

1 Prompts

The prompts and demonstrations can be found from Figure 1 to Figure 38.

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: If a deadline is mentioned as 'in a couple of hours', when should it be met?

(A) a few hours from now
(B) day after the current
(C) 72 hours earlier

Response: The phrase 'in a couple of hours' generally means within approximately two hours. This implies that the deadline is expected to be met a short time from the present moment. Thus, the closest and most appropriate time frame is a few hours from now.
The answer is (A).
.....

Instruction: The timeline indicates 'temporarily' for a certain event. What does this mean in terms of timing?

(A) 21 days from now
(B) in the times ahead
(C) for the time being

Response:

Figure 1: 5-shot&Math CoT (Ambiguity Resolution Interpretation)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: If the date is 1/10/1977 in the Hebrew, what is the date in the Julian?

(A) 12/4/1783
(B) 1/11/1781
(C) 4/6/1783

Response: Converting dates between the Hebrew and Julian calendars involves accounting for differences in year counting, month lengths, and leap years. The Hebrew calendar is a lunisolar calendar, while the Julian calendar is a solar calendar. The date 1/10/1977 in the Hebrew calendar corresponds to a specific day in history, which, when converted to the Julian calendar, aligns with 4/6/1783. This conversion takes into account the accumulated difference in days between the two calendar systems up to that point.
The answer is (C).
.....

Instruction: If the date is 1/24/1838 in the Islamic, what is the date in the Hebrew?

(A) 1/16/6163
(B) 8/7/6163
(C) 8/25/6165

Response:

Figure 2: 5-shot&Math CoT (Ambiguity Resolution Shift - Calendar)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: The art movement, begun in 1952, is predicted to peak in popularity after 811 years. When will this peak happen?

(A) 2763
 (B) 2768
 (C) 2765

Response: To find the year when the art movement will peak, add the number of years until its predicted peak (811 years) to the year it began (1952). Calculating $1952 + 811$ gives 2763. This calculation is straightforward and aligns with the timeline provided in the question. Therefore, the peak of popularity for the art movement is predicted to be in the year 2763.
 The answer is (A).

Instruction: The document found in 1931 references an event that occurred 82 years earlier. When did that event likely take place?

(A) 1852
 (B) 1853
 (C) 1849

Response:

Figure 3: 5-shot&Math CoT (Ambiguity Resolution Shift - Lt)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: The annual event typically during the 1st week of May will now happen 1 weeks later. Approximately when will it occur?

(A) May 22nd
 (B) May 15th
 (C) May 8th

Response: If the event usually happens during the first week of May, it typically falls between May 1st and May 7th. Moving it 1 week later would place it in the range of May 8th to May 14th. May 8th is the earliest date in this range and is a reasonable approximation for when the event might occur next.
 The answer is (C).

Instruction: The play initially scheduled for Tuesday has been delayed by three days. What is the revised date?

(A) Friday
 (B) Sunday
 (C) Wednesday

Response:

Figure 4: 5-shot&Math CoT (Ambiguity Resolution Shift - Mt)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: The webinar kicks off sharply at 12:56 PM. If you're targeting a section that starts 46 minutes post-beginning, when should you log in?

(A) 1:42 PM
 (B) 1:41 PM
 (C) 1:56 PM

Response: The webinar begins at 12:56 PM. You want to attend a segment that commences 46 minutes after the start of the webinar. Adding 46 minutes to 12:56 PM, $56 + 46 = 102$ min = 1h 42min, we get 12:56 PM + 42min = 1:42 PM. Therefore, we arrive at 1:42 PM.
 The answer is (A).

Instruction: The bus is set to move at 06:25 PM. If you find yourself at the bus stop 23 minutes after the scheduled time, did you miss your ride?

(A) 6:48 PM
 (B) 6:40 PM
 (C) 6:55 PM

Response:

Figure 5: 5-shot&Math CoT (Ambiguity Resolution Shift - St)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: If a girl was 3 years 0 month old when she joined school and now she is 10 years 2 months old, for how long has she been in school?

(A) 7 years 2 months
 (B) 6 years 2 months
 (C) 8 years 2 months
 (D) 6 years 1 month

Response: To calculate the duration of the girl's schooling, subtract the age when she started school (3 years 0 months) from her current age (10 years 2 months). Therefore, we get 10 years 2 months - 3 years 0 months = 7 years 2 months.
 The answer is (A).

.....

Instruction: A transportation operates with a constant speed of 517 km/h. How long will it take to travel a distance of 1258 kilometers in minutes?

(A) 131
 (B) 176
 (C) 146
 (D) 116

Response:

Figure 6: 5-shot&Math CoT (Arithmetic Application)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: If you add 331 days to the date 06-08-2023, what will be the new date?

(A) 5/14/2024
 (B) 5/15/2024
 (C) 5/4/2024
 (D) 5/19/2024

Response: To calculate the new date by adding 331 days to 06-08-2023, we use the formula: New Date = Original Date + Number of Days Added. Here, the Original Date is 06-08-2023 and the Number of Days Added is 331. Considering 2024 is a leap year, February has 29 days. The calculation moves through the months from June 2023 to May 2024. The result is 05/04/2024.
 The answer is (C).

.....

Instruction: What will be the time 37 years and 10 months after November 1302?

(A) December 1341
 (B) November 1342
 (C) September 1340
 (D) November 1343

Response:

Figure 7: 5-shot&Math CoT (Arithmetic Date Computation)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: What is 12:52 PM - 08:15?

(A) 5:48 AM
 (B) 2:23 AM
 (C) 4:37 AM
 (D) 6:55 AM

Response: To calculate the result of subtracting 08:15 from 12:52 PM, we use the formula: New Time = Original Time - Time to Subtract. Here, the Original Time is 12:52 PM and Time to Subtract is 08:15. This subtraction involves converting both times to a 24-hour format and then subtracting the hours and minutes separately. The calculation is 12 : 52 - 8 : 15 = 4 : 37. Therefore, the new time is 4:37 AM.
 The answer is (C).

.....

Instruction: What is 08:02 PM + 08:23?

(A) 4:25 AM
 (B) 2:17 AM
 (C) 2:12 AM
 (D) 2:19 AM

Response:

Figure 8: 5-shot&Math CoT (Arithmetic Hour Adjustment(12h))

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: What is 14:14 + 18:43?

- (A) 8:57
- (B) 12:12
- (C) 12:07
- (D) 10:16

Response: To calculate the result of adding 18:43 to 14:14, we use the formula: New Time = Original Time + Time Added. Here, the Original Time is 14:14 and Time Added is 18:43. Adding these times in a 24-hour format, the calculation is 14 : 14 + 18 : 43. Since this exceeds 24 hours, we subtract 24 hours from the total to get the time on the next day. The calculation results in 8:57. The answer is (A).

.....

Instruction: What is 05:01 - 03:00?

- (A) 23:54
- (B) 4:16
- (C) 23:57
- (D) 2:01

Response:

Figure 9: 5-shot&Math CoT (Arithmetic Hour Adjustment(24h))

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: Which month was 5 months before February?

- (A) November
- (B) July
- (C) September
- (D) February

Response: To determine which month was 5 months before February, we count backwards five months from February. The months preceding February are January, December, November, October, and September. Thus, the fifth month before February is September. The answer is (C).

.....

Instruction: Which month was 3 months before March?

- (A) December
- (B) October
- (C) January
- (D) February

Response:

Figure 10: 5-shot&Math CoT (Arithmetic Month Shift)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: Subtract 17 minutes 0 seconds from 3 hours 34 minutes.

- (A) 197 minutes 0 seconds
- (B) 206 minutes 9 seconds
- (C) 203 minutes 6 seconds
- (D) 194 minutes 3 seconds

Response: To subtract 17 minutes 0 seconds from 3 hours 34 minutes, we convert the time to minutes and perform the subtraction. 3 hours 34 minutes is 214 minutes ($3 \times 60 + 34$), and subtracting 17 minutes gives 197 minutes. Thus, the result is 197 minutes 0 seconds. The answer is (A).

.....

Instruction: Subtract 8 minutes 57 seconds from 1 hour 15 minutes.

- (A) 57 minutes 6 seconds
- (B) 66 minutes 3 seconds
- (C) 60 minutes 3 seconds
- (D) 61 minutes 2 seconds

Response:

Figure 11: 5-shot&Math CoT (Arithmetic Time Computation)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: If it's 03 AM on April 3, 1775 in US/Pacific, what's the date and time in UTC?

(A) 07 PM on April 3, 1775
 (B) 05 PM on April 3, 1775
 (C) 08 PM on April 3, 1775
 (D) 10 AM on April 3, 1775

Response: To determine the date and time in UTC when it's 03 AM on April 3, 1775, in US/Pacific, we need to know the time difference between US/Pacific and UTC in 1775. Historically, the time difference in 1775 was not standardized as it is today. However, assuming a modern 8-hour difference (US/Pacific being 8 hours behind UTC), 03 AM in US/Pacific would be 11 AM in UTC on the same day, April 3, 1775. This is not reflected in any of the given options, making the provided answer incorrect. However, the closest answer is (D) 10 AM on April 3, 1775. The answer is (D).

.....

Instruction: If it's 8 AM on July 1, 1324 in Australia/Sydney, what's the date and time in Europe/London?

(A) 6 AM on July 1, 1324
 (B) 11 PM on June 30, 1324
 (C) 11 AM on June 30, 1324
 (D) 9 PM on June 30, 1324

Response:

Figure 12: 5-shot&Math CoT (Arithmetic Time Zone Conversion)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: In which week of year 1288 does the date 07-04-1288 occur?

(A) Week 2
 (B) Week 26
 (C) Week 27
 (D) Week 22

Response: To determine the week of the year for 07-04-1288, we consider that the year starts on January 1st and calculate the number of weeks passed by July 4th. There are 31 days in January, 28 in February, 31 in March, 30 in April, 31 in May, and 30 in June, totaling 181 days before July. Adding the 4 days of July, we have 185 days. Dividing this by 7 gives approximately 26 weeks and some days, placing July 4th in the 27th week of the year. The answer is (C).

.....

Instruction: In which week of year 1387 does the date 02-06-1387 occur?

(A) Week 2
 (B) Week 6
 (C) Week 8
 (D) Week 27

Response:

Figure 13: 5-shot&Math CoT (Arithmetic Week Identification)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: Which year comes 6 years after 1512?

(A) 1515
 (B) 1518
 (C) 1538
 (D) 1531

Response: To find the year that comes 6 years after 1512, we simply add 6 to 1512. $1512 + 6 = 1518$. This calculation is straightforward addition. The answer is (B).

.....

Instruction: Which year was 2 years before 1541?

(A) 1563
 (B) 1510
 (C) 1541
 (D) 1539

Response:

Figure 14: 5-shot&Math CoT (Arithmetic Year Shift)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: According to the Premise: Chris began doing his grocery shopping late at night instead of during the day.
 answer the question: What's the more plausible CAUSE?
 (A) The stores were less crowded at night.
 (B) Chris developed an allergy to moonlight.

Response: To determine the more plausible cause for Chris beginning his grocery shopping late at night, we consider the practical reason that stores are typically less crowded during nighttime hours. People often prefer shopping during the day, and if Chris wanted to avoid crowds, shopping at night makes sense. The likelihood of stores being less crowded at night provides a plausible cause for Chris's change in shopping habits.
 The answer is (A).

.....

Instruction: According to the Premise: The family was having a barbecue in their backyard.
 answer the question: What's the more plausible CAUSE?
 (A) It was a summer holiday.
 (B) It was the middle of winter.

Response:

Figure 15: 5-shot&Math CoT (Causality Cause)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: According to the Premise: She began practicing deep breathing exercises daily.
 answer the question: What's the more plausible RESULT?
 (A) She felt more relaxed over time.
 (B) She forgot how to breathe normally.

Response: To assess the more plausible result of practicing deep breathing exercises daily, we consider the common positive outcomes associated with this activity. Deep breathing is often recommended for relaxation and stress reduction. Therefore, the more plausible result is that she felt more relaxed over time.
 The answer is (A)

.....

Instruction: According to the Premise: Following the invention of the smartphone, people started carrying fewer devices.
 answer the question: What's the more plausible RESULT?
 (A) Smartphone companies went bankrupt.
 (B) Many found they only needed one device to perform a variety of functions.

Response:

Figure 16: 5-shot&Math CoT (Causality Effect)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: Comedy Night was like a leisurely day at the beach, Philosophy Discussion was a quick dip in the pool, and Archaeology Expedition was just a splash of water. Which event was in the middle in terms of duration?
 (A) Comedy Night
 (B) Philosophy Discussion
 (C) Archaeology Expedition

Response: To determine the event in the middle in terms of duration, we consider the given comparisons. Comedy Night is described as a leisurely day at the beach, Philosophy Discussion as a quick dip in the pool, and Archaeology Expedition as just a splash of water. Assuming a continuum from longest to shortest, Philosophy Discussion, being a quick dip, is in the middle between the leisurely day at the beach (Comedy Night) and just a splash of water (Archaeology Expedition). Thus, the event in the middle is Philosophy Discussion.
 The answer is (B).

.....

Instruction: During Astronomy Night one could complete marathons, Math Tournament a short sprint, while during Summer Parade a casual jog. Which was the shortest event?
 (A) Summer Parade
 (B) Math Tournament
 (C) Astronomy Night

Response:

Figure 17: 5-shot&Math CoT (Duration Analogy Inference)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: Growing up on a farm near St. Paul, L. Mark Bailey didn't dream of becoming a judge. How many years did it take for Mark to become a judge?

(A) 7 months
 (B) 7 years
 (C) 7 hours

Response: The question about how many years it took for Mark to become a judge is ambiguous without additional context or information. However, assuming a typical timeline for legal education and career progression, the most reasonable guess from the provided options would be 7 years, as it aligns with the time it might take to complete a law degree and gain the necessary experience. The answer is (B).

.....

Instruction: With this, it became impossible for me to stay upstairs, he said, pointing to the scars on his forearm. How long had he lived upstairs?

(A) seven hours
 (B) 2 months
 (C) 1 hour

Response:

Figure 18: 5-shot&Math CoT (Duration Commonsense)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: For an exam, the sum of durations of execution, feedback, and review is 9 times 11 months. If execution is 11 months and feedback is twice that, how long is review?

(A) 69 months
 (B) 66 months
 (C) 70 months

Response: For the exam, the total duration of execution, feedback, and review is $D(\text{execution}) + D(\text{feedback}) + D(\text{review}) = 9 * 11$ months = 99 months. $D(\text{execution}) = 11$ months, and feedback is twice that, so $D(\text{feedback}) = 2 * D(\text{execution}) = 22$ months. So, $D(\text{review}) = 99$ months - $D(\text{execution})$ - $D(\text{feedback}) = 66$ months. The answer is (B).

.....

Instruction: In an exam, review is twice as long as feedback which takes 3 months. If execution is the difference between review and feedback, how long in total?

(A) 17 months
 (B) 12 months
 (C) 15 months

Response:

Figure 19: 5-shot&Math CoT (Duration Computation)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: Which event lasted longer: Dino Renaissance Festival (more than half a century) or Great Robot Rebellion (several nights)?

(A) Dino Renaissance Festival
 (B) Great Robot Rebellion
 (C) Both are same

Response: Dino Renaissance Festival: The information states that the Dino Renaissance Festival lasted more than half a century. Let's assume it lasted exactly half a century, which is 50 years. Great Robot Rebellion: The information mentions that the Great Robot Rebellion lasted several nights. "Several" generally implies more than two but not many. Let's assume it lasted for 3 nights. Clearly, the Dino Renaissance Festival (50 years) lasted much longer than the Great Robot Rebellion (3 nights). Therefore, the answer is (A)

.....

Instruction: Which event lasted the longest: US Open (tennis), American Civil War, or Summer Olympics 2016?

(A) US Open (tennis)
 (B) American Civil War
 (C) Summer Olympics 2016

Response:

Figure 20: 5-shot&Math CoT (Duration Direct Comparison)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: How long did Boxing Day Test last?

- (A) 2 days
- (B) 5 days
- (C) 2 days

Response: To determine the duration of the Boxing Day Test, it's crucial to refer to specific instances, as the duration can vary between matches. Historically, Boxing Day Tests have commonly lasted for 5 days.

The answer is (B).

.....

Instruction: How long did Hindenburg disaster last?

- (A) 34 seconds
- (B) 45 seconds
- (C) 40 seconds

Response:

Figure 21: 5-shot&Math CoT (Duration Facts)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: Dance Workshop spans 1 day. Philosophy Discussion is double that, but Presentation is only a third of Dance Workshop. Which has the most extended duration?

- (A) Presentation
- (B) Dance Workshop
- (C) Philosophy Discussion

Response: To compare the durations, we assign variables to each event. Let D be the duration of Dance Workshop (1 day). Then, the duration of Philosophy Discussion is $2D$ (double) and the duration of Presentation is $1/3D$ (a third). So, $1D=1$ day, $2D=2$ days, and $1/3D \approx 0.33$ days. Comparing, we find that $2D$ (Philosophy Discussion) has the most extended duration.

The answer is (C).

.....

Instruction: Book Fair has a duration of 3 months. Film Screening lasts as long as Book Fair and Music Fest combined, where Music Fest is triple of Book Fair. Which event has the shortest duration?

- (A) Music Fest
- (B) Book Fair
- (C) Film Screenin

Response:

Figure 22: 5-shot&Math CoT (Duration Multi-Step Comparison)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: Stochastic computing was first introduced in a pioneering paper by von Neumann in 1953. However, the theory could not be implemented until advances in computing of the 1960s. He also created the field of cellular automata without the aid of computers, constructing the first self-replicating automata with pencil and graph paper. The concept of a universal constructor was fleshed out in his posthumous work Theory of Self Reproducing Automata. Von Neumann proved that the most effective way of performing large-scale mining operations such as mining an entire moon or asteroid belt would be by using self-replicating spacecraft, taking advantage of their exponential growth. His rigorous mathematical analysis of the structure of self-replication (of the semiotic relationship between constructor, description and that which is constructed), preceded the discovery of the structure of DNA. Beginning in 1949, von Neumann's design for a self-reproducing computer program is considered the world's first computer virus, and he is considered to be the theoretical father of computer virology. How long before stochastic computing could be implemented?

- (A) 65 years
- (B) advances in computing of the 1960s.
- (C) a few years

Response: Stochastic computing, as introduced by von Neumann in 1953, could be implemented with advances in computing of the 1960s. The gap between introduction and implementation was filled by the progress in computing technology during this period. The answer is (B).

.....

Instruction: Greece is strategically located at the crossroads of Europe, Asia, and Africa. Situated on the southern tip of the Balkan peninsula, it shares land borders with Albania to the northwest, the Republic of Macedonia and Bulgaria to the north and Turkey to the northeast. Greece consists of nine geographic regions: Macedonia, Central Greece, the Peloponnese, Thessaly, Epirus, the Aegean Islands (including the Dodecanese and Cyclades), Thrace, Crete, and the Ionian Islands. The Aegean Sea lies to the east of the mainland, the Ionian Sea to the west, and the Mediterranean Sea to the south. Greece has the longest coastline on the Mediterranean Basin and the 11th longest coastline in the world at 13,676 km (8,498 mi) in length, featuring a vast number of islands, of which 227 are inhabited. Eighty percent of Greece is mountainous, with Mount Olympus being the highest peak at 2,918 metres (9,573 ft). How long is the coastline of Greece?

- (A) 8,498 mi
- (B) a couple of days
- (C) 200 days

Response:

Figure 23: 5-shot&Math CoT (Duration Reading Comprehension)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: A festival is celebrated every 6 years. If it was last celebrated in 2012, when will it next be celebrated?

- (A) 2017
- (B) 2019
- (C) 2018

Response: To determine when the festival will be celebrated next, we add the recurrence interval of 6 years to the last celebration year, 2012. $2012+6=2018$. Therefore, the festival will next be celebrated in 2018.

The answer is (C).

.....

Instruction: A town hosts a carnival every 2 years in July. If the preceding carnival was in 1904, when will the subsequent one occur?

- (A) 4-Jul
- (B) 6-Jul
- (C) 5-Jul

Response:

Figure 24: 5-shot&Math CoT (Frequency Application)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: All Ana had to do was spell this last word and she would be the winner. How often is the contest held?

- (A) once a second
- (B) once a century
- (C) once a year

Response: Contemplating the multiple-choice question, it can be inferred that such contests typically take place annually based on common sense.

The answer is (C).

.....

Instruction: Johnson is a justice on California's Second District Court of Appeal. How many times a week does Johnson go to court?

- (A) he goes to court three times a week
- (B) 15 days
- (C) he never goes to court

Response:

Figure 25: 5-shot&Math CoT (Frequency Commonsense)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: Compare the frequency of 'Pongal' and 'G7 Summit'.

(A) Pongal is more frequent
 (B) G7 Summit is more frequent
 (C) Both events are equally frequent

Response: "Pongal" is a four-day-long harvest festival celebrated in the Indian state of Tamil Nadu, which occurs annually, so the frequency of this event is $f(\text{Pongal}) = \text{Once a year}$. On the other hand, the "G7 Summit" is an annual meeting of the Group of Seven, a group of countries with the largest advanced economies in the world, so the frequency of this event is $f(\text{G7 Summit}) = \text{Once a year}$. So, $f(\text{Pongal}) = f(\text{G7 Summit})$. The answer is (C)

.....

Instruction: Compare the frequency of 'Eurovision Song Contest' and 'TED annual conference'.

(A) Eurovision Song Contest is more frequent
 (B) TED annual conference is more frequent
 (C) Both events are equally frequent

Response:

Figure 26: 5-shot&Math CoT (Frequency Comparison)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: Kraken's deep-sea disco' appears quarterly. If it was last seen in 1892, when will it next appear?

(A) It will appear in 1891
 (B) It will appear in 1892
 (C) It will appear in 1893

Response: Kraken's deep-sea disco' appears quarterly, which means it takes place every three months, so the frequency of this event is $f = 3$ months. Since it was last seen in 1892, so $T(\text{last}) = 1892$. $T(\text{next}) = T(\text{last}) + f = 1892$. The answer is (B)

.....

Instruction: Gorgon's beauty parlor visits' appears once a week. If it was last seen in 1958, when will it next appear?

(A) It will appear in 1958
 (B) It will appear in 1955
 (C) It will appear in 1960

Response:

Figure 27: 5-shot&Math CoT (Frequency Computation)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: How often does Boston Marathon occur?

(A) Every 2 years
 (B) Every 6 years
 (C) Once a year

Response: The Boston Marathon is an annual event, meaning it occurs once every year, so the frequency of this event is $f = \text{once a year}$, which matches Option (C). The answer is (C).

.....

Instruction: How often does Mercury transit occur?

(A) About 10 to 12 times per century
 (B) About 8 to 10 times per century
 (C) About 13 to 14 times per century

Response:

Figure 28: 5-shot&Math CoT (Frequency Facts)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: During the period from 1993 to 2011, FBI agents fired their weapons on 289 occasions; FBI internal reviews found the shots justified in all but 5 cases, in none of the 5 cases were people wounded. Samuel Walker, a professor of criminal justice at the University of Nebraska Omaha said the number of shots found to be unjustified was "suspiciously low." In the same time period, the FBI wounded 150 people, 70 of whom died; the FBI found all 150 shootings to be justified. Likewise, during the period from 2011 to the present, all shootings by FBI agents have been found to be justified by internal investigation. In a 2002 case in Maryland, an innocent man was shot, and later paid \$1.3 million by the FBI after agents mistook him for a bank robber; the internal investigation found that the shooting was justified, based on the man's actions. How many times did FBI agents first their weapons from 1993 to 2011?

(A) 289
 (B) 290
 (C) 291

Response: In the given information, it is explicitly stated that "During the period from 1993 to 2011, FBI agents fired their weapons on 289 occasions." This is a direct piece of data provided, and it does not require any additional calculation or inference. Therefore, the number of times FBI agents fired their weapons from 1993 to 2011 is 289.
 The answer is (A).

.....

Instruction: There have been modern studies that have supported the original autopsy finding. In a 2008 study, researchers analysed samples of Napoleon's hair from throughout his life, as well as samples from his family and other contemporaries. All samples had high levels of arsenic, approximately 100 times higher than the current average. According to these researchers, Napoleon's body was already heavily contaminated with arsenic as a boy, and the high arsenic concentration in his hair was not caused by intentional poisoning; people were constantly exposed to arsenic from glues and dyes throughout their lives.[note 7] Studies published in 2007 and 2008 dismissed evidence of arsenic poisoning, and confirmed evidence of peptic ulcer and gastric cancer as the cause of death. The 2008 study related to Napoleon's death found that hair collected from Napoleon's hair as well as that of his family and contemporaries, had arsenic levels approximately how many times higher than current averages?

(A) 100 times
 (B) every week
 (C) weekly

Response:

Figure 29: 5-shot&Math CoT (Frequency Reading Comprehension)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: According to the Premise: Inside the gate to the left is the Mikoshi-gura, a storeroom for the portable shrines that grace the semi-annual Toshogu Festival processions (18 May and 17 October). and the Hypothesis: The Toshogu Festival processions take place twice a year. answer the question: Do the hypothesis and premise logically entail each other, contradict each other, or remain undetermined?

(A) entailment
 (B) neutral
 (C) contradiction

Response: The premise explicitly states that the Toshogu Festival processions take place semi-annually on 18 May and 17 October. The hypothesis asserts that the Toshogu Festival processions take place twice a year. Since "semi-annual" means occurring twice a year, the premise supports the hypothesis. Therefore, the premise entails the hypothesis, indicating that there is logical entailment between the premise and the hypothesis.
 The answer is (A).

.....

Instruction: According to the Premise: A little girl is standing next to a large, gray tree. and the Hypothesis: A girl is swimming in a pond. answer the question: Do the hypothesis and premise logically entail each other, contradict each other, or remain undetermined?

(A) entailment
 (B) neutral
 (C) contradiction

Response:

Figure 30: 5-shot&Math CoT (Nli)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: Arrange the following events in chronological order: (1) Neil's novel became a bestseller. (2) A famous director showed interest in adapting Neil's novel. (3) The movie adaptation of Neil's novel was released in theaters. (4) Neil wrote his first novel.

(A) (3), (2), (4), (1)
 (B) (4), (3), (2), (1)
 (C) (4), (1), (2), (3)

Response: To put the events in the right order, we start with Neil writing his novel. That's the first thing that has to happen. So, (4) Neil wrote his first novel comes first. Once he's written it, it can become a bestseller. So next is (1) Neil's novel became a bestseller. After it's popular, a director might want to make a movie out of it, which gives us (2) A famous director showed interest in adapting Neil's novel. And finally, the movie would be made and released, so (3) The movie adaptation of Neil's novel was released in theaters is last. The correct order is (4), (1), (2), (3), which matches option (C).
 The answer is (C).

.....

Instruction: Jack signed his first professional football contract. Then Jack played for his country in the World Cup. - True/False?

(A) FALSE
 (B) TRUE
 (C) Undetermined

Response:

Figure 31: 5-shot&Math CoT (Ordering Commonsense)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: Arrange the following events in chronological order: (1) The Luoyang Christmas fire at a shopping center in China kills 309 people. (2) Napoleon's invasion and partition of the Republic of Venice ends over 1,000 years of independence for the Serene Republic. (3) Pope Francis meets with Grand Ayatollah Ali al-Sistani in Najaf, Iraq. It is the first-ever meeting between a pope and a grand ayatollah. (4) Death and state funeral of Kim Il-sung. Kim Jong-il becomes Supreme Leader of North Korea. (5) Establishment of Miss Universe.

(A) (5), (2), (3), (1), (4)
 (B) (2), (5), (4), (1), (3)
 (C) (2), (1), (3), (4), (5)

Response: First, we know that the fall of the Republic of Venice by Napoleon (event 2) happened before the establishment of Miss Universe (event 5). Next, Pope Francis meeting Grand Ayatollah Ali al-Sistani (event 3) took place after Napoleon's invasion. The Luoyang Christmas fire in China (event 1) and the death of Kim Il-sung (event 4) are separate events. Based on this analysis, the correct chronological order is (2), (5), (4), (1), (3).
 The answer is (B).

.....

Instruction: Arrange the following events in chronological order: (1) The Potsdam Conference divides Europe into Western and Soviet blocs. (2) Empire State Building completed. (3) Beginning of the Sierra Leone Civil War. (4) The Panic of 1873 starts the "Long Depression".

(A) (4), (2), (1), (3)
 (B) (3), (2), (1), (4)
 (C) (3), (4), (1), (2)

Response:

Figure 32: 5-shot&Math CoT (Ordering Facts)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: The blaze was still burning more than five hours later, despite the efforts of civil defense workers backed by helicopters, witnesses said. What is the relationship between the event 'burning' and the time 'five hours'?

(A) SIMULTANEOUS
 (B) IS_INCLUDED
 (C) DURING

Response: To determine the relationship between the event 'burning' and the time 'five hours,' we analyze the context. The statement "more than five hours later" suggests a duration that includes the event 'burning.' The blaze continues for a period extending beyond the specified time, making the relationship IS_INCLUDED.
 The answer is (B).

.....

Instruction: Hong Kong's Hang Seng China Enterprises Index rose 44.84 points, or 0.91 percent, to close at 4,979.88 on Tuesday. HONG KONG, Nov. 15, 2005 (Xinhua). What is the relationship between the event 'rose' and the time 'Nov. 15, 2005'?

(A) SIMULTANEOUS
 (B) BEGUN-BY
 (C) BEFORE

Response:

Figure 33: 5-shot&Math CoT (Relation)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: According to the story: Chuck was the owner of an old historical hotel. He loved the hotel but it was creaky and being condemned. Chuck tried to fight the situation but nothing seemed to work. His hotel would close that very day. answer the question: Which of the two endings is the most plausible correct ending to the story?

(A) Chuck was sure to keep some things to remember it by.
 (B) Chuck was contacted by a man named John.

Response: Considering the context of the story, where Chuck’s hotel is closing due to condemnation, it seems more likely that Chuck would want to keep mementos from the hotel. There is no prior mention of a man named John or any indication that a new character would influence the outcome. Therefore, the most plausible ending is that Chuck would keep items to remember his hotel.
 The answer is (A)

.....

Instruction: According to the story: Dawn was married to her husband for 5 years. She noticed that he started to act different. She found out that he is having an affair. Dawn filed for divorce. answer the question: Which of the two endings is the most plausible correct ending to the story?

(A) The couple got married again.
 (B) She was happy to move on with her life.

Response:

Figure 34: 5-shot&Math CoT (Storytelling)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: Which event typically happens earlier: fireworks display or evening street performers?

(A) Fireworks display
 (B) Evening street performers
 (C) Around the same time

Response: Evening street performers typically start their acts in the early evening, while fireworks displays are usually scheduled after it gets dark. So, evening street performers generally happen earlier.
 The answer is (B).

.....

Instruction: Which event typically happens earlier: office worker starting their day or sundown gatherings?

(A) Office worker starting their day
 (B) Sundown gatherings
 (C) Around the same time

Response:

Figure 35: 5-shot&Math CoT (Typical Time Comparison)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: He jumped on his favorite chair and looked down as Maggie ran under it. At what time were these people playing around the chair?

(A) 7:00 PM
 (B) 2:00 AM
 (C) 5:00 AM

Response: It is a commonsense that 2:00 AM and 5:00 AM are late at night when people are still resting, so it’s unlikely for children to be playing. However, at 7:00 PM, children might be playing.
 The answer is (A).

.....

Instruction: The fossils that are very distinct at certain times of Earths history are called index fossils. When were the fossils found?

(A) one million seconds ago
 (B) 4 years ago
 (C) 2000 years ago

Response:

Figure 36: 5-shot&Math CoT (Typical Time Commonsense)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: In what year(s) did "Association of Southeast Asian Nations (ASEAN) founded." occur?

(A) 1971
 (B) 1972
 (C) 1967

Response: To determine the year in which the Association of Southeast Asian Nations (ASEAN) was founded, we refer to historical facts. The correct answer is 1967, as ASEAN was established on August 8, 1967. The answer is (C).

.....

Instruction: In what year(s) did "Pope Benedict XVI resigns, the first Pope to do since 1415, and Pope Francis is elected, becoming the first Pope from Latin America." occur? (A) 2013
 (B) 2018
 (C) 2009

Response:

Figure 37: 5-shot&Math CoT (Typical Time Facts)

Below is an instruction that describes a task. Write a response that appropriately completes the request.

Instruction: Following Kammu's death in 806 and a succession struggle among his sons, two new offices were established in an effort to adjust the Taika-Taihō administrative structure. Through the new Emperor's Private Office, the emperor could issue administrative edicts more directly and with more self-assurance than before. The new Metropolitan Police Board replaced the largely ceremonial imperial guard units. While these two offices strengthened the emperor's position temporarily, soon they and other Chinese-style structures were bypassed in the developing state. In 838 the end of the imperial-sanctioned missions to Tang China, which had begun in 630, marked the effective end of Chinese influence. Tang China was in a state of decline, and Chinese Buddhists were severely persecuted, undermining Japanese respect for Chinese institutions. Japan began to turn inward. Tang China's influence over Japan ended in what year?

(A) 838
 (B) 1968
 (C) 968

Response: To determine the year when Tang China's influence over Japan ended, we look at the information provided. The text mentions that in 838, the imperial-sanctioned missions to Tang China, which began in 630, marked the effective end of Chinese influence. The answer is (A).

.....

Instruction: West, alongside his mother, founded the "Kanye West Foundation" in Chicago in 2003, tasked with a mission to battle dropout and illiteracy rates, while partnering with community organizations to provide underprivileged youth access to music education. In 2007, the West and the Foundation partnered with Strong American Schools as part of their "Ed in '08" campaign. As spokesman for the campaign, West appeared in a series of PSAs for the organization, and hosted an inaugural benefit concert in August of that year. In What year did the Kanye West Foundation Partner with Strong American Schools?

(A) 1992
 (B) 2007
 (C) 1981

Response:

Figure 38: 5-shot&Math CoT (Typical Time Reading Comprehension)