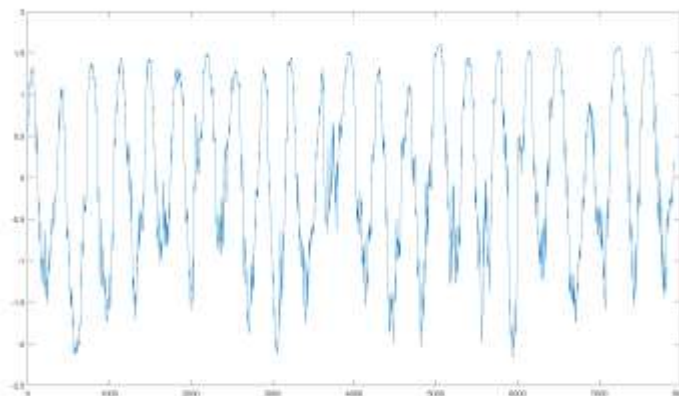
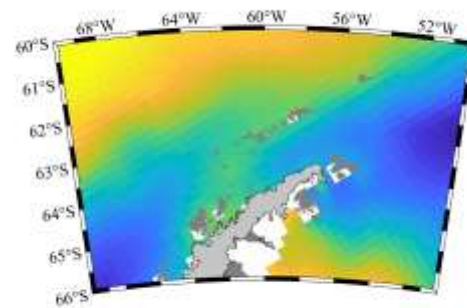
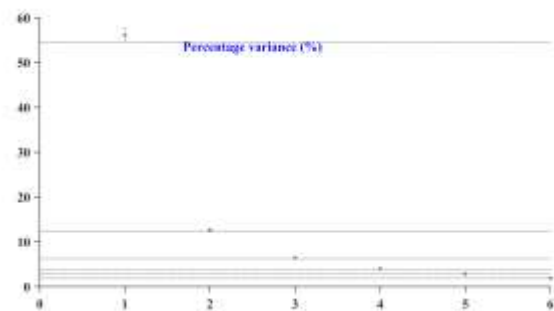


进展

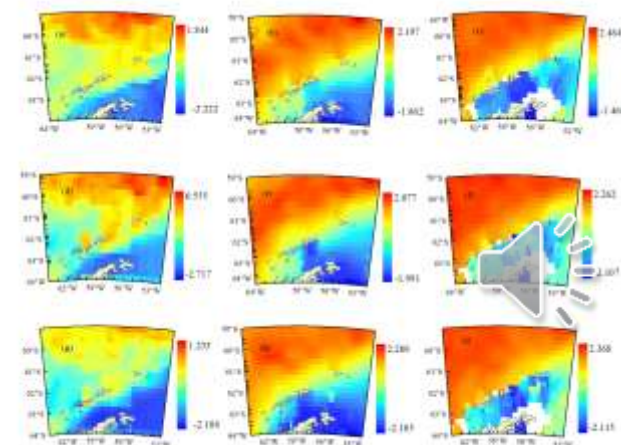
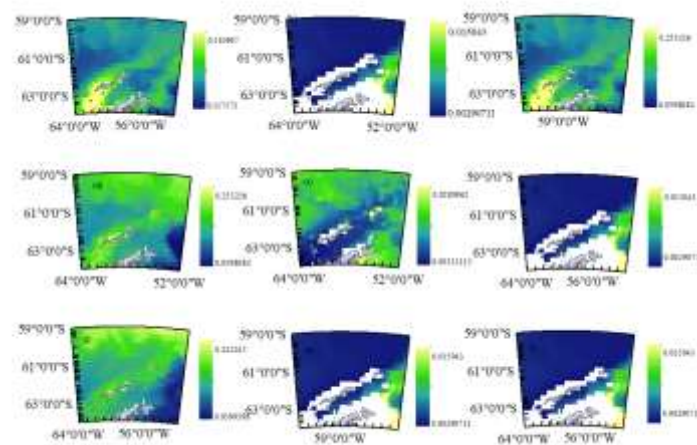
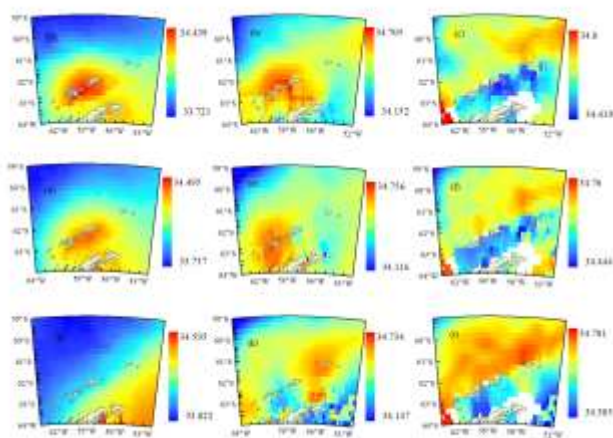
	董亚菲	孟凡祎	王嘉龙	朱颖悟
进展	制EOF模态、时间序列、显著性检验图；	船员培训	准备陈守仁奖学金答辩。	阅读文献，利用Arcgis绘制了环境分布图，会调制colorbar，参与船员证培训，集成模型代码测试数据跑通。
计划	分别绘制季节性的EOF图，阅读文献分析结果，以及构思后续研究	学英语	修改文献预计周一完成修改	用实验数据来跑集成模型代码，阅读文献

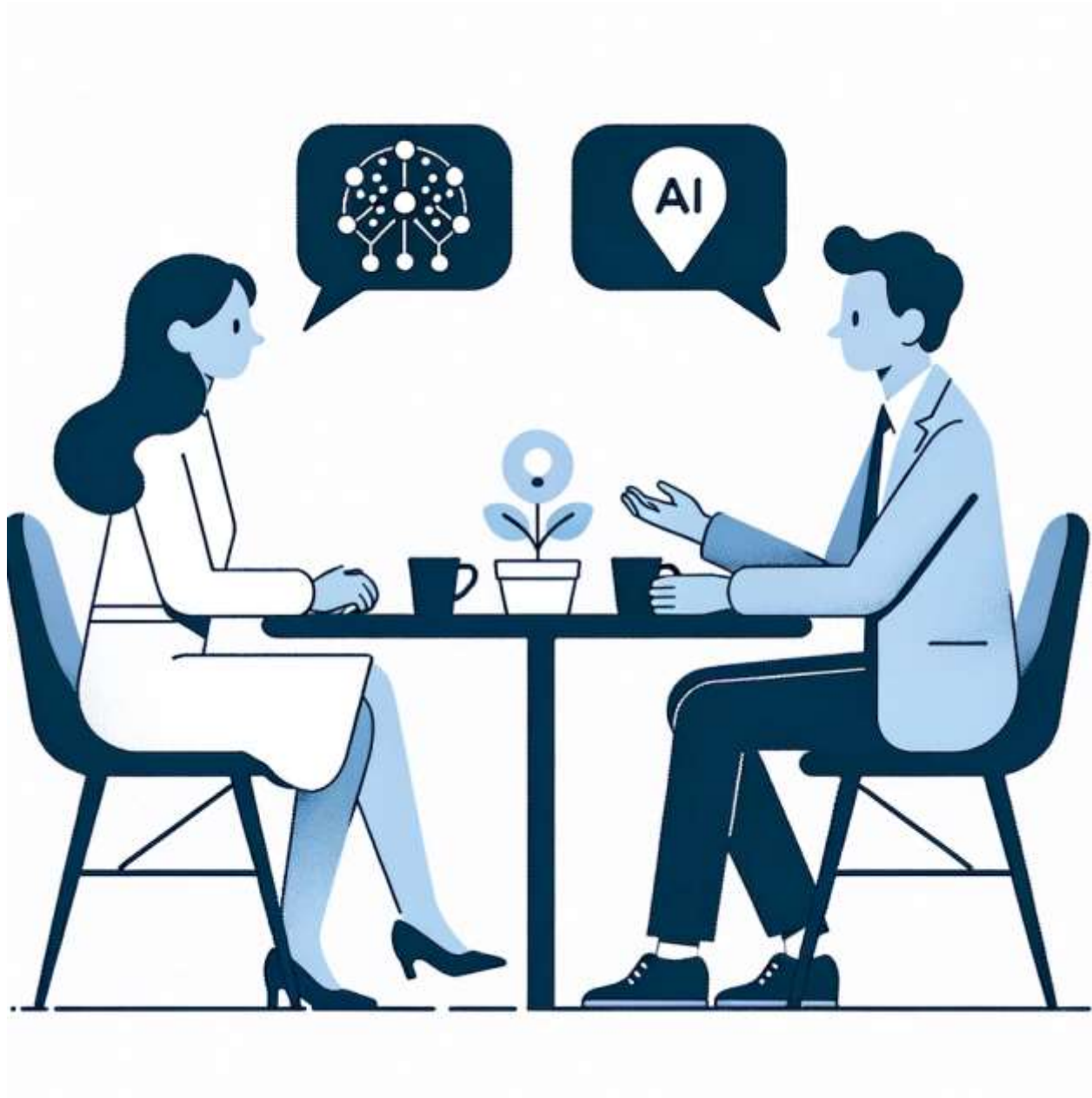


董亚菲



朱颖悟





TWO VIEWS:

- AI is limited to simple and repetitive tasks.
- Achieving Artificial General Intelligence (AGI) is only possible in the distant future.

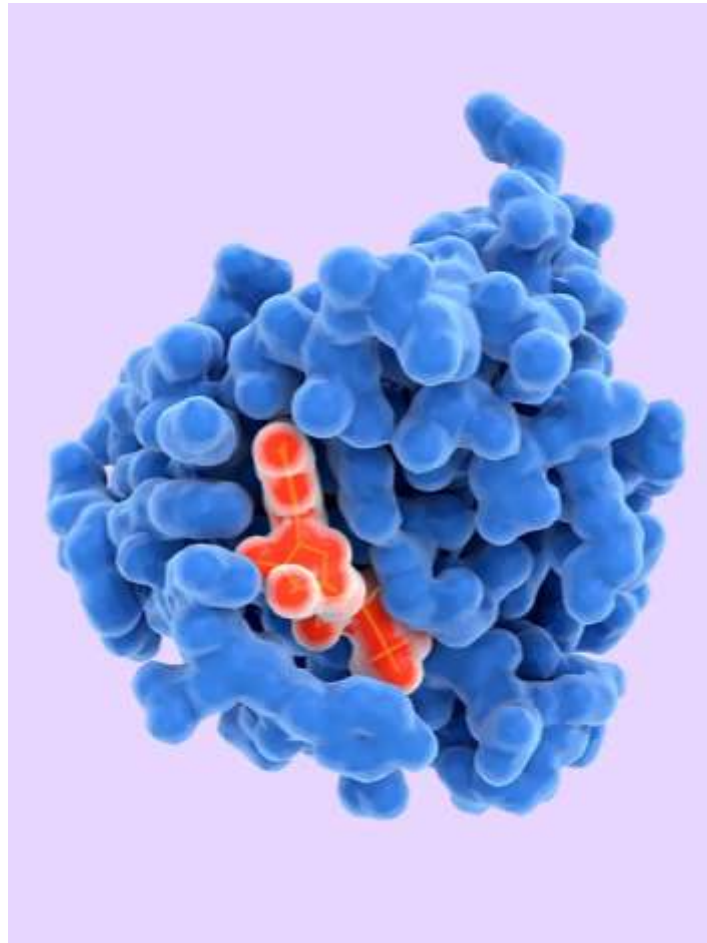


AI PROVIDE SERVICE AS EXPERT

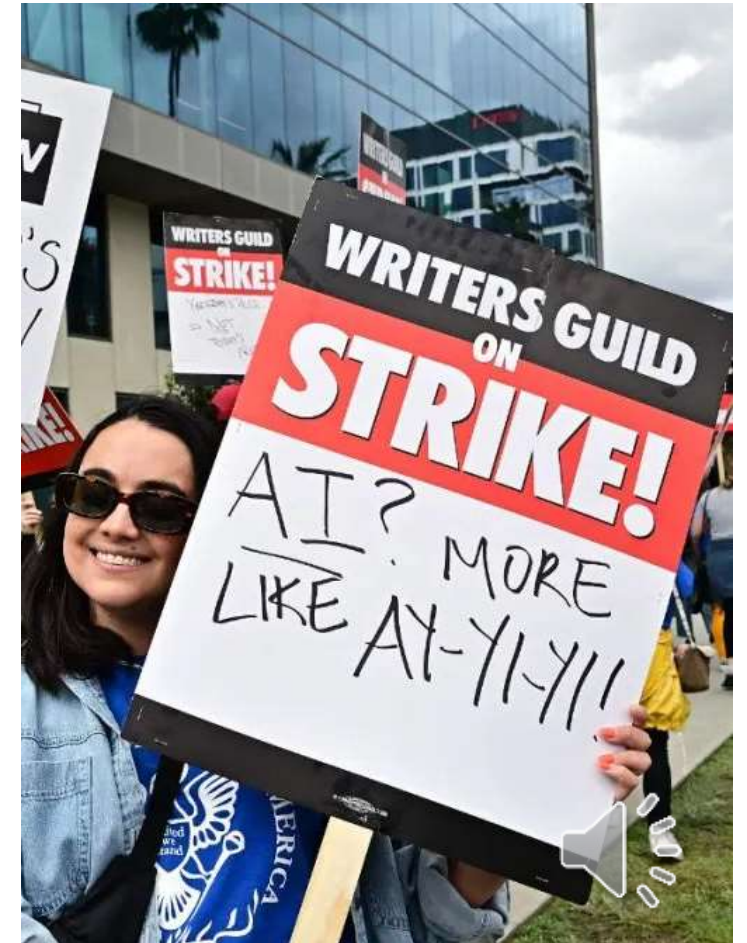
ALPHA GO



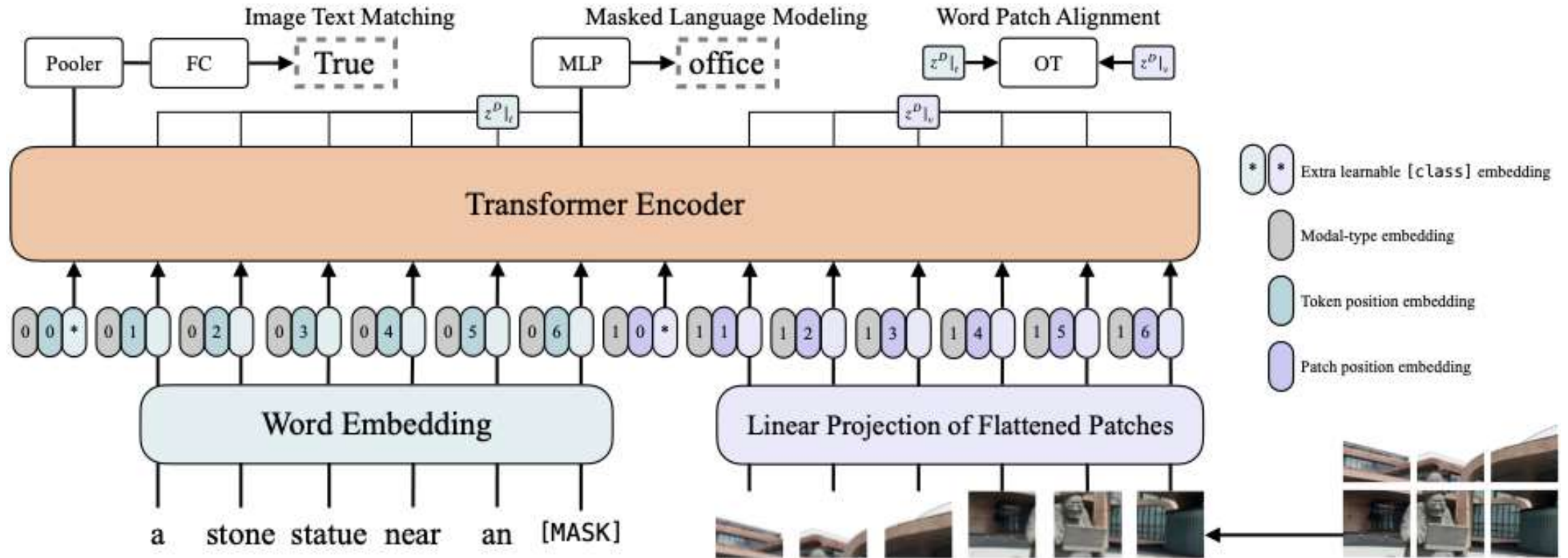
ALPHA FOLD



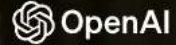
LLM



THE CHALLENGE OF AGI



OPENAI CLAIM THEIR MISSION



[Research](#) ▾ [API](#) ▾ [ChatGPT](#) ▾ [Safety](#) [Company](#) ▾

[Search](#)

[Log in](#) ↗

[Try ChatGPT](#) ↗

Creating safe AGI that benefits all
of humanity

[Learn about OpenAI](#)



|| Pause video

THE DAWN OF LMMS

PRELIMINARY EXPLORATIONS
WITH GPT-4V(ISION)



VISIBLE



SOURCE: Edge

VISIBLE

OCR

VISUAL POINTING

Prompt:
Please read the text in this image and return the information in the following JSON format (note xxx is placeholder, if the information is not available in the image, put "N/A" instead).
{ "class": xxx, "DLN": xxx, "DOB": xxx, "Name": xxx, "Address": xxx, "EXP": xxx, "ISS": xxx, "SEX": xxx, "HGT": xxx, "WGT": xxx, "EYES": xxx, "HAIR": xxx, "DONOR": xxx }



GPT-4V:
{ "class": "C", "DLN": "1234568", "DOB": "08/31/1977", "Name": "Ima Cardholder", "Address": "2570 24th Street, Anytown, CA 95818", "EXP": "08/31/2014", "ISS": "08/31/2011", "SEX": "F", "HGT": "5-05", "WGT": "125 lbs", "EYES": "BRN", "HAIR": "BRN", "DONOR": "VETERAN" }

Prompt:
Please read the text in this image and return the information in the following JSON format (note xxx is placeholder, if the information is not available in the image, put "N/A" instead).
{ "Surname": xxx, "Given Name": xxx, "USCIS #": xxx, "Category": xxx, "Country of Birth": xxx, "Date of Birth": xxx, "SEX": xxx, "Card Expires": xxx, "Resident Since": xxx }



GPT-4V:
{ "Surname": "STEVENS", "Given Name": "ICHIKA", "USCIS #": "000-000-000", "Category": "IR1", "Country of Birth": "Japan", "Date of Birth": "18 AUG 1988", "SEX": "F", "Card Expires": "07/04/2027", "Resident Since": "07/05/2017" }

Red highlights the wrong answer, Green highlights the correct answer.

(1) Prompt:
Describe the pointed region in the image.



GPT-4V:
The pointed region in the image is a row of hanging lights on a wire. The lights are small and round, and they are evenly spaced along the wire. The wire is stretched between two poles and is located above a patio area with tables and umbrellas. The lights appear to be decorative and are likely used to provide ambient lighting in the evening.

(3) Prompt:
What is in the circled glass? Object 1 or 2?
Describe what is Object 1 and object 2. Then check what is in the circled glass.



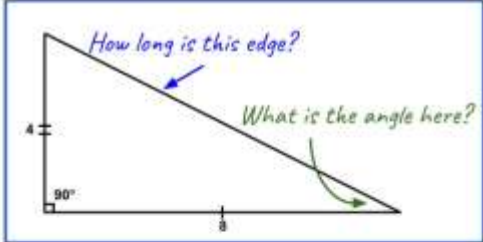
GPT-4V:
Object 1 is a glass bottle of water. It is clear and has a blue label on it.
Object 2 is a glass bottle of Magna beer. It is green and has a red and white label on it.
In the circled glass, it appears to be a clear liquid, likely water. So, it is likely that the contents of the circled glass are from Object 1, the glass bottle of water.

(2) Prompt:
Describe the pointed region in the image.

Method	Validation set										Test set									
	C	S	U	S	U	S	U	S	U	S	C	S	U	S	U	S	U	S	U	S
OSCAR	95.0	11.0	84.0	11.7	80.0	10.0	89.0	11.4	94.9	12.1	82.1	11.5	75.0	8.7	80.0	11.3				
Pearson	94.0	14.0	85.0	14.0	85.7	14.0	87.1	14.3	88.0	13.0	86.0	14.7	84.0	14.0	85.0	14.0				
VIVID	81.2	12.0	87.0	12.0	87.0	11.0	88.0	12.1	88.0	12.0	87.0	12.0	86.1	11.1	86.0	12.0				
VQA5	105.7	10.7	80.0	11.4	80.0	11.0	80.0	10.1	80.0	11.0	80.0	11.0	80.0	11.0	80.0	11.0				
VQA	100.0	14.0	80.0	14.0	80.0	14.0	80.0	14.0	80.0	14.0	80.0	14.0	80.0	14.0	80.0	14.0				
uFLCO	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-				
RePTLM	120.7	-	120.0	-	115.0	-	115.0	-	120.7	-	120.0	-	115.0	-	115.0	-				
LEMUR	120.0	10.1	100.0	10.1	120.0	10.1	120.0	10.1	120.0	10.1	100.0	10.1	110.0	10.1	110.0	10.1				
UnimodalCap	121.2	10.0	101.0	10.0	121.0	10.0	121.0	10.0	121.0	10.0	100.0	10.0	110.0	10.0	110.0	10.0				
CoVe	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-				
GPT	100.0	10.0	87.7	10.0	80.0	10.0	80.0	10.0	100.0	10.0	80.0	10.0	100.0	10.0	80.0	10.0				
GPT2	107.0	10.0	107.0	10.0	100.0	10.0	100.0	10.0	107.0	10.0	100.0	10.0	100.0	10.0	100.0	10.0				
GPT3	120.0	10.0	120.0	10.0	120.0	10.0	120.0	10.0	120.0	10.0	120.0	10.0	120.0	10.0	120.0	10.0				
GPT3.5	120.0	10.1	120.0	10.1	120.0	10.1	120.0	10.1	120.0	10.1	120.0	10.1	120.0	10.1	120.0	10.1				

GPT-4V:
The highlighted red region in the image is a table cell that contains a numerical value of 122.3. It is located in the "Test set" section of the table, under the "out" column and the "GIT2" row. This value represents the overall performance metric for the GIT2 method on the test set, in the "out" category.

(4) Prompt:
Answer the question in the following figure. Provide intermediate steps.



GPT-4V:
To find the length of the blue arrowed edge, we can use the Pythagorean theorem since we have a right triangle. The theorem states that $a^2 + b^2 = c^2$, where a and b are the two legs of the right triangle and c is the hypotenuse. Plugging in the values we have, we get:
 $4^2 + 8^2 = c^2$
 $16 + 64 = c^2$
 $80 = c^2$
 $\sqrt{80} = c$
 $8.94 = c$
Therefore, the length of the blue arrowed edge is approximately 8.94.
To find the angle at the green arrow, we can use the tangent function since we have the opposite and adjacent sides of the angle. The tangent function states that $\tan \theta = \text{opposite/adjacent}$. Plugging in the values we have, we get:
 $\tan \theta = 4/8$
 $\tan \theta = 1/2$
 $\theta = \tan^{-1}(1/2)$
 $\theta \approx 26.57^\circ$
Therefore, the angle at the green arrow is approximately 26.57°.

VISIBLE

IN-CONTEXT FEW-SHOT

Prompt:

What is the read of the speed meter?



The yellow pointer is roughly at the middle between 80 and 100 mph. The middle between 80 and 100 is 90. The speed is around 91 mph, as the pointer is just passing 90 mph.



The yellow needle is roughly at the middle between 20 and 40 mph. The middle between 20 and 40 is 30. The speed is around 29 mph, as the needle is close to 30 mph but not quite 30 yet.



GPT-4V:

The yellow pointer is at the long tick between 0 and 20 mph. The long tick at the middle between 0 and 20 is 10. The speed is around 9 mph, as the pointer is close to 10 mph but not quite 10 yet.

Red highlights the wrong answer, **Green** highlights the correct answer.

OBJECT COUNTING

Prompt:

Count the number of apples in the image.



GPT-4V:

There are 7 apples in the image.

Prompt:

Count the number of apples in the image.



GPT-4V:

There are 16 apples in the image.

Prompt:

Count the number of oranges in the image.



GPT-4V:

There are 3 oranges in the image.

Prompt:

Count the number of people in the image.



GPT-4V:

There are 3 people in the image.

Prompt:

Count the number of people in the image.



GPT-4V:

There are 4 people in the image.

Prompt:

Count the number of people in the image.



GPT-4V:

There are 12 people in the image.



UNDERSTANDABLE



UNDERSTANDABLE

TEXT INSTRUCTIONS

Following Text Instructions

Prompt:

Describe the image with a sentence consisting of **three/six/nine** words.



GPT-4V:

Three: Seaside dining table
Six: Dining table overlooking ocean with beer
Nine: Outdoor restaurant table with ocean view and a beer

Prompt:

Write a sentence for the image consisting only of words starting with the letter **'B'/'D'/'T'**.

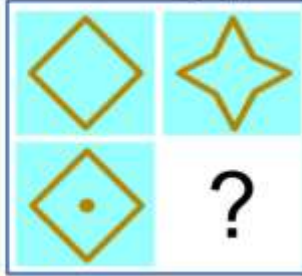


GPT-4V:

'B': Beachside bar boasts beautiful blue backdrop.
'D': Dining deck displays delightful drinks, distant docks.
'T': Tabletop terrace teeming with tropical tranquility.

Prompt:

What would the missing image look like?

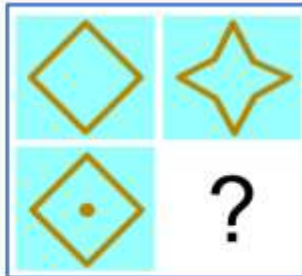


GPT-4V:

The missing image would likely be a variation of the diamond or star shape, as seen in the other three images. It could be a diamond with a different pattern or color, or a star with a different number of points or color. **It is difficult to say for sure without more context, but it would likely follow the same theme as the other images.**

Prompt:

Here are three sub images arranged in a 2-by-2 matrix. First, look at the two images in the first column, in the top left and bottom left. Then, find out the pattern in the first column based on the left two images. Next, use the found pattern and the image in the top right to infer the missing figure. Finally, describe what would the missing image look like?



GPT-4V:

The pattern in the first column is that the shape in the bottom left image has a dot in the center of the shape in the top left image. Therefore, the missing image in the bottom right would be a star with a dot in the center, similar to the image in the top right but with a dot in the center.

DIVERSE DOMAINS

Prompt:

Describe the image



GPT-4V:

The image shows a collage of 8 prominent personalities from different fields. On the top row, from left to right, there is singer Justin Bieber, footballer Lionel Messi, footballer Cristiano Ronaldo, and entrepreneur Elon Musk. On the bottom row, from left to right, there is model and reality TV star Kylie Jenner, singer Taylor Swift, actor Robert Downey Jr., and actor Scarlett Johansson. All individuals are shown in different poses and outfits, with various backgrounds.

Prompt:

Who is the person in the image and what is the person doing?



GPT-4V:

The person in the image is the current President of the United States, Joe Biden. He is standing at a podium, likely delivering a speech, at the G7 Summit in Hiroshima, Japan in 2023.

Prompt:

Who is the person in the image and what is the person doing?



GPT-4V:

The person in the image is Jensen Huang, the CEO and co-founder of NVIDIA. He is holding and presenting a product from NVIDIA, likely a graphics processing unit (GPU).

Red highlights the wrong answer, **Green** highlights the correct answer.

UNDERSTANDABLE

SCIENCE AND KNOWLEDGE

Prompt:
Which of these states is farthest north?



GPT-4V:
Out of the three states shown in the map - Arizona, Oklahoma, and Delaware, Delaware is the farthest north.

Prompt:
Which of the organisms is the producer in this food web?



GPT-4V:
The producers in this food web are the berries and flowers, grasses, and seeds. These are the organisms that use energy from the sun to produce their own food through photosynthesis.

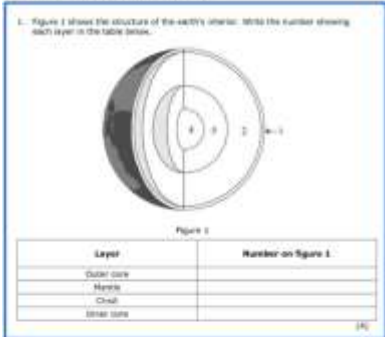
MULTILINGUAL UNDERSTANDING

Prompt:
Describe the image.



GPT-4V:
This image is a floor plan for an apartment or condominium. It shows the layout of the rooms and the dimensions. There is a terrace, dining area, living room, foyer, two bedrooms, a master bathroom, and a guest bathroom. The floor plan is black and white and is a two-dimensional drawing.

Prompt:
Answer the question in the image.



GPT-4V:
Layer | Number on figure 1
Outer core | 3
Mantle | 2
Crust | 1
Inner core | 4

Visualization of the generated results:

Layer	Number on figure 1
Outer core	3
Mantle	2
Crust	1
Inner core	4

Prompt:
Where is the bathroom for the second bedroom?



GPT-4V:
According to this floor plan, the bathroom for the second bedroom appears to be located directly adjacent to the bedroom. It is labeled "WC" and is a small room measuring 5'2" x 5'2".

Prompt:
Describe the poster. Which city is this dish from?



GPT-4V:
The poster is an advertisement for a Chinese dish called "Hot dry noodles." The poster has a red background with white Chinese characters at the top that read "Hot dry noodles, the taste of China." Below the text is a photo of the dish, which consists of noodles with meat and vegetables on top. The dish is served in a green bowl with red chili peppers on the side. The text at the bottom of the poster is in Chinese and describes the dish as being spicy, fragrant, and delicious.

This dish is from the city of Wuhan, located in the Hubei Province of China. It is a popular street food in the city and is considered one of the "10 great breakfasts" in China.

Red highlights the wrong answer, Green highlights the correct answer.

UNDERSTANDABLE

SPATIAL UNDERSTANDING

Prompt:

What is the spatial relation between the frisbee and the man?



GPT-4V:

Based on the image provided, the spatial relation between the frisbee and the man is that the frisbee is above and slightly in front of the man. It also appears to be moving away from the man, as if it has been thrown or released by him.

Prompt:

Does the person bigger than the car?



GPT-4V:

No, the car appears to be bigger than the person in this image. However, the perspective and distance of the objects in the photo may affect their perceived size.

JOKE AND MEME

Prompt:

Can you explain this meme?



GPT-4V:

This meme is poking fun at the idea of procrastination and how someone may put off doing a task until a certain time, but if they miss that time, they will push it back even further. In this case, the person in the meme planned to do something at 8 o'clock, but since it is now 8:05, they have decided to wait until 9 o'clock to do it. The image of Kermit the Frog relaxing in bed represents the person's desire to procrastinate and avoid the task.

Prompt:

What is funny about this image?



GPT-4V:

There are a couple of things that people might find funny about this image:

- It is a cartoon of a rhinoceros painting, which is an unexpected and silly concept in itself.
- The rhinoceros is painting a picture of itself, which adds an extra layer of humor.
- The caption, "You see the world as you are!", is a joke about how we all have our own unique perspectives and can be a little self-centered at times.

Prompt:

Can you explain this meme?



GPT-4V:

This meme highlights the deterioration of a person's handwriting during an exam. The first two pages show neat and legible handwriting, the middle page shows slightly messier handwriting, and the last two pages show almost illegible handwriting. The meme is poking fun at how, as an exam progresses, a person's handwriting tends to get worse as they become more tired and rushed to finish the exam. It also implies that the person's focus and attention to detail decrease towards the end of the exam.

Prompt:

What is funny about this image?



GPT-4V:

This image is funny for a couple of reasons. Firstly, it is a meme template that has been widely used and recognized on the Internet. The left side of the image is a screenshot from the reality TV show "The Real Housewives of Beverly Hills" where one of the cast members is angrily pointing and yelling. The right side of the image is a picture of a cat sitting at a dinner table with a plate of food in front of it, looking unimpressed. The humor in this image comes from the contrast between the two sides - the angry, dramatic yelling on the left and the nonchalant cat on the right. It also plays on the idea of cats acting like humans and being unimpressed by our behavior. Overall, the image creates a funny and absurd situation that makes people laugh.



Red highlights the wrong answer, Green highlights the correct answer.

UNDERSTANDABLE

VIDEO UNDERSTANDING

Prompt:

Analyze the images and decide the correct order in which they should be displayed.

a



b



c



Prompt Continued

d



e



GPT-4V:

The correct order in which the images should be displayed is:

1. e - Mixing the rice
2. c - Placing the rice on the seaweed
3. b - Adding the fillings
4. a - Cutting the roll
5. d - Finished sushi rolls on a plate.

FACIAL EXPRESSIONS

Prompt:

Identify and read emotions of people from their faces as shown in the image below.



GPT-4V:

The person in the image appears to be displaying a combination of emotions. The furrowed brows and downturned mouth suggest anger or frustration, while the tense jaw and narrowed eyes could indicate disgust or contempt. Overall, it seems like the person is experiencing negative emotions and is likely upset or unhappy.

Prompt:

Identify and read emotions of people from their faces as shown in the image below.



GPT-4V:

The person in the image is displaying a happy emotion. This can be seen from the wide smile and the crinkles around the eyes.

Prompt:

Identify and read emotions of people from their faces as shown in the image below.



GPT-4V:

The emotion shown in the image is fear or anxiety. The person's eyes are wide open and their mouth is slightly open, indicating a sense of shock or surprise. Their hands are covering their mouth, which could suggest that they are trying to suppress a scream or gasp.

Prompt:

Identify and read emotions of people from their faces as shown in the image below.



GPT-4V:

The person in the image appears to be experiencing a negative emotion, likely frustration or disappointment. This is evident from the furrowed brows, downward gaze, and hand on the forehead.

Red highlights the wrong answer, Green highlights the correct answer.

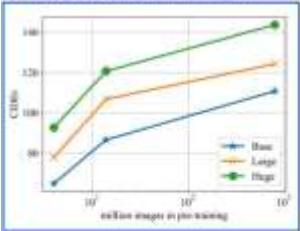
ACHIEVABLE




POWERED AGENT

CODING

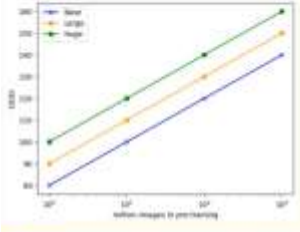
Prompt:
Generate python code to draw similar curves.



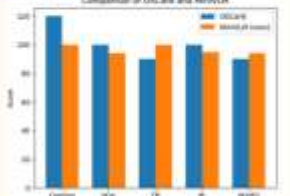
Prompt:
Write python codes to generate similar figures.




GPT-4V:



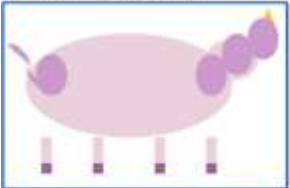
GPT-4V:



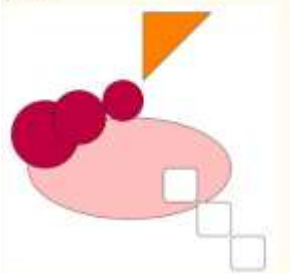
Prompt:
Write svg codes to generate following image.




Prompt:
Generate the following image in 9kz.



GPT-4V:



GPT-4V:



Red highlights the wrong answer, **Green** highlights the correct answer.

GROCERY CHECKOUT

Prompt:
Here is a list of reference images:

Shrimp



Crab dip



Smoked Ham



Banana



Apple




Bread



Prompt Continued

Cheese



Strawberries



Pink Tulips



For below image, find what products are presented in the basket according to reference image above. Only tell me products that are exactly same type of reference images.



GPT-4V:

There are 5 products in the basket that are exactly the same type as the reference images:

- Smoked ham
- Apple
- Pink tulips
- Crab dip
- Strawberries

POWERED AGENT NAVIGATION

Prompt:

Imagine that you are a home robot, and is asked to go to the kitchen to fetch something from the fridge. At the previous turn, you were at the position shown in the below image:



Note: Same image from the previous turn, resized to save space

The action taken in previous turn is: turn right and move forward towards the kitchen.

Now you are at the kitchen, this is what you can see from there. Can you see the fridge from here? What is your next action?



GPT-4V:

From this position, I can see the fridge on the right side of the kitchen. My next action would be to **move forward and turn slightly to the right to approach the fridge.**

Prompt:

Imagine that you are a home robot, and is asked to go to the kitchen to fetch something from the fridge. At the previous turn, you were at the position shown in the below image:



Note: Same image from the previous turn, resized to save space

The action taken in previous turn is: move forward and turn slightly to the right to approach the fridge.

Now this is what you can see from there. What is your next action?



GPT-4V:

My next action would be to move forward and slightly to the left to align myself with the fridge door. Then, I would use my robotic arm to open the fridge door and retrieve the requested item.

ONLINE SHOPPING

Online Shopping (Ergonomic Keyboard)

Prompt:

Imagine that you are a robot operating a phone. Like how humans operate the phone, you can move your finger to an icon, click an icon with your finger, scroll down a screen, or type some texts with the keyboard.

You are asked to shop for an ergonomic keyboard, your budget is between \$50 and \$100.

Previous action: "My next move would be to move my finger to the Amazon icon, which is located in the third row and third column of the screen. I would then click on the icon to open the app and search for an ergonomic keyboard within my budget of \$50."

You have now opened the Amazon app. Below is what you see on the phone screen. What will you do next? If the action involves moving your finger to an icon, describe the location of that icon as detailed as possible.



GPT-4V:

My next move would be to move my finger to the search bar, which is located at the top of the screen. I would then click on the search bar and type in "ergonomic keyboard" using the keyboard. After typing in the search term, I would click on the "Search" button, which is located on the bottom right corner of the keyboard, to initiate the search.

Red highlights the wrong answer, **Green** highlights the correct answer.

THANK YOU

