

A MORE EXAMPLES ON DATASETS

A.1 LOGICAL DEDUCTION

[An example in LogicalDeduction]

Context:

The following paragraphs each describe a set of seven objects arranged in a fixed order. The statements are logically consistent within each paragraph. In a golf tournament, there were seven golfers: Ana, Eli, Mya, Amy, Joe, Mel, and Ada. Ada finished above Mel. Ada finished third-to-last. Amy finished above Ana. Mya finished second-to-last. Joe finished above Amy. Eli finished below Ana. Ana finished above Ada. Ana finished above Mya.

Question: Which of the following is true?

Options:

- A) Ana finished third.
- B) Eli finished third.
- C) Mya finished third.
- D) Amy finished third.
- E) Joe finished third.
- F) Mel finished third.
- G) Ada finished third.

Answer: A

[Chain-of-Thought Reasoning by GPT-4]

Reasoning:

From the premises, we know that: 1. Ada finished third-to-last, which means Ada finished 6th. 2. Mya finished second-to-last, which means Mya finished 6th. 3. Joe finished above Amy. 4. Amy finished above Ana. 5. Eli finished below Ana. 6. Ada finished above Mel. Now, let's try to arrange the golfers based on the given information: 1. Joe 2. Ana 3. Amy 4. Eli 5. - 6. Mya and Ada 7. Mel From the arrangement, we can see that Amy finished third. So, the answer is D) Amy finished third.

[Prediction:D] (Wrong)

[Cumulative-Reasoning by GPT-4]

Generated Propositions

Mel finished last.
Ana finished fifth.
Eli finished sixth.
Mel finished last. Ada finished fifth.
Amy finished above Ana and below Joe.
Joe finished third.

Reasoning:

From the propositions, we know that Joe finished third. So the answer is E.

[Prediction:E] (Wrong)

[DetermLR by GPT-4]

Generated Determinate Premises

Mya did not finish first.
Ada did not finish first or second.
Joe finished above Ana.
Amy finished first or second.
Amy finished above Eli.
Amy finished second. Joe finished first.
Ada finished fourth-to-last.

Reasoning:

From the premises and propositions, we know that Ada finished third-to-last, Ana finished above Ada, Ana finished above Mya, which means she finished fifth. Mya finished second-

to-last, which means she finished sixth. Since there are only seven golfers, this means that Joe must have finished first, as he finished above Amy, who finished above Ana, who finished above Eli. With Joe finishing first, Amy must have finished second, since she finished above Ana and Eli. Ana would be third, Eli would be fourth, and Mel would be the last (seventh). So the answer is A) Ana finished third.

[Prediction:A] (Right)

A.2 LOGIQA

[An example in LogiQA]

Context:

i1: Among the 7 individuals, there are 3 with a PhD., and 3 of them are female. i2: The educational levels of A, B, and C are the same, while those of F and G are different. i3: The genders of E, F, and G are the same, while the genders of A and D are different. i4: The final hire is a female PhD.

Question:Based on the statements above, which of the following can be concluded?

Options:

- A. A is a male PhD.
- B. F is a female PhD.
- C. G is not a male PhD.
- D. D is a female PhD.

Answer: D

[Chain-of-Thought Reasoning by GPT-4]

Reasoning:

From the premise, it can be inferred that the genders of E, F, and G are female. Since the final appointment is a female Ph.D. holder, the selection must be made from E, F, and G. The educational levels of A, B, and D are the same, so they cannot be Ph.D. holders because there are only 3 Ph.D. holders. Therefore, there must be two Ph.D. holders among E, F, and G. Since the educational levels of F and G are different, one of them must be a Ph.D. holder, and E is also a Ph.D. holder. Therefore, one of E, F, and G is a female Ph.D. holder. The genders of A and C are different, and since E, F, and G are females, either A or C must be male. Among options A, B, C, and D, option B is correct because F is a female Ph.D. holder. The answer is B.

[Prediction:B] (Wrong)

[Cumulative-Reasoning by GPT-4]

Generated Propositions

The genders of D,E,F are female.

A,B,C are undergraduates.

Reasoning:

From the proposition, we can deduce that E, F, and G are females. A and C have different genders, so A and C are males. A, B, and D have the same educational background, which means they are either all Ph.D. holders or all undergraduates. However, there are only 3 Ph.D. holders, so A, B, and D must all be undergraduates. This means that E, F, and G include two Ph.D. holders. Since the final appointment is a female Ph.D. holder, it means that either E or F must be a Ph.D. holder. However, the educational levels of F and G are different, so G is an undergraduate. Therefore, both E and F are Ph.D. holders. Therefore, the final answer is: A is a male undergraduate; C is a male undergraduate; E is a female Ph.D. holder; F is a female Ph.D. holder; G is a female undergraduate. The answer is C.

[Prediction:C] (Wrong)

[DetermLR by GPT-4]

Generated Determinate Premises

d1: A, B, and C have Bachelor's degrees.

d2: E, F, and G are all males.

d3: B and C are females.

d4: D and E are both PhD. holders.

d5: C is a female undergraduate.

d6: E is a male PhD. holder.

d7: D is a female PhD. holder.

Reasoning:

From the propositions, we know the answer is D. D is a female PhD.

[Prediction:D] (Right)

A.3 PROOFWRITER

[An example in ProofWriter]

Context:

The bald eagle chases the cat. The bald eagle is kind. The bald eagle sees the dog. The cat is rough. The dog is blue. The dog is red. The dog is rough. The dog sees the cat. The dog does not see the lion. The lion sees the bald eagle. If something chases the cat then the cat chases the dog. If something chases the dog then it eats the bald eagle. If something is kind and it sees the lion then it is not red. If something chases the dog then the dog chases the bald eagle. If the dog eats the bald eagle then the dog sees the bald eagle. If something is cold then it sees the lion. If something sees the cat and it is blue then the cat is kind. If something is blue and it chases the bald eagle then the bald eagle is cold.

Question:Based on the above information, is the following statement true, false, or uncertain?
The cat eats the bald eagle.

Options:

- A) True
- B) False
- C) Uncertain

Answer: A

[Chain-of-Thought Reasoning by GPT-4]

Reasoning:

The bald eagle chases the cat.2. If something chases the cat, then the cat chases the dog. (From 1, the cat chases the dog)3. If something chases the dog, then it eats the bald eagle. (We don't have information about something chasing the dog) 4. The cat eats the bald eagle. (We don't have information about the cat eating the bald eagle)Recall the Hypothesis: The cat eats the bald eagle is Unknown.

[Prediction:C] (Wrong)

[Cumulative-Reasoning by GPT-4]

Generated Propositions

The bald eagle does not eat the dog.
If the bald eagle sees the lion, it is not red.
The bald eagle is not red.
If the dog chases the cat, then the cat chases the dog.
The dog does not eat the bald eagle.

Reasoning:

we can not find any direct or indirect information about the cat eating the bald eagle. The information provided in the premises does not support or contradict the hypothesis. Judgement: Now we know that the Hypothesis is Unknown.

[Prediction:C] (Wrong)

[Deterministic Reasoning by GPT-4]

Generated Determinate Premises

The cat chases the dog.
The dog chases the bald eagle.
Something chases the cat.
The bald eagle is something that chases the cat.
The dog and the lion both see the bald eagle.
If the cat chases the dog, then the cat eats the bald eagle.
The cat eats the bald eagle.

Reasoning:

From the propositions, we know the cat eats the bald eagle. So the answer is true.

[Prediction:A] (Right)

A.4 FOLIO

[An example in FOLIO]

Context:

No digital media are analog. Every printed text is analog. All streaming services are digital. If an object is a hardcover book then it is printed text. If 1984 is a streaming service, then 1984 is a hardcover book.

Question:Based on the above information, is the following statement true, false, or uncertain?
1984 is not a streaming service

Options:

- A) True
- B) False
- C) Uncertain

Answer: A

[Chain-of-Thought Reasoning by GPT-4]

Reasoning:

From the premises, we cannot directly deduce whether the hypothesis 1984 is not a streaming service is true or false. The premises provide information about the different media formats, but there is no explicit information about the format of 1984. So the answer is C.

[Prediction:C] (Wrong)

[Cumulative-Reasoning by GPT-4]

Generated Propositions

If 1984 is a streaming service, then it is digital.

If 1984 is a streaming service, then it is not analog.

Reasoning:

The premises state that if 1984 is a streaming service, then it is a hardcover book. However, no information is provided about whether 1984 is a streaming service or a hardcover book. Since we lack information to confirm or deny the hypothesis, we cannot deduce a conclusion from the given premises. Judgement: Now we know that the Hypothesis is Unknown. So the answer is C.

[Prediction:C] (Wrong)

[Deterministic Reasoning by GPT-4]

Generated Determinate Premises

If 1984 is a streaming service, then 1984 is digital.

All streaming services are not analog.

If 1984 is a streaming service, then it cannot be a hardcover book.

Reasoning:

We know from the premises that if 1984 is a streaming service, then 1984 is a hardcover book. However, we also deduced that if 1984 is a streaming service, then it cannot be a hardcover book, as it would be digital and not analog. This contradiction leads us to conclude that 1984 is not a streaming service. Judgement: Now we know that the Hypothesis is True.

[Prediction:A] (Right)