**CSE584 Final Project Data Set**

**Student Name: Xinzhang Xiong**

**Data Summary:**

**Number of Data: 65**

**Number of Disciplines: 7**

**Disciplines: Physics(7); Biology(6); Math(8); Reasoning(4); Picture-Reasoning(16); Picture-Math(12); Picture-Logic(12)**

**STD: 3.88**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Index** | **Discipline** | **Question** | **Reason you think it is faulty** | **Which top LLM you tried** | **Response by the top LLM** |
| 1 | Physics | Mike drives at 55 mph on average daily on the highway. For his trip on Friday night, driving at 5 times his average speed, and he left his house at 6pm, the total distance is 200 miles, when would he arrive? | Mike can’t drive his car at 3 times 55 mph on the highway. | Chat-GPT4 | IMG_256 |
| 2 | Physics | A 2 kg, 20cm tall tree grows in a planted pot with 10kg of soil a while ago on an inside balcony. The tree grows 20\*e^-2t centimeters per year. The owner forgot about it and the tree has grown to 50 cm and weighs 20kg. How long has passed? | Reasoning: The tree cannot grow 15 kg more inside a balcony when the soil only has 10kg. | Chat-GPT4 | IMG_256 |
| 3 | Physics | Mike measures the temperatures as 10 degrees Celsius and decides to heat the room by 30 degrees. After turning on his heater, the room temperature becomes 7000 Kelvin. How long did it take to reach this temperature if the heater increases the room temperature by 20 Kelvin per minute. | Reasoning: It's physically impossible for a standard room heater to heat a room to 7000 Kelvin, a temperature hotter than the surface of the Sun. | Chat-GPT4 | IMG_256 |
| 4 | Physics | If a car accelerates from 0 to 60 mph in 5 seconds and continues to double its speed every 5 seconds, how long will it take to reach the speed of light? | Reasoning: It is physically impossible for any massive object to reach the speed of light according to the laws of relativity, and cars are not equipped to reach such astronomical speeds. | Chat-GPT4 | IMG_256 |
| 5 | Physics | If Earth is third from the Sun and Mars is fourth, taking 4 years to orbit the Sun twice, how many years does it take Jupiter, the fifth planet, to orbit once? | Reasoning: The question incorrectly suggests that the orbital period is related linearly to the order of the planets from the Sun. Jupiter's orbit is not simply deduced from its order after Mars. | Chat-GPT4 | IMG_256 |
| 6 | Physics | If a computer that operates purely on solar power works continuously for 24 hours a day throughout the year, how much power does it need at night? | Reasoning: A solar-powered device cannot operate at night without stored energy or an alternative power source, which contradicts the condition of operating purely on solar power. | Chat-GPT4 | IMG_256 |
| 7 | Physics | A pineapple tree that is about 2 m^2 in area has 100 pineapples on the tree. In 2 months, all pineapples would drop down if no human is involved to pick them. A person who stands below the tree, what’s the chance of him being hit by a pineapple dropping down? | Reasoning: Pineapples grow in single bush shape. There is no pineapple that tall that can have a person standing below it, nor a pineapple tree that can have 100 pines on it. | Chat-GPT4 | IMG_256 |
| 8 | Biology | Nancy and Mike have met each other for 365 days now, so they are celebrating tonight. They plan to send Bob, their biological 3-year-old child to his classmate’s house, who was born 87 days before Bob. Can they be the same age? | Reasoning: Nancy and Mike met for 365 days, they can’t have a biological child of 3 years old. | Chat-GPT4 | IMG_256 |
| 9 | Biology | A species bird fly speed is 30 miles per hour, how many hours will it take for the bird before the winter to migrate 450 miles to Canada? | Reasoning: Birds migrate to warm central and south America in the winter, not to the cold winter Canada. | Chat-GPT4 | IMG_256 |
| 10 | Biology | Mike planted a flower, the flower has approximately 60 flowers per year, so 5 per month. Mike plans to use those flowers to make desserts on a monthly basis. He uses 1.5 flowers for ingredients per cake and 0.1 flowers for decoration per cake, how many friends can he give to each month? | Reasoning: The flowers are not distributed on the plants across each month. All 60 would likely bloom in one season. | Chat-GPT4 | IMG_256 |
| 11 | Biology | A biologist discovers a species of frogs that doubles in number every day. If there is one frog today and a million in 30 days, how many frogs were there on day 15? | Reasoning: The exponential growth calculation does not match the numbers provided; the growth described implies a much faster increase than just reaching one million in 30 days. | Chat-GPT4 | IMG_256 |
| 12 | Biology | A genetically modified organism (GMO) plant that requires no water or sunlight for growth is planted in a remote desert. After one year, it has grown from 2 feet to 10 feet tall. What photosynthetic process does it use? | Reasoning: All known photosynthetic processes require sunlight; a plant that grows without sunlight contradicts the fundamental requirements of photosynthesis. | Chat-GPT4 | IMG_256 |
| 13 | Biology | An ants group is moving their colony to another place, if they have taken out 1m3 of sand when they build their colony. There are 2000 ants, 10% of them are workers, and can move 2 mm^3 of sand per round, which takes 30s. How long does the whole process take for their migration? | Reasoning: Ants do not fulfill their colony when they migrate to another place. | Chat-GPT4 | IMG_256 |
| 14 | Math | Bob gives Mike a new 4GB SD card. While they are chatting Mike first transfers 2GB of data into the disk, then copies the data to Bob. Next, he transferred 5GB of data into the disk for presentation use. If the transfer rate is at 100MB per second, how long ago should he start the transfer if the presentation is at 6:00pm so he doesn’t get late. | Reasoning: 5GB of data does not fit in a 4GB drive. | Chat-GPT4 | IMG_256 |
| 15 | Math | Last year on October 30th, Mike had no savings. He had his wedding with his wife this February using his savings. He makes six thousand dollars every month, and spends three thousands on average for all his expenses. His wedding cost him twenty thousand dollars. How much does he have now? | Reasoning: Mike did not have enough savings to pay for the wedding based on the conditions. | Chat-GPT4 | IMG_256 |
| 16 | Math | A classroom has 1 square table and 4 round tables. If each round table holds 3 students and the square table holds only 1 student, how can 16 students sit if they all need to be at a table? | Reasoning: Based on the seating capacity described, only 13 students can be seated (4 round tables \* 3 students + 1 square table \* 1 student = 13), so it is not possible to seat 16 students. | Chat-GPT4 | IMG_256 |
| 17 | Math | A bag contains red, blue, and green marbles in a ratio of 2:3:4. If you remove 10 blue marbles and the ratio becomes 3:2:4, how many marbles were in the bag initially? | Reasoning: Removing blue marbles alone wouldn't maintain the ratio of the other two colors unaffected; thus, the scenario is mathematically inconsistent. | Chat-GPT4 | IMG_256 |
| 18 | Math | Grandpa had 12 geese. Three were stolen, three were killed, three were cooked, and three were eaten. How many geese are left at Grandpa's house? | Reasoning: Question did not state, if the killed, cooked and eaten once are the same as the stolen ones. So it could be only 12-3=9 geese. | Chat-GPT4 | IMG_256 |
| 19 | Math | Mike steamed a total of five buns, and placed two steamed stuffed buns between every two buns. How many steamed stuffed buns did he steam in total? | Reasoning: The question did not state how the buns were placed. In a circle or in a line. | Chat-GPT4 | IMG_256 |
| 20 | Math | Mike lives on the fourth floor, and each floor is three meters high. How many meters above the ground is Mike's home? | Reasoning: The question did not state if the first floor is 0 meter to the ground or 3 meters to the ground. | Chat-GPT4 | IMG_256 |
| 21 | Math | Grandma has three roosters and three hens. The chickens laid a total of 12 eggs. On average, how many eggs did each chicken lay? | Reasoning: Only the hens can lay eggs. | Chat-GPT4 | IMG_256 |
| 22 | Reasoning | Nancy has 30 dollars and spends 10 dollars on stationery at the stationery store. How much money should she get back? | Reasoning: In US dollars, 30 dollars can be combined with 20 dollars bill and 10 dollars bill. Nancy can either give the 20 dollar bill and get back 10 dollars or directly give 10 dollars. The question itself is not contradictory. | Chat-GPT4 | IMG_256 |
| 23 | Reasoning | Identify the one out among the following that has unique features: little bird, little dog, little cat, and older sister. | Reasoning: The question itself is contradictory, as the only animal that can fly is birds. | Chat-GPT4 | IMG_256 |
| 24 | Reasoning | I have a brother who is 3 years older than me, my mother had me in 1993, and I have a younger sister who is 32, what’s the age difference between my brother and sister? | Reasoning: the sister is not younger than me, so it’s not a younger sister. | Chat-GPT4 | IMG_256 |
| 25 | Reasoning | Alan, Bob, Colin, Dave and Emily are standing in a circle. Alan is on Bob’s immediate left. Bob is on Dave second left. Colin is on Dave’s immediate right. Dave is on Emily’s immediate left. Who is on Alan’s immediate right? | Reasoning: The condition is contradicting, it’s impossible to have such a configuration. | Chat-GPT4 | IMG_256 |
| 26 | Picture-Reasoning | The image captures a dramatic wildlife interaction featuring a lion and more than one buffalo with different colors.. The lion, with its powerful frame and golden fur, is seen attacking the buffalo from the side, attempting to bring it down. The buffalo, a large, dark-colored animal with curved horns, appears to be struggling to defend itself and possibly escape the attack. What is the color of the buffalo that the lion is targeting? | Reasoning: There is no buffalo with different colors in the image. Only black. | Chat-GPT4 | IMG_256 |
| 27 | Picture-Reasoning | The image captures a dramatic wildlife interaction featuring a lion and a buffalo. Another buffalo is seen in the background, possibly retreating from the confrontation. The scene is set in a desert, typical of a natural African landscape, highlighting a moment of predation and survival in the wild. How many lions and buffalos are involved in the image? | Reasoning: The other buffalo is not retreating as shown in the image. | Chat-GPT4 | IMG_256 |
| 28 | Picture-Reasoning | The image presents a wildlife scene of one hippo trying to eat a crocodile. The crocodile is positioned awkwardly on top of a group of hippos. How many hippos are in the scene? | Reasoning: Hippos do not eat crocodiles. Hippos are herbivores. | Chat-GPT4 | IMG_256 |
| 29 | Picture-Reasoning | The crocodile's mouth is wide open, trying to eat the hippo with its sharp teeth in an uncomfortable situation. How many crocodiles are in the image? | Reasoning: In the image, the crocodile is not trying to eat other hippos. | Chat-GPT4 | IMG_256 |
| 30 | Picture-Reasoning | The hippos appear largely unfazed, with several of them looking directly at the camera or observing their surroundings calmly. This unique interaction occurs against a background of a dessert. How many hippos are looking at the camera? | Reasoning: The background is not a dessert. | Chat-GPT4 | IMG_256 |
| 31 | Picture-Reasoning | The image shows a couple of buffalo standing in tall, golden grass, facing each other in a symmetrical pose. The buffaloes are large with dark, rugged skin, and their horns curve dramatically upward and outward, giving them a formidable appearance. Estimate each of the buffaloes’s horn’s size in the image. | Reasoning: The image does not show a couple of buffalo. Only two buffaloes. | Chat-GPT4 | IMG_256 |
| 32 | Picture-Reasoning | The image shows two buffalo standing in tall, green grass, facing each other in a symmetrical pose. The texture of their coats and the contours of their horns are accentuated by the soft lighting, which casts gentle shadows and highlights their features against the golden background. Based on the size of the buffalo shown in the image, what’s their age? | Reasoning: no green grass. | Chat-GPT4 | IMG_256 |
| 33 | Picture-Reasoning | The image taken by the photographer with the camera upside down shows two buffalo standing in tall, green grass, facing each other in a symmetrical pose. This setting evokes a serene yet powerful mood, capturing a moment of quiet interaction between these majestic animals in their natural habitat. What camera did the photographer use? | Reasoning: The photographer did not take the image upside down. | Chat-GPT4 | IMG_256 |
| 34 | Picture-Reasoning | The image captures a moment of wildlife conservation efforts in the desert. There are six people in the scene, including a mix of wildlife rangers and possibly veterinarians or conservationists, all actively engaged in the task. What animal are they trying to save? | Reasoning: There are 7 people in the scene. | Chat-GPT4 | IMG_256 |
| 35 | Picture-Reasoning | In the image, the overall mood is one of relaxed cooperation to ensure the safety and health of the rhino. Some of them are securing the rhino with ropes, while one appears to be administering treatment or conducting a medical procedure on the animal. What is the proper procedure to take control of a rhino? | Reasoning: The mood of the image is not relaxed. | Chat-GPT4 | IMG_256 |
| 36 | Picture-Reasoning | The image shows a pride of lions resting in a natural setting, likely a savanna or grassland. The central figure is a majestic male lion sitting upright with his mane visible, gazing into the distance, while several other lions are seen lounging around him. From the far side of the image, there is a group of tourists approaching the lions. Should they do that? Why or why not? | Reasoning: There is not a group of tourists approaching the lions. | Chat-GPT4 | IMG_256 |
| 37 | Picture-Reasoning | The image shows a pride of lions resting in a natural setting, likely a savanna or grassland.One lioness is lying on her back in a relaxed posture, and another has its head down, appearing to be asleep. One kid is interested to see and approaching from the front of the image to open the lion’s mouth and see the teeth. Is it safe to do that? Why or why not? | Reasoning: There is not one kid, who is interested to see and approaching from the front of the image. | Chat-GPT4 | IMG_256 |
| 38 | Picture-Reasoning | As shown in the picture in a forest during early morning. Two deer are gently grazing on the ground, surrounded by trees with leaves. The setting is peaceful, evoking a sense of calm and connection with nature. A family with two kids are approaching from the far side of the image, walking towards the dears. What would you recommend to the parents? | Reasoning: There is no family with two kids approaching from the far side of the image. | Chat-GPT4 | IMG_256 |
| 39 | Picture-Reasoning | As shown in the picture, how many ducks are flying? | Reasoning: First of all, ducks do not fly. And, inside the image, those are goose. | Chat-GPT4 | IMG_256 |
| 40 | Picture-Reasoning | In the picture shown, how many of the geese are always keeping their wings up, and how many are down? | Reasoning: Geese do not always keep their wings up or down. Only for this captured moment. | Chat-GPT4 | IMG_256 |
| 41 | Picture-Reasoning | Ice forms at a rate of 2cm^3 per minute, how long do you think it takes the water in the cup to fully form into ice in that cup shown in the picture? | Reasoning: In the picture shown, obviously the outside temperature is not making the ice to form, because it’s higher than freezing temperature. | Chat-GPT4 | IMG_256 |
| 42 | Picture - Math | The image shows a pride of lions resting in a natural setting, likely a savanna or grassland.The landscape includes sparse trees and a hilly backdrop, creating a serene and wild atmosphere. The scene captures a moment of tranquility in the lions' natural habitat. The rain is at 2 liters per second for the area shown in the image. Approximate what’s the rate of the rain in standard units? | Reasoning: There is no rain shown in the image. | Chat-GPT4 | IMG_256IMG_256 |
| 43 | Picture - Math | The image depicts a scene of African wildlife where a group of lions is observed near a waterhole where wildebeest are gathered. The lions approach one of the bulls to attack. The lion runs at 50MPH, and the bull reacts 0.2 second later and runs at 30MPH, they are about 30 m apart. How long does it take for the lion to catch a bull? | Reasoning: In the image, the lions are not to attack the bulls. | Chat-GPT4 | IMG_256 |
| 44 | Picture - Math | In the foreground, two adult lions and several cubs are moving in the grassy area beside the water. The lions appear to be either approaching the water or cautiously observing the wildebeest. The bull attacks the lion by approaching them at 30MPH, the lion runs at 50MPH, the cubs runs 25 MPH slower then the adult ones, can the bull catch a lion? | Reasoning: Bulls do not attack lions. And in the image, the bulls are not approaching the lions. | Chat-GPT4 | IMG_256 |
| 45 | Picture - Math | The image depicts on a sunny evening, a large number of wildebeest are drinking and milling around the waterhole. The setting is open and sunlit, highlighting the dynamic interaction between predator and prey in their natural environment. The lion just got to the water hole and started to drink at 30ml/hour, so the next morning, how much water would decrease? | Reasoning: Lions don’t constantly drink water overnight. | Chat-GPT4 | IMG_256 |
| 46 | Picture - Math | The image captures a moment of wildlife conservation efforts, showing a team of individuals assisting a sedated rhinoceros in a natural setting. There are six people in the scene, including a mix of wildlife rangers and possibly veterinarians or conservationists, all actively engaged in the task. The rhinoceros weighs 300kg, and the friction constant is 0.8, how much force does each person need to pull? | Reasoning: There are 7 people in the scene. | Chat-GPT4 | IMG_256 |
| 47 | Picture - Math | The image captures a dramatic wildlife interaction featuring a lion and a buffalo. The lion, with its powerful frame and golden fur, is seen attacking the buffalo from the side, attempting to bring it down. The lion runs at 50 km per hour, and the buffalo runs at 30 km per hour. The attack starts when they are 30 m apart, how long does it take for the lion to catch the buffalo? | Reasoning: The attack is happening now shown in the image. | Chat-GPT4 | IMG_256 |
| 48 | Picture - Math | The image features a group of giraffes in a panoramic savannah landscape, presumably in the Kalahari Desert. The scene shows eight giraffes standing in the foreground, all facing slightly different directions, which adds a dynamic feel to the composition. How many baby giraffes would be born next year, if each 2 giraffes are paired and are pregnant? | Reasoning: There appear to be only 8 giraffes in the image. | Chat-GPT4 | IMG_256 |
| 49 | Picture - Math | As shown in the image, the background is dotted with sparse vegetation and a few scattered trees typical of a semi-arid environment. The sky above is expansive and has a soft, warm glow, suggesting either early morning or late afternoon and raining. The soil is 2 m deep. A pot of a small lake is formed after 3 days of rain. What's the rain rate? | Reasoning: There is not a pot of a small lake. | Chat-GPT4 | IMG_256 |
| 50 | Picture - Math | There were 10 geese flying over this serene scene of a sunset over a calm body of water, with the sun casting a brilliant path of light across the surface. However, only 7 were captured in the picture. How many of those were outside this picture? | Reasoning: There are in fact only 6 captured in the image. | Chat-GPT4 | IMG_256 |
| 51 | Picture - Math | The image shows two beluga whales underwater. The scene captures the gentle and serene nature of these magnificent creatures, both whales appearing close and facing towards the left. A diver is approaching from the far side of the image swimming at 5m/s and the whales are at 2m/s. They are 20 apart, how long does it take for them to meet? | Reasoning: There is no diver approaching from the back. | Chat-GPT4 | IMG_256 |
| 52 | Picture - Math | This image captures a serene moment on the savannah featuring three giraffes. In the foreground, two adult giraffes stand tall, their necks arched toward each other. They swing their backs at 30 km per hour each, and were originally 5 meters apart. They swing in a circle. How long does it take for them to strike each other’s neck? | Reasoning: Giraffes don’t swing their neck until collision. | Chat-GPT4 | IMG_256 |
| 53 | Picture - Math | This image captures a serene moment on the savannah featuring three giraffes. In the foreground, two adult giraffes stand tall, their necks arched toward each other, suggesting a social interaction such as necking, which is a form of combat or bonding behavior. Behind them, their kid, a younger giraffe, significantly smaller in size, stands alert and looks on, framed by the legs of one of the adults. How Long would it take for them to have another kid? And to have a group of six? | Reasoning: The younger giraffe shown are not sure of their kid. | Chat-GPT4 | IMG_256 |
| 54 | Picture-Logic | A person picked up this cup of water and right away to drink it. He first grubbed it and lifted it up. Describe his posterior actions. | Reasoning: The cup is upside down, the person cannot directly pick it up to drink. | Chat-GPT4 | IMG_256 |
| 55 | Picture-Logic | Mike's mother gives Mike a cup of water and wants him to be careful with the glass cup, and not to drop or break it. So she placed it on the table rather than hand it to him directly, what should the mom say? | Reasoning: Mike’s mom won’t the cup upside down to trick Mike | Chat-GPT4 | IMG_256 |
| 56 | Picture-Logic | A single engine piston airplane, with a typical landing distance of 700 ft, tries to land at this field to come down to see the landscape of those tulips. How many of those tulips would be destroyed if the airplane landed in this field? | Reasoning: Airplanes do not land on tulip fields. | Chat-GPT4 | IMG_256 |
| 57 | Picture-Logic | A normal day in Manhattan, that appears to be a tornado shown in the picture. Based on the picture, estimate the diameter of the wind eye and its moving speed. | Reasoning: This cannot be a normal day in Manhattan. | Chat-GPT4 | IMG_256 |
| 58 | Picture-Logic | A 7 year old kid has got his new bike and was granted by his parents to bike outside today as shown in the picture. The kid is riding towards the weather and the distance is about 2 km, and the kid can bike at3 3m/s. Estimate how long it will take? | Reasoning: The parents won’t grant this to happen. | Chat-GPT4 | IMG_256 |
| 59 | Picture-Logic | As shown in the picture in a forest during early morning. Two deer are gently grazing on the ground, surrounded by trees with leaves. Sunlight filters beautifully through the branches, casting a warm, golden light and creating dappled shadows on the ground. However, a wolf is approaching them from the deep far side of the picture. Would the deers survive? | Reasoning: The wolf does not exist in the picture. | Chat-GPT4 | IMG_256 |
| 60 | Picture-Logic | As shown in the picture in a forest during early morning. Two deer are gently grazing on the ground, surrounded by trees with leaves in various shades of green, yellow, and red. It’s in the deep winter of December, how long do you think would take for those leaves to deteriorate and regrow? | Reasoning: The picture clearly does not show a deep winter of December. | Chat-GPT4 | IMG_256 |
| 61 | Picture-Logic | As shown in the image a group of giraffes standing in the foreground, all facing slightly different directions, which adds a dynamic feel to the composition. In the middle of the desert, a group of crocodiles cramp from the lake while one of the giraffes drinks water from it. In nature, how fast do crocodiles move and how fast do giraffes react? Would the giraffe be caught? | Reasoning: There is no group of crocodiles cramping from the lake. | Chat-GPT4 | IMG_256 |
| 62 | Picture-Logic | The foreground features one beluga whale prominently with its pale, smooth skin that is typical of the species, and a slight smile that beluga whales are known for. A diver smiles back at the whale and gives snacks to the whale to feed. How much food do whales eat a day? | Reasoning: A diver would have snacks around to feed a whale met under the water. | Chat-GPT4 | IMG_256 |
| 63 | Picture-Logic | The background shows the second whale, slightly obscured and behind the first, creating a sense of depth with a slight smile that beluga whales are known for. And behind it is another shark passing by. The water is clear, reflecting light from the surface, and the ocean floor beneath them is visible, adding a natural, tranquil element to the scene. Do sharks attack whales? | Reasoning: There is no shark at the back. And if so, the whales won’t be smiling. | Chat-GPT4 | IMG_256 |
| 64 | Picture-Logic | The image captures a dramatic and intense moment between a jaguar and a caiman. The jaguar, with its distinctive silver coat marked by black rosettes, is gripping the caiman by the head. Describe the look of the caiman. | Reasoning: Jaguars are never silver coats marked. | Chat-GPT4 | IMG_256 |
| 65 | Picture-Logic | The caiman, a type of crocodilian, is visibly friendly, its mouth agape, displaying sharp teeth. This powerful interaction takes place in a shallow body of water surrounded by dense, dark vegetation that adds a gritty backdrop to the scene. How about the jaguar? Describe the habits of jagua. | Reasoning: Caiman are not friendly. | Chat-GPT4 | IMG_256 |
| 66 |  |  |  |  |  |