

Vision Language Models with Explainable Bottleneck Layers

Anh Totti Nguyen



@anh_ng8

anhnguyen.me



Long-term Research Goals

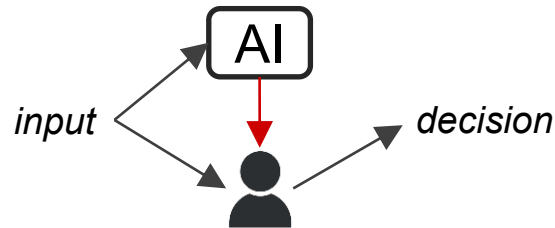
Trustworthy AI

1. Build AIs that are accurate in edge cases
(and common cases)

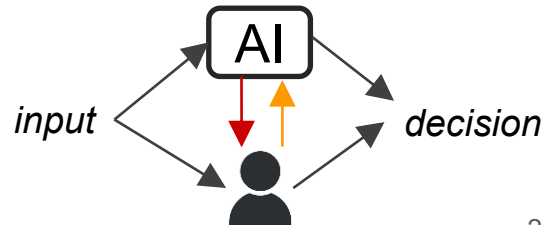


Explainable AI

2. Build AIs that maximize human-AI team accuracy

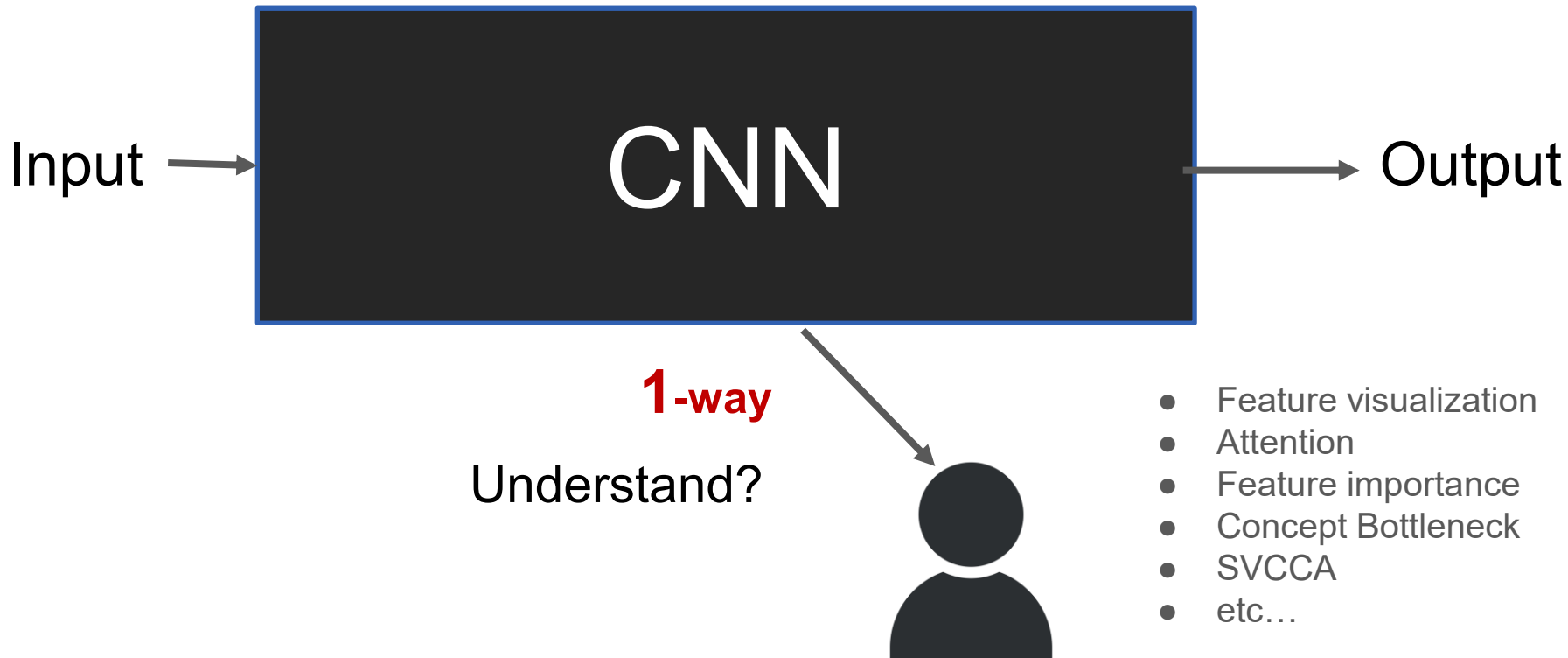


3. Build AIs that humans can debug and edit
(their decision-making process)

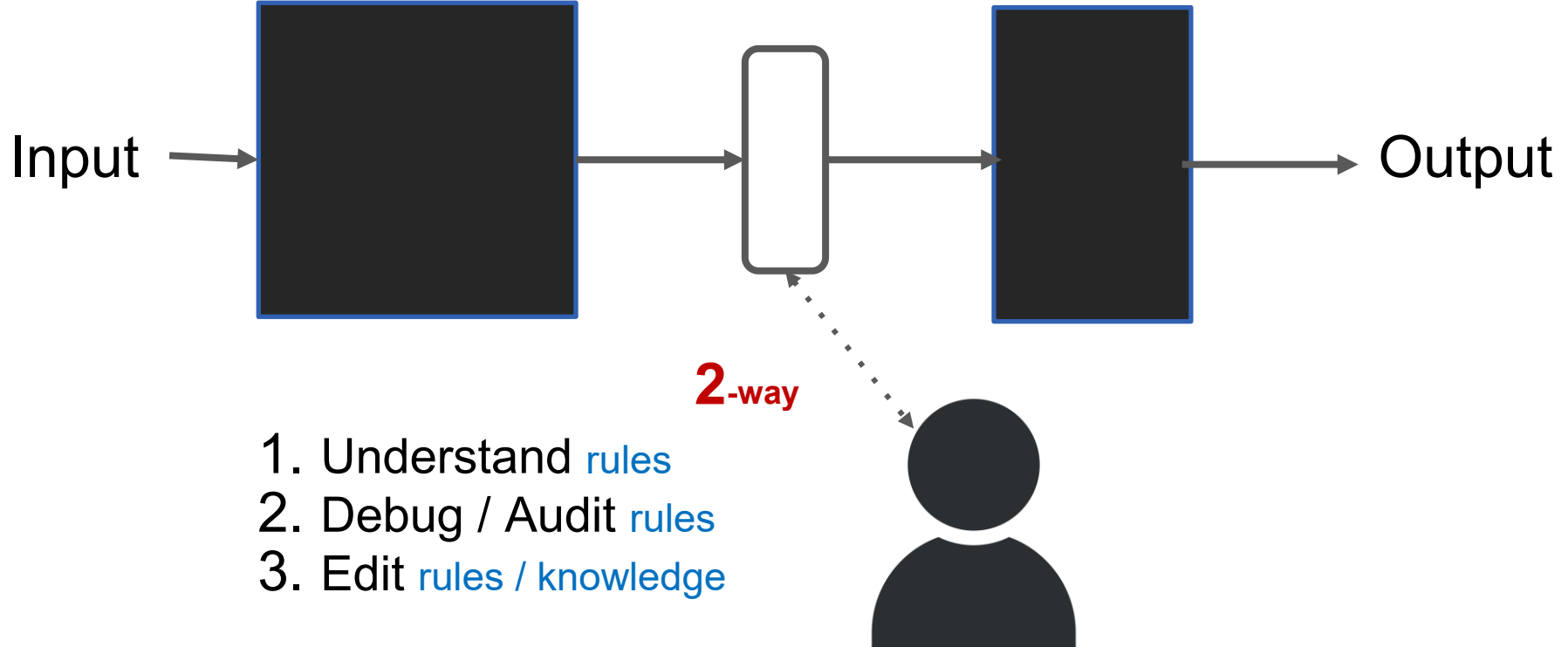


Traditional AI

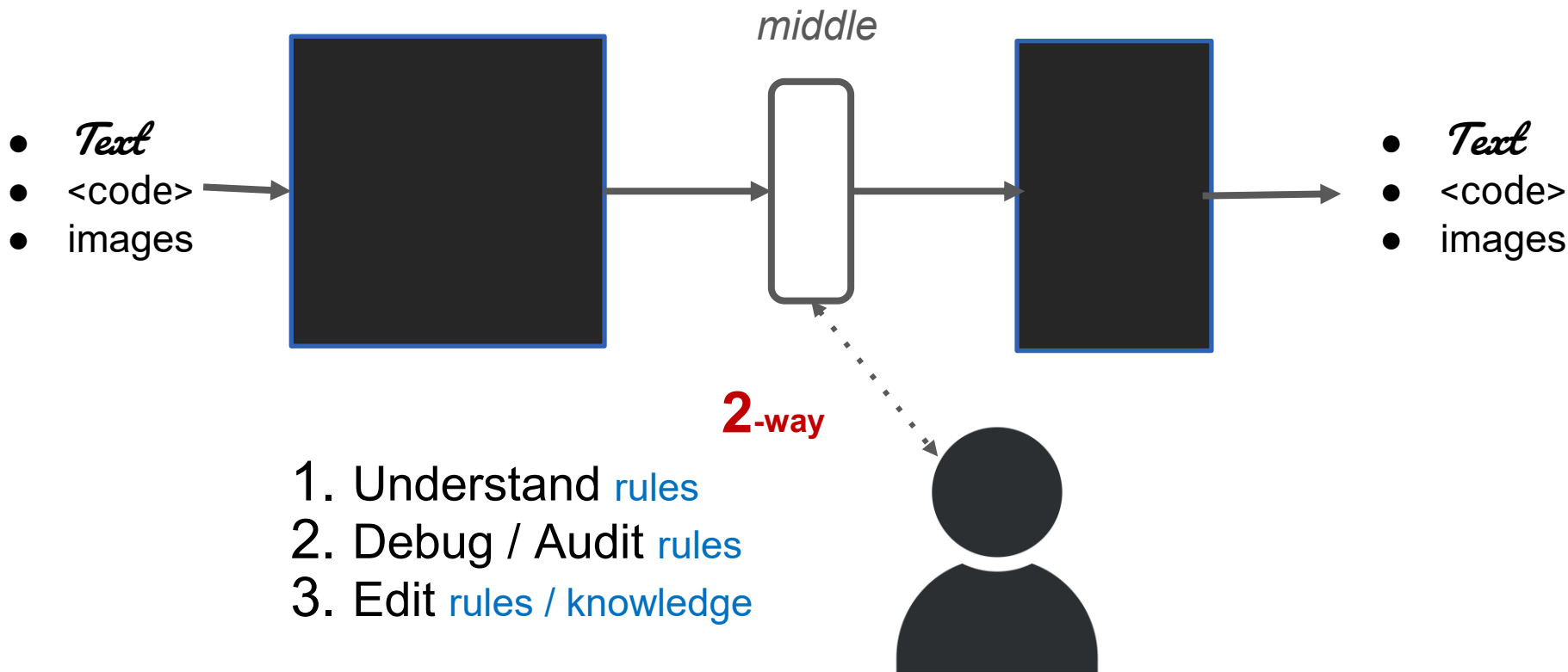
Causal analysis? Intervention?



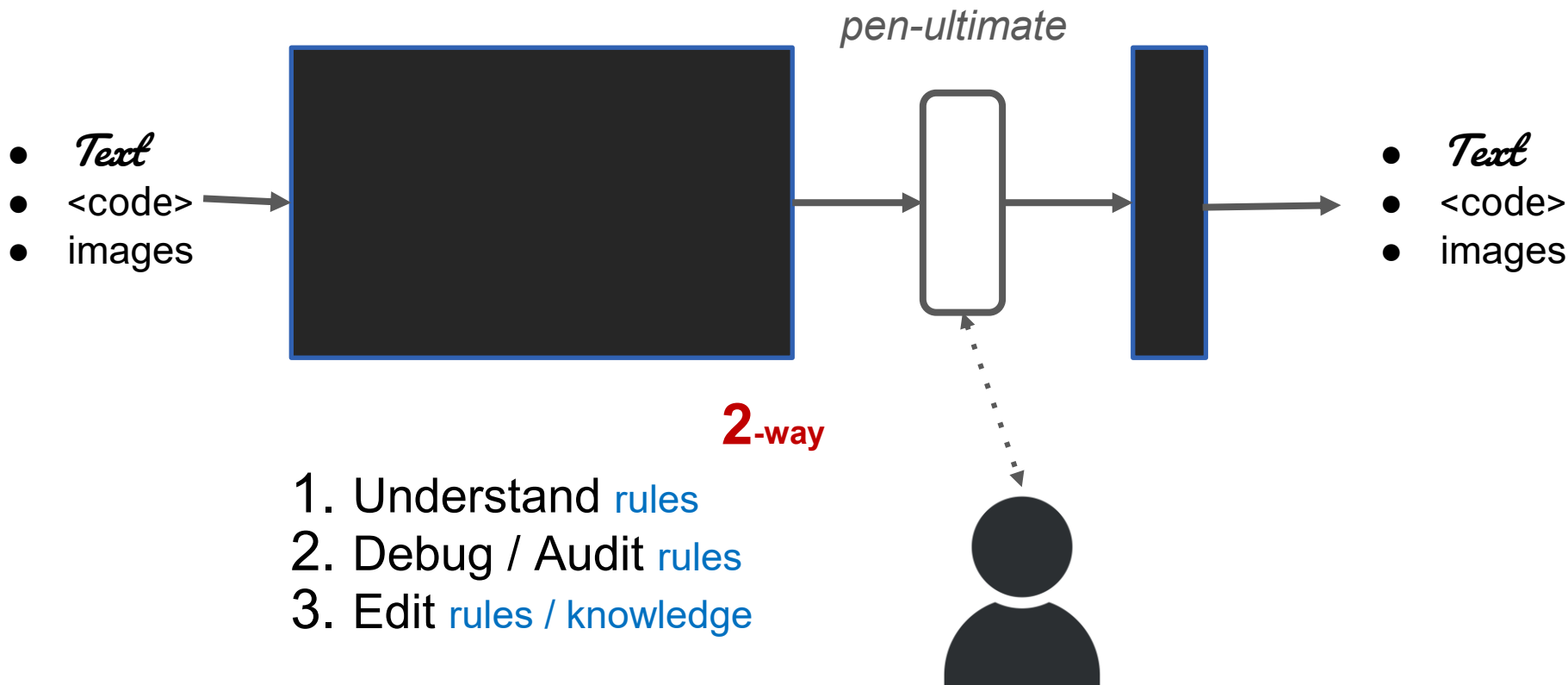
Bottleneck XAI



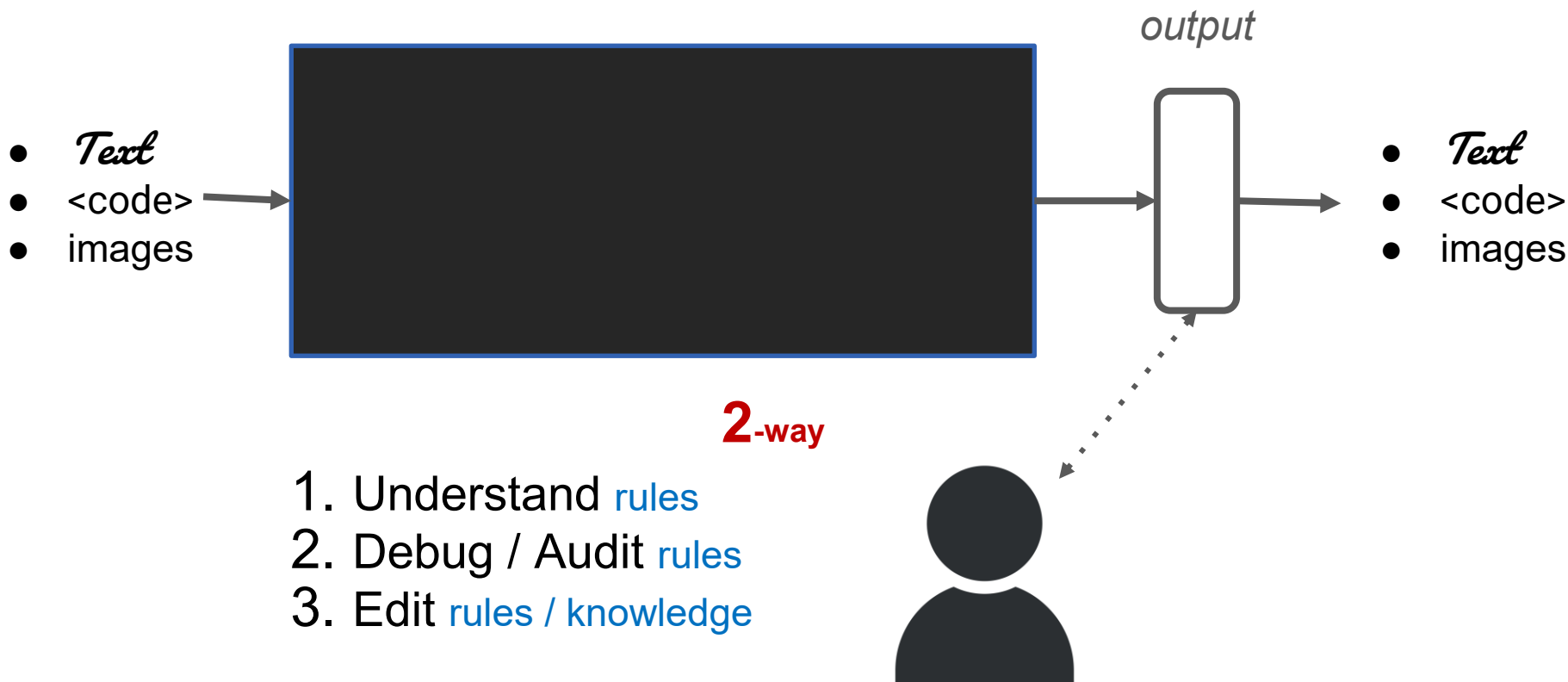
Multimodal Bottleneck XAI



Bottleneck XAI



Bottleneck XAI



How to make

Vision Language Models

explainable?

Llama 3.1



ChatGPT

...

LLM



smart but ...

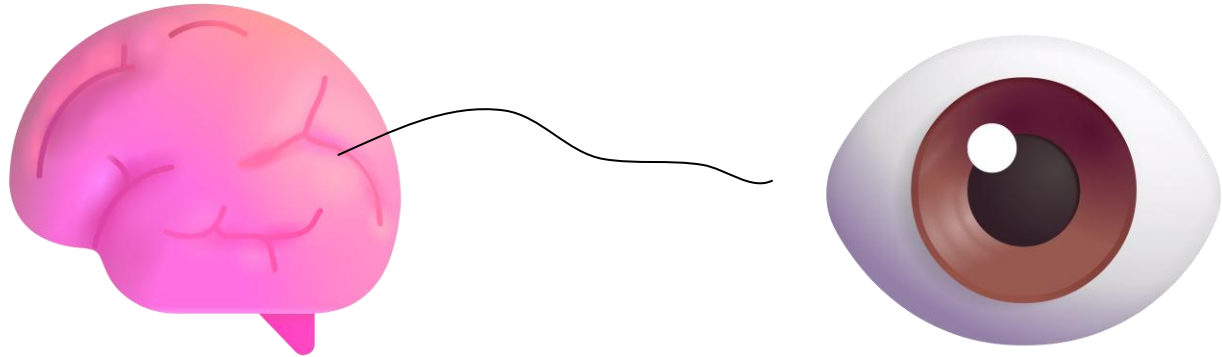
Llama 3.1



ChatGPT

...

LLM



smart but ... ***without eyes***

Llama 3.1



ChatGPT

...



GPT-4(Vision)

Claude 3 Sonnet



LLaVA



Gemini 1.5 Pro

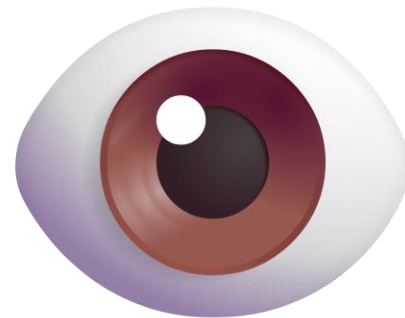
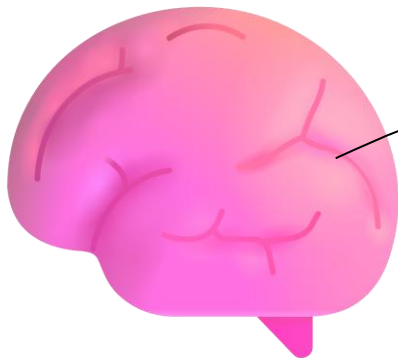


GPT-o1

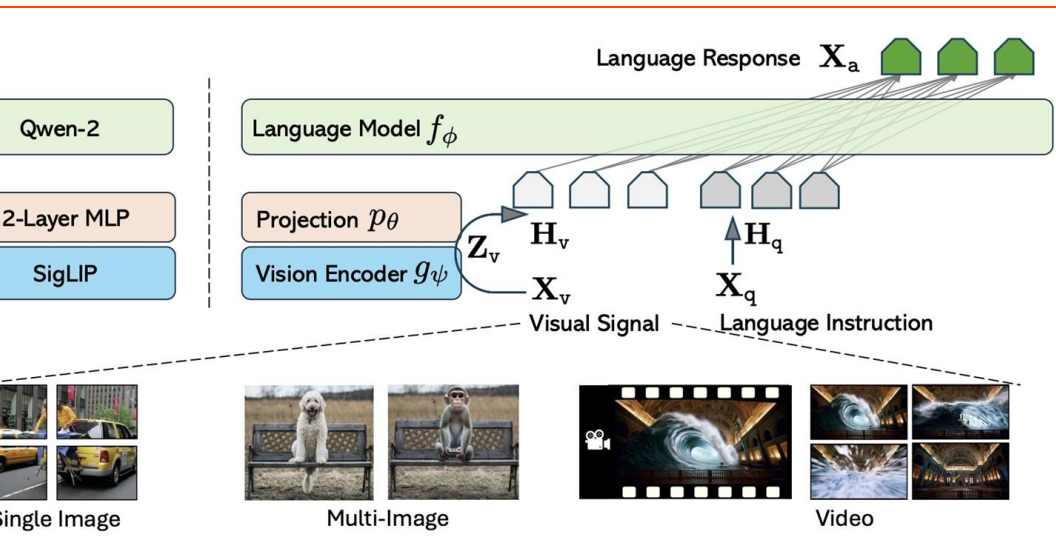
9 / 2023 -

12 / 2024 -

LLM



smart but *without eyes*



GPT-4(Vision)

9 / 2023 -

Claude 3 Sonnet



LLaVA



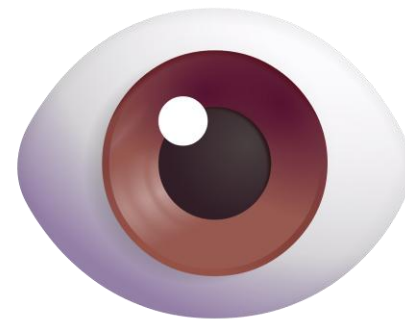
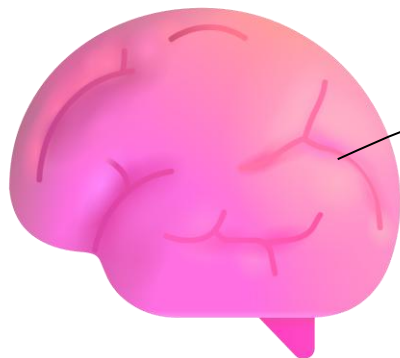
Gemini 1.5 Pro



GPT-o1

12 / 2024 -

LLM



smart but *without eyes*

A: 2010

Europe Sees China, Not U.S., as Leading Economic Power

Median across 5 European nations (France, Germany, Poland, Spain, UK) that name each as world's leading economic power



Q9: Which year shows the tiniest difference in values between China and US being seen as leading economic power across all the years?

Eval Sets	GPT-4o	GPT-4T 2024-04-09	Gemini 1.0 Ultra
			
MMMU (%) (val)	69.1	63.1	59.4
MathVista (%) (testmini)	63.8	58.1	53.0
AI2D (%) (test)	94.2	89.4	79.5
ChartQA (%) (test)	85.7	78.1	80.8
DocVQA (%) (test)	92.8	87.2	90.9



Vision understanding evals - GPT-4o achieves state-of-the-art performance on visual perception benchmarks. All vision evals are 0-shot, with MMMU, MathVista, and ChartQA as 0-shot CoT.

Question: ...The graph shown is compiled from data collected by Gallup <image 1>. Find the probability that the selected Emotional Health Index Score is between 80.5 and 82?

Options:

- (A) 0
 (B) 0.2142
(C) 0.3571
 (D) 0.5



Europe Sees China, Not U.S., as Leading Economic Power

Median across 5 European nations (France, Germany, Poland, Spain, UK) that name each as world's leading economic power



A: 2010

Q9: Which year shows the tiniest difference in values between China and US being seen as leading economic power across all the years?

Eval Sets	GPT-4o	GPT-4T 2024-04-09	Gemini 1.0 Ultra
			
MMMU (%) (val)	69.1	63.1	59.4
MathVista (%) (testmini)	63.8	58.1	53.0
AI2D (%) (test)	94.2	89.4	79.5
ChartQA (%) (test)	85.7	78.1	80.8
DocVQA (%) (test)	92.8	87.2	90.9



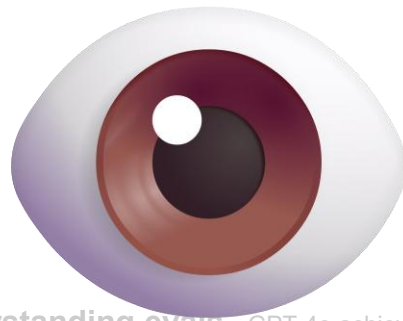
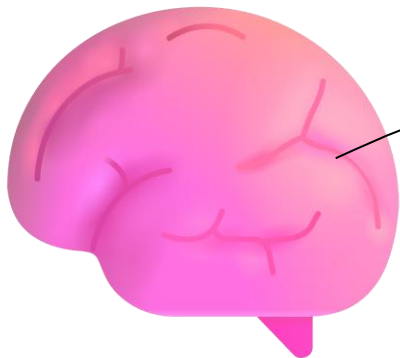
Vision understanding evals - GPT-4o achieves state-of-the-art performance on visual perception benchmarks. All vision evals are 0-shot, with MMMU, MathVista, and ChartQA as 0-shot CoT.

But,

But,

they have... blindspot!

LLM

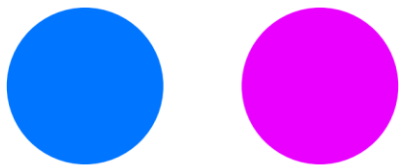


Vision understanding evals - GPT-4o achieves state-of-the-art performance on visual perception benchmarks. All vision evals are 0-shot, with MMMU, MathVista, and ChartQA as 0-shot CoT.

Are the two circles **touching**? Answer with only Yes/No.



Gemini 1.5 Pro:



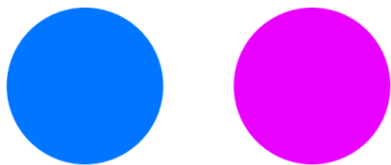
No



Are the two circles **touching**? Answer with only Yes/No.



Gemini 1.5 Pro:

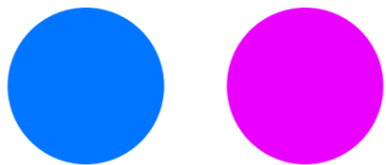


No



Are the two circles **touching**? Answer with only Yes/No.

✦ Gemini 1.5 Pro:

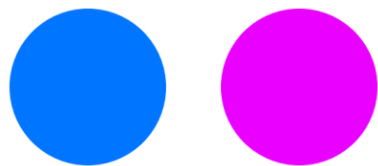


No



Are the two circles **touching**? Answer with only Yes/No.

✦ Gemini 1.5 Pro:

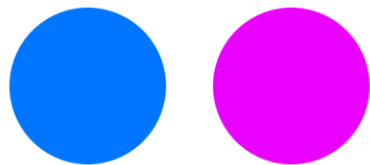


No



Are the two circles **touching**? Answer with only Yes/No.

✦ Gemini 1.5 Pro:



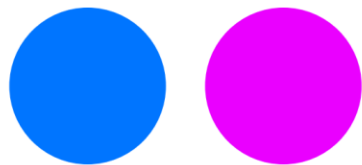
No



Are the two circles **touching**? Answer with only Yes/No.



Gemini 1.5 Pro:

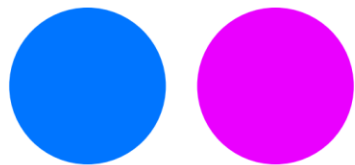


No



Are the two circles **touching**? Answer with only Yes/No.

✦ Gemini 1.5 Pro:

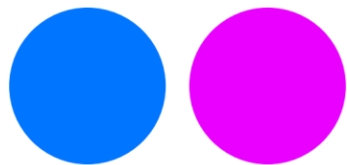


No



Are the two circles **touching**? Answer with only Yes/No.

✦ Gemini 1.5 Pro:

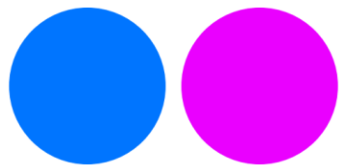


Yes !

Are the two circles **touching**? Answer with only Yes/No.



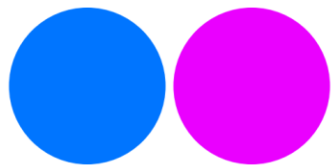
Gemini 1.5 Pro:



Yes !

Are the two circles **touching**? Answer with only Yes/No.

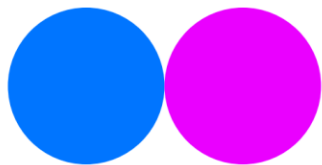
✦ Gemini 1.5 Pro:



Yes !

Are the two circles **touching**? Answer with only Yes/No.

✦ Gemini 1.5 Pro:

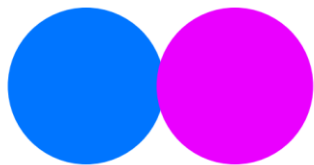


Yes



Are the two circles **touching**? Answer with only Yes/No.

✦ Gemini 1.5 Pro:

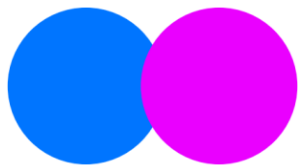


Yes



Are the two circles **touching**? Answer with only Yes/No.

✦ Gemini 1.5 Pro:



Yes



Are the two circles **touching**? Answer with only Yes/No.



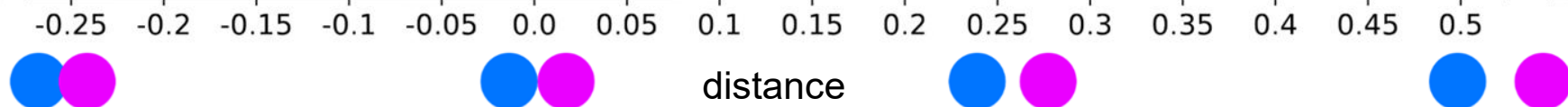
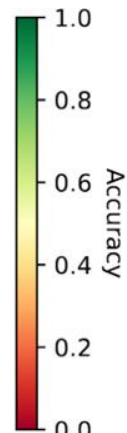
Pooyan
Rahmazadehgervi



Logan
Bolton

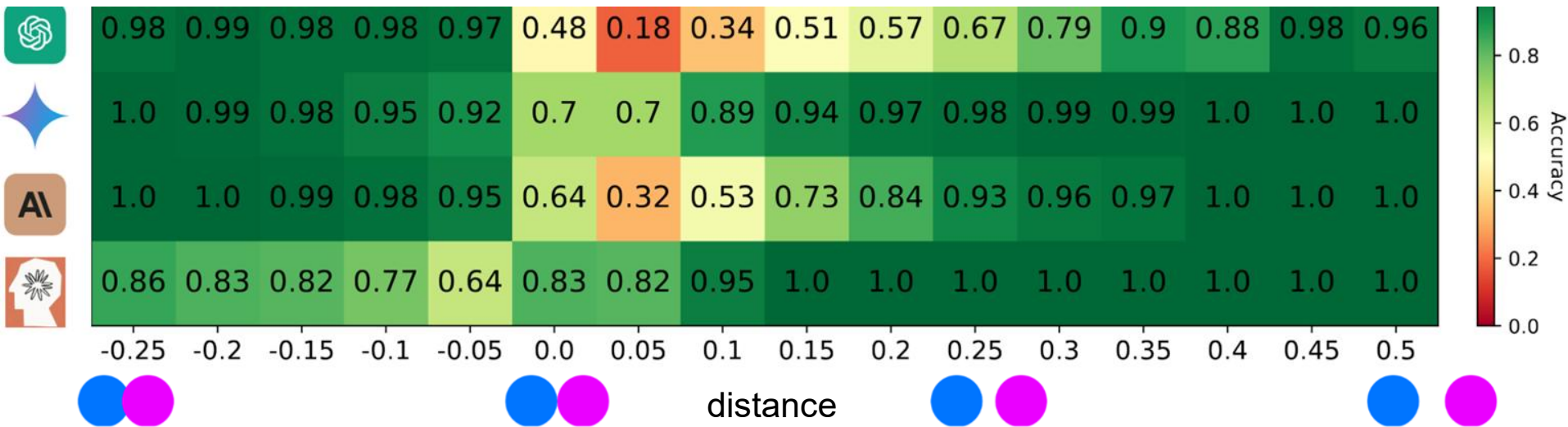


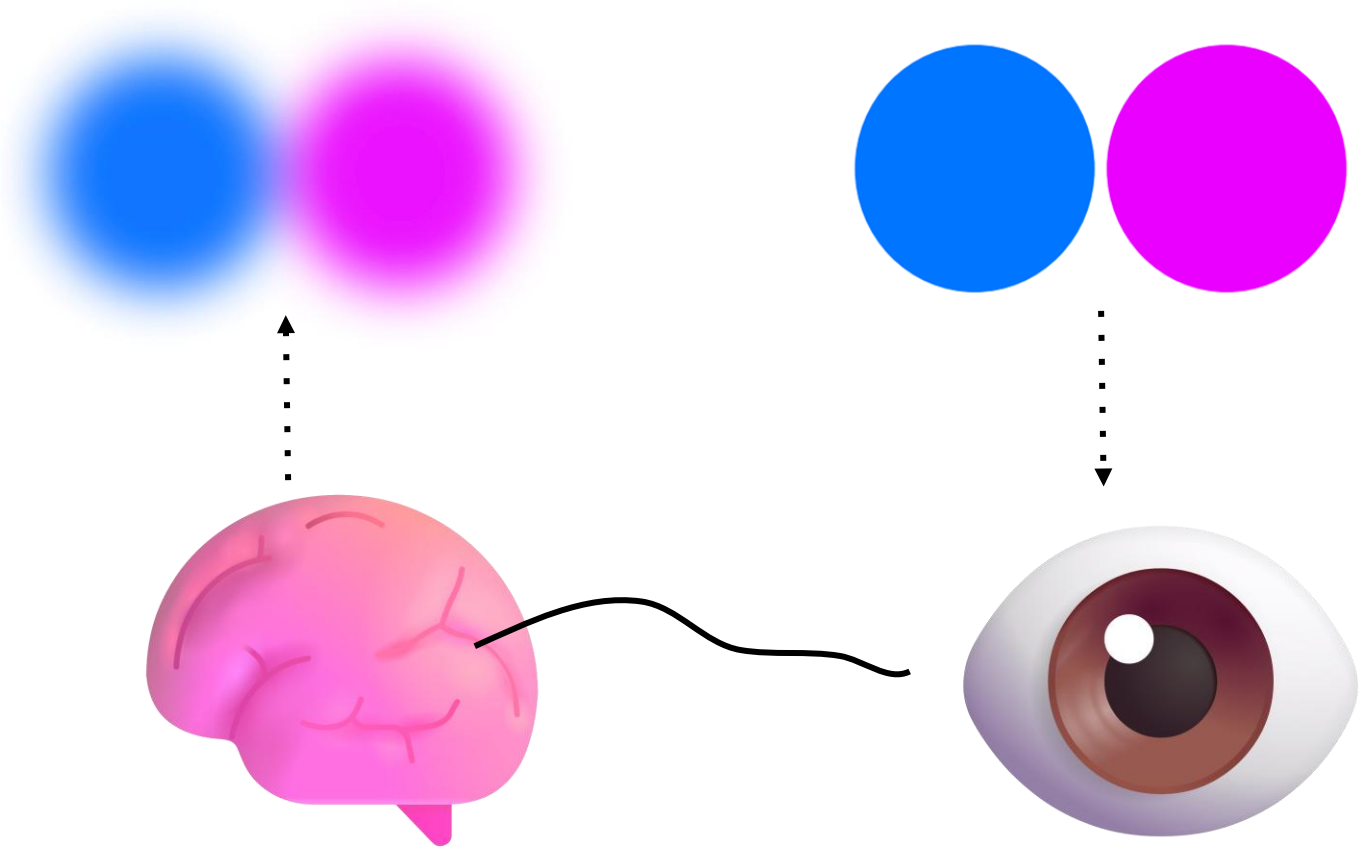
Mohammad Reza
Taesiri



Are the two circles **touching**? Answer with only Yes/No.

VLMs: Not sure when two
circles are close together

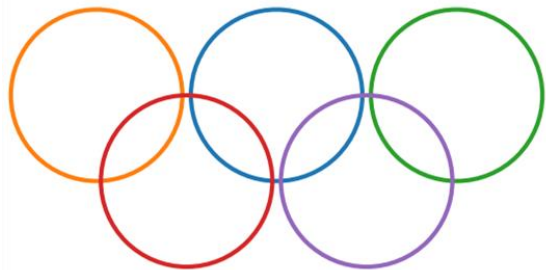




VLMs myopic?

Vision Language Models are Blind. ACCV 2024

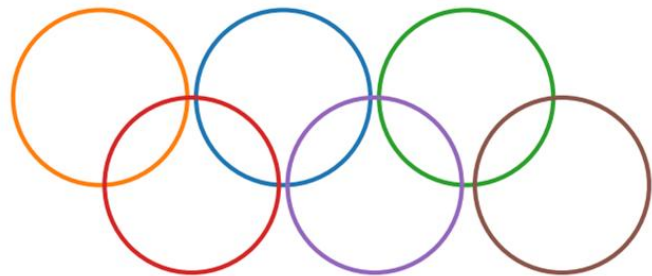
How many circles are in the image? Answer with only the number in numerical format.



✦ Gemini 1.5 Pro:

?

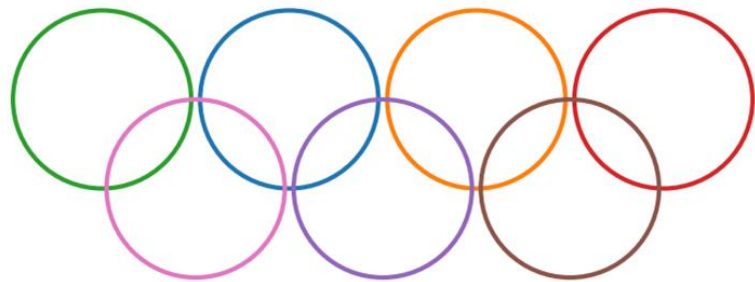
How many circles are in the image? Answer with only the number in numerical format.



✦ Gemini 1.5 Pro:

?

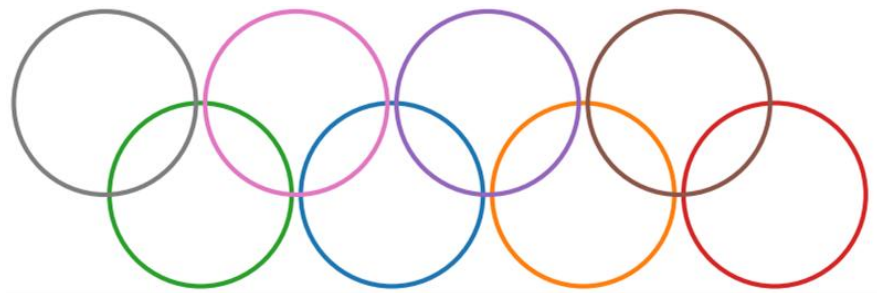
How many circles are in the image? Answer with only the number in numerical format.



✦ Gemini 1.5 Pro:

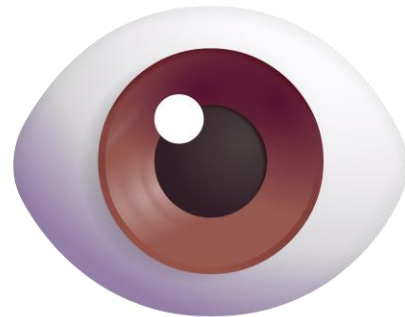
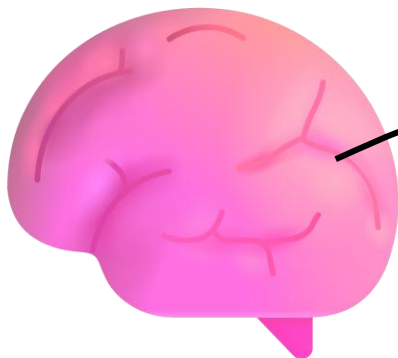
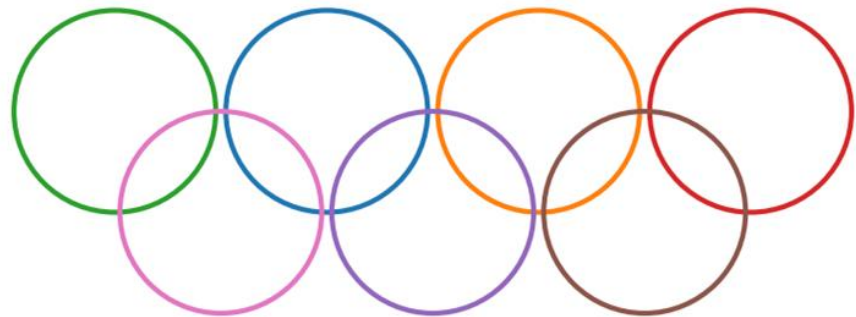
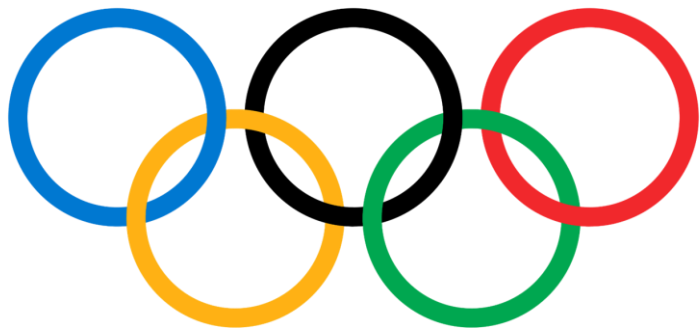
?

How many circles are in the image? Answer with only the number in numerical format.

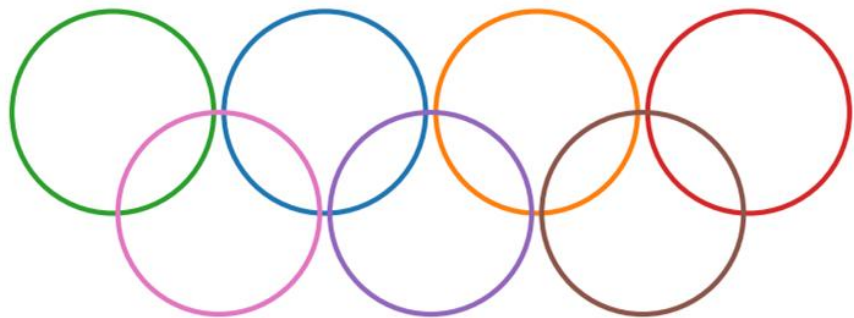


✦ Gemini 1.5 Pro:

?



VLMs biased towards Olympic logo?

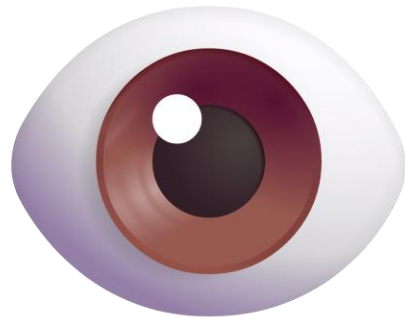


(a) Circles ○○○

	0.83	0.38	0.46	0.22	0.18
	1.0	0.0	0.0	0.0	0.01
	1.0	0.26	0.04	0.18	0.01
	0.88	0.7	0.67	0.57	0.51
	5	6	7	8	9

99.74%

predicts "5"



VLMs biased towards Olympic logo?

No VLMs can solve **BlindTest** at 100% accuracy

Model	Size		 							Task mean
Random		33.33	50.00	5.77	20.00	20.00	25.00	4.55	33.33	24.00
GPT-4o	n/a	41.61	75.91	74.23	41.25	20.21	55.83	39.58	53.19	50.23
Gemini-1.5	n/a	66.94	93.62	83.29	20.25	24.17	87.08	39.39	53.13	58.48
Sonnet-3	n/a	43.41	86.46	72.06	29.79	1.87	65.00	36.17	31.11	45.73
Sonnet-3.5	n/a	75.36	90.82	87.88	66.46	77.71	92.08	74.26	58.19	77.84

58%
accuracy

No VLMs can solve **BlindTest** at 100% accuracy

Model	Size									Task mean
Random		33.33	50.00	5.77	20.00	20.00	25.00	4.55	33.33	24.00
GPT-4o	n/a	41.61	75.91	74.23	41.25	20.21	55.83	39.58	53.19	50.23
Gemini-1.5	n/a	66.94	93.62	83.29	20.25	24.17	87.08	39.39	53.13	58.48
Sonnet-3	n/a	43.41	86.46	72.06	29.79	1.87	65.00	36.17	31.11	45.73
Sonnet-3.5	n/a	75.36	90.82	87.88	66.46	77.71	92.08	74.26	58.19	77.84
LLaVA OneVision-qwen2-ov	72B	45.83	90.92	44.71	20.00	11.74	87.07	8.95	58.06	45.92
LLaVA OneVision-qwen2-si	72B	45.33	83.48	38.14	20.00	11.46	57.50	10.23	48.06	38.41
LLaVA OneVision-qwen2-ov	7B	48.17	83.93	42.79	20.00	7.29	42.92	21.02	47.22	39.17
LLaVA OneVision-qwen2-si	7B	44.50	84.67	40.22	20.00	7.29	58.75	14.01	55.00	40.00
LLaVA OneVision-qwen2-ov	0.5B	17.28	75.07	9.78	12.50	9.58	20.42	0.38	5.56	18.82
LLaVA OneVision-qwen2-si	0.5B	33.14	73.21	6.25	27.29	2.50	14.58	1.13	26.11	23.03
InternVL-2	8B	47.28	91.00	57.69	20.00	13.96	28.33	7.57	60.28	40.76
Phi-3.5-vision-instruct	4.2B	37.78	83.63	16.51	18.75	11.46	32.50	11.74	19.72	29.01
Mean		45.55	84.39	47.79	26.36	16.60	53.50	22.03	42.97	42.28

ChatGPT 5. No extended reasoning

ChatGPT 5 ▾

↑ Share



Count the legs of this animal. Answer with a number in curly brackets, e.g., {9}.

{4}



Vision Language Models are Biased. ICLR 2026



An
Vo



Khai-Nguyen
Nguyen

Count the legs of this animal. Answer with a number in curly brackets, e.g., {9}.



o4-mini

answer: 4 ×

Vision Language Models are Biased. ICLR 2026

```
2025-06-12 13:12:05.577499 [CODE] (node: 80b513699-43e7-494a-b826-5a314c532182) from PIL import Image
import matplotlib.pyplot as plt
img = Image.open('/mnt/data/image_1749733918161.png')
plt.imshow(img)
plt.axis('off')
2025-06-12 13:12:08.761760.5, -0.5) Aggregate Result: status: success run_id: 28842b0f-5ab7-46fd-918d-db7146c7f887 start_time: 1749733942.1730304
1749733933.6227942 code: import matplotlib.pyplot as plt
```

Run : 24 seconds

```
img = Image.open('/mnt/data/image_1749733918161.png')
plt.imshow(img)
plt.axis('off') end_time: 1749733933.6227942 final_expression_output: (-0.5, 1151.5, 760.5, -0.5) in_kernel_exception: None
system_exception: None messages: [{"message_type": "image", "time": 1749733933.570759, "sender": "server", "image_payload": None,
'image_url': 'file-service://file-GIER2isn2p3EY64hfjktYR', 'width': 1438, 'height': 964}] jupyter_messages: [{"msg_type": 'status',
'parent_header': {'msg_id': '442969cb-41434b4aed91f12512385924_3_1', 'version': '5.3'}, 'content': 'execution_state: 'busy'}], {'msg_type':
'execute_input', 'parent_header': {'msg_id': '442969cb-41434b4aed91f12512385924_3_1', 'version': '5.3'}, 'content': 'execute_result',
'parent_header': {'msg_id': '442969cb-41434b4aed91f12512385924_3_1', 'version': '5.3'}, 'content': 'data': {'text/plain': '(-0.5, 1151.5, 760.5, -
0.5)'}}, {'msg_type': 'display_data', 'parent_header': {'msg_id': '442969cb-41434b4aed91f12512385924_3_1', 'version': '5.3'}, 'content': 'data':
{'text/plain': '<Figure size 2000x1200 with 1 Axes>', 'image/vnd.openai.file-service-2.png', 'url': 'file-service://file-GIER2isn2p3EY64hfjktYR',
'image_width': 1438, 'image_height': 964, 'bytes': '2367698'}, 'image/vnd.openai.file-service.png': 'file-service://file-
GIER2isn2p3EY64hfjktYR'}], {'msg_type': 'status', 'parent_header': {'msg_id': '442969cb-41434b4aed91f12512385924_3_1', 'version': '5.3'},
'content': 'execution_state: 'idle'}}] timeout_triggered: None
```

2025-06-12 13:12:13.649898 [THOUGHT] The image shows a cow walking. From the side view, the cow has four legs, but the visibility depends on the angle. On the near side, two legs are visible: the front and rear left. The front right is partially visible. On the far side, the front and hind left legs might be faintly spotted behind the body. In most side views, typically only one leg from the far side is hard to see, but in this picture, we see parts of both. So, all four legs are visible with some hidden details.

```
2025-06-12 13:12:13.689116 [CODE] (node: 3e817297-12a9-40e2-9610-bfa0bd89d03) import numpy as np
# crop area around front legs
h, w = np.array(img).shape[:2]
crop = img.crop((w*0.5, h*0.5, w*1.0, h*1.0))
plt.imshow(crop)
plt.axis('off')
```

2025-06-12 13:12:17.903386 [THOUGHT] In this image, I see four legs — two at the front and two at the back. The front right leg seems extended forward, while the front left (on the far side) is positioned behind it. The back legs are also visible: the hind right leg is located near the udder, with the far side hind left leg partially hidden. The visibility of the legs can be tricky since the cow is in motion, but it looks like all four legs are visible, just in different positions.

2025-06-12 13:12:17.903386 [THOUGHT] Okay, the cow seems to have four visible legs. From left to right in the image: the leftmost leg is the near side hind leg, followed by the hind left leg (near the tail). Next up, the hind right leg is slightly visible, with the udder to the right of it. The near side hind leg is behind the udder. Then, moving to the front, the far side front leg is mostly blocked by the body, but I think it's visible, with another leg further to the right. So, all four legs are visible, but a bit tricky due to positioning.

2025-06-12 13:12:17.903386 [THOUGHT] From the image, I can see two front legs and two back legs. The near side front leg is extended, while the far side front leg is partially behind the body, so it's not fully visible. On the back side, both hind legs are visible, with one partly hidden behind the udder and the other more fully visible to the left. The hooves of both front legs seem to be visible, with a bit of overlap. I'll crop the image to focus on the front leg region for further clarity.

```
2025-06-12 13:12:22.173030 [TOOL EXEC] (node: 780d080a-a0e5-4587-a402-9fdcd4d05682) Output: <<ImageDisplayed>> (-0.5, 575.5,
380.5, -0.5) Aggregate Result: status: success run_id: 28842b0f-5ab7-46fd-918d-db7146c7f887 start_time: 1749733942.1730304
update_time: 1749733944.475133 code: import numpy as np
```

```
# crop area around front legs
h, w = np.array(img).shape[:2]
crop = img.crop((w*0.5, h*0.5, w*1.0, h*1.0))
plt.imshow(crop)
plt.axis('off') end_time: 1749733944.475133 final_expression_output: (-0.5, 575.5, 380.5, -0.5) in_kernel_exception: None system_exception:
None messages: [{"message_type": "image", "time": 1749733944.4242198, "sender": "server", "image_payload": None, "image_url": "file-
service://file-JhZ7edbMEVAKKsvZRLTDwq", "width": 1436, "height": 964}] jupyter_messages: [{"msg_type": 'status', 'parent_header': {'msg_id':
'442969cb-41434b4aed91f12512385924_3_1', 'version': '5.3'}, 'content': 'execution_state: 'busy'}], {'msg_type': 'execute_input',
'parent_header': {'msg_id': '442969cb-41434b4aed91f12512385924_3_1', 'version': '5.3'}, 'content': 'execute_result', 'parent_header':
{'msg_id': '442969cb-41434b4aed91f12512385924_3_1', 'version': '5.3'}, 'content': 'data': {'text/plain': '(-0.5, 575.5, 380.5, -0.5)'}},
{'msg_type': 'display_data', 'parent_header': {'msg_id': '442969cb-41434b4aed91f12512385924_3_1', 'version': '5.3'}, 'content': 'data':
{'text/plain': '<Figure size 2000x1200 with 1 Axes>', 'image/vnd.openai.file-service-2.png', 'url': 'file-service://file-
JhZ7edbMEVAKKsvZRLTDwq', 'image_width': 1436, 'image_height': 964, 'bytes': '2367698'}, 'image/vnd.openai.file-service.png': 'file-
service://file-JhZ7edbMEVAKKsvZRLTDwq'}], {'msg_type': 'status', 'parent_header': {'msg_id': '442969cb-41434b4aed91f12512385924_3_1', 'version': '5.3'},
'content': 'execution_state: 'idle'}}] timeout_triggered: None
```



How many legs does this animal have? Answer with a number in curly brackets, e.g., {9}.

Ground Truth: 3 Bias: 2



How many legs does this animal have? Answer with a number in curly brackets, e.g., {9}.

Ground Truth: 3 Bias: 2



How many legs does this animal have? Answer with a number in curly brackets, e.g., {9}.

Ground Truth: 5 Bias: 4



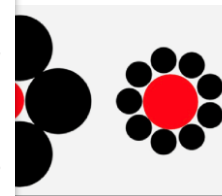
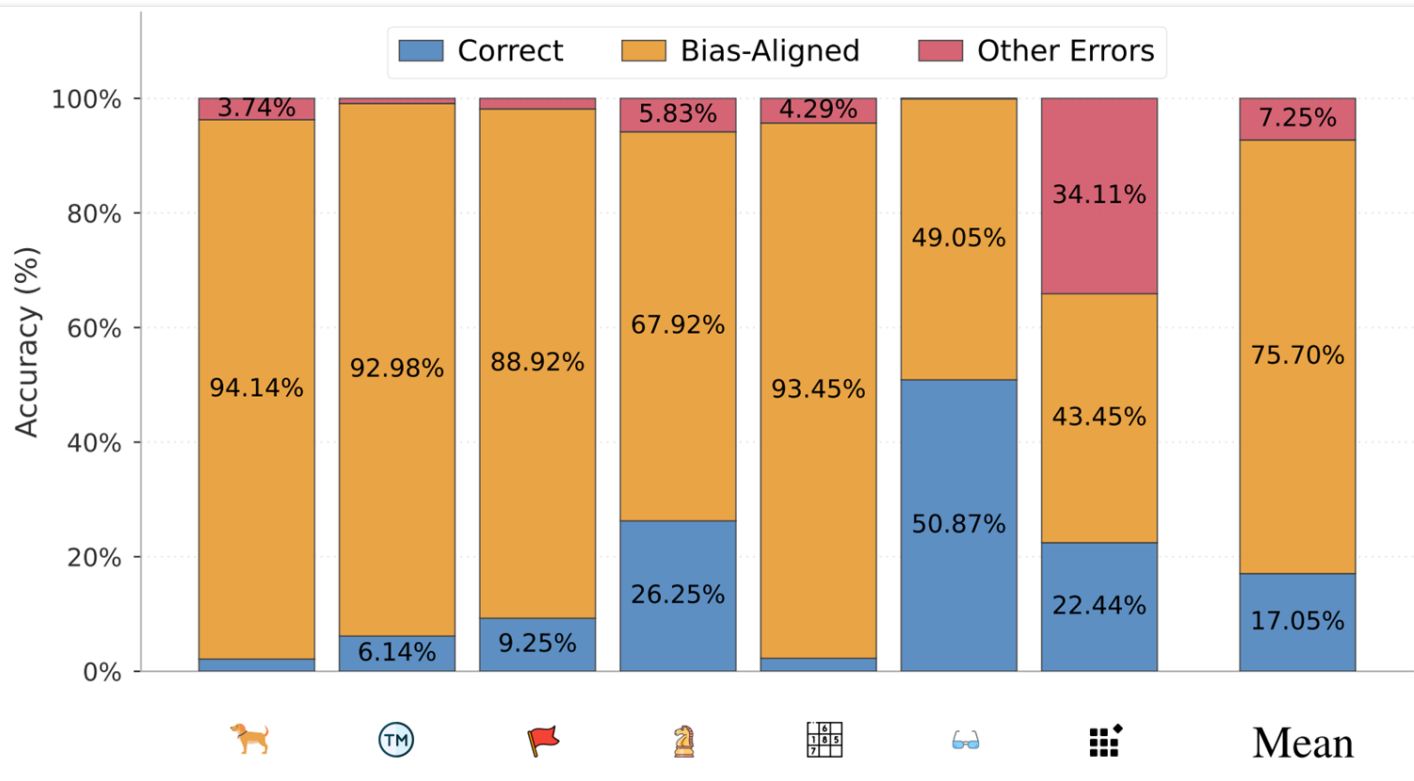
How many legs does this animal have? Answer with a number in curly brackets, e.g., {9}.

Ground Truth: 5 Bias: 4

VLMs are Biased in 7 tasks over 7 domains

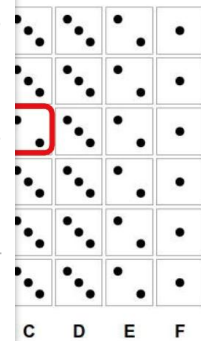


Flags



Are the two red circles of equal size?

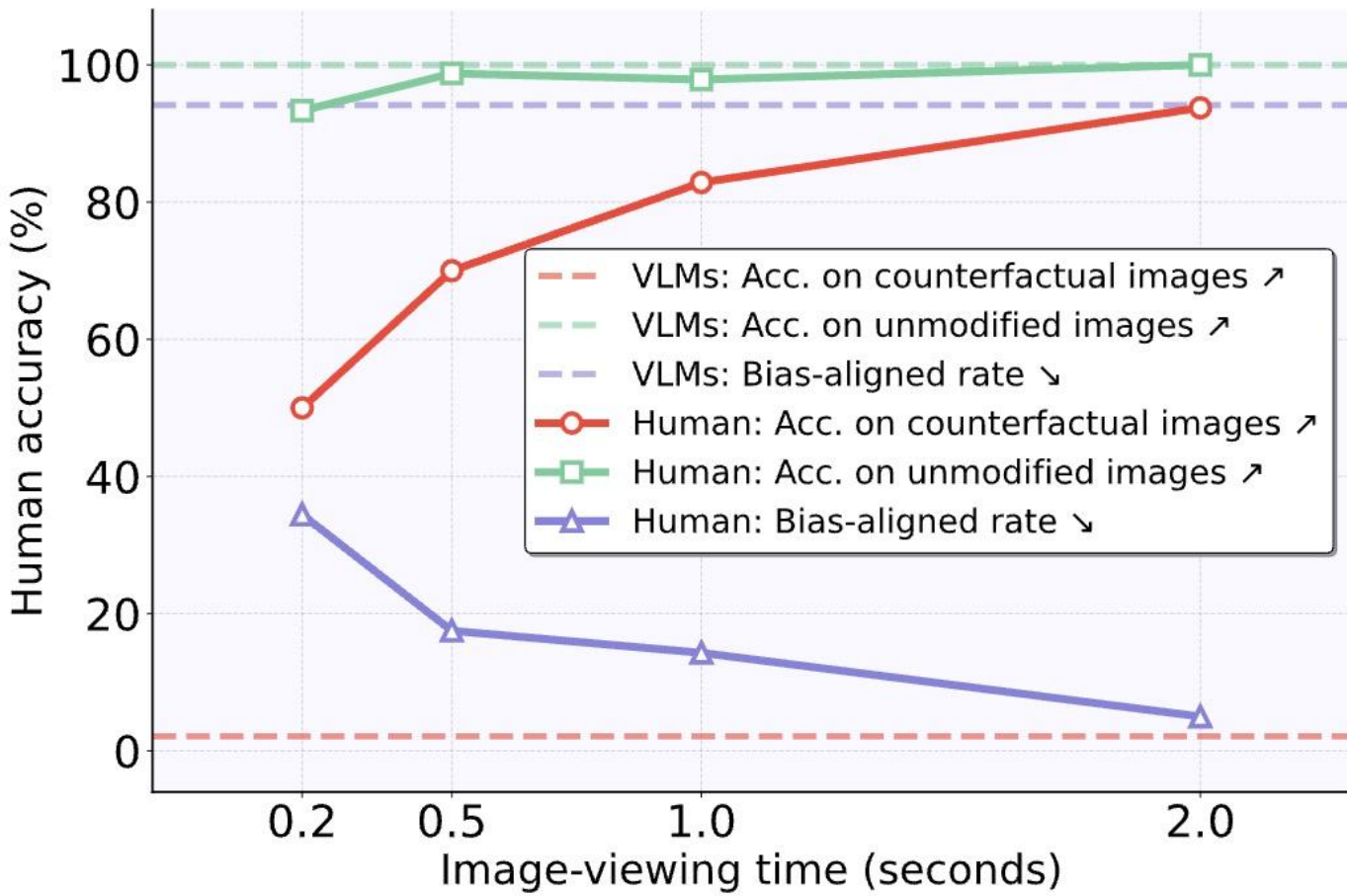
Counted grid



Count the dots in cell C3

count the stars

Vision Language Models are more biased than humans



78 humans: **93.75%**

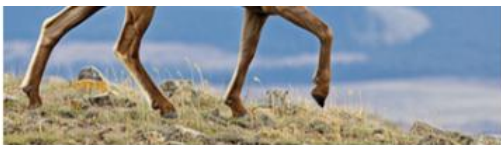
5 VLMs: **2.12%**



How many legs does this animal have? Answer with a number in curly brackets, e.g., {9}.

count legs

Tt



Prompt

Question: Based on the examples provided above, please answer the following question: How many legs does this animal have? Answer with a number in curly brackets, e.g., {9}. After providing your answer, briefly explain your reasoning and describe what you observed in the examples.

Prompt



 o4-mini: {4} Explanation: In the two examples, both the giraffe and the bighorn sheep are four-legged mammals (**the "5-legged" label on the giraffe was clearly mistaken**). Tigers, like other large cats, also have one pair of forelegs and one pair of hindlegs, for a total of four. **✗**

Ground truth: 5 **✓**

Lucas Beyer (bl16) 
@giffmana · Follow

Oh wow, this VLM benchmark is pure evil, and I love it!

"Vision Language Models are Biased" by @an_vo12, @taesiri, @anh_ng8, et al.

Also really good idea to have one-click copy-paste of images and prompts, makes trying it super easy.



2:27 PM · Aug 8, 2025



929 Reply Copy link

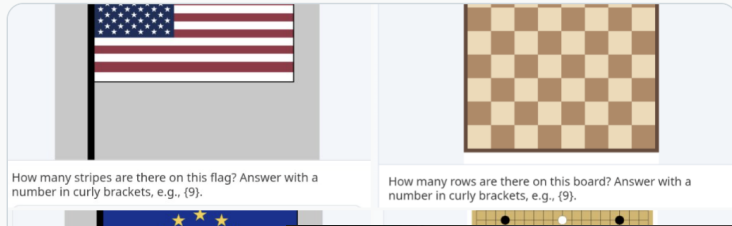
Read 32 replies



 (((yoav' (J)(J))) 
@yoavgo · Follow



beautiful adversarial dataset playing exactly on the soft-spot of VLMs.



An Vo @_vo_an
Our latest work shows that (GPT-4o, Gemini Pro) fail at counting legs due to

See simple cases where VLMs fail to count them.

Think your VLM can do better. [vlmsarebiased.github.io/#examples](https://github.com/yoavgo/vlmsarebiased)

1/n
#ICML2025



This 'chicken' test shows why AI could be so deadly in war



Sky News
9.27M subscribers

Subscribe

4.8K



Share

Ask

Vision Language Models are Biased. ICLR 2026

Lucas Beyer (bl16) ✓
@giffmana · Follow

Oh wow, this VLM benchmark is pure evil, and I love it!

"Vision Language Models are Biased" by @an_vo12, @taesiri, @anh_ng8, et al.

Also really
and prom



2:27 PM · Aug 8, 2025

929 Reply Copy link

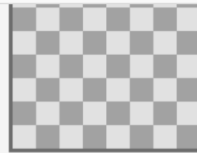
Read 32 replies

🔗 (((yoav' (J)(J))) ✓
@yoavgo · Follow

beautiful adversarial dataset playing exactly on the soft-spot of VLMs.



How many stripes are there on this flag? Answer with a



How many rows are there on this board? Answer with a

It's important to debug and intervene on VLMs

Pro) fail at counting legs due

See simple cases where VLM

Think your VLM can do be
vmlsarebiased.github.io/#exa

1/n
#ICML2025



This 'chicken' test shows why AI could be so deadly in war



Sky News
9.27M subscribers

Subscribe

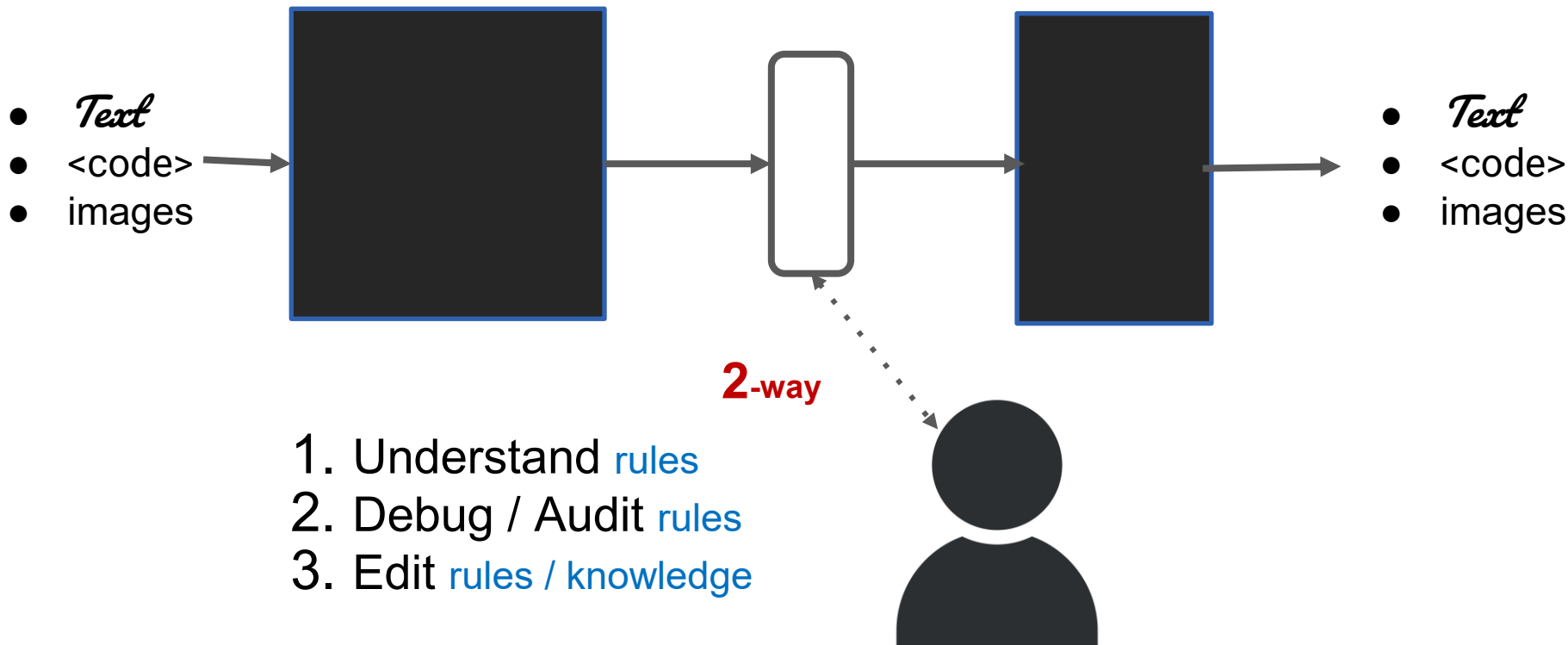
4.8K



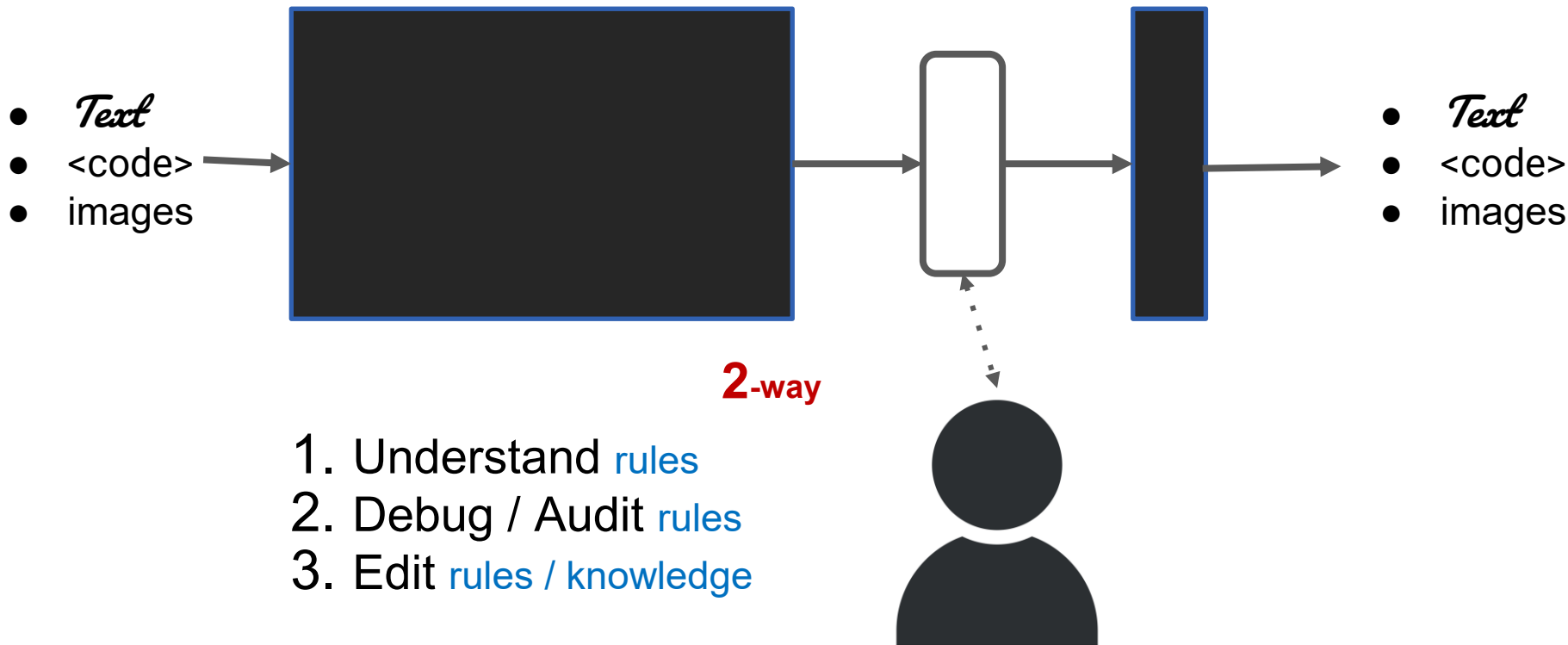
Share

Ask

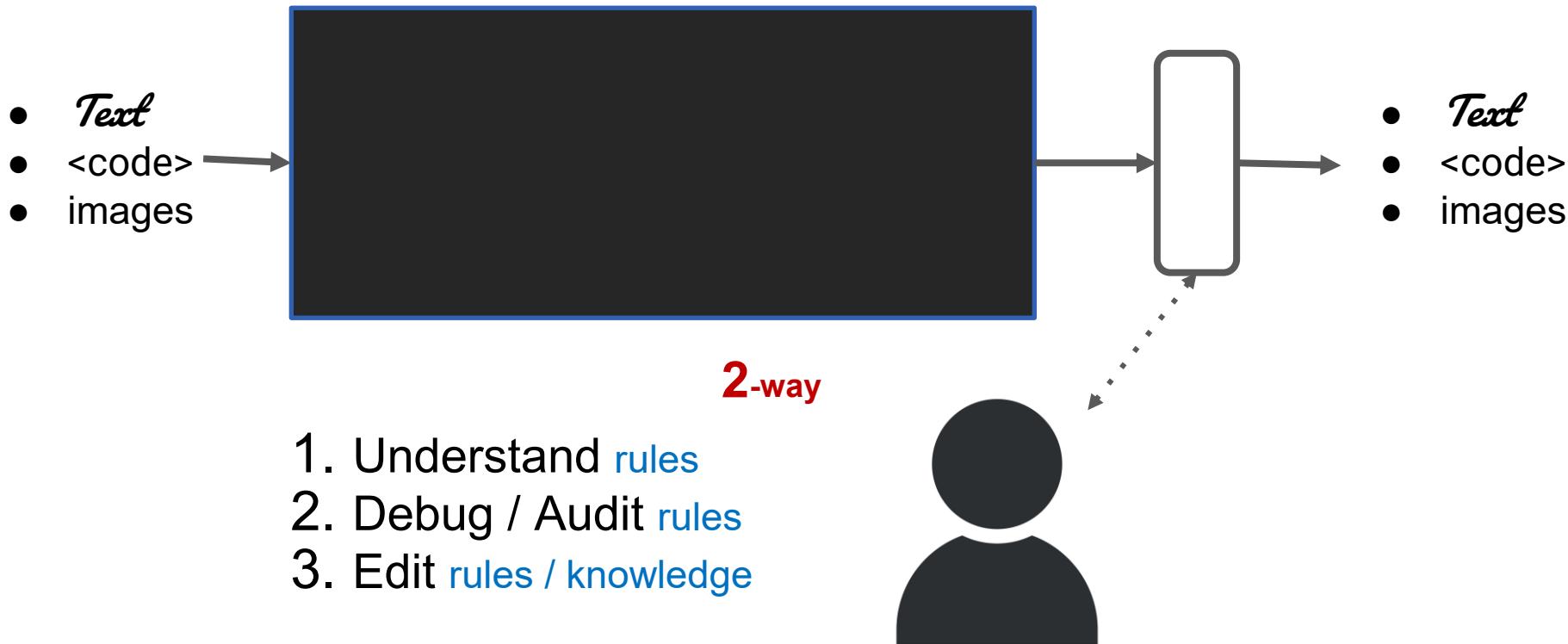
Bottleneck XAI



Bottleneck XAI

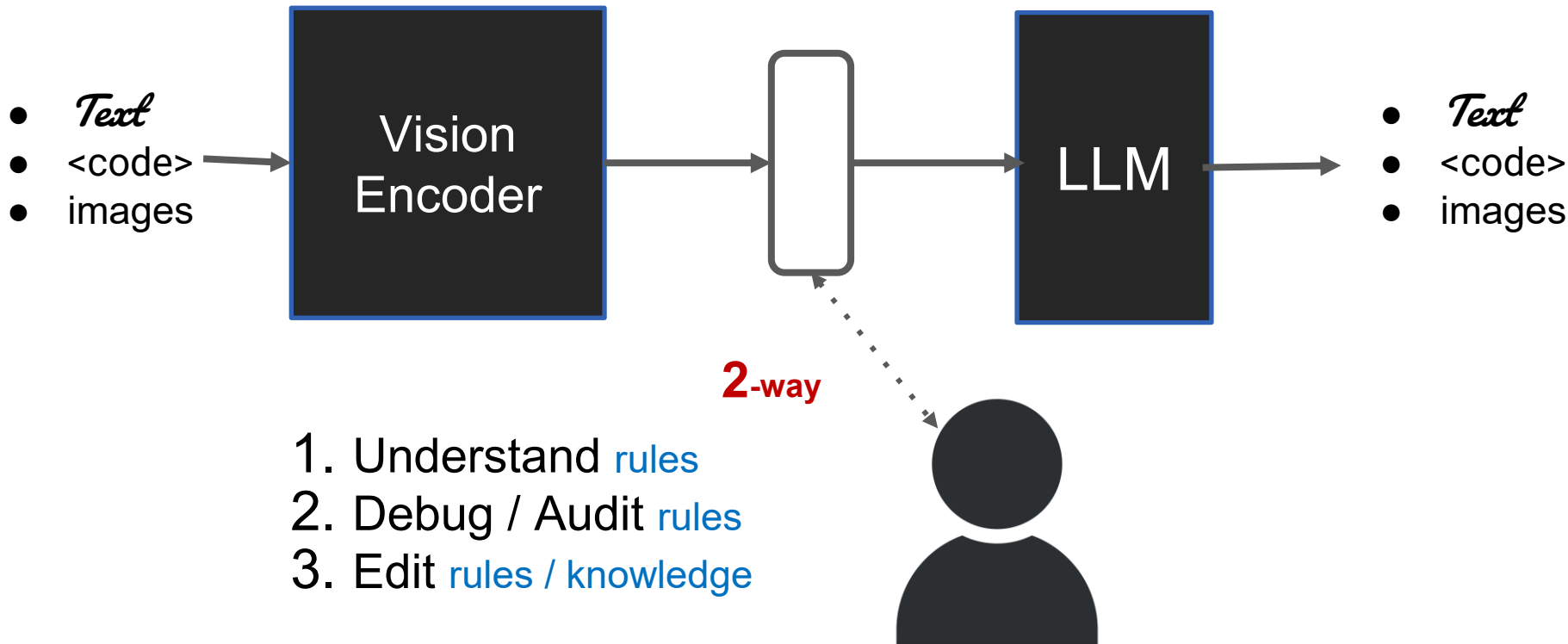


Bottleneck XAI



1. Transformer Attention Bottleneck

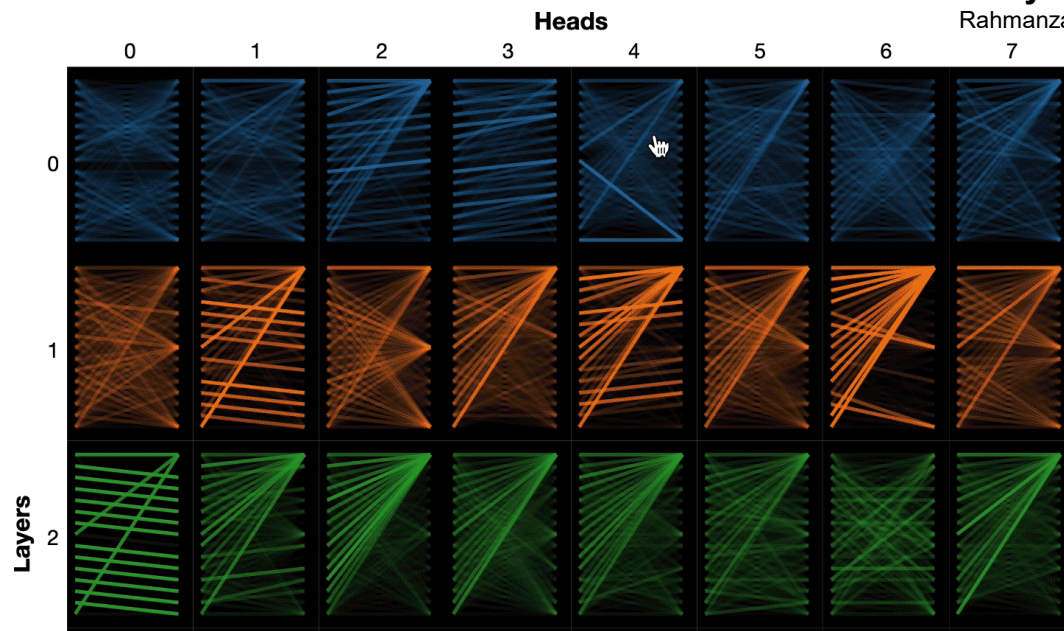
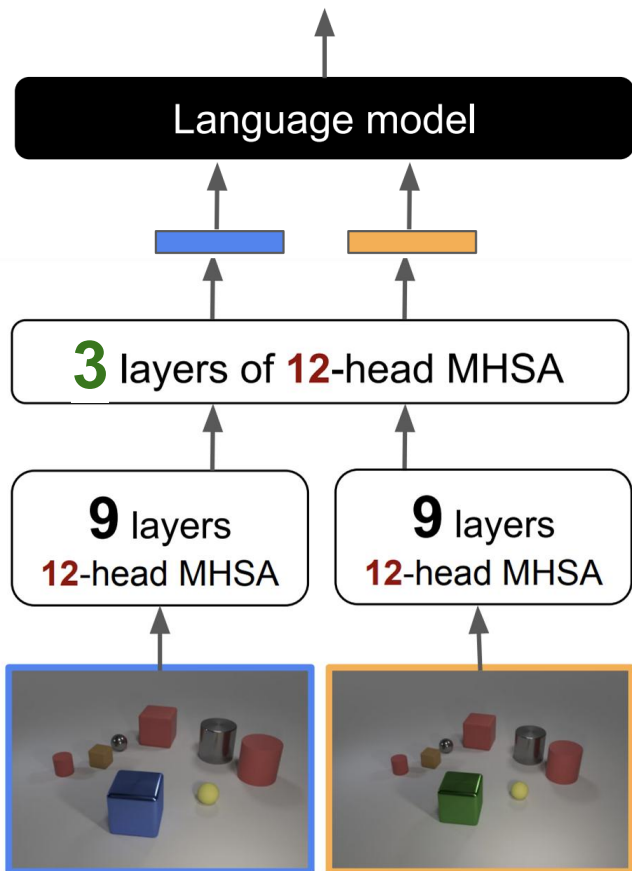
TAB: Transformer Attention Bottlenecks enable User Intervention and Debugging in Vision-Language Models. ICCV 2025



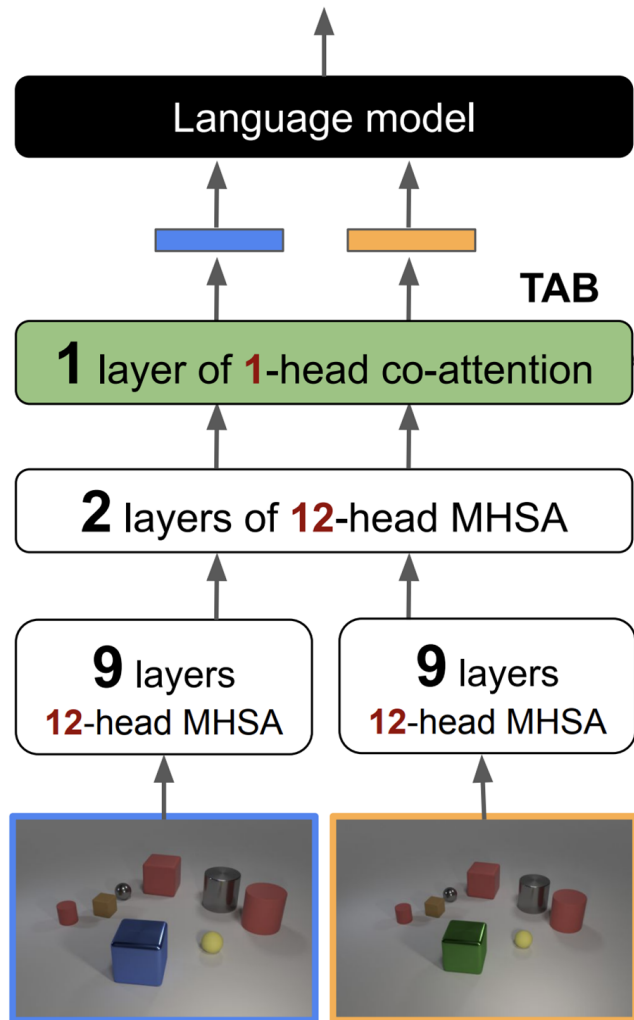


Pooyan
Rahmanzadehgervi

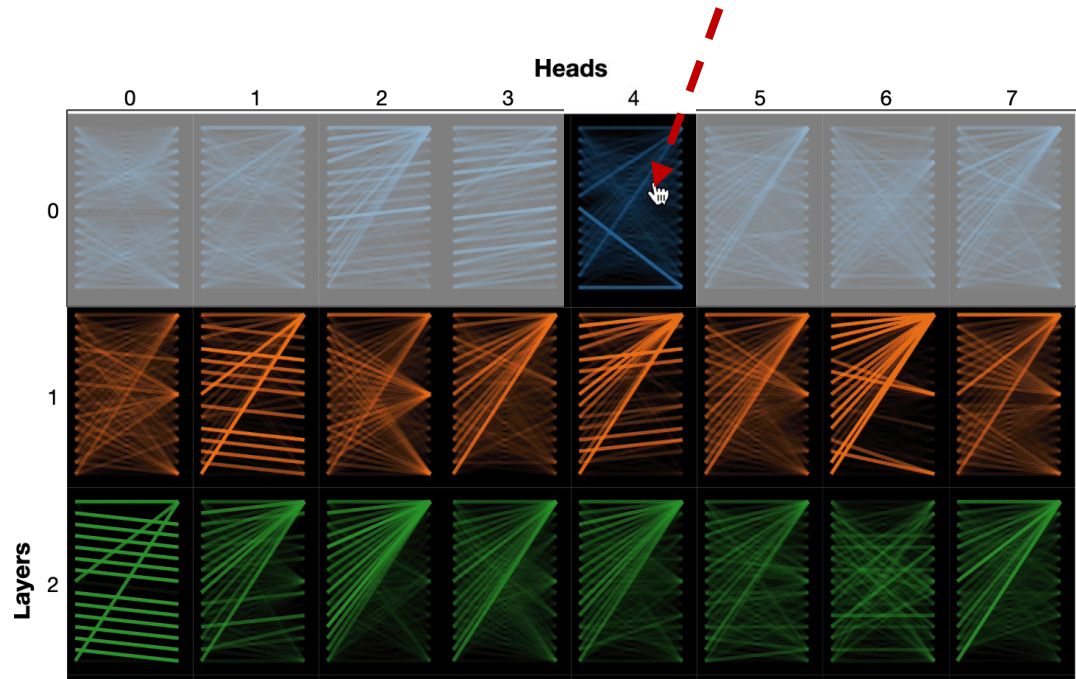
“The large blue cube turned green”



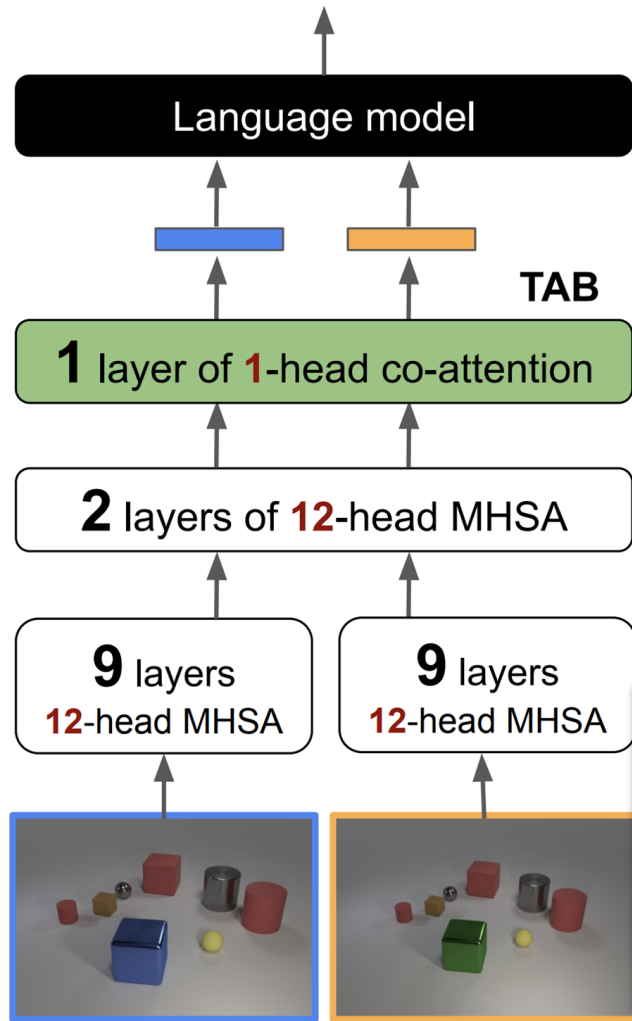
“The large blue cube turned green”



Bottleneck

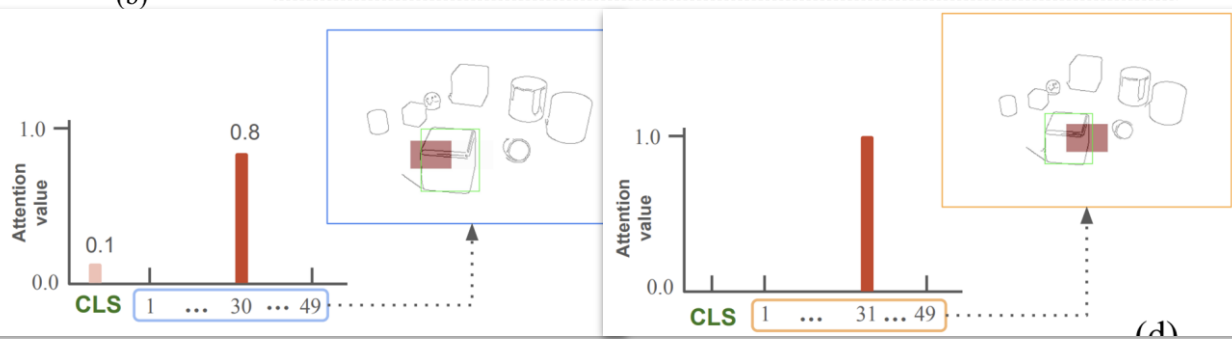
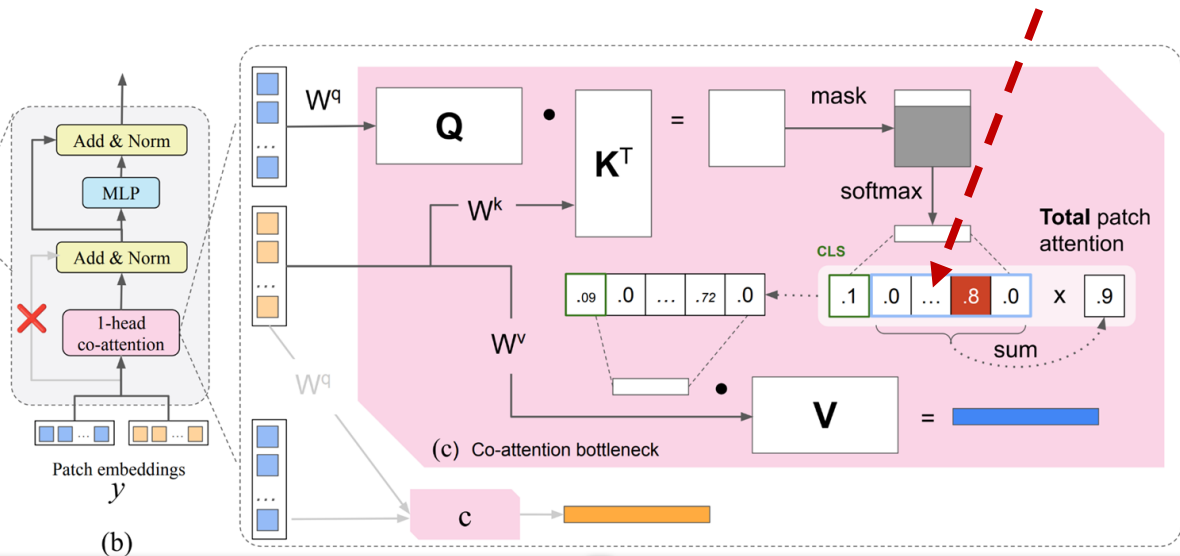


“The large blue cube turned green”

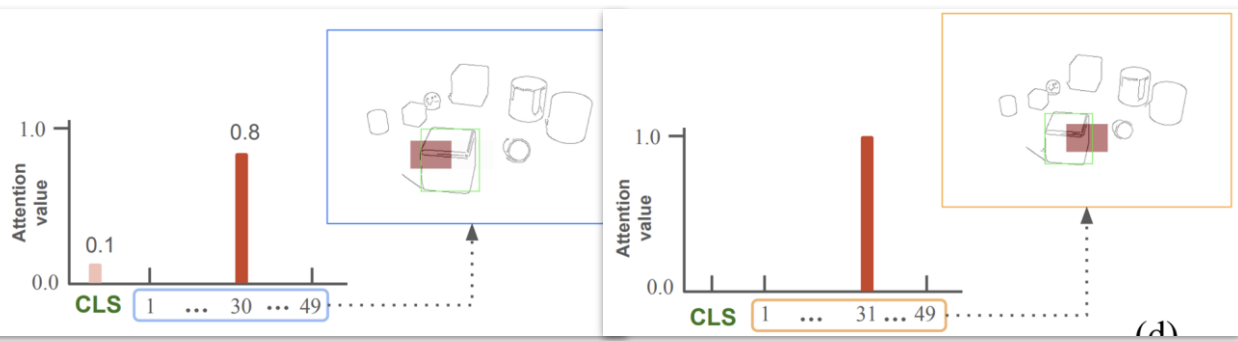
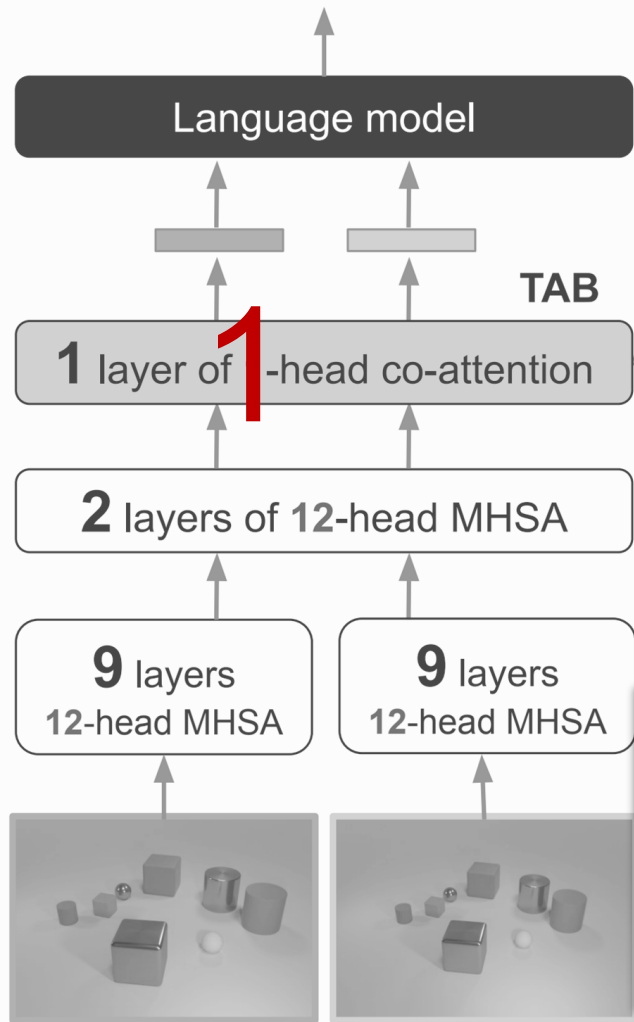


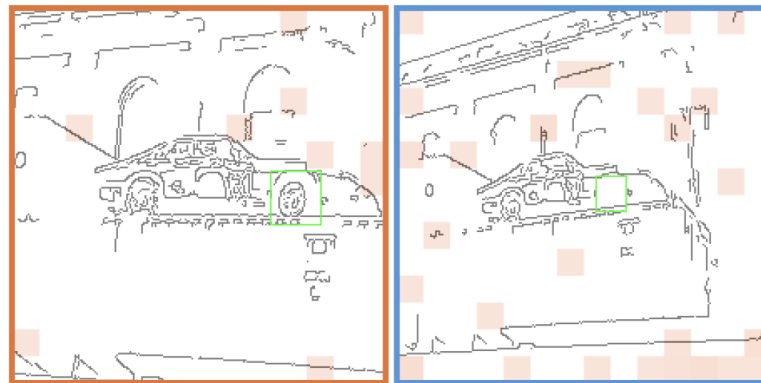
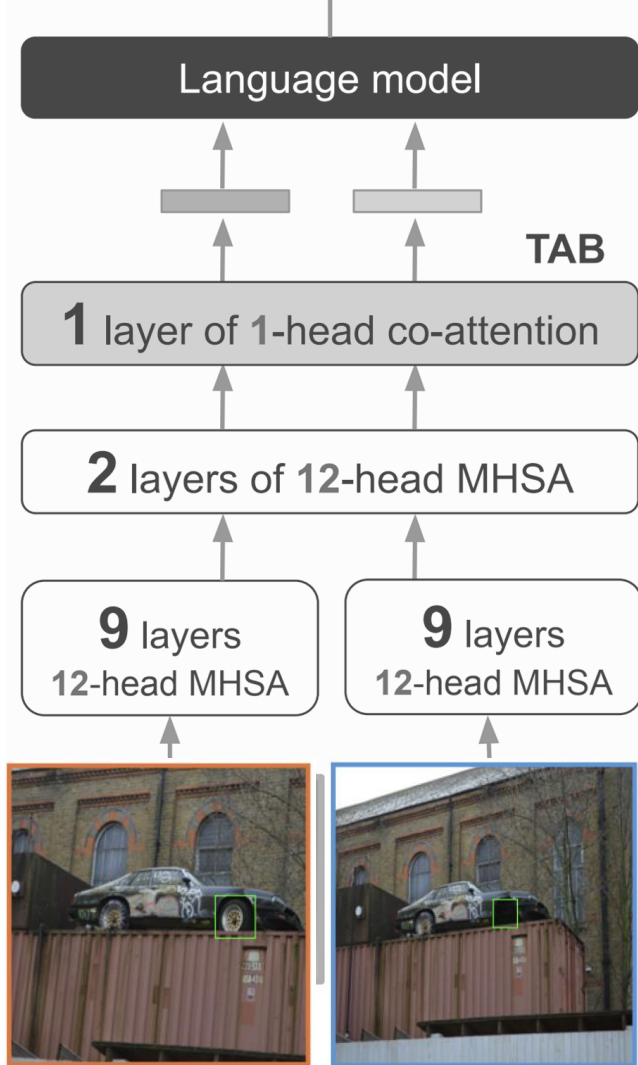
- gating
- renormalization

Editable bottleneck



“The large blue cube turned green”

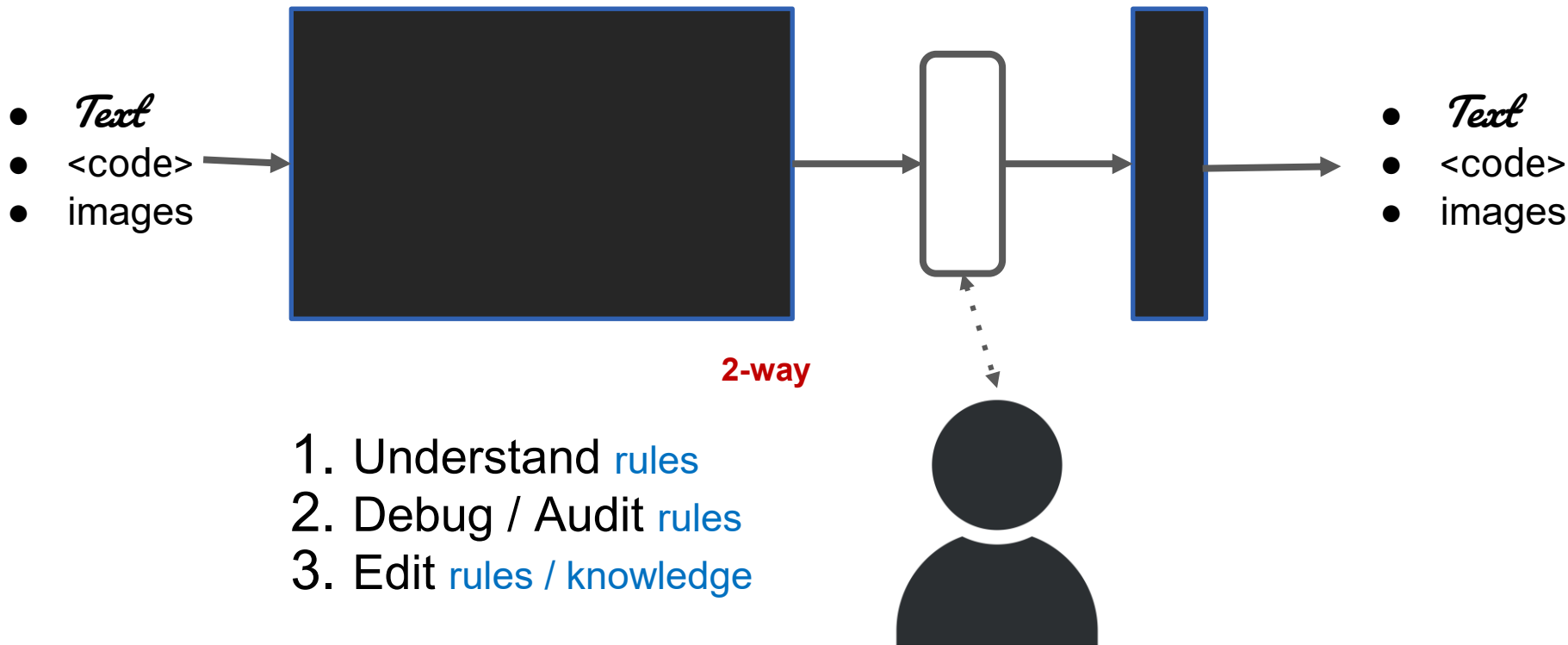




(b) MHSA attention maps
CLIP4IDC: the window is missing ❌

2. PEEB

PEEB: **P**art-based Image Classifiers with an **E**xplainable and **E**ditable Language **B**ottleneck. NAACL 2024



Input image



green back
long, pointed beak
yellow or red belly
...
vibrant red throat

(a) textual concept explanations
operate at the image level

Input image



green back
long, pointed beak
yellow or red belly
...
vibrant red throat

(a) textual concept explanations
operate at the image level

Input image

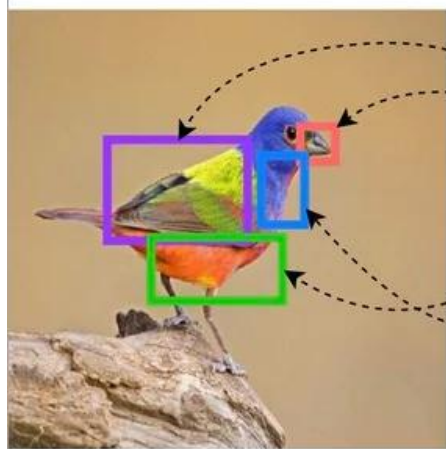


Part prototypes



Editable Bottleneck

(b) part-based prototypes represent
image patches and not editable by humans



Text descriptors

back: vibrant green coloring
beak: conical, silver-gray
belly: rich red hue
...
throat: bright red plumage



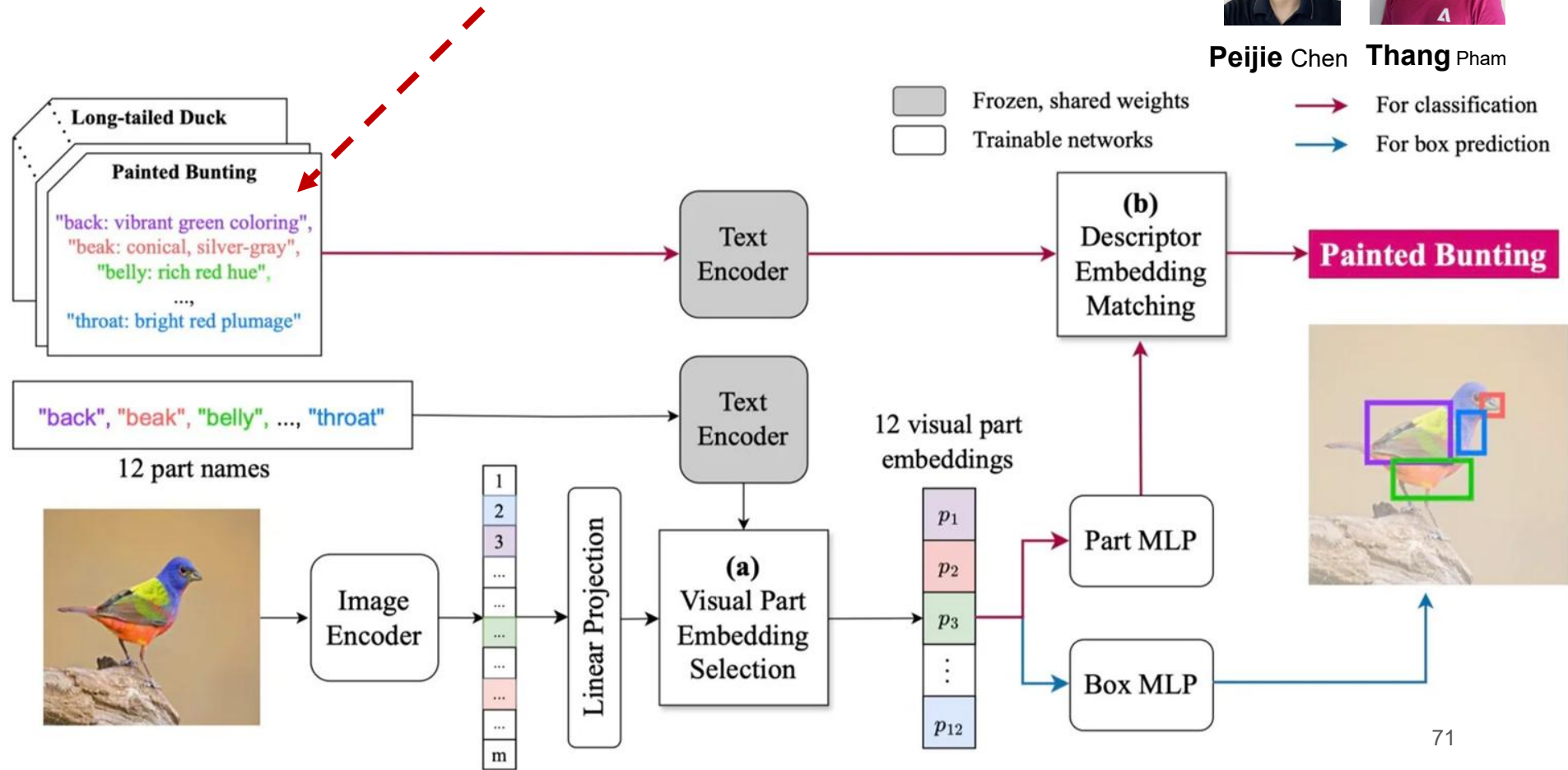
Painted bunting
0.72

(c) PEEB explanations pair up each detected object part with a textual descriptor

Editable Bottleneck



Peijie Chen Thang Pham



American Goldfinch



Predict

American Goldfinch 0.0405

- crown: black on males, dull green in females 0.10
- forehead: vibrant yellow in males, pale green in females 0.45
- nape: olive-green with a hint of yellow 0.66
- eyes: small, round, and black 0.87
- beak: short, conical, and pale pink-orange 0.62
- throat: bright yellow in males, light green-yellow in females 0.61
- breast: bright yellow in males, duller yellow in females 0.35
- belly: pale white to light yellow 0.57
- back: olive-green with black streaks 0.58
- wings: black with vivid white markings 0.30
- legs: pinkish-orange and slender 0.50

Rese

Black Tern X 0.0337

Please use ";" to separate the de
format of [part name]: {descript

class name: custom;
crown: black on males, du
forehead: vibrant yellow in
nape: olive-green with a h
eyes: small, round, and bla
beak: short, conical, and p
throat: bright yellow in ma
breast: bright yellow in m
belly: pale white to light y
back: olive-green with bla
wings: black with vivid wh
legs: pinkish-orange and s
tail: black with white edge

crown: S##*DSJJA
forehead: SJJ@(\$747373

American Goldfinch



Predict

American Goldfinch 0.0405

crown: black on males, dull green in females 0.10
 forehead: vibrant yellow in males, pale green in females 0.45
 nape: olive-green with a hint of yellow 0.66
 eyes: small, round, and black 0.87
 beak: short, conical, and pale pink-orange 0.62
 throat: bright yellow in males, light green-yellow in females 0.61
 breast: bright yellow in males, duller yellow in females 0.35
 belly: pale white to light yellow 0.57
 back: olive-green with black streaks 0.58
 wings: black with vivid white markings 0.30
 legs: pinkish-orange and slender 0.50
 tail: black with white edges

Reset

Black Tern X 0.0337

Please use ";" to separate the descriptions.
 format of {part name}: {description}

class name: custom;
 crown: black on males, dull green in females;
 forehead: vibrant yellow in males, pale green in females;
 nape: olive-green with a hint of yellow;
 eyes: small, round, and black;
 beak: short, conical, and pale pink-orange;
 throat: bright yellow in males, light green-yellow in females;
 breast: bright yellow in males, duller yellow in females;
 belly: pale white to light yellow;
 back: olive-green with black streaks;
 wings: black with vivid white markings;
 legs: pinkish-orange and slender;
 tail: black with white edges

crown: S##*DSJJA

forehead: SJJ@{(\$/747373



Predict

American Goldfinch 0.0405

crown: black on males, dull green in females 0.10
 forehead: vibrant yellow in males, pale green in females 0.45
 nape: olive-green with a hint of yellow 0.66
 eyes: small, round, and black 0.87
 beak: short, conical, and pale pink-orange 0.62
 throat: bright yellow in males, light green-yellow in females 0.61
 breast: bright yellow in males, duller yellow in females 0.35
 belly: pale white to light yellow 0.57
 back: olive-green with black streaks 0.58
 wings: black with vivid white markings 0.30
 legs: pinkish-orange and slender 0.50
 tail: black with white edges

Reset Descriptions

Black Tern X 0.0337

Please use ";" to separate the descriptions for each part, and keep the format of {part name}: {descriptions}

class name: cŭstom;
 crown: black on males, dull green in females;
 forehead: vibrant yellow in males, pale green in females;
 nape: olive-green with a hint of yellow;
 eyes: small, round, and black;
 beak: short, conical, and pale pink-orange;
 throat: bright yellow in males, light green-yellow in females;
 breast: bright yellow in males, duller yellow in females;
 belly: pale white to light yellow;
 back: olive-green with black streaks;
 wings: black with vivid white markings;
 legs: pinkish-orange and slender;
 tail: black with white edges

crown: S##*DSJJA

forehead: SJJ@(\$747373

tail: \$*(@)@336

0.07

0.22

New Thread



Home

Discover

Library

find gifts for boys on...
why do you think the...
look at all the gifts und...
find gifts for boys on...
what is the spring brea...
find me a boxing gloves...

Try Pro

Upgrade for image upload, smarter AI, and more Pro Search.

Answer

The American Goldfinch (*Spinus tristis*) can be identified by examining specific parts of its anatomy, each offering distinctive features that aid in recognition:

- Crown:** The top of the head; in males during the breeding season, it is black, while females and non-breeding males have a more subdued color.
- Forehead:** In breeding males, the forehead is also black, contrasting with the bright yellow of the crown.
- Nape:** The back of the neck; coloration here is generally consistent with the back, which can be olive or yellow depending on the season and sex.
- Eyes:** Small and dark, surrounded by feather coloration that contrasts with the face, enhancing their visibility.



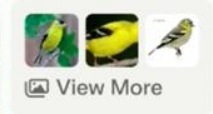
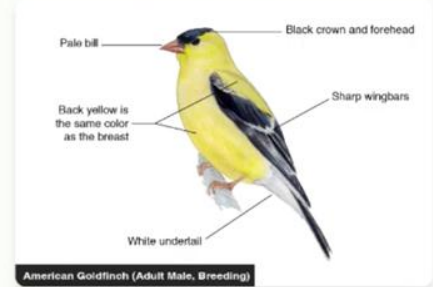
Ask follow-up



Pro



- Throat:** Matches the general body color; bright yellow in breeding males and more olive or tan in females and non-



View More



Editable Bottleneck



Predict

American Goldfinch 0.0405

crown: black on males, dull green in females 0.10
forehead: vibrant yellow in males, pale green in females 0.45
nape: olive-green with a hint of yellow 0.66
eyes: small, round, and black 0.87
beak: short, conical, and pale pink-orange 0.62
throat: bright yellow in males, light green-yellow in females 0.61
breast: bright yellow in males, duller yellow in females 0.35
belly: pale white to light yellow 0.57
back: olive-green with black streaks 0.58
wings: black with vivid white markings 0.30
legs: pinkish-orange and slender 0.50
tail: black with white edges

Reset Descriptions

Black Tern X 0.0337

Please use ";" to separate the descriptions for each part, and keep the format of {part name}: {descriptions}

class name: American Goldfinch v2;
crown: The top of the head; in males during the breeding season, it is black, while females and non-breeding males have a more subdued color;
forehead: vibrant yellow in males, pale green in females;
nape: olive-green with a hint of yellow;
eyes: small, round, and black;
beak: short, conical, and pale pink-orange;
throat: bright yellow in males, light green-yellow in females;
breast: bright yellow in males, duller yellow in females;
belly: pale white to light yellow;
back: olive-green with black streaks;
wings: black with vivid white markings;
legs: pinkish-orange and slender;
tail: black with white edges

crown: S##*DSJJA

0.07



Predict 

American Goldfinch 0.0405



processing | 0.5/3.0s

crown: black on males, dull green in females

forehead: vibrant yellow in males, pale green in females

nape: olive-green with a hint of yellow

eyes: small, round, and black

beak: short, conical, and pale pink-orange

throat: bright yellow in males, light green-yellow in females

breast: bright yellow in males, duller yellow in females

belly: pale white to light yellow

back: olive-green with black streaks

underparts: white with black streaks

0.30

processing | 0.5/3.0s

0.45

0.66

0.87

0.62

0.61

0.36

0.57

0.58

Reset Descriptions

Black Tern X 0.0337



processing | 0.5/3.0s

processing | 0.5/3.0s



Edited



Predict



American Goldfinch v2 0.0527

crown: The top of the head; in males during the breeding season, it is black, while females and non-breeding males have a more subdued color

0.42

forehead: vibrant yellow in males, pale green in females

0.45

nape: olive-green with a hint of yellow

0.66

eyes: small, round, and black

0.87

beak: short, conical, and pale pink-orange

0.62

throat: bright yellow in males, light green-yellow in females

0.61

breast: bright yellow in males, duller yellow in females

0.35

belly: pale white to light yellow

0.57

back: olive-green with black streaks

0.58

wings: black with vivid white markings

0.30

legs: pinkish-orange and slender

Reset Descriptions

American Goldfinch v2 0.0527

crown: The top of the head; in males during the breeding season, it is black, while females and non-breeding males have a more subdued color

0.42

forehead: vibrant yellow in males, pale green in females

0.45

nape: olive-green with a hint of yellow

0.66

eyes: small, round, and black

0.87

beak: short, conical, and pale pink-orange

0.62

throat: bright yellow in males, light green-yellow in females

0.61

breast: bright yellow in males, duller yellow in females

0.35

belly: pale white to light yellow

0.57

back: olive-green with black streaks

0.58

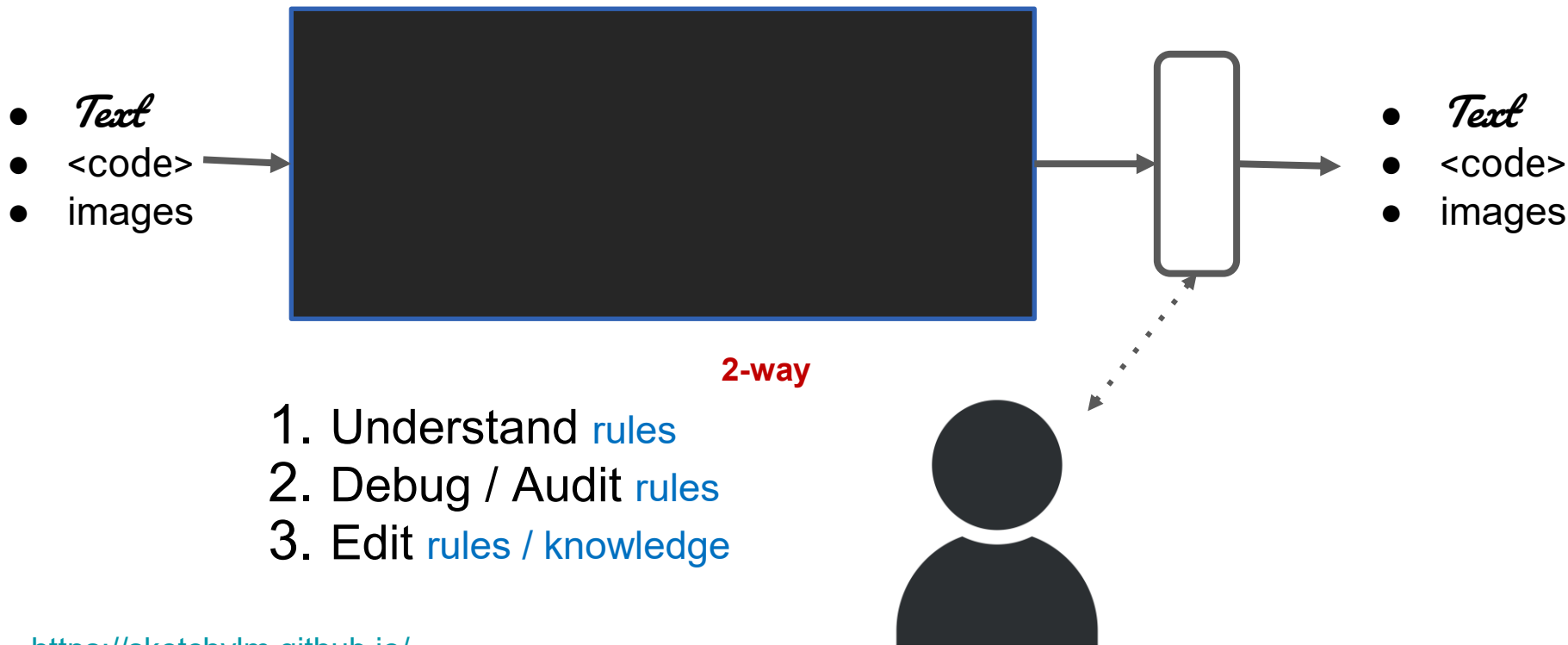
wings: black with vivid white markings

0.30

legs: pinkish-orange and slender

3. SketchVLM

SketchVLM: Vision-Language Models Can Annotate Images to Explain Thoughts and Guide Users. arXiv 2026





(a)

how do I check my car's oil level?



(a)

how do I check my car's oil level?



ChatGPT 5.2

Thought for 52s

In your photo:

Oil fill cap: the black cap that says “dexos” on the engine cover (upper left).

Oil dipstick: the yellow handle near the front-center, tucked just under the front edge of the plastic cover.

How to check it: Park on level ground and shut the engine off.

Best: check first thing in the morning (cold).

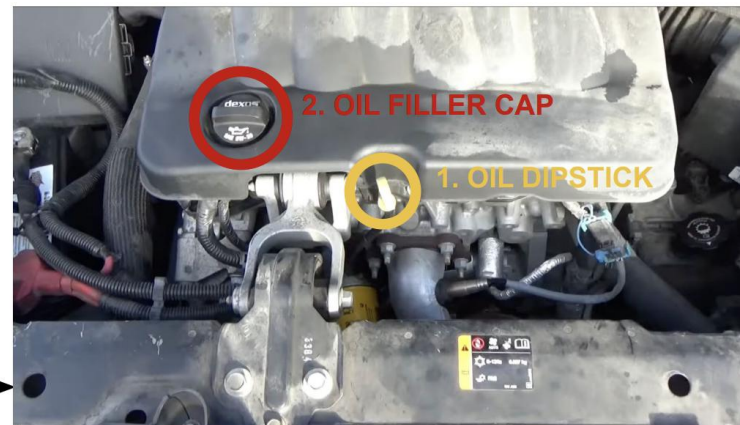
Pull the dipstick (yellow handle), wipe it clean with a paper towel or rag

Reinsert fully until it seats, then pull it again...



(a)

how do I check my car's oil level?



To check your car's oil, follow these steps:

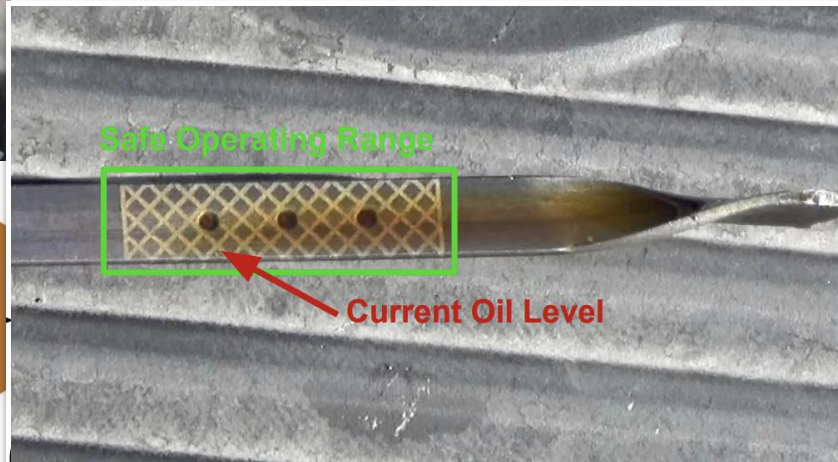
1. Locate the oil dipstick, which usually has a brightly colored handle (often yellow or orange). In this image, it is the yellow handle near the center...

2. Locate the oil filler cap



Is this enough
oil?

2



Yes, the oil level appears to be sufficient. On this dipstick, the crosshatched area represents the safe operating range ...

the yellow handle near the center...

2. Locate the oil filler cap

Editable Bottleneck



How do I check my car's oil?

1



Sketch
VLM

System Prompt

Here's a list of annotations
you can use:



```
"annotations": [  
  {  
    "type": "text",  
    "content": "1. OIL  
DIPSTICK",  
    "x": 515,  
    "y": 444,  
    "color": "#f1c232",  
    "fontSize": 25  
  },  
  {  
    "type": "circle",  
    "cx": 496,  
    "cy": 444,  
    "r": 25,  
    ...  
  }  
]
```

```
"answer": "To check your  
car's oil..."
```



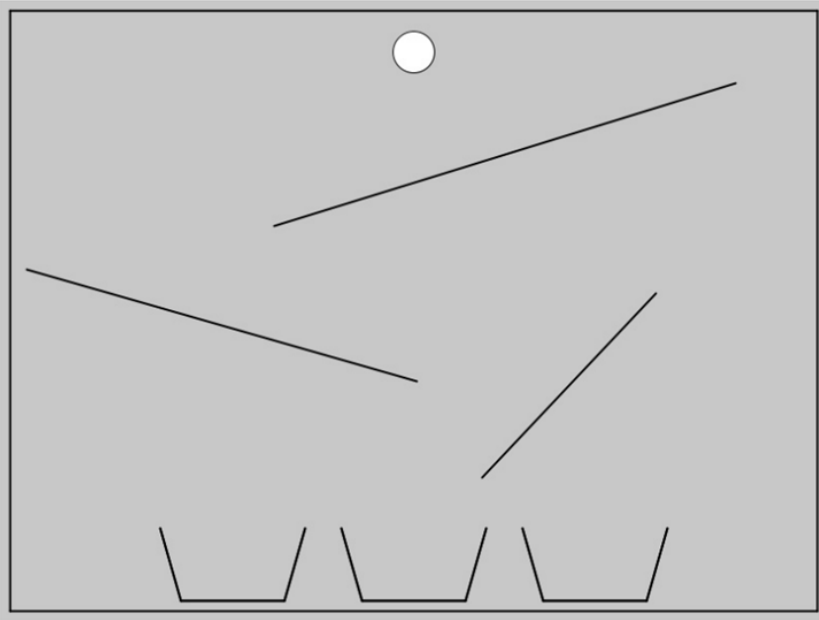
To check your car's oil, follow these steps:

1. Locate the oil dipstick, which usually has a brightly colored handle (often yellow or orange). In this image, it is the yellow handle near the center...

2. Locate the oil filler cap

Input

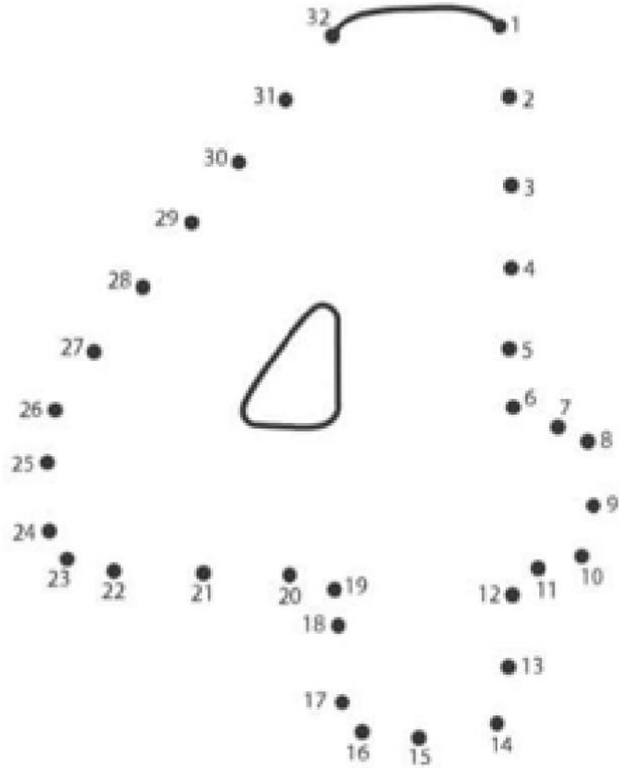
Ball Drop



*Which bucket will the
ball fall into?*

Input

Connect-the-Dots



*Connect the dots in
order.*



Brandon
Collins



Logan
Bolton

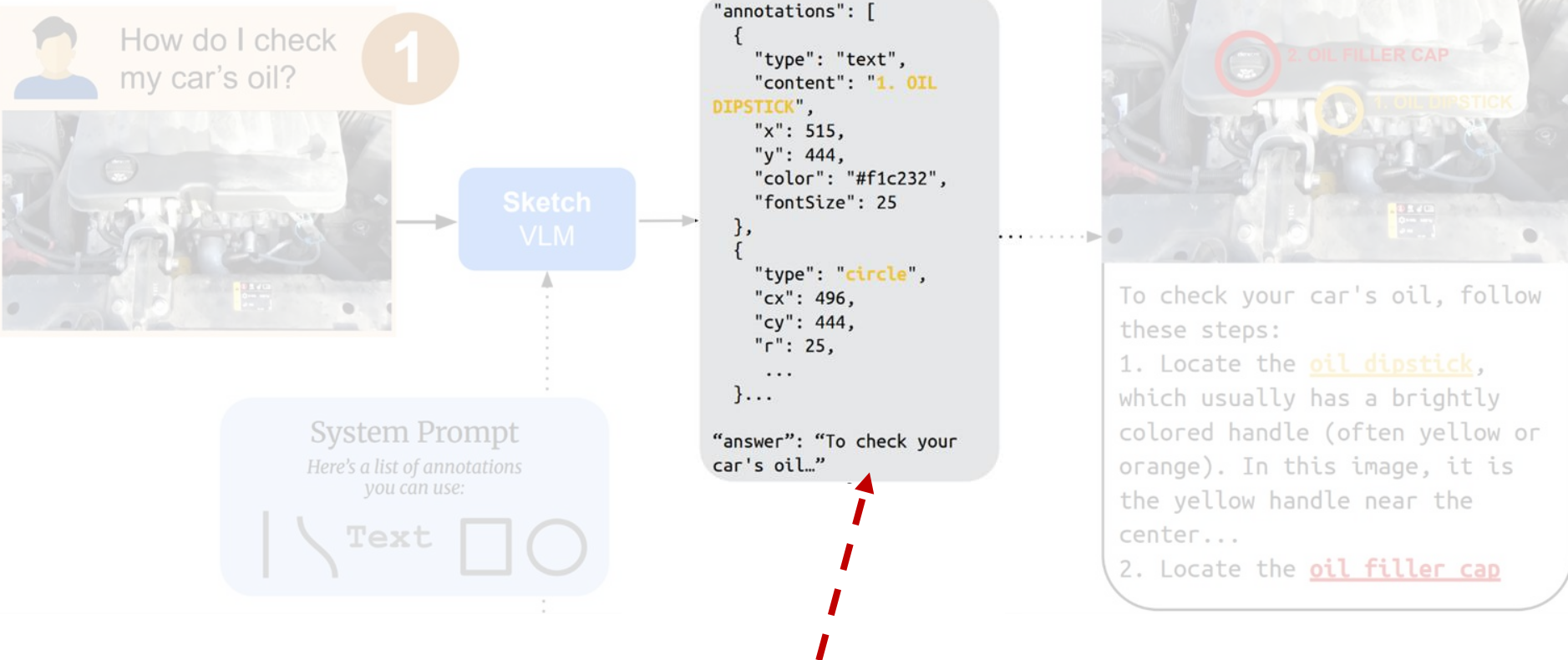


Hung Nguyen

How many people are there?



Edit and keep talking!

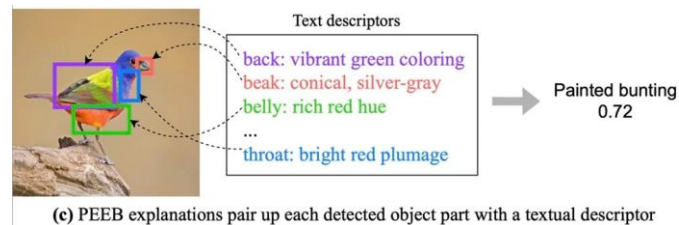
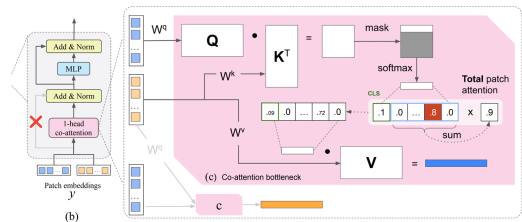


Annotating can also improve accuracy!

Model	VPCT video physics	 ball drop	 maze trace	 counting	 labeling	 draw shapes		 connect dots		
								RMSE	Order	Acc%
✦ sketchVLM	96.0 ± 1.4	79.7 ± 2.8	98.0 ± 1.7	94.5	60.3	58.8	55.4	5.92	99.0	
✦	89.3 ± 2.2	83.8 ± 3.4	99.3 ± 0.8	93.0	64.1	63.1	59.8	†	†	
 sketchVLM	70.0 ± 2.9	68.5 ± 2.2	92.8 ± 2.5	75.4	20.4	18.7	11.2	46.69	74.0	
	63.5 ± 2.5	66.0 ± 4.7	92.3 ± 2.3	72.0	19.1	22.4	15.4	†	†	
 + ✦	63.0	62.6	93.3 ± 3.0	91.7	†	†	†	†	39.0	
	37.0	35.9	50.8 ± 1.5	48.6	—	—	†	198.74	9.0	
	27.0	30.3	62.5 ± 2.1	68.1	†	†	†	†	0.0	

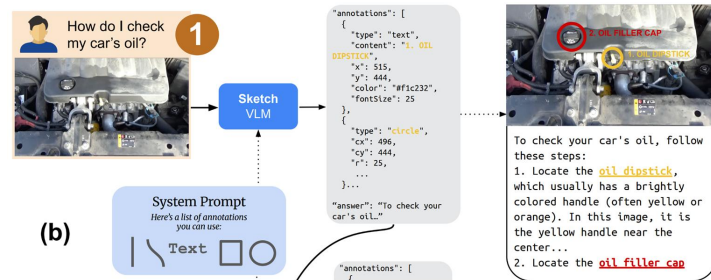
† Model is unable to output this format.

Conclusion



1. VLMs have **poor vision** and are heavily **biased**
2. Insert **editable** bottleneck at:

- Attention layer: TAB (ICCV 2025)
- Penultimate layer: PEEB (NAACL 2024)
- Output layer: SketchVLM (arXiv 2026)
 - Highlighted Chain of Thoughts (TMLR 2026)
 - iCoT (UIST 2026)



✓ Causal analysis: How does editing X change output Y ?

✓ Help user **intervene** in AI decisions?