

Supplementary Materials

1. Dataset statistics

To provide a more comprehensive overview of INS-MMBench, we present descriptive statistics from four perspectives: question format, input image quantity, number of options, and answer distribution, as illustrated in Figure 1.

Question format: INS-MMBench includes two types of questions: multiple-choice visual questions and free-text visual questions, comprising 85.38% and 14.62% of the dataset, respectively. Overall, multiple-choice questions dominate the benchmark.

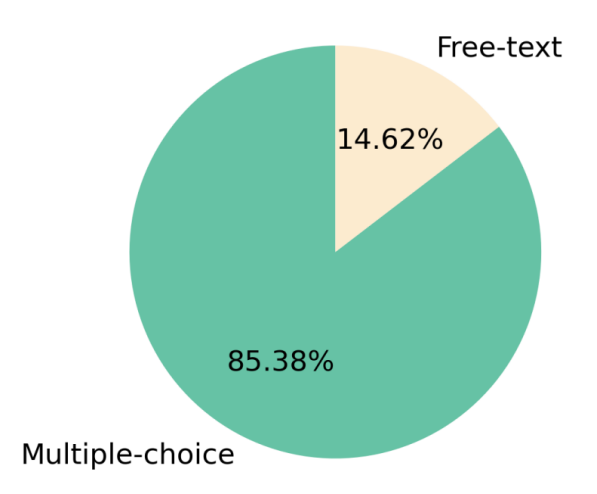
Input image quantity: The majority of questions involve a single input image, accounting for 91.61% of the dataset. A smaller subset (8.39%) requires multi-image inputs, ranging from 2 to 5 images.

Number of options: Among multiple-choice visual questions, 78.08% provide four answer options. Questions with only two or three options—often due to limited label sets (*e.g.*, yes/no tasks)—represent 16.62% and 5.30% of the dataset, respectively.

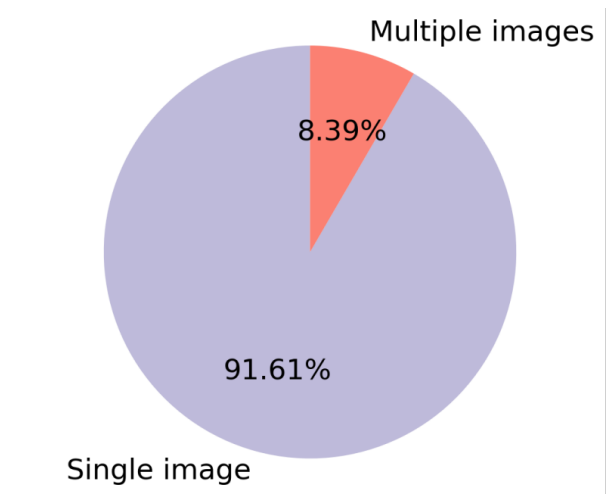
Answer distribution: The distribution of correct answers across options A, B, C, and D is 27.94%, 31.41%, 21.09%, and 19.56%, respectively, indicating a relatively balanced distribution. However, the slightly higher proportions for options A and B can be attributed to the presence of questions with fewer answer choices.

2. Cases

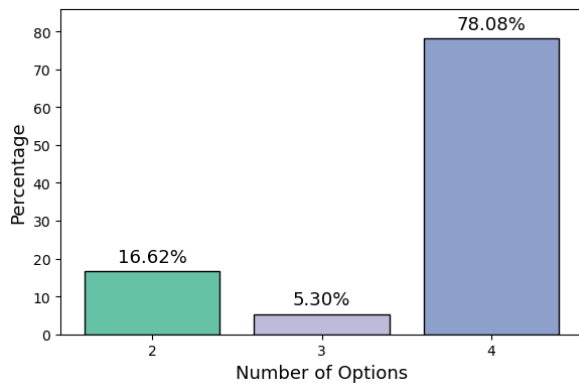
To provide a comprehensive overview of the task settings and diverse evaluation formats in INS-MMBench—including multiple-choice questions and multi-step reasoning—we curate representative sample cases for each core task. Additionally, we include detailed responses from representative LVLMs, such as GPT-4o, Gemini 1.5 Flash, and Qwen-VL-Max.



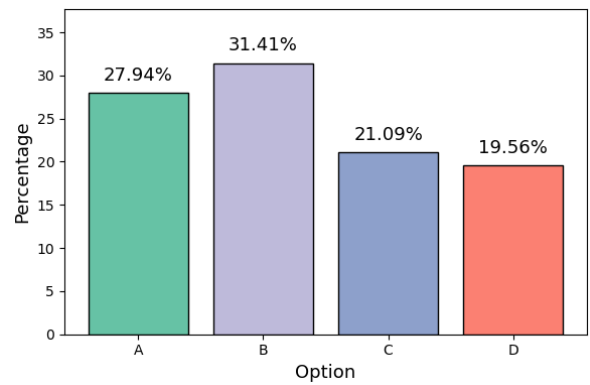
(a) Question format



(b) Input image quantity



(c) Number of options



(d) Answer distribution

Figure 1. Descriptive statistics of INS-MMBench across four dimensions: (a) Question format, (b) Input image quantity, (c) Number of options, and (d) Answer distribution.

License plate recognition

Prompt:

What is the license plate number of the vehicle in the image?

Options:

- A. NEWJOB
- B. NEWLOB
- C. N3WJOB
- D. MEWJOB

Correct answer: A



GPT-4o:

The license plate number of the vehicle in the image is:
A. NEWJOB



Gemini 1.5 Flash:

The answer is A. The license plate number is NEWJOB.



Qwen-VL-Max:

The license plate number of the vehicle in the image is "NEWJOB".

Figure 2. An example of the license plate recognition task and the corresponding responses from GPT-4o, Gemini 1.5 Flash, and Qwen-VL-Max.

Vehicle milage reading

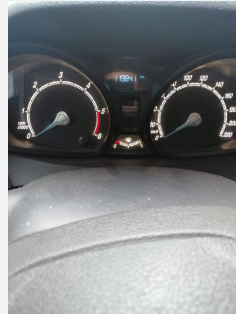
Prompt:

What is the mileage reading of the vehicle in the image?

Options:

- A. 119841
- B. 118941
- C. 119481
- D. 191841

Correct answer: A



GPT-4o:

The mileage reading of the vehicle in the image is 119841.

So, the correct answer is:

A. 119841



Gemini 1.5 Flash:

The answer is A. 119841.



Qwen-VL-Max:

The mileage reading of the vehicle in the image is option B, 118941 kilometers.

Figure 3. An example of the vehicle mileage reading task and the corresponding responses from GPT-4o, Gemini 1.5 Flash, and Qwen-VL-Max.

Vehicle warning indicator recognition

Prompt:

Which indicator icons are lit on the vehicle dashboard in the image?

Options:

- A. Door open, Low Fuel
- B. Check Engine, High Beam
- C. Tire Pressure, Battery
- D. Engine Oil, ABS



Correct answer: A



GPT-4o:

The correct answer is:

- A. Door open, Low Fuel



Gemini 1.5 Flash:

The correct answer is C. Tire Pressure, Battery.

The image shows two indicator icons lit up on the dashboard: a tire pressure icon and a battery icon.



Qwen-VL-Max:

The correct answer is A. Door open, Low Fuel. In the image of the vehicle dashboard, there are two indicator icons that are lit up. The first icon indicates that one or more doors are open, and the second icon indicates that the fuel level is low. These symbols serve as important reminders for the driver to ensure safety and proper vehicle maintenance.

Figure 4. An example of the vehicle warning indicator recognition task and the corresponding responses from GPT-4o, Gemini 1.5 Flash, and Qwen-VL-Max.

Vehicle make and model identification

Prompt:

What is the make and model of the vehicle in the image?

Options:

- A. Jaguar XKR-S Convertible 2012
- B. Bentley Continental GT Convertible 2012
- C. Rolls-Royce Phantom Drophead Coupe Convertible 2012U
- D. Aston Martin DB9 Volante Convertible 2012



Correct answer: C



GPT-4o:

The vehicle in the image is a Rolls-Royce Phantom Drophead Coupe Convertible 2012. So, the correct answer is:
C. Rolls-Royce Phantom Drophead Coupe Convertible 2012



Gemini 1.5 Flash:

The answer is C. Rolls-Royce Phantom Drophead Coupe Convertible 2012.



Qwen-VL-Max:

The make and model of the vehicle in the image is C. Rolls-Royce Phantom Drophead Coupe Convertible 2012.

Figure 5. An example of the vehicle make and model identification task and the corresponding responses from GPT-4o, Gemini 1.5 Flash, and Qwen-VL-Max.

Vehicle modification detection

Prompt:

Please determine whether the vehicle in the image has been modified?

Options:

A. Yes

B. No



Correct answer: A



GPT-4o:

A. Yes



Gemini 1.5 Flash:

A. Yes.



Qwen-VL-Max:

A. Yes

The vehicle in the image appears to have been modified with aftermarket wheels and a lowered suspension, which is evident by its stance on the ground. The modifications give the car a distinct look that deviates from its original factory appearance.

Figure 6. An example of the vehicle modification detection task and the corresponding responses from GPT-4o, Gemini 1.5 Flash, and Qwen-VL-Max.

Incar driving behavior detection

Prompt:

What is the behavior of the driver in the image?

Options:

- A. Texting - left
- B. Adjusting mirror - right
- C. Eating - right
- D. Talking on the phone - left



Correct answer: A



GPT-4o:

A. Texting - left



Gemini 1.5 Flash:

The answer is A. Texting - left. The driver is holding a phone in his left hand and looking at it.



Qwen-VL-Max:

The correct answer is D. Talking on the phone - left. In the image, the driver is sitting in the car and holding his phone to his ear, indicating that he is talking on the phone. He is not texting or adjusting the mirror as suggested by options A and B respectively. Option C, eating, does not apply since there is no visible food or action related to eating in the image.

Figure 7. An example of the incar driving behavior detection task and the corresponding responses from GPT-4o, Gemini 1.5 Flash, and Qwen-VL-Max.

Vehicle damage part detection

Prompt:

What are the damaged parts of the vehicle in the image?

Options:

- A. Bonnet-dent
- B. Door-scratch
- C. Bumper-dent
- D. Bonnet-crack



Correct answer: A



GPT-4o:

C. Bumper-dent



Gemini 1.5 Flash:

The answer is A. Bonnet-dent.
The image shows a dent on the bonnet of the vehicle.



Qwen-VL-Max:

The correct answer is A. Bonnet-dent. In the image, there is a blue BMW car with visible damage on its front end. The hood (bonnet) has a dent and some scratches, while one of the grills appears to be missing or broken. The bumper also seems to have some dents and scratches. The license plate holder is bent, and the license plate itself is slightly crooked.

Figure 8. An example of the vehicle damage part detection task and the corresponding responses from GPT-4o, Gemini 1.5 Flash, and Qwen-VL-Max.

Vehicle damage type detection

Prompt:

What are the damaged types of the vehicle in the image?

Options:

- A. Broken Headlight
- B. Dented Door
- C. Scratched Paint
- D. Tire Flat



Correct answer: D



GPT-4o:

The correct answer is:
D. Tire Flat



Gemini 1.5 Flash:

The answer is D. Tire Flat. The tire is clearly deflated and the car is resting on the rim.



Qwen-VL-Max:

The correct answer is D. Tire Flat. In the image, there is a white car with a flat tire on the front left side. The tire appears to be completely deflated and rests against the curb. There is no visible damage to the headlight or door, nor any noticeable scratches on the paint. Therefore, the only visible damage to the vehicle in this image is the flat tire.

Figure 9. An example of the vehicle damage type detection task and the corresponding responses from GPT-4o, Gemini 1.5 Flash, and Qwen-VL-Max.

Vehicle damage severity detection

Prompt:

What is the damaged severity level of the vehicle in the image?

Options:

- A. severe
- B. moderate
- C. no_accident
- D. minor

Correct answer: D



GPT-4o:

The correct answer is:

D. minor

The vehicle in the image has some visible damage, particularly around the front left fender and headlight area, but it does not appear to be severe.



Gemini 1.5 Flash:

The answer is B. moderate.

The vehicle has visible damage to the front bumper and fender, but the overall structure of the vehicle appears intact. This suggests moderate damage.



Qwen-VL-Max:

The correct answer is D. minor. Based on the image, there appears to be some damage to the front bumper and possibly the headlights of the white car. However, the overall damage seems to be relatively minor compared to more severe accidents or collisions. Therefore, option D, minor, would be the most appropriate description for the damage severity level of the vehicle in the image.

Figure 10. An example of the vehicle damage severity detection task and the corresponding responses from GPT-4o, Gemini 1.5 Flash, and Qwen-VL-Max.

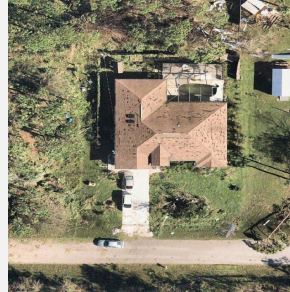
Roof condition assessment

Prompt:

What is the condition of the roof of the house in the image?

Options:

- A. Damage
- B. Not damage



Correct answer: A



GPT-4o:

A. Damage



Gemini 1.5 Flash:

A. Damage



Qwen-VL-Max:

A. Damage

Figure 11. An example of the roof condition assessment task and the corresponding responses from GPT-4o, Gemini 1.5 Flash, and Qwen-VL-Max.

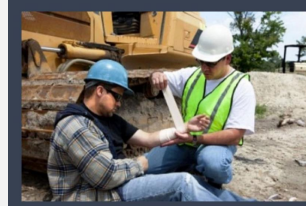
Workplace risk assessment

Prompt:

What's the risk condition of the workplace in the image?

Options:

- A. There is risk of not wearing helmet in the image
- B. There is risk of not wearing helmet and vest in the image
- C. There is risk of not wearing vest in the image
- D. There is no risk in the image



Correct answer: C



GPT-4o:

- A. There is risk of not wearing helmet in the image



Gemini 1.5 Flash:

There is risk of not wearing vest in the image



Qwen-VL-Max:


C. There is risk of not wearing vest in the image.
In the image, one worker is seen wearing a hard hat and helping another worker who appears to be injured. However, neither of them is wearing a safety vest, which poses a risk on the construction site. Therefore, option C is the correct answer.


Figure 12. An example of the workplace risk assessment task and the corresponding responses from GPT-4o, Gemini 1.5 Flash, and Qwen-VL-Max.

House fire detection


Prompt:
Is there fire in the house in the image?
Options:
A. no fire
B. fire

Correct answer: B






GPT-4o:
A. no fire



Gemini 1.5 Flash:
A. no fire



Qwen-VL-Max:
Option A: No fire.
There is no evidence of a fire in the image. The room appears to be well-lit and there are no signs of smoke or flames.

Figure 13. An example of the house fire detection task and the corresponding responses from GPT-4o, Gemini 1.5 Flash, and Qwen-VL-Max.

House damage type detection

Prompt:

What type of damage is done to the house in the image?

Options:

- A. Combined Damage
- B. Flexural
- C. Shear



Correct answer: A



GPT-4o:

The damage shown in the image appears to be characterized by diagonal cracks and a pattern that suggests a sliding or shearing motion. This type of damage is typically associated with shear forces. The correct answer is:
C. Shear



Gemini 1.5 Flash:

The answer is A. Combined Damage.
The image shows cracks in the wall, which are a sign of both flexural and shear damage. Flexural damage occurs when a material is bent or deformed, while shear damage occurs when a material is subjected to a force that is parallel to its surface. In this case, the cracks are likely caused by a combination of these two types of damage.



Qwen-VL-Max:

The correct answer is A. Combined Damage.
Based on the image provided, it appears that there is combined damage to the house's exterior wall. The wall shows signs of cracking and peeling, which could be due to a combination of factors such as age, weathering, or structural issues. The presence of a metal mesh suggests that repairs have been attempted, but the damage still remains visible. Therefore, option A, Combined Damage, best describes the type of damage seen in the image.

Figure 14. An example of the house damage type detection task and the corresponding responses from GPT-4o, Gemini 1.5 Flash, and Qwen-VL-Max.

House damage level detection

Prompt:

What is the level of damage to the house in the image?

Options:

- A. Undamage
- B. Minor Damage
- C. Moderate Damage
- D. Heavy Damage

Correct answer: C



GPT-4o:

The correct answer is:

C. Moderate Damage

The image shows significant cracks and peeling on the exterior wall, indicating structural issues that are more than minor but not extensive enough to be classified as heavy damage.



Gemini 1.5 Flash:

The answer is C. Moderate Damage.

The image shows a large crack running across the wall of the house, indicating a significant structural issue. However, the house is still standing and the crack doesn't appear to have caused a complete collapse. Therefore, it's not heavy damage, but it's more than just minor damage.



Qwen-VL-Max:


The correct answer based on the image provided is:\n\nC. Moderate Damage\n\nThe image shows a pink building with visible cracks and peeling paint, indicating some degree of structural damage or wear over time. The extent of the damage appears to be moderate rather than minor or heavy.

Figure 15. An example of the house damage level detection task and the corresponding responses from GPT-4o, Gemini 1.5 Flash, and Qwen-VL-Max.

Fall detection

Prompt:
Does the person in the picture fall down?
Options:
A. fallen
B. not fallen

Correct answer: A



✓

GPT-4o:
A. fallen

✓

Gemini 1.5 Flash:
A. fallen

✗

Qwen-VL-Max:
B. not fallen

Figure 16. An example of the fall detection task and the corresponding responses from GPT-4o, Gemini 1.5 Flash, and Qwen-VL-Max.

Health device reading

Prompt:

What is the reading of the blood pressure monitor in the image?

Options:

- A. SYS:130,DIA:70,PULSE:60
- B. SYS:110,DIA:65,PULSE:55
- C. SYS:120,DIA:75,PULSE:50
- D. SYS:120,DIA:67,PULSE:55



Correct answer: D



GPT-4o:

The correct answer is:
D. SYS:120,DIA:67,PULSE:55



Gemini 1.5 Flash:

D. SYS:120,DIA:67,PULSE:55



Qwen-VL-Max:

The correct answer is D. SYS:120,DIA:67,PULSE:55.

Figure 17. An example of the health device reading task and the corresponding responses from GPT-4o, Gemini 1.5 Flash, and Qwen-VL-Max.

Medical image abnormality recognition

Prompt:

what is the primary abnormality in this image?

Options:

- A. Osteomyelitis
- B. Osteogenesis Imperfecta
- C. Osteoarthritis
- D. Osteosarcoma



Correct answer: B



GPT-4o:

The primary abnormality in the image is consistent with Osteogenesis Imperfecta (Option B). This condition is characterized by bones that break easily, often with little or no apparent cause. The X-ray shows multiple fractures and bone deformities, which are typical features of Osteogenesis Imperfecta.



Gemini 1.5 Flash:

The answer is D. Osteosarcoma. Osteosarcoma is a type of bone cancer that typically affects children and young adults. It is characterized by the formation of malignant tumors in the bone. The image shows a large, destructive tumor in the humerus, which is a classic finding of osteosarcoma.



Qwen-VL-Max:

B. Osteogenesis Imperfecta

Figure 18. An example of the medical image abnormality detection task and the corresponding responses from GPT-4o, Gemini 1.5 Flash, and Qwen-VL-Max.

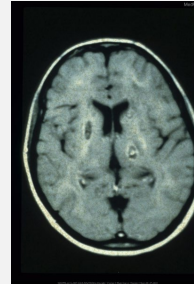
Medical image part recognition

Prompt:

what is the organ system in this image?

Options:

- A. Kidney, Ureter, Bladder
- B. Liver, Gallbladder, Pancreas
- C. Stomach, Duodenum, Jejunum
- D. Lung, Mediastinum, Pleura



Correct answer: D



GPT-4o:

The organ system in this image is the brain, which is part of the central nervous system. Therefore, none of the provided options are correct.



Gemini 1.5 Flash:

The answer is D. Lung, Mediastinum, Pleura. This is a brain scan, which is part of the nervous system.



Qwen-VL-Max:

The correct answer is B. Liver, Gallbladder, Pancreas. In the provided image, we can see an MRI scan of the liver, gallbladder, and pancreas. The image shows these organs in detail, allowing for evaluation of their structure and potential abnormalities.

Figure 19. An example of the medical image part recognition task and the corresponding responses from GPT-4o, Gemini 1.5 Flash, and Qwen-VL-Max.

Field image crop type identification

Prompt:

What type of crops are in the image?

Options:

- A. Jute
- B. Sisal
- C. Flax
- D. Hemp



Correct answer: A



GPT-4o:

D. Hemp



Gemini 1.5 Flash:

The answer is A. Jute.

Jute is a natural fiber obtained from the stem of the jute plant. It is a long, soft, and shiny fiber that is used to make a variety of products, including rope, twine, burlap, and canvas. The image shows a field of jute plants, which are characterized by their tall, slender stems and green leaves.



Qwen-VL-Max:

The correct answer is C. Flax. The image shows flax plants growing in a field.

Figure 20. An example of the field image crop type identification task and the corresponding responses from GPT-4o, Gemini 1.5 Flash, and Qwen-VL-Max.

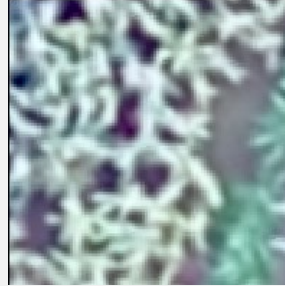
Satellite image crop type identification

Prompt:

What type of crops are in the image?

Options:

- A. Barley
- B. Sorghum
- C. Wheat
- D. Maize



Correct answer: D



GPT-4o:

The image appears to show a close-up of a crop, but the details are not clear enough to definitively identify the type of crop. However, based on the general appearance, it resembles the structure of wheat. So, the correct answer is:
C. Wheat



Gemini 1.5 Flash:

The answer is C. Wheat.
Wheat crops are characterized by their long, slender leaves and tall stalks. The image shows these features, making it likely that the crops are wheat.



Qwen-VL-Max:

Based on the image provided, it is difficult to determine the exact type of crop as the image is blurry and lacks detail. However, based on the given options, the most likely answer would be C. Wheat, as it is a common cereal grain that is often grown in large fields and has a similar appearance to the作物 in the image. However, without more information or a clearer image, it is impossible to确定ly identify the specific type of crop.

Figure 21. An example of the satellite image crop type identification task and the corresponding responses from GPT-4o, Gemini 1.5 Flash, and Qwen-VL-Max.

Crop growth stage recognition

Prompt:

What is the growth stage of the crops in the image?

Options:

- A. Booting
- B. Ripening
- C. Tillering
- D. Flowering

Correct answer: A



GPT-4o:

The growth stage of the crops in the image appears to be "C. Tillering." This stage is characterized by the development of multiple shoots from the base of the plant, which is evident in the dense and lush green appearance of the crop.



Gemini 1.5 Flash:

The answer is C. Tillering.

The crops in the image are in the tillering stage. This is the stage where the plant is producing new shoots from the base of the stem. The shoots are called tillers. The tillering stage is important for the plant because it allows it to produce more leaves and stems, which will eventually produce more grain.



Qwen-VL-Max:

Based on the image, it appears that the crops are at the tillering stage, which is when the plants are developing side shoots and branching out to form a dense canopy. Therefore, the correct answer is C. Tillering.

Figure 22. An example of the crop growth stage recognition task and the corresponding responses from GPT-4o, Gemini 1.5 Flash, and Qwen-VL-Max.

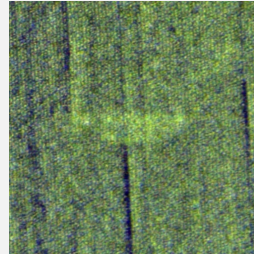
Farmland damage type detection

Prompt:

What type of damage has occurred on the farmland?

Options:

- A. Double plant
- B. Lodging
- C. Erosion
- D. Pest infestation



Correct answer: A



GPT-4o:

B. Lodging



Gemini 1.5 Flash:

The answer is C. Erosion.

Erosion is the process of soil and rock being moved from one place to another by wind, water, or ice. The image shows a pattern of vertical lines, which are likely caused by erosion.



Qwen-VL-Max:

Based on the image provided, it appears that there is lodging damage on the farmland. The plants appear to be leaning or fallen over, which is characteristic of lodging damage caused by strong winds or other factors. Therefore, the correct answer is B. Lodging.

Figure 23. An example of the farmland damage detection task and the corresponding responses from GPT-4o, Gemini 1.5 Flash, and Qwen-VL-Max.

Auto Insurance Claim Processing



Prompt:

You are an expert in auto insurance claim processing, here is an image of a vehicle and basic information: cost of vehicle is \$47800.0, the expired date of insurance is 2024-04-04, the min and max coverage of the vehicle is \$1195.0 and \$12028.0. Based on the textual and visual information, please provide a step-by-step evaluation of the vehicle's condition and make an insurance claim decision. Follow these steps:

1. Determine whether the vehicle is damaged (Yes/No).
2. If damaged, assess the severity of the damage (Minor, Moderate, Severe).
3. Estimate the expected repair cost based on the damage.
4. Determine whether the damage qualifies for an insurance claim (Yes/No), with a short and brief explanation.
5. If eligible for a claim, calculate the final claimable amount, with a short and brief explanation.

please strictly output in this format:

1. Yes/No
2. Minor/Moderate/Severe
3. Cost Amount
4. Yes/No. Reason: <Reason>
5. Final Claim Amount. Reason: <Reason>

Correct answer: 1. Yes 2. Minor 3. 648 4.NO 5. 0



GPT-4o:

1. Yes
2. Minor
3. \$1,200
4. Yes. Reason: The damage is minor and falls within the coverage limits.
5. Final Claim Amount: \$1,200. Reason: The estimated repair cost is within the maximum coverage limit.

Figure 24. An example of the auto insurance claim processing task and the corresponding response from GPT-4o.

Auto Insurance Accident Liability



Prompt:

You are an expert in auto insurance claim processing. You are provided with a sequence of continuous dashcam images. Based on the visual information provided by these images, please perform a step-by-step evaluation as follows:

1. Weather: Identify the prevailing weather conditions in the images. (sunny, rainy, snowy, foggy)
2. Scene: Determine the type of scene where the vehicle is operating. (highway, tunnel, mountain, urban, rural)
3. Linear: Identify the road layout. (arterials, curve, intersection, T-junction, ramp)
4. Accident Occurred: Determine whether an accident occurred. (Yes/No)
5. Accident Cause: If an accident occurred, provide a brief one-sentence description of the accident cause.
6. Main Responsible Party: If an accident occurred, identify the primary responsible party. (Ego-car, Pedestrian, Other car, Both)

Please strictly output in the following format:

1. Weather: sunny/rainy/snowy/foggy
2. Scene: highway/tunnel/mountain/urban/rural
3. Linear: arterials/curve/intersection/T-junction/ramp
4. Accident Occurred: Yes/No.
5. Accident Cause: one sentence brief description
6. Main Responsible Party: Ego-car/Pedestrian/Other car/Both

Correct answer: 1. sunny 2. urban 3. T-junction 4. Yes
5. Ego-car does not notice the pedestrians when turning
6. Ego-car

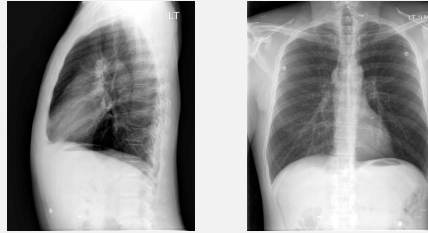


GPT-4o:

1. Weather: sunny
2. Scene: urban
3. Linear: arterials
4. Accident Occurred: No

Figure 25. An example of the auto insurance accident liability determination task and the corresponding response from GPT-4o.

Health Insurance Risk Assessment



Prompt:

You are an expert in health insurance underwriting. Here is X-ray images and its associated radiology report.

report: Cardiomedastinal silhouette and pulmonary vasculature are within normal limits. Lungs are clear. No pneumothorax or pleural effusion. Evidence of prior granulomatous disease. No acute osseous findings.

Based on the visual and textual information provided, please perform a step-by-step evaluation as follows:

1. Scan Region: Determine the scanned region in the X-ray image.
2. Scan Result Match: Determine whether the scan result (findings on the image) matches the findings described in the report (Yes/No).
3. Health Risk Severity: Assess the overall health risk based on the X-ray image and report. Classify the risk as no risk, mild risk, or severe risk.
4. Underwriting Evaluation: Based on your evaluation, decide the appropriate underwriting outcome (Declined, Standard, or Higher Premium). Provide a brief explanation.

Please strictly output in the following format:

1. Scan Region: <Scan Region>
2. Scan Result Match: Yes/No.
3. Health Risk Severity: no risk/mild risk/severe risk.
4. Underwriting Evaluation: Declined/Standard/Higher Premium. Reason: <Reason>

Correct answer: 1. Chest 2. yes 3. mild risk 4.Higher Premium



GPT-4o:

1. Scan Region: Chest
2. Scan Result Match: Yes.
3. Health Risk Severity: mild risk.
4. Underwriting Evaluation: Higher Premium. Reason: Evidence of prior granulomatous disease, which could potentially indicate a past infection or other underlying issue, may warrant closer monitoring and suggests a mildly elevated health risk.

Figure 26. An example of the health insurance risk assessment task and the corresponding response from GPT-4o.

Property Insurance Risk Management



Prompt:

You are an expert in property insurance assessment. Here are two satellite images of the same area. Based on the visual information provided by the images, please perform a step-by-step evaluation of the property damage and make an insurance claim decision.

Please follow these steps:

1. Determine whether a disaster occurred: Analyze the before and after images to decide if a disaster has taken place (Yes/No).
2. Identify the disaster type: If a disaster is detected, specify the type (e.g., flood, earthquake, hurricane, fire, tsunami, volcano etc.).
3. Count the number of houses: Based on the satellite images, provide an concrete count of the houses.
4. Estimate the number of damaged houses: Based on the changes observed in the satellite images, provide an concrete count of the houses that have suffered damage.
5. Insurance Claim Decision: Based on your evaluation, decide the appropriate claim decision (No compensation, Partial compensation, or Full compensation). Provide a short and brief explanation.

Please strictly output in this format:

1. Disaster Occurred: Yes/No
2. Disaster Type: <type>
3. House Count: <number>
4. Damaged House Count: <number>
5. Insurance Decision: <No Compensation/Partial Compensation/Full Compensation>. Reason: <Reason>

Correct answer: 1. Yes 2. hurricane 3. 7 4. 7 5. Full Compensation



GPT-4o:

1. Disaster Occurred: Yes
2. Disaster Type: Flood
3. House Count: 3
4. Damaged House Count: 3
5. Insurance Decision: Full Compensation. Reason: All houses are submerged in water, indicating complete damage due to flooding.

Figure 27. An example of the property insurance risk management task and the corresponding response from GPT-4o.

Agricultural Insurance Claim Processing



Prompt:

You are an expert in agricultural insurance claim processing, here is an image of a crop field and corresponding insurance policy information:

Insured Subject: This contract covers the main crops planted by the policyholder in the designated agricultural area during the insurance period, with the actual yield and crop quality serving as the insured subject.

Scope of Insurance Coverage:

1. **Natural Disaster Liability:** Crop damage or yield reduction directly caused by natural disasters such as heavy rain, floods, drought, frost, hail, typhoon, etc.

2. **Pest and Disease Liability:** Crop losses caused by sudden and severe pest or disease outbreaks confirmed by agricultural experts. For mild pest or disease damage where the loss does not meet the minimum compensation standard stipulated in the contract, only partial compensation is provided; for severe pest or disease damage that results in large-scale crop damage or a significant decline in yield, full or maximum compensation is provided in accordance with the contract.

3. **Other Accident Liability:** Crop losses directly caused by other special factors such as force majeure (e.g., fire, explosion) that, upon investigation and confirmation, meet the compensation conditions, will be compensated in accordance with the contract terms.

Based on the textual and visual information, please provide a step-by-step evaluation of the crop's condition and determine an insurance claim decision. Follow these steps:

1. **Identify the Crop Species:** Analyze the image and determine the specific crop species.

2. **Estimate the Number of Fruit Clusters:** Based on the image, estimate the number of fruit clusters present.

3. **Detect Pest/Disease Presence:** Determine whether there is any pest or disease infection in the crop (Yes/No).

4. **Insurance Claim Decision:** Based on your evaluation and the insurance policy, decide on the claim outcome: No Compensation, Partial Compensation, or Full Compensation. Provide a short and brief explanation.

Please strictly output in this format:

1. Crop species: <species>

2. Fruit cluster count: <a specific number>

3. Pest/Disease infection: Yes/No

4. Insurance decision: No Compensation/Partial Compensation/Full Compensation. Reason: <Reason>

Correct answer: 1. Grape 2. 1 3. Yes 4. Partial Compensation



GPT-4o:

1. Crop species: Grape

2. Fruit cluster count: 0

3. Pest/Disease infection: Yes

4. Insurance decision: Partial Compensation.

Reason: There is visible evidence of pest or disease infection with damaged and discolored leaves, indicating mild to moderate damage to the grape crop. Since the image indicates pest or disease presence but no fruit clusters are visible, partial compensation is appropriate as per the policy's stipulation for mild damage.

Figure 28. An example of the agricultural insurance claim processing task and the corresponding response from GPT-4o.