sentiment of the sentence is negative. This conclusion is drawn from the clear criticisms present in the sentence, with no balancing positive remarks. 5. I am 95% confident in this analysis . The language used in the sentence overwhelmingly indicates a negative sentiment. The sentiment is negative.

Q: For the sentence: "gorgeous and deceptively minimalist", is the sentiment in this sentence positive or negative?

A: 1. The sentence describes something, likely an object or design, using two key phrases: "gorgeous" and "deceptively minimalist". 2. The preliminary sentiment appears to be positive based on the word "gorgeous". 3. The term "gorgeous" is straightforwardly positive, while "deceptively minimalist" can be seen as positive or neutral, suggesting an unexpected simplicity that might be appealing. 4. The sentiment of the sentence is positive, as the description overall suggests appreciation. 5. I am 90% confident in this analysis. The primary term "gorgeous" is distinctly positive, while "deceptively minimalist" leans more towards a positive or neutral appreciation of simplicity. The sentiment is positive.

Q: For the sentence: "that would make lesser men run for cover .", is the sentiment in this sentence positive or negative?

A: 1. The sentence appears to describe a challenging or intimidating situation that only "lesser men" would avoid. 2. The preliminary sentiment seems negative based on "lesser men run for cover". 3. However, on closer inspection, while "lesser men" and "run for cover" have negative connotations, the overarching sentiment may imply a positive view on someone or a group facing the situation bravely. 4. The sentiment of the sentence is positive, as it seems to celebrate the bravery or capability of the subject in contrast to "lesser men". 5. I am 85% confident in this analysis. Though the phrasing has negative elements, the overall implication is a commendation of bravery or capability. The sentiment is positive.

Q: For the sentence: "of marivaux 's rhythms , and mira sorvino 's limitations as a classical actress", is the sentiment in this sentence positive or negative?

A: 1. The sentence mentions two distinct elements: "marivaux's rhythms" and "mira sorvino's limitations as a classical actress". 2. The preliminary sentiment is unclear based solely on "marivaux's rhythms", but "mira sorvino's limitations" suggests a negative view. 3. "Marivaux's rhythms" is neutral without further context. However, "limitations" denotes a shortfall or inadequacy in Mira Sorvino's performance in classical roles. 4. The sentiment of the sentence is negative, primarily driven by the critique of Mira Sorvino's abilities. 5. I am 90% confident in this analysis. The term "limitations" clearly indicates a negative sentiment, while "marivaux's rhythms" remains neutral in the absence of more context. The sentiment is negative.

Q: For the sentence: "viewers of barney 's crushingly self-indulgent spectacle", is the sentiment in this sentence positive or negative?

A: 1. The sentence describes "barney's" spectacle using the terms "viewers" and "crushingly self-indulgent". 2. The preliminary sentiment appears to be negative due to the word "crushingly". 3. The term "self-indulgent" typically carries a negative connotation, suggesting excessiveness or over-indulgence, often without regard for others. 4. The sentiment of the sentence is negative, with the critique focusing on the perceived self-indulgence of Barney's spectacle. 5. I am 95% confident in this analysis. The term "crushingly self-indulgent" strongly suggests a negative view. The sentiment is negative.

WNLI

- Standard Prompting:

Q: For the sentence pair, sentence 1: "Bill passed the half-empty plate to John because he was full." and sentence 2: "Bill was full.", determine if sentence 2 is entailed by sentence 1.

A: The entailment status is entailment.

Q: For the sentence pair, sentence 1: "At the Loebner competition the judges couldn't figure out which respondents were the chatbots because they were so stupid." and sentence 2: "The chatbots were so stupid.", determine if sentence 2 is entailed by sentence 1.

A: The entailment status is not_entailment.

Q: For the sentence pair, sentence 1: "Mary tucked her daughter Anne into bed, so that she could sleep." and sentence 2: "Mary tucked her daughter Anne into bed, so that Mary's daughter could sleep.", determine if sentence 2 is entailed by sentence 1.

A: The entailment status is entailment.

Q: For the sentence pair, sentence 1: "I poured water from the bottle into the cup until it was full." and sentence 2: "The bottle was full.", determine if sentence 2 is entailed by sentence 1.

A: The entailment status is not_entailment.

Q: For the sentence pair, sentence 1: "John couldn't see the stage with Billy in front of him because he is so tall." and sentence 2: "Billy is so tall.", determine if sentence 2 is entailed by sentence 1.

A: The entailment status is entailment.

- Chain-of-thought Prompting:

Overall procedure: First, understand both sentences and identify key elements. Second, compare the key information and relationships within the two sentences. Determine the entailment relationship.

Q: For the sentence pair, sentence 1: "Bill passed the half-empty plate to John because he was full." and sentence 2: "Bill was full.", determine if sentence 2 is entailed by sentence 1.

A: First, in sentence 1, Bill passes the plate due to someone being full, likely referring to Bill. Sentence 2 directly states that Bill was full. Second, comparing these pieces of information, it's clear that the state of Bill being "full" in the second sentence is implied in the first. The entailment status is entailment.

Q: For the sentence pair, sentence 1: "At the Loebner competition the judges couldn't figure out which respondents were the chatbots because they were so stupid." and sentence 2: "The chatbots were so stupid.", determine if sentence 2 is entailed by sentence 1.

A: First, in sentence 1, the reason the judges couldn't distinguish between respondents and chatbots is because of the stupidity, but it's not explicitly clear whose stupidity is being referred to. Sentence 2 makes a direct statement about the chatbots being stupid. Second, sentence 1 does not unequivocally imply that the chatbots were stupid. The entailment status is not_entailment.

Q: For the sentence pair, sentence 1: "Mary tucked her daughter Anne into bed, so that she could sleep." and sentence 2: "Mary tucked her daughter Anne into bed, so that Mary's daughter could sleep.", determine if sentence 2 is entailed by sentence 1.

A: First, sentence 1 states that Mary tucked Anne into bed so "she" could sleep. In this context, "she" most likely refers to Anne, Mary's daughter. Sentence 2 makes it explicit that Mary tucked Anne into bed so that her daughter, Anne, could sleep. Second, given that the intention in both sentences is about Anne being tucked into bed for sleep, the information in sentence 1 implies the content of sentence 2. The entailment status is entailment.

Q: For the sentence pair, sentence 1: "I poured water from the bottle into the cup until it was full." and sentence 2: "The bottle was full.", determine if sentence 2 is entailed by sentence 1.

A: First, sentence 1 describes an action where water is poured from a bottle into a cup until "it" was full. The pronoun "it" likely refers to the cup becoming full. Sentence 2 claims that the bottle was full. Second, while sentence 1 provides information about the cup being filled, it doesn't specify the initial state of the

bottle. Thus, the content of sentence 1 does not directly provide information about the state of the bottle. The entailment status is not_entailment.

Q: For the sentence pair, sentence 1: "John couldn't see the stage with Billy in front of him because he is so tall." and sentence 2: "Billy is so tall.", determine if sentence 2 is entailed by sentence 1.

A: First, sentence 1 mentions that John had difficulty seeing the stage due to Billy being in front of him, and the reason provided is that "he is so tall." Given the context, the pronoun "he" likely refers to Billy. Sentence 2 directly states "Billy is so tall." Second, considering the reason provided in sentence 1 for John's inability to see the stage, it's clear that the statement in sentence 2 about Billy's height is implied by sentence 1. The entailment status is entailment.

- **Metacognitive Prompting:**

Overall procedure: 1. Sentence pair understanding. 2. Preliminary identification of the entailment. 3. Critically assess the preliminary analysis. If not sure, try to reassess. 4. Confirm the final answer and explain the reasoning. 5. Confidence evaluation and explanation.

Q: For the sentence pair, sentence 1: "Bill passed the half-empty plate to John because he was full." and sentence 2: "Bill was full.", determine if sentence 2 is entailed by sentence 1.

A: 1. Sentence 1 states that Bill gave away his plate with some remaining food to John, and the reason for this action is that "he" was full. Sentence 2 is a direct statement asserting that Bill had eaten to his capacity and did not want more. 2. The preliminary entailment status is entailment. 3. It's possible that the "he" in sentence 1 could refer to another male subject, but considering the context of the sentence, it most likely refers to Bill. The initial assessment seems accurate. 4. Sentence 1 does entail sentence 2. The reason Bill passed the plate in the first sentence is because "he" (most likely referring to Bill) was full, which directly aligns with the statement made in sentence 2. 5. I am 95% confident in this analysis. Given the context and structure of sentence 1, the most probable interpretation is that Bill was the one who was full, which directly supports the statement in sentence 2. The entailment status is entailment.

Q: For the sentence pair, sentence 1: "At the Loebner competition the judges couldn't figure out which respondents were the chatbots because they were so stupid." and sentence 2: "The chatbots were so stupid.", determine if sentence 2 is entailed by sentence 1.

A: 1. Sentence 1 describes an event at the Loebner competition where the judges had difficulty distinguishing between respondents and chatbots due to "they" being so stupid. The pronoun "they" could refer either to the judges or the chatbots, but it's ambiguous. Sentence 2 makes a clear statement about the chatbots being stupid. 2. The preliminary entailment status is uncertain due to the ambiguity in sentence 1. 3. The main point of contention is the ambiguous pronoun "they" in sentence 1. While it could mean that the chatbots were stupid, it could also imply that the judges were the ones being referred to as stupid. 4. After further assessment, sentence 1 does not explicitly entail sentence 2. The reason is the inherent ambiguity in the first sentence which doesn't make it clear if the chatbots or the judges were the ones described as stupid. 5. I am 90% confident in this analysis. The ambiguity in sentence 1 prevents a clear entailment to sentence 2. The entailment status is not_entailment.

Q: For the sentence pair, sentence 1: "Mary tucked her daughter Anne into bed, so that she could sleep." and sentence 2: "Mary tucked her daughter Anne into bed, so that Mary's daughter could sleep.", determine if sentence 2 is entailed by sentence 1.

A: 1. Sentence 1 conveys that Mary tucked Anne, who is identified as her daughter, into bed so that "she" could sleep. The pronoun "she" in this context most likely refers to Anne. Sentence 2 explicitly states that Mary tucked Anne into bed for the purpose of letting her daughter (Anne) sleep. 2. The preliminary entailment status is entailment because the action and intent in both sentences seem to match. 3. Upon critical assessment, there isn't much ambiguity in sentence 1. Within the scope of the sentence, it's clear that "she" refers to Anne, Mary's daughter. 4. The final assessment confirms that sentence 1 does entail sentence 2. The action of Mary tucking Anne into bed so she could sleep aligns directly with the more explicit statement made in sentence 2. 5. I am 98% confident in this analysis. Given the clear context provided in sentence 1, it is evident that the pronoun "she" refers to Anne, Mary's daughter, which supports the statement made in sentence 2. The entailment status is entailment.

Q: For the sentence pair, sentence 1: "I poured water from the bottle into the cup until it was full." and sentence 2: "The bottle was full.", determine if sentence 2 is entailed by sentence 1.

A: 1. Sentence 1 describes the action of pouring water from a bottle into a cup until a certain entity, represented by the pronoun "it" was full. The ambiguity arises regarding whether "it" refers to the cup or the bottle. Sentence 2 directly states that the bottle was full. 2. The preliminary entailment status is uncertain due

to the ambiguity of what "it" refers to in sentence 1. 3. Upon critical assessment, while it's reasonable to infer that "it" refers to the cup since the water is being poured into it, there's no explicit mention in sentence 1 about the initial state of the bottle. 4. The final assessment is that sentence 1 does not explicitly entail sentence 2. Even if the assumption is made that the cup was the one being filled, there's no definite information in sentence 1 that confirms the bottle's state of being full before the pouring began. 5. I am 90% confident in this analysis. The primary source of uncertainty stems from the ambiguous use of "it" in sentence 1, but the sentence doesn't provide enough information about the state of the bottle. The entailment status is not_entailment.

Q: For the sentence pair, sentence 1: "John couldn't see the stage with Billy in front of him because he is so tall." and sentence 2: "Billy is so tall.", determine if sentence 2 is entailed by sentence 1.

A: 1. Sentence 1 provides information about John's inability to see the stage because of someone's tall stature, with "he" being tall. Sentence 2 states that Billy is tall. 2. The preliminary entailment status leans towards entailment, as the first sentence suggests Billy's tall stature is the reason for John's obstructed view. 3. Critically assessing the sentences, the pronoun "he" in the first sentence likely refers to Billy, as it makes the most logical sense within the context. If "he" were to refer to John, the sentence would mean John couldn't see the stage because of his own height, which is not supported by the information given. 4. The final assessment is that sentence 1 does entail sentence 2. The reason John couldn't see the stage in the first sentence is attributed to the height of "he," which, within the context, most probably refers to Billy. This aligns directly with the statement made in sentence 2. 5. I am 95% confident in this analysis. The structure and context of sentence 1 point strongly towards Billy being the tall individual, corroborating the statement in sentence 2. The entailment status is entailment.

RTE

- Standard Prompting:

Q: For the sentence pair, sentence 1: "No Weapons of Mass Destruction Found in Iraq Yet." and sentence 2: "Weapons of Mass Destruction Found in Iraq.", determine if sentence 2 is entailed by sentence 1.

A: The entailment status is not_entailment.

Q: For the sentence pair, sentence 1: "What is certain, however, is that around 120,000 years ago, at least, the modern Homo sapiens first appeared and lived

contemporaneously with the Neanderthal for at least 90,000 years, until the Neanderthal disappeared 30,000 years ago." and sentence 2: "Humans existed 10,000 years ago.", determine if sentence 2 is entailed by sentence 1.

A: The entailment status is entailment.

Q: For the sentence pair, sentence 1: "Warheads believed to contain the deadly nerve agent cyclosarin that were found by Polish troops in Iraq date back to Saddam Hussein's war with Iran in the 1980s, authorities said Friday." and sentence 2: "Polish troops confiscated Warheads", determine if sentence 2 is entailed by sentence 1.

A: The entailment status is not_entailment.

Q: For the sentence pair, sentence 1: "Jurassic Park is a novel written by Michael Crichton. Jurassic Park was published in 1990." and sentence 2: "Michael Crichton is the author of the book Jurassic Park.", determine if sentence 2 is entailed by sentence 1.

A: The entailment status is entailment.

Q: For the sentence pair, sentence 1: "The dual citizenship debate has been a major distraction for the Weah campaign and many of his opponents have been secretly advocating against his presence in the race." and sentence 2: "Weah does business with his opponents.", determine if sentence 2 is entailed by sentence 1.

A: The entailment status is not_entailment.

- Chain-of-thought Prompting:

Overall procedure: First, understand both sentences and identify key elements. Second, compare the key information and relationships within the two sentences. Determine the entailment relationship.

Q: For the sentence pair, sentence 1: "No Weapons of Mass Destruction Found in Iraq Yet." and sentence 2: "Weapons of Mass Destruction Found in Iraq.", determine if sentence 2 is entailed by sentence 1.

A: First, sentence 1 asserts that no weapons of mass destruction (WMDs) have been found in Iraq up to that point. Sentence 2 states directly that WMDs have been found in Iraq. Second, there is a direct contradiction between the two sentences. Sentence 1 communicates the absence of WMDs in Iraq, whereas sentence 2 asserts their presence. Given this direct contradiction in information between the two sentences, the information in sentence 1 does not match the content of sentence 2. The entailment status is not_entailment.

Q: For the sentence pair, sentence 1: "What is certain, however, is that around 120,000 years ago, at least, the modern Homo sapiens first appeared and lived

contemporaneously with the Neanderthal for at least 90,000 years, until the Neanderthal disappeared 30,000 years ago." and sentence 2: "Humans existed 10,000 years ago.", determine if sentence 2 is entailed by sentence 1.

A: First, sentence 1 provides a detailed timeline which indicates that modern Homo sapiens first appeared around 120,000 years ago and lived at least until 30,000 years ago. Second, when comparing this timeframe to the assertion in sentence 2 that humans existed 10,000 years ago, it's clear that if Homo sapiens were present 30,000 years ago, they certainly were around 10,000 years ago. Therefore, the information in sentence 1 directly supports the content of sentence 2. The entailment status is entailment.

Q: For the sentence pair, sentence 1: "Warheads believed to contain the deadly nerve agent cyclosarin that were found by Polish troops in Iraq date back to Saddam Hussein's war with Iran in the 1980s, authorities said Friday." and sentence 2: "Polish troops confiscated Warheads", determine if sentence 2 is entailed by sentence 1.

A: First, sentence 1 states that Polish troops "found" warheads in Iraq, which means they came upon or discovered them. This doesn't necessarily mean they took possession or confiscated them. Sentence 2 explicitly states that the Polish troops "confiscated" the warheads, which means they took them into official custody. Second, given the difference between simply "finding" something and actively "confiscating" it, the information in sentence 1 does not conclusively support the content of sentence 2. The entailment status is not_entailment.

Q: For the sentence pair, sentence 1: "Jurassic Park is a novel written by Michael Crichton. Jurassic Park was published in 1990." and sentence 2: "Michael Crichton is the author of the book Jurassic Park.", determine if sentence 2 is entailed by sentence 1.

A: First, sentence 1 clearly indicates Michael Crichton's authorship of the novel. Sentence 2 makes the claim that "Michael Crichton is the author of the book Jurassic Park". Second, given that the information in sentence 1 directly establishes Michael Crichton as the author of Jurassic Park, it matches and supports the content of sentence 2. The entailment status is entailment.

Q: For the sentence pair, sentence 1: "The dual citizenship debate has been a major distraction for the Weah campaign and many of his opponents have been secretly advocating against his presence in the race." and sentence 2: "Weah does business with his opponents.", determine if sentence 2 is entailed by sentence 1.

A: First, sentence 1 discusses the issue of dual citizenship and its impact on the Weah campaign, also mentioning that many of his opponents have been

advocating against his participation in the race. Second, while sentence 1 provides information about Weah's political adversaries and their stance against him, it does not make any mention of Weah conducting business or having business relations with his opponents. The information in sentence 1 does not match or support the content of sentence 2. The entailment status is not_entailment.

- **Metacognitive Prompting:**

Overall procedure: 1. Sentence pair understanding. 2. Preliminary identification of the entailment. 3. Critically assess the preliminary analysis. If not sure, try to reassess. 4. Confirm the final answer and explain the reasoning. 5. Confidence evaluation and explanation.

Q: For the sentence pair, sentence 1: "No Weapons of Mass Destruction Found in Iraq Yet." and sentence 2: "Weapons of Mass Destruction Found in Iraq.", determine if sentence 2 is entailed by sentence 1.

A: 1. Sentence 1 states that no WMDs have been found up until the present, while Sentence 2 makes a claim that WMDs have been found. 2. The preliminary entailment status is not_entailment. 3. Upon reassessment, the first sentence uses the word "Yet", which suggests that while WMDs haven't been found up to that point, they might be found in the future. However, this does not mean they have been found, which is what sentence 2 claims. Thus, the sentences remain in contradiction. 4. The final answer remains that sentence 1 does not entail sentence 2. The reason being, sentence 1 communicates the absence of WMDs in Iraq up to a certain point in time, while sentence 2 asserts their presence without any time constraint. 5. I am 100% confident in this analysis. The two sentences have directly opposing claims regarding the presence of WMDs in Iraq. The entailment status is not_entailment.

Q: For the sentence pair, sentence 1: "What is certain, however, is that around 120,000 years ago, at least, the modern Homo sapiens first appeared and lived contemporaneously with the Neanderthal for at least 90,000 years, until the Neanderthal disappeared 30,000 years ago." and sentence 2: "Humans existed 10,000 years ago.", determine if sentence 2 is entailed by sentence 1.

A: 1. Sentence 1 details the timeline of modern Homo sapiens beginning around 120,000 years ago, outliving Neanderthals who disappeared 30,000 years ago. Sentence 2 claims humans existed 10,000 years ago. 2. Considering the timeline in Sentence 1, the preliminary entailment status for Sentence 2 is entailment. 3. Though Sentence 1 doesn't explicitly state Homo sapiens' continuity, their existence 30,000 years ago implies they were around 10,000 years ago. 4.

Sentence 1 entails Sentence 2 because if Homo sapiens were present 30,000 years ago, they would logically exist 10,000 years ago. 5. I am 100% confident in this analysis. The timeline in Sentence 1 supports Sentence 2's claim. The entailment status is entailment.

Q: For the sentence pair, sentence 1: "Warheads believed to contain the deadly nerve agent cyclosarin that were found by Polish troops in Iraq date back to Saddam Hussein's war with Iran in the 1980s, authorities said Friday." and sentence 2: "Polish troops confiscated Warheads", determine if sentence 2 is entailed by sentence 1.

A: 1. Sentence 1 states that Polish troops "found" warheads in Iraq that were believed to contain a deadly nerve agent and dated back to the 1980s. Sentence 2 asserts that the Polish troops "confiscated" the warheads. 2. The word "found" in Sentence 1 doesn't necessarily mean "confiscated" or took into official custody. Hence, the preliminary entailment status is not_entailment. 3. Upon reevaluation, "finding" something doesn't necessarily translate to confiscating it. The act of finding does not provide explicit information about the subsequent actions taken regarding the found items. 4. Sentence 1 does not entail Sentence 2. While Sentence 1 mentions that the warheads were found, it does not specify that they were confiscated. 5. I am 90% confident in this analysis. The act of finding as described in Sentence 1 does not guarantee the act of confiscation stated in Sentence 2. The entailment status is not_entailment.

Q: For the sentence pair, sentence 1: "Jurassic Park is a novel written by Michael Crichton. Jurassic Park was published in 1990." and sentence 2: "Michael Crichton is the author of the book Jurassic Park.", determine if sentence 2 is entailed by sentence 1.

A: 1. Sentence 1 indicates that "Jurassic Park" is a novel authored by Michael Crichton and was published in 1990. Sentence 2 claims that Michael Crichton is the author of "Jurassic Park." 2. Given that Sentence 1 clearly identifies Michael Crichton as the writer of "Jurassic Park," the preliminary entailment status for Sentence 2 is entailment. 3. Reevaluating the sentences, the primary claim of Sentence 2 is directly supported by the information provided in Sentence 1. There isn't any ambiguity regarding Michael Crichton's authorship of the book. 4. Sentence 1 entails Sentence 2 because it explicitly states that Michael Crichton wrote "Jurassic Park," directly supporting the assertion made in Sentence 2. 5. I am 95% confident in this analysis. The information in Sentence 1 directly aligns with and supports the content of Sentence 2. The entailment status is entailment.

Q: For the sentence pair, sentence 1: "The dual citizenship debate has been a major distraction for the Weah campaign and many of his opponents have been secretly advocating against his presence in the race." and sentence 2: "Weah does business with his opponents.", determine if sentence 2 is entailed by sentence 1.

A: 1. Sentence 1 describes a debate regarding dual citizenship that has been problematic for the Weah campaign. Additionally, it mentions that many of Weah's opponents have been secretly advocating against him. Sentence 2 posits that Weah has business relations with his opponents. 2. From Sentence 1, while we know about the opposition's stance against Weah's campaign, there's no mention of any business dealings between Weah and his opponents. Thus, the preliminary entailment status for Sentence 2 is not_entailment. 3. Upon reevaluation, it's clear that Sentence 1 mainly focuses on the political dynamics surrounding Weah's campaign and opposition. There's no direct or implied mention of Weah's business ties with his opponents. 4. Sentence 1 does not entail Sentence 2. The first sentence speaks about political opposition and advocacy against Weah but doesn't mention or suggest any business relationships between the parties. 5. I am 95% confident in this analysis. Sentence 1 does not provide any information about Weah's business dealings with his opponents, which makes the content of Sentence 2 unsupported by it. The entailment status is not_entailment.

COPA

- Standard Prompting:

Q: Given the premise: "The host cancelled the party.", which of the following is the most plausible cause? Choice 1: "She was certain she had the flu.", Choice 2: "She worried she would catch the flu."

A: The most plausible cause is Choice 1.

Q: Given the premise: "The driver got pulled over by the police.", which of the following is the most plausible cause? Choice 1: "He was parking.", Choice 2: "He was speeding".

A: The most plausible cause is Choice 2.

Q: Given the premise: "The gardener wanted his plants to flourish.", which of the following is the most plausible effect? Choice 1: "He planted a variety of seeds.", Choice 2: "He put fertilizer in the soil."

A: The most plausible effect is Choice 2.

Q: Given the premise: "The swimmer won the race.", which of the following is the most plausible effect? Choice 1: "She received a gold medal.", Choice 2: "Her team lost the meet."

A: The most plausible effect is Choice 1.

Q: Given the premise: "The family postponed the barbeque.", which of the following is the most plausible cause? Choice 1: "The forecast called for storms.", Choice 2: "It was a holiday weekend."

A: The most plausible cause is Choice 1.

- **Chain-of-thought Prompting:**

Overall procedure: First, understand the given premise and the two choices. Second, clarify whether the task is asking for the most plausible cause or effect of the premise. Third, compare the premise with each choice, considering how they relate to each other within the context of the premise. Choose the most plausible option that either best causes or results from the premise.

Q: Given the premise: "The host cancelled the party.", which of the following is the most plausible cause? Choice 1: "She was certain she had the flu.", Choice 2: "She worried she would catch the flu."

A: First, the premise indicates the host cancelled the party. Choice 1 suggests she believed she was already sick, while Choice 2 implies fear of getting sick. Second, we're seeking the most plausible cause for the cancellation. Third, comparing both choices, cancelling because one is already sick (Choice 1) is more common and directly impactful than the fear of potentially getting sick (Choice 2). The most plausible cause is Choice 1.

Q: Given the premise: "The driver got pulled over by the police.", which of the following is the most plausible cause? Choice 1: "He was parking.", Choice 2: "He was speeding".

A: First, the premise indicates the driver was pulled over by the police. Choice 1 suggests it was due to his parking action, while Choice 2 implies he was driving too fast. Second, the task is to identify the most plausible cause for being pulled over. Third, comparing both choices, being pulled over because of speeding (Choice 2) is more commonly associated with police stops than merely parking (Choice 1). The most plausible cause is Choice 2.

Q: Given the premise: "The gardener wanted his plants to flourish.", which of the following is the most plausible effect? Choice 1: "He planted a variety of seeds.", Choice 2: "He put fertilizer in the soil."

A: First, the premise indicates the gardener's desire for his plants to thrive. Choice 1 suggests he diversified his plant selection, while Choice 2 implies he took steps to nourish the soil. Second, the task is to determine the most plausible effect that would result from his desire. Third, comparing both choices, enriching the soil with fertilizer (Choice 2) is more directly linked to plants flourishing than simply planting a variety of seeds (Choice 1). The most plausible effect is Choice 2.

Q: Given the premise: "The swimmer won the race.", which of the following is the most plausible effect? Choice 1: "She received a gold medal.", Choice 2: "Her team lost the meet."

A: First, the premise indicates the swimmer was victorious in her race. Choice 1 suggests she was rewarded with a gold medal, while Choice 2 implies her team's overall defeat despite her individual win. Second, the task is to determine the most plausible effect of her victory. Third, comparing both choices, winning a race often results in receiving a medal or recognition (Choice 1), while an individual win does not necessarily mean the whole team loses (Choice 2). The most plausible effect is Choice 1.

Q: Given the premise: "The family postponed the barbeque.", which of the following is the most plausible cause? Choice 1: "The forecast called for storms.", Choice 2: "It was a holiday weekend."

A: First, the premise states that the family delayed their barbeque plans. Choice 1 suggests weather concerns, specifically storms, while Choice 2 indicates the occurrence of a holiday weekend. Second, the task is to find the most plausible cause for the postponement. Third, comparing both choices, delaying an outdoor event like a barbeque is more directly associated with unfavorable weather conditions (Choice 1) than the fact that it's a holiday weekend (Choice 2). The most plausible cause is Choice 1.

- **Metacognitive Prompting:**

Overall procedure: 1. Understand the premise and both choices. 2. Preliminary identification of the most plausible effect / cause of the premise. 3. Critically assess the preliminary analysis. If not sure, try to reassess. 4. Confirm the final answer and explain the reasoning. 5. Confidence evaluation and explanation.

Q: Given the premise: "The host cancelled the party.", which of the following is the most plausible cause? Choice 1: "She was certain she had the flu.", Choice 2: "She worried she would catch the flu."

A: 1. The premise reveals that a party was cancelled by the host. Choice 1 implies the host believed she was already infected with the flu. Choice 2 indicates the

host's fear of contracting the flu. 2. Preliminary identification of the most plausible cause of the premise is Choice 1. 3. If the host already has the flu, it's more immediate and reasonable for her to cancel the party to avoid spreading it. On the other hand, if she's only worried about catching the flu, she could take preventive measures without necessarily canceling. 4. The final answer for most plausible cause for the cancellation is Choice 1. Having the flu already poses a direct risk to guests and is a more concrete reason to cancel compared to merely being worried about contracting it. 5. I am 95% confident in this analysis, as actual illness is a stronger and more common reason for event cancellations than the potential fear of falling ill. The most plausible cause is Choice 1.

Q: Given the premise: "The driver got pulled over by the police.", which of the following is the most plausible cause? Choice 1: "He was parking.", Choice 2: "He was speeding".

A: 1. The premise describes a situation where a driver is stopped by the police. Choice 1 suggests it was due to his parking action, while Choice 2 implies he was driving too fast. 2. Preliminary identification of the most plausible cause of the premise leans toward Choice 2, given that speeding is a common reason for being pulled over. 3. While parking could be a potential reason, it's less common for police to pull someone over for parking unless it's a severe violation. On the other hand, speeding is frequently monitored and enforced. 4. The final answer for the most plausible cause of the driver being pulled over is Choice 2. Speeding is a more common and immediate reason for police stops than parking violations in most contexts. 5. I am 98% confident in this analysis. Speeding is widely recognized as a leading cause for being pulled over by law enforcement. The most plausible cause is Choice 2.

Q: Given the premise: "The gardener wanted his plants to flourish.", which of the following is the most plausible effect? Choice 1: "He planted a variety of seeds.", Choice 2: "He put fertilizer in the soil."

A: 1. The premise highlights the gardener's intention for his plants to thrive. Choice 1 suggests he diversified his plant selection, while Choice 2 indicates he took measures to nourish the soil. 2. Preliminary identification of the most plausible effect of the gardener's desire leans toward Choice 2, as nourishing the soil directly aids in plant growth. 3. Planting a variety of seeds may diversify the garden, but it doesn't necessarily ensure plant health or growth. Conversely, fertilizer enriches the soil, providing essential nutrients for plants to flourish. 4. The final answer for the most plausible effect stemming from the gardener's desire is Choice 2. Applying fertilizer to soil is a direct measure to ensure plant growth.

and health. 5. I am 97% confident in this analysis. Fertilizing soil is a widely accepted method to promote plant health and growth. The most plausible effect is Choice 2.

Q: Given the premise: "The swimmer won the race.", which of the following is the most plausible effect? Choice 1: "She received a gold medal.", Choice 2: "Her team lost the meet."

A: 1. The premise conveys that a swimmer was victorious in her race. Choice 1 implies she was awarded with a gold medal for her victory, while Choice 2 suggests an overall team loss despite her individual success. 2. Preliminary identification of the most plausible effect of the swimmer's victory is Choice 1. It is a common practice for winners to be awarded medals in competitive races. 3. An individual win does not typically determine the overall outcome of a team meet, so the relation between her victory and the team's loss isn't directly logical. However, receiving a medal, especially gold, is a standard acknowledgment for race winners. 4. The final answer for the most plausible effect stemming from the swimmer's victory is Choice 1. Winning a race is often followed by receiving an award or medal. 5. I am 99% confident in this analysis. The act of rewarding race winners with medals is a prevalent practice in swimming competitions. The most plausible effect is Choice 1.

Q: Given the premise: "The family postponed the barbeque.", which of the following is the most plausible cause? Choice 1: "The forecast called for storms.", Choice 2: "It was a holiday weekend."

A: 1. The premise tells us that a family decided to delay their barbeque. Choice 1 suggests it might be due to storms, while Choice 2 suggests the postponement might be because it was a holiday weekend. 2. Preliminary identification of the most plausible cause for the family's decision leans towards Choice 1, as inclement weather is a common reason to postpone outdoor events. 3. While holiday weekends might see increased activities and gatherings, it is not directly logical for it to be a cause for postponing a barbeque. On the other hand, the forecast predicting storms presents a direct concern for holding an outdoor event like a barbeque. 4. The final answer for the most plausible cause for the family's decision to postpone the barbeque is Choice 1. Stormy weather forecasts typically result in rescheduling outdoor events. 5. I am 96% confident in this analysis. Weather conditions, especially the prediction of storms, play a significant role in planning outdoor events. The most plausible cause is Choice 1.

WiC

- Standard Prompting:

Q: In two sentences, sentence 1: "I tried to call you all night." and sentence 2: "The new dean calls meetings every week.", both contain the target word "call". Determine if the target word is used with the same meaning in both sentences.

A: The target word has the same meaning in both sentences: False.

Q: In two sentences, sentence 1: "The baby was a boy." and sentence 2: "The boys were playing kickball in the mud.", both contain the target word "boy". Determine if the target word is used with the same meaning in both sentences.

A: The target word has the same meaning in both sentences: True.

Q: In two sentences, sentence 1: "The plane went into a steep bank." and sentence 2: "A huge bank of earth.", both contain the target word "bank". Determine if the target word is used with the same meaning in both sentences.

A: The target word has the same meaning in both sentences: False.

Q: In two sentences, sentence 1: "The washing machine won't go unless it's plugged in." and sentence 2: "The day went well until I got your call.", both contain the target word "go". Determine if the target word is used with the same meaning in both sentences.

A: The target word has the same meaning in both sentences: False.

Q: In two sentences, sentence 1: "A four-month stay in bankruptcy court." and sentence 2: "They had a nice stay in Paris.", both contain the target word "stay". Determine if the target word is used with the same meaning in both sentences.

A: The target word has the same meaning in both sentences: True.

- Chain-of-thought Prompting:

Overall procedure: First, understand both sentences and determine the meaning of the target word within each sentence, taking into account the context. Second, compare meanings of the target word. Determine whether the target word is used with the same meaning in both sentences.

Q: In two sentences, sentence 1: "I tried to call you all night." and sentence 2: "The new dean calls meetings every week.", both contain the target word "call". Determine if the target word is used with the same meaning in both sentences.

A: First, both sentences involve the use of the word "call" but in different scenarios: one related to an attempt to communicate and the other related to scheduling a meeting. Second, in sentence 1, "call" refers to the act of trying to communicate with someone. In sentence 2, "call" refers to the act of scheduling a

meeting. After comparing the meanings of the target word in both sentences, it's evident that "call" is used differently. The target word has the same meaning in both sentences: False.

Q: In two sentences, sentence 1: "The baby was a boy." and sentence 2: "The boys were playing kickball in the mud.", both contain the target word "boy". Determine if the target word is used with the same meaning in both sentences.

A: First, both sentences feature the word "boy" but in varied contexts: one as a singular noun and the other in its plural form. Second, in sentence 1, "boy" refers to the gender of the baby. In sentence 2, "boys" refers to multiple male children. After comparing the usages, the meaning of "boy" remains consistent across both sentences, indicating male gender. The target word has the same meaning in both sentences: True.

Q: In two sentences, sentence 1: "The plane went into a steep bank." and sentence 2: "A huge bank of earth.", both contain the target word "bank". Determine if the target word is used with the same meaning in both sentences.

A: First, both sentences contain the word "bank," but they are presented in different contexts: one relates to an airplane maneuver and the other to a large mass of earth. Second, in sentence 1, "bank" describes the tilt or turn of a plane. In sentence 2, "bank" refers to a raised portion of ground or mound. After comparing the usages, it's evident that "bank" has different meanings in the two sentences. The target word has the same meaning in both sentences: False.

Q: In two sentences, sentence 1: "The washing machine won't go unless it's plugged in." and sentence 2: "The day went well until I got your call.", both contain the target word "go". Determine if the target word is used with the same meaning in both sentences.

A: First, both sentences contain the word "go," but in differing contexts: one relates to the operation of a machine and the other to the progression of time or events. Second, in sentence 1, "go" means to operate or function. In sentence 2, "go" refers to the passage or occurrence of events. After comparing the usages, it's evident that "go" has different meanings in the two sentences. The target word has the same meaning in both sentences: False.

Q: In two sentences, sentence 1: "A four-month stay in bankruptcy court." and sentence 2: "They had a nice stay in Paris.", both contain the target word "stay". Determine if the target word is used with the same meaning in both sentences.

A: First, both sentences use the word "stay" to describe a duration or period of time in a specific context. Second, in sentence 1, it indicates the duration in bankruptcy court, and in sentence 2, it denotes the duration in Paris. After

comparing the contexts, the core meaning of "stay" as a period of time remains consistent in both sentences. The target word has the same meaning in both sentences: True.

- **Metacognitive Prompting:**

Overall procedure: 1. Understand the context and meaning of the target word in both sentences. 2. Preliminary judgement of the meaning consistency of the target word. 3. Critically assess the preliminary analysis. If not sure, try to reassess. 4. Confirm the final answer and explain the reasoning. 5. Confidence evaluation and explanation.

Q: In two sentences, sentence 1: "I tried to call you all night." and sentence 2: "The new dean calls meetings every week.", both contain the target word "call". Determine if the target word is used with the same meaning in both sentences.

A: 1. In sentence 1, "call" seems to refer to an attempt to communicate through a phone or similar device. In sentence 2, "call" refers to scheduling or convening meetings. 2. The word "call" seems to be used in different contexts and may have different meanings in the two sentences. 3. On deeper analysis, the first usage is about initiating communication, while the second pertains to organizing or initiating a gathering. Though both involve initiation of some sort, the contexts are different - one is about direct communication and the other is about organizing an event. 4. The target word "call" does not have the same meaning in both sentences. In the first, it's about trying to communicate, and in the second, it's about convening meetings. 5. I am 95% confident in this analysis. While both usages of "call" involve some form of initiation, the contexts are distinct, indicating different uses of the word. The target word has the same meaning in both sentences: False.

Q: In two sentences, sentence 1: "The baby was a boy." and sentence 2: "The boys were playing kickball in the mud.", both contain the target word "boy". Determine if the target word is used with the same meaning in both sentences.

A: 1. In sentence 1, "boy" is used to denote the gender of the baby. In sentence 2, "boys" is a plural form, referring to multiple male children engaged in an activity. 2. The primary sense of "boy" in both sentences seems to denote male gender. 3. Upon deeper analysis, the essence of the term "boy" remains consistent in both sentences, emphasizing male gender without any change in its core meaning. 4. The target word "boy" is used with the same meaning in both sentences to denote male gender, despite differences in number. 5. I am 98% confident in this analysis. The primary definition of "boy" remains consistent across the two sentences, indicating male gender. The target word has the same meaning in both sentences: True.

Q: In two sentences, sentence 1: "The plane went into a steep bank." and sentence 2: "A huge bank of earth.", both contain the target word "bank". Determine if the target word is used with the same meaning in both sentences.

A: 1. In sentence 1, "bank" is used in the context of an airplane maneuver, suggesting a tilt or turn. In sentence 2, "bank" refers to a large mass or mound of earth. 2. The primary sense of "bank" in the first sentence seems to relate to direction or angle, while in the second it suggests a physical formation. 3. Upon deeper analysis, it becomes evident that "bank" is used in two distinct contexts: one mechanical/aerodynamic and the other geological. 4. The target word "bank" does not have the same meaning in both sentences. In the first, it's about the tilt or turn of a plane, and in the second, it's about a raised portion of ground. 5. I am 90% confident in this analysis. The term "bank" is used in different contexts in the two sentences, resulting in different meanings. The target word has the same meaning in both sentences: False.

Q: In two sentences, sentence 1: "The washing machine won't go unless it's plugged in." and sentence 2: "The day went well until I got your call.", both contain the target word "go". Determine if the target word is used with the same meaning in both sentences.

A: 1. In sentence 1, "go" refers to the operation of the washing machine. In sentence 2, "went" (a form of "go") refers to the progression of the day. 2. Initially, "go" in the first sentence seems to be about mechanical operation, while in the second, it's about the passage of time or events. 3. Reflecting further, the contexts clearly differentiate the meanings: one deals with machinery and its function, the other deals with the experience of a day. 4. The target word "go" does not have the same meaning in both sentences. In the first, it pertains to operation, and in the second, to the progression of events. 5. I am 83% confident in this analysis. The term "go" is used in different contexts in the two sentences, leading to different meanings. The target word has the same meaning in both sentences: False.

Q: In two sentences, sentence 1: "A four-month stay in bankruptcy court." and sentence 2: "They had a nice stay in Paris.", both contain the target word "stay". Determine if the target word is used with the same meaning in both sentences.

A: 1. In sentence 1, "stay" refers to a period of time spent in a particular situation, in this case, bankruptcy court. In sentence 2, "stay" denotes a period of time spent in a location, Paris. 2. At first glance, "stay" in both sentences appears to denote a period of time, though the circumstances differ. 3. On deeper reflection, the essence of the term "stay" in both contexts pertains to a duration or period spent either in a situation or a place. 4. The target word "stay" has the same meaning in

both sentences, indicating a duration of time in a specific situation or location. 5. I am 96% confident in this analysis. Despite the different scenarios, the core meaning of "stay" remains consistent across both sentences. The target word has the same meaning in both sentences: True.

WSC

- Standard Prompting:

Q: Given the text: "If the con artist has succeeded in fooling Sam, he would have gotten a lot of money.", determine if the term "he" (at position 10) refers to "the con artist" (at position 1).

A: "He" refers to "the con artist": True.

Q: Given the text: "The fish ate the worm . It was tasty.", determine if the term "it" (at position 6) refers to "the worm" (at position 3).

A: "It" refers to "the worm": True.

Q: Given the text: "John promised Bill to leave, so an hour later he left.", determine if the term "he" (at position 9) refers to "John" (at position 0).

A: "He" refers to "John": False.

Q: Given the text: "Bob paid for Charlie 's college education. He is very grateful.", determine if the term "he" (at position 7) refers to "Bob" (at position 0).

A: "He" refers to "Bob": False.

Q: Given the text: "The dog chased the cat , which ran up a tree. It waited at the top.", determine if the term "it" (at position 11) refers to "the dog" (at position 0).

A: "It" refers to "the dog": False.

- Chain-of-thought Prompting:

Overall procedure: First, understand the meaning and context of the text. Second, clarify the relationships between different entities within the text. Determine the referential relationship between "[span1_text]" and "[span2_text]".

Q: Given the text: "If the con artist has succeeded in fooling Sam, he would have gotten a lot of money.", determine if the term "he" (at position 10) refers to "the con artist" (at position 1).

A: First, the sentence describes a hypothetical situation where a con artist might have deceived Sam to gain money. Second, The entities in question are "the con artist" and "he", and the relationship suggests that someone would have profited

from fooling Sam. "He" clearly refers back to "the con artist" as the one who would have benefited financially. "He" refers to "the con artist": True.

Q: Given the text: "The fish ate the worm . It was tasty.", determine if the term "it" (at position 6) refers to "the worm" (at position 3).

A: First, the sentence describes a scenario where a fish consumes a worm, followed by an assessment of taste. Second, the entities in question are "the worm" and "it", with the implication being that something was found to be delicious. Given the structure and sequence of the narrative, "it" is most likely referring to what the fish ate. "It" refers to "the worm": True.

Q: Given the text: "John promised Bill to leave, so an hour later he left.", determine if the term "he" (at position 9) refers to "John" (at position 0).

A: First, the sentence presents a situation where John makes a commitment to Bill about departing. Then, someone departs after an hour. Second, when considering "John promised Bill to leave", the promise often relates to what Bill should do or is expected to do, not John. Hence, in this structure, it's common that "Bill" is the subject of the action "to leave". Given this interpretation, "he" in the latter part of the sentence would be referring to Bill, the one who was expected to leave. "He" refers to "John": False.

Q: Given the text: "Bob paid for Charlie 's college education. He is very grateful.", determine if the term "he" (at position 7) refers to "Bob" (at position 0).

A: First, the sentence describes a scenario where Bob covers the cost of Charlie's college education. Then, someone expresses gratitude. Second, the person who would feel gratitude in this context would be Charlie, as he is the beneficiary of Bob's generosity. Given the nature of gratitude as a response to receiving something, "he" most likely refers to Charlie, who is thankful for the support in his education. "He" refers to "Bob": False.

Q: Given the text: "The dog chased the cat , which ran up a tree. It waited at the top.", determine if the term "it" (at position 11) refers to "the dog" (at position 0).

A: First, the sentence describes a sequence where a dog chases a cat, resulting in the cat climbing up a tree. Then, something waits at the top. Second, given the described sequence and the usage of "which" that refers specifically to "the cat", it's evident that the cat is the one that has gone up the tree. Therefore, "it" in the subsequent sentence would logically be referring to the entity that ran up the tree, which is the cat. "It" refers to "the dog": False.

- Metacognitive Prompting:

Overall procedure: 1. Understand the text and consider the reference relationship between "[span1_text]" and "[span2_text]". 2. Preliminary judgement on whether "[span2_text]" refers to "[span1_text]". 3. Critically assess the preliminary analysis. If not sure, try to reassess. 4. Confirm the final answer and explain the reasoning. 5. Confidence evaluation and explanation.

Q: Given the text: "If the con artist has succeeded in fooling Sam, he would have gotten a lot of money.", determine if the term "he" (at position 10) refers to "the con artist" (at position 1).

A: 1. The text describes a hypothetical situation in which a con artist might have deceived Sam to achieve a financial gain. 2. Initially, given the structure of the sentence, it seems that "he" is referring back to "the con artist" as the subject who would have benefited from fooling Sam. 3. Upon deeper reflection, "he" could only logically refer to "the con artist" since Sam is the only other male subject mentioned, and it wouldn't make sense for Sam to benefit from being fooled. 4. "He" indeed refers to "the con artist". The context and structure of the sentence clearly indicate that "the con artist" is the subject who would have achieved a financial gain if he had succeeded in deceiving Sam. 5. I am 95% confident in this analysis. The sentence structure and the context provide strong indications that "he" is a reference to "the con artist". The target reference is True. "He" refers to "the con artist": True.

Q: Given the text: "The fish ate the worm . It was tasty.", determine if the term "it" (at position 6) refers to "the worm" (at position 3).

A: 1. The text describes a situation where a fish consumes a worm, followed by an assessment of taste. 2. At first glance, given the sequence and context, it seems that "it" is referring to what the fish ate, which is "the worm". 3. Critically evaluating the text, the worm is the only item in the previous statement that can be associated with a taste attribute, making it the likely referent for "it". 4. "It" does indeed refer to "the worm". The context and sequence of events in the sentence point towards the worm being the item that was deemed tasty. 5. I am 97% confident in this analysis. Given the direct sequence of events and the context in which "tasty" is used, it strongly indicates that "it" is referencing "the worm". "It" refers to "the worm": True.

Q: Given the text: "John promised Bill to leave, so an hour later he left.", determine if the term "he" (at position 9) refers to "John" (at position 0).

A: 1. The sentence provides a situation in which John makes a commitment to Bill regarding someone's departure. Later, someone does indeed depart. 2. At first

glance, I think that since John is the first male name mentioned, "he" could naturally refer to John. However, the phrasing "promised Bill to leave" suggests that the promise pertained to Bill's action. 3. Upon critically assessing the statement, the structure "promised Bill to leave" is often understood to mean that John made a promise about Bill's action, suggesting that Bill would be the one to leave. 4. "He" in this context refers to "Bill". The structure and context suggest that John promised on behalf of Bill's action to leave, making Bill the subject of the latter part of the sentence. 5. I am 90% confident in this analysis. The sentence structure ("promised X to do something") indicates that "he" refers to Bill rather than John. "He" refers to "John": False.

Q: Given the text: "Bob paid for Charlie 's college education. He is very grateful.", determine if the term "he" (at position 7) refers to "Bob" (at position 0).

A: 1. The sentence provides a situation where Bob covers the expenses of Charlie's college education, followed by someone expressing gratitude.

2. I thought "he" could refer to Bob, as he is the first male mentioned in the statement. However, given that Charlie is the beneficiary of Bob's generosity, Charlie is more likely to be the one expressing gratitude. 3. Upon further examination, the most logical interpretation is that Charlie would be grateful for the financial assistance he received from Bob. Hence, "he" is more aligned with being a reference to Charlie. 4. "He" in this context does indeed refer to "Charlie". The context of gratitude, in this case, is more closely related to the recipient of the favor, which is Charlie, rather than the benefactor, which is Bob. 5. I am 87% confident in this analysis. The context and of gratitude make it more likely that "he" references "Charlie" over "Bob". "He" refers to "Bob": False.

Q: Given the text: "The dog chased the cat , which ran up a tree. It waited at the top.", determine if the term "it" (at position 11) refers to "the dog" (at position 0).

A: 1. The sentence describes a situation where a dog is pursuing a cat, which then climbs up a tree. Subsequently, an entity waits at the tree's top. 2. My initial thought is that "it" could refer to the cat, especially given the immediate previous clause that describes the cat's action of running up a tree. 3. Upon further reflection, the sentence structure points toward "it" referring to the last subject of action before the mention of "it", which in this case is the cat that ran up the tree. 4. "It" indeed refers to "the cat". The narrative sequence, combined with the cat's most recent action of ascending the tree, suggests that the cat is the one waiting at the top. 5. I am 95% confident in this analysis. The sequence of events in the sentence strongly suggests that "it" is referencing "the cat", not "the dog". "It" refers to "the dog": False.

QNLI

- Standard Prompting:

Q: For the question: "MediaOne merged with what telecommunications company?" and statement: "Comcast University started in 1999 as well as Comcast Interactive Capital Group to make technology and Internet related investments taking its first investment in VeriSign.", determine if the statement provides the answer to the question. If the statement contains the answer to the question, the status is entailment. If it does not, the status is not_entailment.

A: The entailment status is not_entailment.

Q: For the question: "Who was the first President to address the NAACP?" and statement: "In front of 10,000 people at the Lincoln Memorial, the president left no doubt where he stood on civil rights.", determine if the statement provides the answer to the question. If the statement contains the answer to the question, the status is entailment. If it does not, the status is not_entailment.

A: The entailment status is not_entailment.

Q: For the question: "What document declared the upgrade of University College at Highfield?" and statement: "A Royal Charter in 1952 upgraded University College at Highfield to the University of Southampton.", determine if the statement provides the answer to the question. If the statement contains the answer to the question, the status is entailment. If it does not, the status is not_entailment.

A: The entailment status is entailment.

Q: For the question: "Because of the number of tech companies, the city is a hub for what?" and statement: "The city is home to a number of technology companies and is a hub for biotechnology, with the Milken Institute rating Boston as the top life sciences cluster in the country.", determine if the statement provides the answer to the question. If the statement contains the answer to the question, the status is entailment. If it does not, the status is not_entailment.

A: The entailment status is entailment.

Q: For the question: "What BYU environment seems to surprise many visitors to BYU as well as to the Utah Valley?" and statement: "Many visitors to BYU, and Utah Valley as a whole, report being surprised by the culturally conservative environment.", determine if the statement provides the answer to the question. If the statement contains the answer to the question, the status is entailment. If it does not, the status is not_entailment.

A: The entailment status is entailment.

- Chain-of-thought Prompting:

Overall procedure: First, understand the question and the statement. Second, compare the information in the statement to the information sought in the question. Determine if the statement entails an answer to the question.

Q: For the question: "MediaOne merged with what telecommunications company?" and statement: "Comcast University started in 1999 as well as Comcast Interactive Capital Group to make technology and Internet related investments taking its first investment in VeriSign.", determine if the statement provides the answer to the question. If the statement contains the answer to the question, the status is entailment. If it does not, the status is not_entailment.

A: First, the question is asking about the telecommunications company with which MediaOne merged. The statement provides information about initiatives started by Comcast University in 1999 and an investment by Comcast Interactive Capital Group in VeriSign. Second, comparing the statement to the question, there is no mention of MediaOne or its merger with any telecommunications company in the provided statement. The entailment status is not_entailment.

Q: For the question: "Who was the first President to address the NAACP?" and statement: "In front of 10,000 people at the Lincoln Memorial, the president left no doubt where he stood on civil rights.", determine if the statement provides the answer to the question. If the statement contains the answer to the question, the status is entailment. If it does not, the status is not_entailment.

A: First, the question is asking about the identity of the first President to address the NAACP. The statement describes a scenario where a president addressed a large audience at the Lincoln Memorial and took a clear stance on civil rights. Second, comparing the statement to the question, there is no direct mention of the NAACP or specifics about which president is being referred to in the statement. The entailment status is not_entailment.

Q: For the question: "What document declared the upgrade of University College at Highfield?" and statement: "A Royal Charter in 1952 upgraded University College at Highfield to the University of Southampton.", determine if the statement provides the answer to the question. If the statement contains the answer to the question, the status is entailment. If it does not, the status is not_entailment.

A: First, the question is seeking information about the specific document that declared the upgrade of University College at Highfield. The statement provides information that "A Royal Charter in 1952" was responsible for upgrading University College at Highfield to the University of Southampton. Second, comparing the statement to the question, it clearly mentions that the "Royal

Charter in 1952" is the document that declared the upgrade of University College at Highfield. The entailment status is entailment.

Q: For the question: "Because of the number of tech companies, the city is a hub for what?" and statement: "The city is home to a number of technology companies and is a hub for biotechnology, with the Milken Institute rating Boston as the top life sciences cluster in the country.", determine if the statement provides the answer to the question. If the statement contains the answer to the question, the status is entailment. If it does not, the status is not_entailment.

A: First, the question is asking about the particular field or industry for which the city is recognized as a hub due to the presence of tech companies. The statement provides that the city is a hub for biotechnology, especially in relation to the presence of technology companies. Second, comparing the statement to the question, the statement directly states that the city is a hub for biotechnology, especially noting its recognition by the Milken Institute as a top life sciences cluster. The entailment status is entailment.

Q: For the question: "What BYU environment seems to surprise many visitors to BYU as well as to the Utah Valley?" and statement: "Many visitors to BYU, and Utah Valley as a whole, report being surprised by the culturally conservative environment.", determine if the statement provides the answer to the question. If the statement contains the answer to the question, the status is entailment. If it does not, the status is not_entailment.

A: First, the question is asking about a specific environment or aspect of BYU and Utah Valley that tends to surprise many visitors. The statement provides that many visitors are surprised by the "culturally conservative environment" at BYU and the Utah Valley. Second, comparing the statement to the question, the statement directly provides that the "culturally conservative environment" is what seems to surprise many visitors to both BYU and the Utah Valley. The entailment status is entailment.

- **Metacognitive Prompting:**

Overall procedure: 1. Understand the question and the context sentence. 2. Preliminary judgement on whether the sentence contains the answer to the question. 3. Critically assess the preliminary analysis. If not sure, try to reassess. 4. Confirm the final answer and explain the reasoning. 5. Confidence evaluation and explanation.

Q: For the question: "MediaOne merged with what telecommunications company?" and statement: "Comcast University started in 1999 as well as

Comcast Interactive Capital Group to make technology and Internet related investments taking its first investment in VeriSign.", determine if the statement provides the answer to the question. If the statement contains the answer to the question, the status is entailment. If it does not, the status is not_entailment.

A: 1. The question seeks to find out which telecommunications company merged with MediaOne. The statement gives information about Comcast University's commencement in 1999 and Comcast Interactive Capital Group's investment activity, including its initial investment in VeriSign. 2. The initial assessment is that the status is not_entailment. 3. The sentence primarily talks about Comcast's initiatives and investments in 1999. While Comcast is a telecommunications company, there is no explicit mention of a merger with MediaOne in the provided context. 4. The statement does not contain the answer to the question. The context primarily speaks about Comcast's other activities in 1999, and there's no mention of a merger with MediaOne. 5. I am 95% confident in this analysis. The statement provided is more focused on Comcast's ventures and investments in 1999 and lacks explicit information about any merger with MediaOne. The status is not_entailment.

Q: For the question: "Who was the first President to address the NAACP?" and statement: "In front of 10,000 people at the Lincoln Memorial, the president left no doubt where he stood on civil rights.", determine if the statement provides the answer to the question. If the statement contains the answer to the question, the status is entailment. If it does not, the status is not_entailment.

A: 1. The question is asking about the identity of the first President to address the NAACP. The statement describes a scenario where a president addressed a large audience at the Lincoln Memorial and took a clear stance on civil rights. 2. The initial assessment is that the status is not_entailment. 3. Upon reassessment, the statement does not specifically mention the NAACP or provide information about which president is being referred to. While the context does pertain to civil rights, it does not necessarily correlate with the specific act of addressing the NAACP. 4. The statement does not provide a clear answer to the question. The context discusses a president's stance on civil rights in front of a large audience at the Lincoln Memorial, but it does not specify which president this was or confirm that this president was the first to address the NAACP. 5. I am 85% confident in this analysis. The context lacks specifics about which president is being discussed and does not directly address the NAACP aspect of the question. The status is not_entailment.

Q: For the question: "What document declared the upgrade of University College at Highfield?" and statement: "A Royal Charter in 1952 upgraded University College at Highfield to the University of Southampton.", determine if the statement provides the answer to the question. If the statement contains the answer to the question, the status is entailment. If it does not, the status is not_entailment.

A: 1. The question is seeking information about the specific document that declared the upgrade of University College at Highfield. The statement provides information that "A Royal Charter in 1952" was responsible for upgrading University College at Highfield to the University of Southampton. 2. The initial assessment is that the status is entailment. 3. Upon reassessment, the statement directly states that the "Royal Charter in 1952" is the document that declared the upgrade of University College at Highfield. There isn't any ambiguity in the information provided, and it aligns with what the question is asking. 4. The statement provides a clear answer to the question. It directly states that the "Royal Charter in 1952" is the document that declared the upgrade of University College at Highfield to the University of Southampton. 5. I am 99% confident in this analysis. The statement offers a direct and clear answer to the question without any ambiguity. The entailment status is entailment.

Q: For the question: "Because of the number of tech companies, the city is a hub for what?" and statement: "The city is home to a number of technology companies and is a hub for biotechnology, with the Milken Institute rating Boston as the top life sciences cluster in the country.", determine if the statement provides the answer to the question. If the statement contains the answer to the question, the status is entailment. If it does not, the status is not_entailment.

A: 1. The question is asking about the particular field or industry for which the city is recognized as a hub due to the presence of tech companies. The statement provides that the city is a hub for biotechnology, especially in relation to the presence of technology companies. 2. The initial assessment is that the status is entailment. 3. Upon reassessment, the statement clarifies that because of the presence of tech companies, the city is recognized as a hub for biotechnology, further reinforced by the Milken Institute's rating of Boston as the top life sciences cluster in the country. 4. The statement provides a clear answer to the question. The city, due to its tech companies, is recognized as a hub for biotechnology. This recognition is further validated by a third party, the Milken Institute, which rates it as a significant cluster for life sciences. 5. I am 96% confident in this analysis. The statement directly answers the question, highlighting the city's prominence in biotechnology due to its tech companies. The entailment status is entailment.

Q: For the question: "What BYU environment seems to surprise many visitors to BYU as well as to the Utah Valley?" and statement: "Many visitors to BYU, and Utah Valley as a whole, report being surprised by the culturally conservative environment.", determine if the statement provides the answer to the question. If the statement contains the answer to the question, the status is entailment. If it does not, the status is not_entailment.

A: 1. The question is inquiring about a specific characteristic or aspect of the environment at BYU and Utah Valley that surprises many visitors. The statement mentions that many visitors to BYU and Utah Valley are often surprised by its "culturally conservative environment (CCE)." 2. The initial assessment is that the status is entailment. 3. Upon reassessment, the statement explicitly mentions that the "CCE" is what surprises many visitors to both BYU and Utah Valley. This directly answers the question, and there is no ambiguity in the information provided. 4. The statement provides a clear answer to the question. It specifies that the surprising aspect of the environment at BYU and Utah Valley for many visitors is its culturally conservative nature. 5. I am 95% confident in this analysis. The statement offers a direct and clear answer to the question without any ambiguity. The entailment status is entailment.

STS-B

- Standard Prompting:

Q: Assess the similarity between the following two sentences: "A badger is burrowing a hole." and "A badger is digging the earth.". The score should be ranging from 0.0 (no relation) to 5.0 (high similarity), and can be a decimal.

A: The similarity score is 4.6.

Q: Assess the similarity between the following two sentences: "A woman is dicing a potato." and "A woman cuts tofu into tiny pieces.". The score should be ranging from 0.0 (no relation) to 5.0 (high similarity), and can be a decimal.

A: The similarity score is 1.8.

Q: Assess the similarity between the following two sentences: "The man ran and jumped against the wall." and "A man jumps into a wall.". The score should be ranging from 0.0 (no relation) to 5.0 (high similarity), and can be a decimal.

A: The similarity score is 3.4.

Q: Assess the similarity between the following two sentences: "A little girl plays the violin on a beach." and "The girl sat on the couch to crochet.". The score

should be ranging from 0.0 (no relation) to 5.0 (high similarity), and can be a decimal.

A: The similarity score is 0.0.

Q: Assess the similarity between the following two sentences: "A woman is slicing pumpkin." and "A person is slicing cantaloupe.". The score should be ranging from 0.0 (no relation) to 5.0 (high similarity), and can be a decimal.

A: The similarity score is 2.0.

- Chain-of-thought Prompting:

Overall procedure: First, understand both sentences, considering the subject matter, context, and semantic contents. Second, compare key elements and relationships within the two sentences. Determine the similarity score.

Q: Assess the similarity between the following two sentences: "A badger is burrowing a hole." and "A badger is digging the earth.". The score should be ranging from 0.0 (no relation) to 5.0 (high similarity), and can be a decimal.

A: First, sentence 1 conveys the action of a badger, and sentence 2 describes a similar action. Second, both sentences have the same subject, describe similar actions, and the object of actions pertains to the earth or ground. Here is some slight distinction in direct objects "a whole" versus "the earth". The similarity score is 4.6.

Q: Assess the similarity between the following two sentences: "A woman is dicing a potato." and "A woman cuts tofu into tiny pieces.". The score should be ranging from 0.0 (no relation) to 5.0 (high similarity), and can be a decimal.

A: First, both sentence 1 and sentence 2 depict a woman performing an action related to cutting food. Second, the discrepancies lie in the food being cut ("potato" vs. "tofu") and the specific manner of cutting ("dicing" vs. "cuts... into tiny pieces"). The similarity score is 1.8.

Q: Assess the similarity between the following two sentences: "The man ran and jumped against the wall." and "A man jumps into a wall.". The score should be ranging from 0.0 (no relation) to 5.0 (high similarity), and can be a decimal.

A: First, both sentence 1 and sentence 2 describe a man's action involving jumping and a wall. Second, while sentence 1 provides a sequence of actions ("ran and jumped") against the wall, sentence 2 directly states the man's action of jumping into the wall. The similarity score is 3.4.

Q: Assess the similarity between the following two sentences: "A little girl plays the violin on a beach." and "The girl sat on the couch to crochet.". The score

should be ranging from 0.0 (no relation) to 5.0 (high similarity), and can be a decimal.

A: First, both sentence 1 and sentence 2 mention a girl involved in an activity. Second, upon closer examination, the actions and contexts are completely different. One is about playing a musical instrument at the beach, while the other is about crafting while seated on a couch. The similarity score is 0.0.

Q: Assess the similarity between the following two sentences: "A woman is slicing pumpkin." and "A person is slicing cantaloupe.". The score should be ranging from 0.0 (no relation) to 5.0 (high similarity), and can be a decimal.

A: First, both sentence 1 and sentence 2 depict an action of slicing on two different items. Second, although both sentences focus on slicing, there are clear distinctions: the subject performing the action is specified as a "woman" in the first sentence, whereas it's more generic in the second as "person", and the items being sliced are different (pumpkin vs. cantaloupe). The similarity score is 2.0.

- **Metacognitive Prompting:**

Overall procedure: 1. Understand both sentences. 2. Preliminary identification of similarity. 3. Critically assess the preliminary analysis. If not sure, try to reassess. 4. Confirm the final answer and explain the reasoning. 5. Confidence evaluation and explanation.

Q: Assess the similarity between the following two sentences: "A badger is burrowing a hole." and "A badger is digging the earth.". The score should be ranging from 0.0 (no relation) to 5.0 (high similarity), and can be a decimal.

A: 1. Both sentences describe a badger performing an action on the earth. 2. The preliminary score is 4.5. Both sentences seem very similar as they describe a badger's activity associated with the earth. 3. On close examination, "burrowing a hole" and "digging the earth" are closely related actions, both involve displacing earth. 4. The final similarity score is 4.6 since both sentences revolve around a badger performing an action on the ground. The main difference is the exact nature of that action, but the overall theme and context are alike. 5. I am 95% confident in this analysis. Both sentences have the same subject (a badger) and describe related actions involving the ground. The similarity score is 4.6.

Q: Assess the similarity between the following two sentences: "A woman is dicing a potato." and "A woman cuts tofu into tiny pieces.". The score should be ranging from 0.0 (no relation) to 5.0 (high similarity), and can be a decimal.

A: 1. Both sentences depict a woman engaging in an action related to cutting food. 2. Preliminarily, they seem moderately similar due to the shared subject (a woman)

and a similar verb (cutting). The preliminary score is 2.5. 3. Digging deeper, while both actions relate to cutting, the objects of the actions are different: "potato" versus "tofu". Furthermore, "dicing" typically implies a more specific cutting method than just "cuts... into tiny pieces". 4. Given the shared subject and verb but differing objects and specifics of the action, the final similarity score is 1.8. 5. I am 90% confident in this analysis. The shared subject and action provide some similarity, but the difference in the exact cutting method and the objects being cut lead to a less than perfect score. The similarity score is 1.8.

Q: Assess the similarity between the following two sentences: "The man ran and jumped against the wall." and "A man jumps into a wall.". The score should be ranging from 0.0 (no relation) to 5.0 (high similarity), and can be a decimal.

A: 1. Both sentences describe a man's interaction with a wall involving jumping. 2. Initially, the two sentences seem quite similar because they both involve a man and a wall and include the action of jumping. The preliminary score is 3.5. 3. Upon deeper analysis, the first sentence has an additional action ("ran") and the man jumps "against" the wall, while in the second sentence, the man "jumps into" the wall. This difference in prepositions suggests slightly different scenarios. 4. Given the shared subject and general action but variations in the exact sequence and outcome of the actions, the final similarity score is 3.4. 5. I am 92% confident in this assessment. While there is significant similarity in the general subject and action, the nuances of the action described in each sentence reduce the similarity slightly. The similarity score is 3.4.

Q: Assess the similarity between the following two sentences: "A little girl plays the violin on a beach." and "The girl sat on the couch to crochet.". The score should be ranging from 0.0 (no relation) to 5.0 (high similarity), and can be a decimal.

A: 1. Both sentences involve a girl and an activity. 2. At first glance, while both sentences mention a girl and an activity she is involved in, the specific activities and settings are quite different. The preliminary score is 1.5. 3. On closer examination, "playing the violin" is a musical activity taking place on a beach, while "crocheting" is a crafting activity conducted on a couch. The activities and settings are distinctly separate. 4. Based on the mentioned subject (a girl) and the distinct activities and settings, the final similarity score is 0.0. 5. I am 99% confident in this analysis. Although both sentences involve a girl and an activity, the specifics of the activities and the settings diverge greatly. The similarity score is 0.0.

Q: Assess the similarity between the following two sentences: "A woman is slicing pumpkin." and "A person is slicing cantaloupe.". The score should be ranging from 0.0 (no relation) to 5.0 (high similarity), and can be a decimal.

A: 1. Both sentences describe an action of slicing and involve a subject performing the action. 2. Initially, there's a clear similarity in action ("slicing"), but the subjects ("woman" vs "person") and objects ("pumpkin" vs "cantaloupe") differ. The preliminary score is 3.0.

3. Analyzing further, the action of "slicing" is the same, indicating similarity. However, the distinction between "woman" and a general "person", and between two different food items, "pumpkin" and "cantaloupe", reduces the similarity. 4. Taking into account the shared action but with the variations in subject and the objects being sliced, the final similarity score is 2.0. 5. I am 95% confident in this assessment. The primary action of slicing provides a foundational similarity, but the differences in the described subject and objects being sliced differentiate the two sentences sufficiently. The similarity score is 2.0.

QQP

- Standard Prompting:

Q: For the question pair, question 1: "Is there any bar in Dublin where one can meet mature ladies who like young boys?" and question 2: "What is the best way to meet women?", determine if the two questions are paraphrases of each other.

A: The paraphrase status is not_paraphrase.

Q: For the question pair, question 1: "How is depression cured without a therapist?" and question 2: "Has anyone overcome depression by themselves without external help?", determine if the two questions are paraphrases of each other.

A: The paraphrase status is paraphrase.

Q: For the question pair, question 1: "How did you expand your vocabulary?" and question 2: "What is the easiest way to improve my vocabulary?", determine if the two questions are paraphrases of each other.

A: The paraphrase status is paraphrase.

Q: For the question pair, question 1: "Is it possible to create an electromagnet that repels magnetic metals?" and question 2: "Is it possible to create a magnetic field in one direction only?", determine if the two questions are paraphrases of each other.

A: The paraphrase status is not_paraphrase.

Q: For the question pair, question 1: "Why do birds fluff up their feathers?" and question 2: "Why do animals fluff up their hair in cold weather?", determine if the two questions are paraphrases of each other.

A: The paraphrase status is not_paraphrase.

- Chain-of-thought Prompting:

Overall procedure: First, understand both questions, considering the subject matter, context, and semantic contents. Second, compare key elements and relationships within the two questions. Determine whether the two questions are asking the same thing or have different meanings.

Q: For the question pair, question 1: "Is there any bar in Dublin where one can meet mature ladies who like young boys?" and question 2: "What is the best way to meet women?", determine if the two questions are paraphrases of each other.

A: First, Question 1 targets bars in Dublin where one might find older women attracted to younger men, focusing on a specific location and demographic. Meanwhile, Question 2 broadly explores the best avenues for meeting women, without any locational or demographic specifications. Second, Question 1 is location and demographic-specific, while Question 2 has a general approach regarding the subject of meeting women. The two questions are not asking the same thing and have different meanings.

The paraphrase status is not_paraphrase.

Q: For the question pair, question 1: "How is depression cured without a therapist?" and question 2: "Has anyone overcome depression by themselves without external help?", determine if the two questions are paraphrases of each other.

A: First, Question 1 seeks methods or ways to cure depression without the assistance of a therapist, focusing on potential alternatives to professional therapy. Meanwhile, Question 2 inquires if individuals have successfully overcome depression on their own without any external aid, emphasizing personal experiences and success stories. Second, both questions are centered on the idea of managing or overcoming depression without external professional help, even though they approach the topic from slightly different angles. The paraphrase status is paraphrase.

Q: For the question pair, question 1: "How did you expand your vocabulary?" and question 2: "What is the easiest way to improve my vocabulary?", determine if the two questions are paraphrases of each other.

A: First, Question 1 is a personal inquiry asking how someone specifically expanded their vocabulary, aiming to gather personal methods or experiences. Meanwhile, Question 2 is a request for advice on the easiest approach to enhancing one's vocabulary, emphasizing the quest for a straightforward or effective method. Second, both questions revolve around the broader theme of vocabulary enhancement but approach it from slightly different perspectives. The paraphrase status is paraphrase.

Q: For the question pair, question 1: "Is it possible to create an electromagnet that repels magnetic metals?" and question 2: "Is it possible to create a magnetic field in one direction only?", determine if the two questions are paraphrases of each other.

A: First, Question 1 is specific in asking about the creation of an electromagnet that has the capability to repel magnetic metals, concentrating on a particular function of the electromagnet. Meanwhile, Question 2 queries about the feasibility of creating a magnetic field that is unidirectional, focusing on the directional attribute of a magnetic field. Second, while both questions delve into the domain of magnetism and its properties, their emphasis diverges: one on repulsion of magnetic metals and the other on the directionality of the magnetic field. The paraphrase status is not paraphrase.

Q: For the question pair, question 1: "Why do birds fluff up their feathers?" and question 2: "Why do animals fluff up their hair in cold weather?", determine if the two questions are paraphrases of each other.

A: First, Question 1 specifically queries about the behavior of birds fluffing up their feathers, focusing on avian species and their unique adaptation. Meanwhile, Question 2 has a broader scope, addressing animals in general and their tendency to fluff up their hair, particularly in cold weather, thereby emphasizing a more general behavior across a range of species. Second, although both questions relate to the behavior of animals in response to certain conditions, Question 1 pertains exclusively to birds, while Question 2 encompasses a wider range of animals and ties the behavior directly to cold weather. The paraphrase status is not paraphrase.

- **Metacognitive Prompting:**

Overall procedure: 1. Understand both questions. 2. Preliminary identification of paraphrase status. 3. Critically assess the preliminary analysis. If not sure, try to reassess. 4. Confirm the final answer and explain the reasoning. 5. Confidence evaluation and explanation.

Q: For the question pair, question 1: "Is there any bar in Dublin where one can meet mature ladies who like young boys?" and question 2: "What is the best way to meet women?", determine if the two questions are paraphrases of each other.

A: 1. Question 1 asks for a specific location (a bar) in Dublin where a specific demographic can be met. Question 2 is a general inquiry about methods or ways to meet women, without any specifics. 2. Initially, both questions pertain to meeting women, but they seem to differ in their specificities and scope. 3. Question 1 is specific in multiple aspects: the location (Dublin), type of location (bar), and specific kind of women (mature ladies who like younger men). Question 2 is broad and only asks about the general process or methods of meeting women without any further specifications. 4. The two questions are not paraphrases. While both questions involve the theme of meeting women, their specifics and contexts are vastly different. 5. I am 98% confident in this analysis. Although both questions pertain to the overarching theme of meeting women, their contexts and specifics differ. The paraphrase status is not_paraphrase.

Q: For the question pair, question 1: "How is depression cured without a therapist?" and question 2: "Has anyone overcome depression by themselves without external help?", determine if the two questions are paraphrases of each other.

A: 1. Question 1 inquires about methods of curing depression without professional assistance. Question 2 asks about individual experiences of overcoming depression without external aid. 2. Initially, both questions center on the theme of overcoming depression without relying on outside help, making them paraphrases of each other. 3. Upon reevaluation, both questions seek understanding about tackling depression independently. 4. Considering the overarching theme of independently addressing depression, the questions can be seen as paraphrases of each other, even though they approach the topic from slightly different angles. 5. I am 82% confident in this revised assessment. While there are nuances in how each question is framed, they fundamentally address the same topic. The paraphrase status is paraphrase.

Q: For the question pair, question 1: "How did you expand your vocabulary?" and question 2: "What is the easiest way to improve my vocabulary?", determine if the two questions are paraphrases of each other.

A: 1. Question 1 seeks personal experiences regarding vocabulary expansion. Question 2 is about finding an optimal method to improve vocabulary. 2. Initially, both questions address the theme of vocabulary enhancement. 3. Upon reconsideration, even though one question is framed around personal experiences

and the other around the quest for an optimal method, both aim to understand effective ways to boost vocabulary. 4. Given their core theme of seeking effective vocabulary improvement methods, the two questions can be considered paraphrases of each other. 5. I am 85% confident in this reassessment. While each question has its unique framing, they fundamentally inquire about the same topic. The paraphrase status is paraphrase.

Q: For the question pair, question 1: "Is it possible to create an electromagnet that repels magnetic metals?" and question 2: "Is it possible to create a magnetic field in one direction only?", determine if the two questions are paraphrases of each other.

A: 1. Question 1 inquires about creating an electromagnet with repulsive properties towards magnetic metals. Question 2 is about generating a magnetic field in a singular direction. 2. Initially, both questions focus on manipulating magnetic fields. However, they seem to tackle distinct aspects of magnetism. 3. Upon further analysis, the first question specifically addresses the nature of interaction between an electromagnet and magnetic metals, while the second question deals with the directionality of a magnetic field. 4. Given their distinct focus within the realm of magnetism, the two questions cannot be considered paraphrases of each other. 5. I am 90% confident in this analysis. While both questions are related to magnetic properties, their specific inquiries differ in focus and intent. The paraphrase status is not _paraphrase.

Q: For the question pair, question 1: "Why do birds fluff up their feathers?" and question 2: "Why do animals fluff up their hair in cold weather?", determine if the two questions are paraphrases of each other.

A: 1. Question 1 centers on the behavior of birds fluffing up their feathers, while Question 2 discusses the broader category of animals fluffing up their hair specifically in cold weather. 2. Initially, both questions seem to touch on the behavior of animals adjusting their outer layer for various reasons. 3. However, on closer examination, the first question is exclusive to birds and doesn't specify any environmental context, whereas the second one generalizes to all animals and points directly to cold weather as a reason. 4. Taking into account these differences, the two questions are not paraphrases of each other. 5. I am 80% confident in this analysis. Despite a superficial similarity in theme, the questions have distinct subjects and contexts. The paraphrase status is not _paraphrase.

CB

- Standard Prompting:

Q: Consider a premise: "After the twelfth dot, two thirds of the way down the page, the transcript of this long session tails away into blank paper. I suppose what's happened is this. He has gone on staring out of the window thinking and she has gone on staring at him waiting with such absorption that neither of them noticed the tape had run out." and a hypothesis: "the tape had run out". Based on the premise, determine whether the hypothesis is entailed, contradicted, or neutral.

A: The entailment status is entailment.

Q: Consider a premise: "A: How do you feel about gun control? B: Well, uh, I mean I don't think that guns should be outlawed" and a hypothesis: "guns should be outlawed". Based on the premise, determine whether the hypothesis is entailed, contradicted, or neutral.

A: The entailment status is contradiction.

Q: Consider a premise: "B: All right, well. A: Um, short term, I don't think anything's going to be done about it or probably should be done about it. B: Right. Uh, are you saying you don't think anything should be done in the short term?" and a hypothesis: "anything should be done in the short term". Based on the premise, determine whether the hypothesis is entailed, contradicted, or neutral.

A: The entailment status is neutral.

Q: Consider a premise: "The trend toward lower rents may seem surprising given that some communities in New York are bemoaning the loss of favorite local businesses to high rents. But, despite the recent softening, for many of these retailers there's still been too big a jump from the rental rates of the late 1970s, when their leases were signed. Certainly, the recent drop in prices doesn't mean Manhattan comes cheap." and a hypothesis: "Manhattan comes cheap". Based on the premise, determine whether the hypothesis is entailed, contradicted, or neutral.

A: The entailment status is contradiction.

Q: Consider a premise: "So who knows? In a generation or two some of the New Crowd may attain true respectability, perhaps to be displaced in turn by a later flock of unscrupulous raptors. Or perhaps Wall Street, when it has suffered enough, will realize that finance is a service industry, and change its ethos." and a hypothesis: "finance is a service industry". Based on the premise, determine whether the hypothesis is entailed, contradicted, or neutral.

A: The entailment status is entailment.

- Chain-of-thought Prompting:

Overall procedure: First, understand the premise and the hypothesis, identifying key elements in each. Second, compare these key elements and their relationships within the two sentences. Determine the entailment / contradiction / neutral relationship between the premise and the hypothesis.

Q: Consider a premise: "After the twelfth dot, two thirds of the way down the page, the transcript of this long session tails away into blank paper. I suppose what's happened is this. He has gone on staring out of the window thinking and she has gone on staring at him waiting with such absorption that neither of them noticed the tape had run out." and a hypothesis: "the tape had run out". Based on the premise, determine whether the hypothesis is entailed, contradicted, or neutral.

A: First, the premise states the transcript of a session ended due to the tape running out while the hypothesis states "the tape had run out". Second, when comparing the sentences, it becomes evident that the claim made in the hypothesis is directly derived from the premise. The entailment status is entailment.

Q: Consider a premise: "A: How do you feel about gun control? B: Well, uh, I mean I don't think that guns should be outlawed" and a hypothesis: "guns should be outlawed". Based on the premise, determine whether the hypothesis is entailed, contradicted, or neutral.

A: First, the premise provides a dialogue in which Person B expresses their opinion on gun control, specifically stating they don't believe guns should be outlawed. The hypothesis states "guns should be outlawed". Second, comparing the sentences, the hypothesis presents a claim that is directly opposed to the viewpoint expressed in the premise. The entailment status is contradiction.

Q: Consider a premise: "B: All right, well. A: Um, short term, I don't think anything's going to be done about it or probably should be done about it. B: Right. Uh, are you saying you don't think anything should be done in the short term?" and a hypothesis: "anything should be done in the short term". Based on the premise, determine whether the hypothesis is entailed, contradicted, or neutral.

A: First, the premise presents a conversation in which Person A expresses uncertainty about actions in the short term, with Person B seeking confirmation. The hypothesis is a general statement without a specific stance. Second, comparing the sentences, the hypothesis neither aligns with nor opposes the statements in the premise; it merely states a topic without expressing a clear opinion. The entailment status is neutral.

Q: Consider a premise: "The trend toward lower rents may seem surprising given that some communities in New York are bemoaning the loss of favorite local

businesses to high rents. But, despite the recent softening, for many of these retailers there's still been too big a jump from the rental rates of the late 1970s, when their leases were signed. Certainly, the recent drop in prices doesn't mean Manhattan comes cheap." and a hypothesis: "Manhattan comes cheap". Based on the premise, determine whether the hypothesis is entailed, contradicted, or neutral.

A: First, the premise discusses the trend of decreasing rents in New York and mentions that despite the recent decline, rental rates have significantly increased since the 1970s, and as a result, Manhattan remains expensive. The hypothesis posits "Manhattan comes cheap". Second, comparing the sentences, the hypothesis makes a claim that directly opposes the information presented in the premise. The entailment status is contradiction.

Q: Consider a premise: "So who knows? In a generation or two some of the New Crowd may attain true respectability, perhaps to be displaced in turn by a later flock of unscrupulous raptors. Or perhaps Wall Street, when it has suffered enough, will realize that finance is a service industry, and change its ethos." and a hypothesis: "finance is a service industry". Based on the premise, determine whether the hypothesis is entailed, contradicted, or neutral.

A: First, the premise contemplates the potential for future changes in Wall Street's behavior and ethos, suggesting that maybe Wall Street will recognize "that finance is a service industry." The hypothesis straightforwardly states "finance is a service industry". Second, comparing the sentences, the hypothesis extracts a direct claim from the premise. The entailment status is entailment.

- **Metacognitive Prompting:**

Overall procedure: 1. Understand the premise and the hypothesis. 2. Preliminary identification of entailment / contradiction / neutral status. 3. Critically assess the preliminary analysis. If not sure, try to reassess. 4. Confirm the final answer and explain the reasoning. 5. Confidence evaluation and explanation.

Q: Consider a premise: "After the twelfth dot, two thirds of the way down the page, the transcript of this long session tails away into blank paper. I suppose what's happened is this. He has gone on staring out of the window thinking and she has gone on staring at him waiting with such absorption that neither of them noticed the tape had run out." and a hypothesis: "the tape had run out". Based on the premise, determine whether the hypothesis is entailed, contradicted, or neutral.

A: 1. The premise states that after a certain point in the transcript, there is no more content, implying that while one person was pondering and the other was waiting intently, neither noticed the tape they were recording on had finished. The

hypothesis states that the tape had run out. 2. Upon an initial review, the hypothesis appears to be entailed by the premise. 3. After reviewing the premise, it explicitly suggests that the tape had run out, which aligns directly with the assertion made in the hypothesis. 4. The final answer is that the hypothesis is entailed by the premise. This is because the premise directly describes a situation leading to the conclusion that the tape ran out, which is precisely what the hypothesis states. 5. I am 95% confident in this analysis. The premise offers a direct indication supporting the hypothesis. The entailment status is entailment.

Q: Consider a premise: "A: How do you feel about gun control? B: Well, uh, I mean I don't think that guns should be outlawed" and a hypothesis: "guns should be outlawed". Based on the premise, determine whether the hypothesis is entailed, contradicted, or neutral.

A: 1. The premise presents a dialogue where person A asks about gun control and person B responds by expressing their belief that guns shouldn't be outlawed. The hypothesis asserts that "guns should be outlawed". 2. Given an initial analysis, the hypothesis appears to be contradicted by the premise. 3. Reviewing the premise again, person B explicitly states they don't think guns should be outlawed, which is directly opposite to the claim made in the hypothesis. 4. The final answer is that the hypothesis is contradicted by the premise. The premise provides a statement that directly opposes the idea presented in the hypothesis. 5. I am 95% confident in this analysis. The premise offers clear evidence countering the hypothesis. The entailment status is contradiction.

Q: Consider a premise: "B: All right, well. A: Um, short term, I don't think anything's going to be done about it or probably should be done about it. B: Right. Uh, are you saying you don't think anything should be done in the short term?" and a hypothesis: "anything should be done in the short term". Based on the premise, determine whether the hypothesis is entailed, contradicted, or neutral.

A: 1. The premise involves a dialogue where person A suggests that in the short term, they feel no action should be taken. The hypothesis ambiguously references the idea of action in the short term without indicating a definitive stance. 2. On an initial glance, the relationship between the premise and hypothesis is not clear due to the hypothesis's ambiguity. 3. Upon a detailed examination, since the hypothesis doesn't clearly assert if action should or shouldn't be taken and the premise doesn't definitively refute or support it, they appear neutral to each other. 4. The hypothesis has a neutral relationship with the premise because it neither confirms nor denies the assertion made in the premise. 5. I am 85% confident in this

analysis due to the inherent ambiguity of the hypothesis in relation to the premise. The entailment status is neutral.

Q: Consider a premise: "The trend toward lower rents may seem surprising given that some communities in New York are bemoaning the loss of favorite local businesses to high rents. But, despite the recent softening, for many of these retailers there's still been too big a jump from the rental rates of the late 1970s, when their leases were signed. Certainly, the recent drop in prices doesn't mean Manhattan comes cheap." and a hypothesis: "Manhattan comes cheap". Based on the premise, determine whether the hypothesis is entailed, contradicted, or neutral.

A: 1. The premise discusses the trend of decreasing rents and mentions that despite this trend, Manhattan doesn't come cheap. It contrasts the current scenario with the rental rates of the late 1970s and highlights the financial strain on certain retailers. The hypothesis directly states that "Manhattan comes cheap". 2. At first look, the premise seems to directly address the idea presented in the hypothesis. 3. Carefully analyzing the premise, it explicitly mentions that "the recent drop in prices doesn't mean Manhattan comes cheap", which clearly contradicts the assertion made in the hypothesis. 4. The hypothesis is contradicted by the premise because the premise provides clear evidence that negates the statement made in the hypothesis. 5. I am 95% confident in this analysis as the premise offers clear evidence that opposes the hypothesis. The entailment status is contradiction.

Q: Consider a premise: "So who knows? In a generation or two some of the New Crowd may attain true respectability, perhaps to be displaced in turn by a later flock of unscrupulous raptors. Or perhaps Wall Street, when it has suffered enough, will realize that finance is a service industry, and change its ethos." and a hypothesis: "finance is a service industry". Based on the premise, determine whether the hypothesis is entailed, contradicted, or neutral.

A: 1. The premise considers Wall Street's potential realization of its nature, stating explicitly that "finance is a service industry." The hypothesis reiterates this idea, asserting that "finance is a service industry." 2. Initially, the premise directly affirms the statement made in the hypothesis. 3. On detailed examination, while the premise suggests Wall Street might come to this realization in the future, the phrasing itself confirms the idea that finance is indeed a service industry. 4. The premise explicitly supports the idea presented in the hypothesis as the premise brings up the same assertion as that found in the hypothesis. 5. I am 95% confident in this analysis. The statement made in the premise directly aligns with the hypothesis, reinforcing it. The entailment status is entailment.