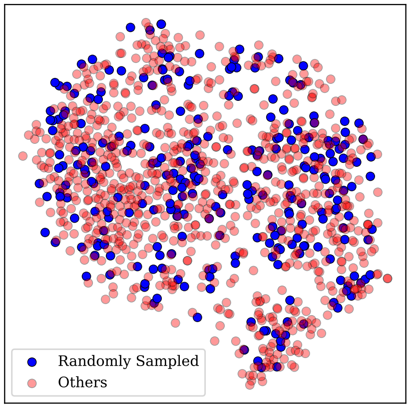
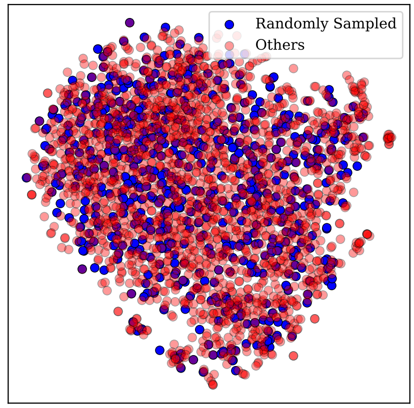
VLM4Bio: A Benchmark Dataset to Evaluate Pretrained Vision-Language Models for Trait Discovery from Biological Images Supplementary Materials

Contents

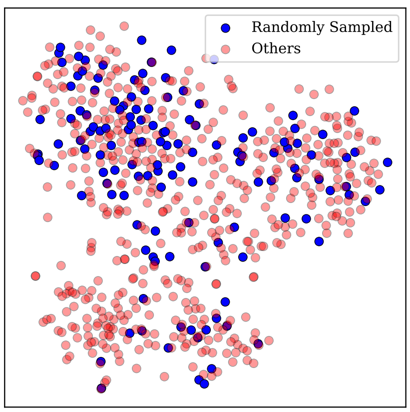
38th Conference on Neural Information Processing Systems (NeurIPS 2024) Track on Datasets and Benchmarks.



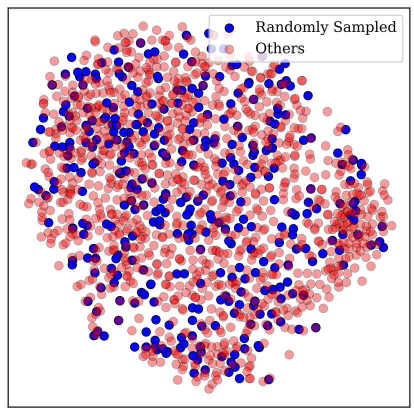
(a) *Lepomis humilis*



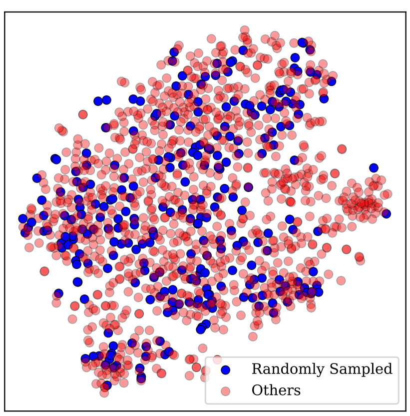
(b) *Lepomis cyanellus*



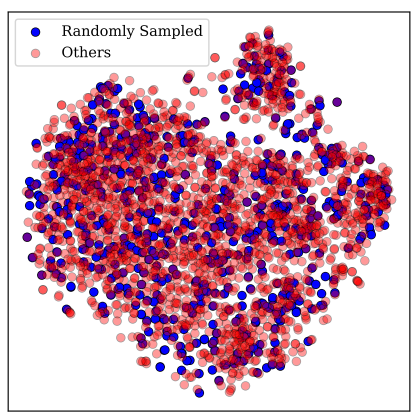
(c) *Noturus gyrinus*



(d) *Lepomis megalotis*



(e) *Phenacobius mirabilis*



(f) *Lepomis macrochirus*

Figure 5: t-SNE plots to illustrate the effectiveness of random sampling with the majority species in the Fish-10K dataset. Randomly sampled images are shown as blue dots, while the remaining data points are represented by red dots. Subcaptions display the scientific names of the corresponding species. To generate the vector representation of the images, we leverage a VGG19 pretrained on the ImageNet dataset.

A Dataset Preprocessing

We collected images of three taxonomic groups of organisms: fish, birds, and butterflies, each containing around $10K$ images. Images for fish (**Fish-10K**) were curated from the larger image collection, FishAIR [1], which contains images from the Great Lakes Invasive Network Project (GLIN) [2]. These images originate from various museum collections such as INHS [3], FMNH [4], OSUM [5], JFBM [6], UMMZ [7] and UWZM [8]. We created the Fish-10K dataset by randomly sampling $10K$ images and preprocessing the images to crop and remove the background.

To ensure diversity within the Fish-10K dataset, we applied a targeted sampling strategy in the source collection, FishAIR [1]. Specifically, we retained all images of species with fewer than 200 images, considering these as minority or rare classes. Random sampling was applied only to the majority species—those with more than 200 images per class. To assess the potential sampling bias among the majority species, we generated feature vectors for each image in Fish-10K using a pretrained VGG-19 model. In Figure 5, we present species-wise t-SNE plots of these feature vectors for several majority species. Our analysis shows that the distribution of sampled images closely mirrors the distribution of images that were not included in the dataset (denoted as "others" in the plot). This suggests that our random sampling approach provides a sufficiently accurate representation of the original distribution for the majority species. For consistency, we leverage GroundingDINO [9] to crop the fish body from the background and Segment Anything Model (SAM) [10] to remove the background. The Fish-10K dataset contains images of specimens preserved in museum collections with artificial backgrounds with imaging artifacts that are not typical for large-scale computer vision datasets. Moreover, these backgrounds can introduce unexpected bias. Hence, we removed the backgrounds using SAM to create a controlled environment for our experiments.

We curated the images for butterflies (**Butterfly-10K**) from the Jiggins Heliconius Collection dataset [11], which has images collected from various sources [1]. We carefully sampled $10K$ images for Butterfly-10K from the entire collection to ensure the images capture unique specimens and represent a diverse set of species by adopting the following two steps. **First**, the butterfly images show various angles, including dorsal and ventral views, forewing dorsal and ventral views, and hindwing dorsal and ventral views. To ensure consistency, we only selected images with dorsal view and removed all images of hybrid species. **Second**, we further filtered the dataset based on the unique specimen ID to ensure no specimen was repeated more than once. For species with more than 2000 images, we performed random sampling (no sampling was performed for species with sizes less than 2000). We ensure each species has a minimum of 20 images and no more than 2,000 images. The Butterfly-10K dataset contains a significant number of images of *Heliconius melpomene* and *Heliconius erato* species. We utilized the subspecies information of these two species to create a hard dataset for analyzing the impact of answer choices on VLM performance, as described in Section 5.1.

The images for birds (**Bird-10K**) are obtained from the CUB-200-2011 [37] dataset by taking 190 species for which the common name to scientific name mapping is available. This results in a fairly balanced dataset with around $11K$ images in total.

The scientific names for the images of Fish-10K and Butterfly-10K were obtained directly from their respective sources. For Bird-10K, we obtained the scientific names from the iNatLoc500 [38] dataset. We curated around $31K$ question-answer pairs in both open and multiple-choice (MC) question formats for evaluating species classification tasks. The species-level trait presence/absence matrix for Fish-10K was manually curated with the help of biological experts co-authored in this paper. We leveraged the Phenoscape knowledge [39] base with manual annotations to procure the presence-absence trait matrix. For Bird-10K, we obtained the trait matrix from the attribute annotations provided along with CUB-200-2011. We constructed approximately $380K$ question-answer pairs for trait identification tasks.

For grounding and referring VQA tasks, the ground truths were manually annotated with the help of expert biologists on our team. We manually annotated bounding boxes corresponding to the traits of 500 fish specimens and 500 bird specimens, which are subsets of the larger Fish-10K and Bird-10K datasets, respectively. We used the CVAT tool [40] for annotation. The task-specific question formats with the default prompts are provided in Section I.

[1]Sources: [12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36]

Statistics	Datasets														
	Fish-10K	*Bird-10K*	*Butterfly-10K*	*Fish-500*	*Bird-500*	*Fish-Easy*	*Fish-Medium*	*Bird-Easy*	*Bird-Medium*	*Butterfly-Easy*	*Butterfly-Medium*	*Butterfly-Hard*	*Fish-Prompting*	*Bird-Prompting*	*Butterfly-Prompting*
Images	10,347	11,092	10,013	500	492	200	200	200	200	200	200	200	500	500	500
Species	495	188	60	60	47	51	10	50	10	50	10	1	25	37	25
Genera	178	114	27	18	33	10	1	10	1	10	1	1	12	30	10
Traits	10	28	-	8	5	-	-	-	-	-	-	-	-	-	-

Table 7: Statistics of the VLM4Bio dataset.

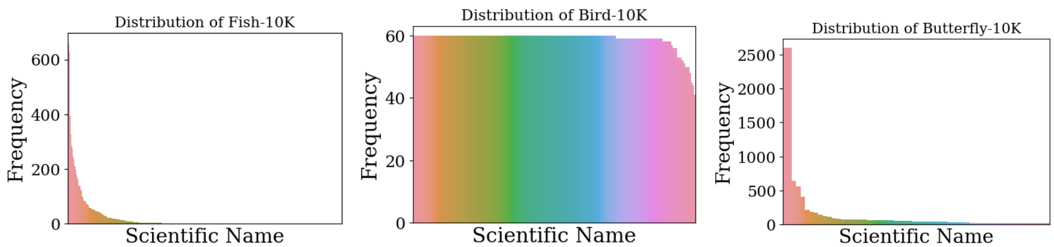


Figure 6: Dataset Distribution of Fish-$10K$, Bird-$10K$, and Butterfly-$10K$.

B Links to Access the Dataset and Its Metadata

We provide a GitHub link `https://github.com/imageomics/VLM4Bio` and an accessible Hugging Face link `https://huggingface.co/datasets/imageomics/VLM4Bio` to access the dataset and its metadata.

C Dataset Availability and Maintanance

The VLM4Bio dataset and metadata are available in a Hugging Face repository. To access the VLM4Bio dataset, please visit `https://huggingface.co/datasets/imageomics/VLM4Bio`. Long-term support and maintenance of the dataset will be provided by our team. We have published a code repository for dataset preprocessing, including tasks such as downloading the dataset, reading images and metadata, cropping images, and running the evaluation experiments presented in the VLM4Bio paper. To access the VLM4Bio code repository, please visit `https://github.com/imageomics/VLM4Bio`.

D Data Licenses

VLM4Bio dataset is licensed as Creative Commons Attribution 4.0 International. The images of the corresponding organisms are licensed as follows:

1. Fish Dataset License: CC BY-NC.
2. All the bird images are sourced from the CUB-200-2011 dataset; CalTech indicates that they do not own the copyrights to these images and that their use is restricted to non-commercial research and educational purposes.
3. Butterfly Dataset License: Creative Commons Attribution 4.0 International.

We provide image-specific licenses in the dataset card `https://huggingface.co/datasets/imageomics/VLM4Bio#licensing-information`. We have hosted the dataset on HuggingFace (DOI: 10.57967/hf/3393).

E Data Distribution and Key Statistics

Table 7 provides the key statistics for the datasets, including the number of images, species, genera, and traits present in each one. We are examining the Zero-shot accuracy of the VLMs on Fish-10K, Bird-10K, and Butterfly-10K for Species Classification and Trait Identification tasks, Fish-500 and Bird-500 for Trait Grounding, Trait Referring and Trait Counting, and easy, medium, hard, prompting datasets for analyzing the role of answer choices, VLM reasoning and hallucination tests. From Figure 6, it is clear that Fish-10K and Butterfly-10K are imbalanced, with a bias toward some species that are more common in our environment (such as *Heliconius erato* and *Heliconius melpomene* for Butterflies). The imbalance in Fish-10K and Butterfly-10K reflects the natural imbalance in the occurrence and observation of species in museum collections. Due to the scarcity of images for the rare species, it is difficult to increase their representation to avoid imbalance. As a result, we have included many under-represented species in the Fish and Butterfly datasets to report performance on the rare classes. In contrast, the Bird-10K dataset is well-balanced, with most species having 60 images. The easy, medium, hard, and prompting datasets are also balanced, which ensures a comprehensive evaluation of the zero-shot performance of the competing VLMs.

F Traits Considered for the Task of Trait Identification

Fish Traits		Bird Traits			
		Color		Pattern	Measurements
. Eye . Head . Mouth . Barbel . Dorsal fin	. Pectoral fin . Pelvic fin . Anal fin . Two dorsal fins . Adipose fin	. Bill-color . Crown-color . Eye-color . Forehead-color . Nape-color . Primary-color . Throat-color . Back-color	. Belly-color . Breast-color . Leg-color . Under-tail-color . Underparts-color . Upper-tail-color . Upperparts-color . Wing-color	. Head-pattern . Back-pattern . Breast-pattern . Wing-pattern . Tail-pattern . Belly-pattern	. Bill-length . Bill-shape . Shape . Size . Tail-shape . Wing-shape

Figure 7: Trait list for Trait Identification task.

Figure 7 shows the Fish traits and Bird traits used for evaluating the VLM's performance in the identification task. For fishes, we considered 10 binary (presence/absence) traits which include the *eye, head, mouth, barrel, dorsal fin, pectoral fin, pelvic fin, anal fin, and adipose fin*. We generated MC questions for the presence of each trait in an image (with two options: yes or no). Whereas for birds, we considered 28 traits covering their color, pattern, and measurements (size and shape of regions) in a multiple-choice format.

G Traits Considered for the Tasks of Trait Grounding and Referring

To evaluate the VLM performance in Grounding and Referring, we identified 8 traits for fish and 5 traits for birds. Specifically, we manually annotated the *dorsal fin, adipose fin, caudal fin, anal fin, pelvic fin, pectoral fin, head, and eye* of the 500 fish specimens. Similarly, for birds, we annotated the *beak, head, eye, wings, and tail*. Trait grounding and referring tasks are carried out using the Fish-500 and Bird-500 datasets.

H VLM Baselines

We consider the following VLM baselines to evaluate the performance on VLM4Bio dataset: (1) GPT-4V(ision) [41], which is a proprietary VLM from OpenAI, that uses a generative pre-trained transformer model capable of understanding and generating both text and visual contents, (2) LLaVA-v1.5 (7B/13B) [42], which builds on top of the Vicuna LLM [43] by linearly projecting the visual embedding into the word embedding space. The LLaVA model has two different variants with 7B and 13B parameters, respectively, that depend on the size of the base Vicuna model, (3) COG-VLM [44], which performs a simple concatenation of the image and the text modalities, and uses trainable visual layers in the text-based transformer blocks, (4) MiniGPT-4 (Vicuna 7B/13B) [45], which is similar to LLaVA as it is built on top of the Vicuna model and linearly projects the visual embeddings for better understanding. Similar to LLaVA, MiniGPT-4 is available in two variants depending on the type the base Vicuna model (Vicuna 7B/13B), (5) BLIP-FLAN-T5-XL/XXL [46], which utilizes an effective pre-training strategy that relies on bootstrapping from frozen-pretrained CLIP encoders

Task	Prompt Format
Species Classification	*<image>* What is the scientific name of the *<organism>* shown in the image? *<options>* Write the answer after writing the answer is: .
Trait Identification	*<image>* Is there *<trait>* visible in the *<organism>* shown in the image? *<options>* Write the answer after writing the answer is: .
Trait Grounding	*<image>* What is the bounding box coordinates of the *<trait>* in the fish shown in the image? *<options>* Write the answer after writing the answer is: .
Trait Referring	*<image>* What is the trait of the *<organism>* that corresponds to the bounding box region *<coordinates>* in the image? *<options>* Write the answer after writing the answer is: .
Trait Counting	*<image>* How many unique *<trait>* are visible in the *<organism>* shown in the image? *<options>* Write the answer after writing the answer is: .
Contextual Prompting	*<image>* Each biological species has a unique scientific name composed of two parts: the first for the genus and the second for the species within that genus. What is the scientific name of the *<organism>* shown in the image? *<options>* Write the answer after writing the answer is: .
Dense Caption Prompting	*<image> <dense caption>.* Use the above dense caption and the image to answer the following question. What is the scientific name of the *<organism>* shown in the image? *<options>* Write the answer after writing the answer is: .
Chain-of-Thought Prompting	*<image>* What is the scientific name of the *<organism>* shown in the image? *<options>* Please consider the following reasoning to formulate your answer. ***<reasoning>***. Write the answer after writing the answer is: .
False Confidence Test (FCT)	*<image>* What is the scientific name of the *<organism>* shown in the image? *<options>* Chosen Answer: *<suggested answer>*. Please provide: 1) Whether the chosen answer is correct (True/False). 2) The correct answer.
None of the Above Test (NOTA)	*<image>* What is the scientific name of the *<organism>* shown in the image? *<options: A) _ B) _ C) _ D) None of the above.>* Write the answer after writing the answer is: .

Figure 8: Prompts Templates used for Evaluation. There will be no `<options>` for Open set questions.

and LLMS by using a querying transformer block (available as two variants: XL and XXL), and (6) Instruct-BLIP (Vicuna 7B/13B) [47], which performs finetuning on BLIP-2 with visual-instruction tuning data to improve zero-shot capabilities of BLIP-2 (available as two variants depending on the Vicuna model: Vicuna 7B/13B).

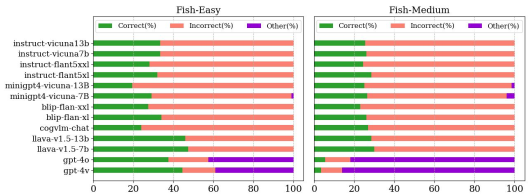
I Prompts to Evaluate VLM performance

In order to ensure a fair comparison of the VLM responses to different types of questions in our dataset, we used the same question prompt for all the models across the various scientific tasks. It's worth noting that each model may perform differently with different prompts. However, for the sake of simplicity in our evaluation, we opted for a consistent prompt for all the models. The prompts specific to each task are displayed in Figure 8.

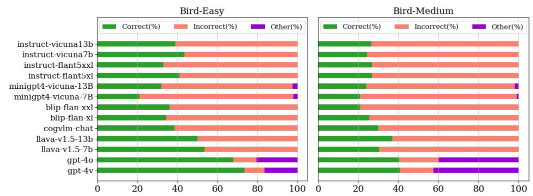
J Error Analyses for VLM Responses

We categorize the VLM responses into 3 categories: (1) *Correct (%)*: where the scientific name is accurately predicted, (2) *Incorrect (%)*: where the scientific name is incorrect, and (3) *Other (%)*: a special category for instances where the model abstains from providing a scientific name.

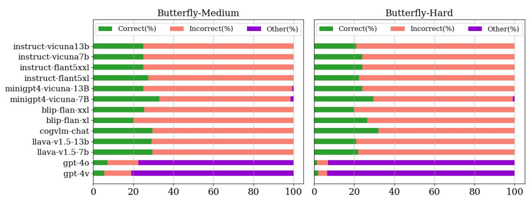
Figure 9a, 9b and 9c show the distribution of errors of different VLMs on Fish-Easy and Fish-Medium, Bird-Easy and Bird-Medium, and Butterfly-Medium and Butterfly-Hard datasets respectively using stacked-bar plots showing the three categories of VLM predictions. GPT-4V, for instance, shows a



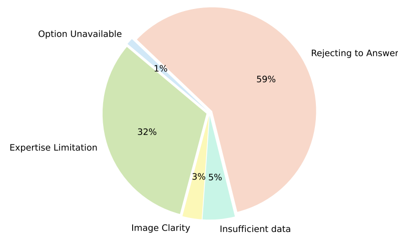
(a) Error Analysis for Fish-Easy and -Medium.



(b) Error Analysis for Bird-Easy and -Medium.



(c) Error Analysis for Butterfly-Medium and -Hard.



(d) Categories for 250 annotated GPT-4V "Other" responses.

Figure 9: Analysis of errors for the pretrained VLM responses.

reduced rate of incorrect responses but a higher incidence of "Other" responses for these datasets, which include apologetic expressions, admissions of inability to precisely visualize the organism, and disclaimers regarding prediction without sufficient expert data and guidance.

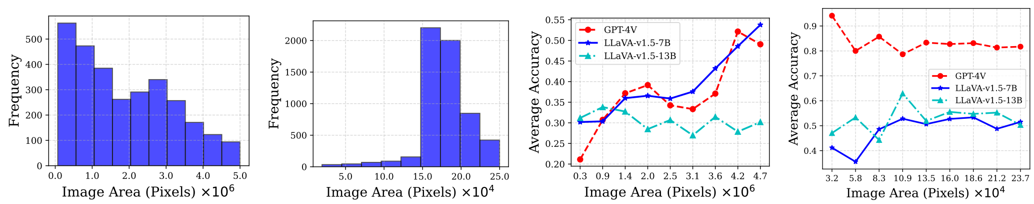
To further analyze the type of errors happening in the other (%) category of VLM predictions, we manually examined 250 randomly selected "Other" GPT-4V responses for the task of fish species classification (MC question type) to generate the pie-chart of error categories shown in Figure 9d. We can see that a majority of the "Other" responses belong to the category: *Rejecting to Answer* (59%), where the GPT-4V states that it is unable to provide an answer, sometimes stating the reason that it cannot answer based on a single image. We also observe a large fraction of *Expertise Limitation* responses where GPT-4V states that an expert taxonomist is needed to answer the question and its capabilities do not include recognizing or confirming species based on visual data. The next major type of "Other" responses are *Insufficient Data*, where GPT-4V states that it requires additional data to answer the question, e.g., taxonomic information or habitat information. The other error categories include *Image Clarity* issues and *Option Unavailable* (i.e., GPT-4V could not find a suitable option from the list of options provided in the prompt).

K Comparing Pre-trained VLMs with a Biologically Fine-tuned Model

Dataset	**Question type**	**Models**				
		gpt-4v	*llava v1.5-7b*	*cogvlm chat*	*CLIP*	*BioCLIP*
Species Classification						
Fish-10K	Open	1.01	2.32	0.11	0.57	1.24
	MC	35.91	40.20	31.72	42.45	50.65
Bird-10K	Open	17.40	1.45	0.86	7.74	67.12
	MC	82.58	50.32	44.73	45.78	93.93
Butterfly-10K	Open	0.04	0.05	0.01	5.33	15.95
	MC	28.91	50.24	36.45	45.60	62.32

Table 8: Zero-shot accuracy comparison of VLM baselines (in % ranging from 0 to 100) with BioCLIP for the species classification task. Results are color-coded as Best, and Worst.

We compare the large pretrained VLMs and BioCLIP [48], a state-of-the-art foundation model for species classification. Furthermore, we include the simple CLIP model pretrained with OpenAI weights [49] to evaluate the zero-shot classification performance. Our evaluation was carried out on the Fish-10K, Bird-10K, and Butterfly-10K datasets, and the results are presented in Table 8. We can see that BioCLIP significantly outperforms large pretrained VLMs on the Bird-10K and Butterfly-10K datasets, suggesting that BioCLIP may have been trained on images that are similar to the organisms present in these datasets. However, as noted in the paper, BioCLIP is not trained on fish images, and hence, the performance of large VLMs is similar to that of BioCLIP on Fish-10K images. We can also see that despite BioCLIP's ability to effectively select the correct scientific name from a smaller set of options in multiple-choice (MC) questions, its performance significantly declines when asked to choose the scientific name from a larger set of open questions. From our observation, it is noteworthy that fine-tuning biological images with scientific names can help improve the overall accuracy of species classification, suggesting directions for future research in this area.



(a) Fish-10K (b) Bird-10K (c) Fish-10K (d) Bird-10K

Figure 10: Distribution of image resolutions for Fish-10K and Bird-10K are shown in Figures (a) and (b), respectively. The average score over image resolution for the GPT-4V, LLaVA-v1.5-7B, and LLaVA-v1.5-13B models on Fish-10K and Bird-10K are presented in Figures (c) and (d). We conduct the experiment in the context of the Species Classification task with Multiple-Choice (MC) questions.

L Analyzing Effects of Image Resolution on VLM Performance

To investigate the effect of image resolution on VLM performance, we perform additional experiments summarized in Figure 10 of the attached pdf. In this Figure, we show distribution plots for the Fish-10K and Bird-10K datasets with variations in the image resolutions and their impact on the species classification performance (MC question format) for GPT-4V, LLaVA-1.5-7B, and LLaVA-1.5-13B. All the images of the Butterfly-10K have the exact resolution (500×333); hence, they were not included in the experiment. From Figure 10c, it is clear that image resolution is influential on the VLM performance for the Fish-10K dataset since higher resolution helps in recognizing the details of the biological traits and correct species. However, for Figure 10d, the VLM performances do not vary significantly with the image resolution for the Bird-10K dataset. A potential reason is that the bird dataset is a subset of the CUB dataset, and we hypothesize that the pre-trained VLMs may have seen images with resolutions similar to those in the Bird-10K dataset during training, leading to this behavior.

M Case Studies for Effects of Prompting on VLM Performance

M.1 No Prompting

1. No Prompting. GPT-4o Correct prediction. Refer to Figure 39.
2. No Prompting. GPT-4o Incorrect prediction. Refer to Figure 40.
3. No Prompting. COG-VLM Correct prediction. Refer to Figure 41.
4. No Prompting. COG-VLM Incorrect prediction. Refer to Figure 42.

M.2 Contextual Prompting

1. Contextual Prompting. GPT-4o Correct prediction. Refer to Figure 43.

2. Contextual Prompting. GPT-4o Incorrect prediction. Refer to Figure 44.
3. Contextual Prompting. LLaVa-13B Correct prediction. Refer to Figure 45.
4. Contextual Prompting. LLaVa-13B Incorrect prediction. Refer to Figure 46.

M.3 Dense Caption

1. Dense Captions in Prompts. GPT-4o Correct prediction. Refer to Figure 47.
2. Dense Captions in Prompts. GPT-4o Incorrect prediction. Refer to Figure 48.
3. Dense Captions in Prompts. LLaVa-7B Correct prediction. Refer to Figure 49.
4. Dense Captions in Prompts. LLaVa-7B Incorrect prediction. Refer to Figure 50.

M.4 Chain-Of-Thought Prompting

1. Chain-Of-Thought Prompting. GPT-4o Correct prediction. Refer to Figure 51.
2. Chain-Of-Thought Prompting. GPT-4o Incorrect prediction. Refer to Figure 52.
3. Chain-Of-Thought Prompting. LLaVa-13B Correct prediction. Refer to Figure 53.
4. Chain-Of-Thought Prompting. LLaVa-13B Incorrect prediction. Refer to Figure 54.

N Case Studies for Reasoning Hallucination Tests

N.1 False Confidence Test (FCT)

1. FCT test on Fish dataset. GPT-4o Correct prediction. Refer to Figure 55.
2. FCT test on Fish dataset. LLaVa-13B Incorrect prediction. Refer to Figure 56.
3. FCT test on Bird dataset. GPT-4o Correct prediction. Refer to Figure 57.
4. FCT test on Bird dataset. LLaVa-13B Incorrect prediction. Refer to Figure 58.
5. FCT test on Butterfly dataset. GPT-4o Correct prediction. Refer to Figure 59.
6. FCT test on Butterfly dataset. LLaVa-13B Incorrect prediction. Refer to Figure 60.

N.2 None of The Above (NOTA) Test

1. NOTA test on Fish dataset. GPT-4o Correct prediction. Actual species name is *Esox Americanus*. Refer to Figure 61.
2. NOTA test on Fish dataset. LLaVa-13B Incorrect prediction. Actual species name is *Esox Americanus*. Refer to Figure 62.
3. NOTA test on Bird dataset. GPT-4o Correct prediction. Actual species name is *Corvus Albicollis*. Refer to Figure 63.
4. NOTA test on Bird dataset. Blip-Flan-XL Incorrect prediction. Actual species name is *Corvus Albicollis*. Refer to Figure 64.
5. NOTA test on Butterfly dataset. GPT-4o Incorrect prediction. Actual species name is *Batesia Hypochlora*. Refer to Figure 65.
6. NOTA test on Butterfly dataset. Blip-Flan-XL Correct prediction. Actual species name is *Batesia Hypochlora*. Refer to Figure 66.

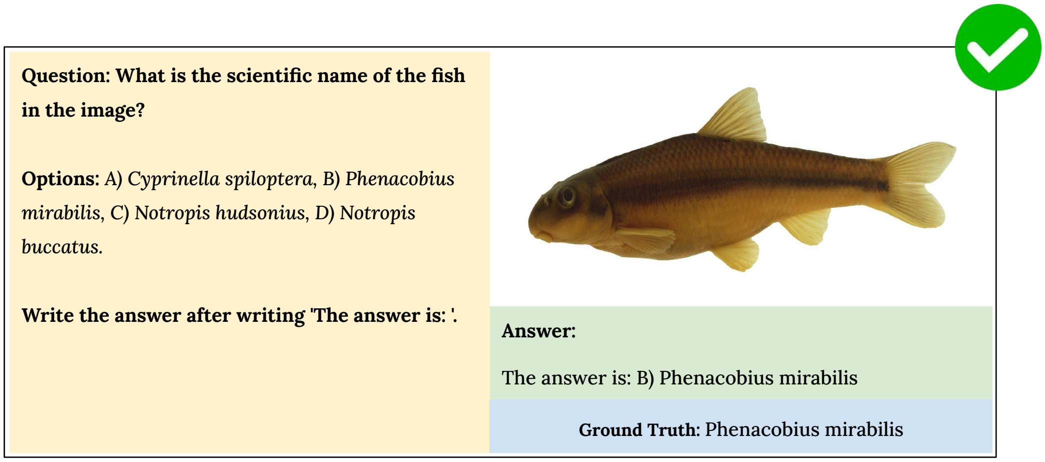


Figure 11: No Prompting. GPT-4o Correct prediction. Section M.1.

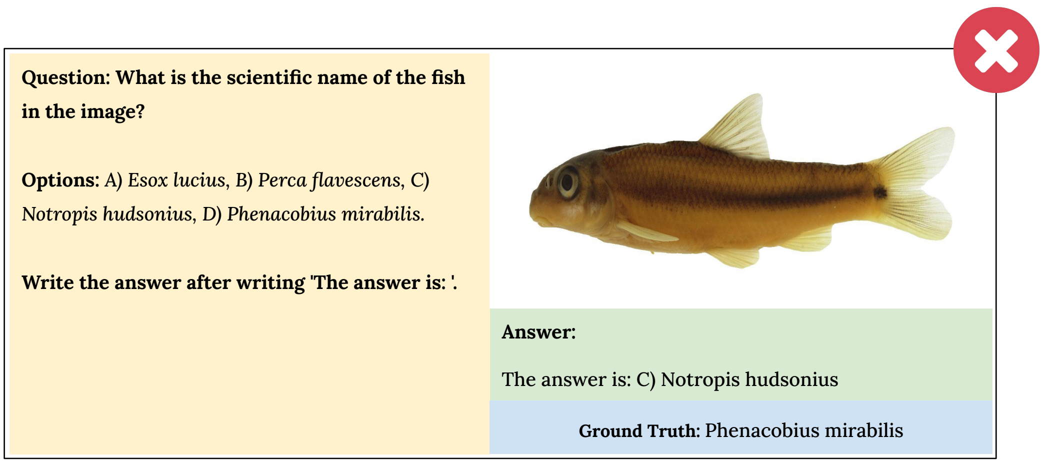


Figure 12: No Prompting. GPT-4o Incorrect prediction. Section M.1.

References

[1] fishair.org. Fish-air. *fishair.org*.

[2] Great Lakes Invasive Network Project (GLIN). https://greatlakesinvasives.org/portal/index.php.

[3] Biodiversity occurrence data published by: INHS Collections Data (accessed through the INHS Collections Data Portal, biocoll.inhs.illinois.edu/portal, 2024-06-04).

[4] FMNH Field Museum of Natural History (Zoology) Fish Collection.Field Museum.https://fmipt.fieldmuseum.org/ipt/resource?r=fmnh_fishes.

[5] Daly M and Johnson N. Ohio State University Fish Division (OSUM). Museum of Biological Diversity, The Ohio State University, February 2018.

[6] JFBM Bell Atlas. 2022. http://bellatlas.umn.edu/index.php.

[7] UMMZ University of Michigan Museum of Zoology, Division of Fishes. https://ipt.lsa.umich.edu/resource?r=ummz_fish.

[8] University of Wisconsin-Madison Zoological Museum - Fish. http://zoology.wisc.edu/uwzm/.

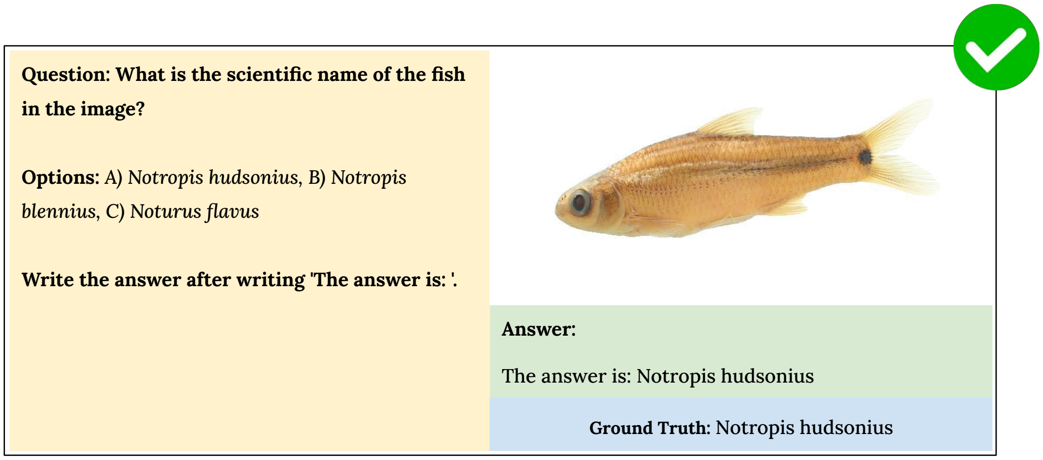


Figure 13: No Prompting. COG-VLM Correct prediction. Section M.1.

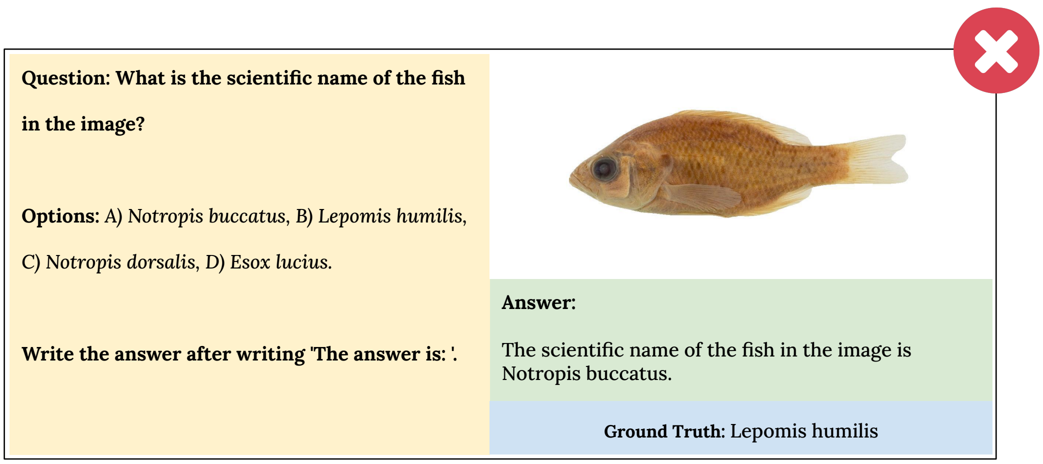


Figure 14: No Prompting. COG-VLM Incorrect prediction. Section M.1.

[9] Shilong Liu, Zhaoyang Zeng, Tianhe Ren, Feng Li, Hao Zhang, Jie Yang, Chunyuan Li, Jianwei Yang, Hang Su, Jun Zhu, et al. Grounding dino: Marrying dino with grounded pre-training for open-set object detection. *arXiv preprint arXiv:2303.05499*, 2023.

[10] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023.

[11] Christopher Lawrence and Elizabeth G. Campolongo. Heliconius collection (cambridge butterfly), 2024.

[12] Gabriela Montejo-Kovacevich, Eva van der Heijden, Nicola Nadeau, and Chris Jiggins. Cambridge butterfly wing collection batch 10, November 2020.

[13] Patricio A. Salazar, Nicola Nadeau, Gabriela Montejo-Kovacevich, and Chris Jiggins. Sheffield butterfly wing collection - Patricio Salazar, Nicola Nadeau, Ikiam broods batch 1 and 2, November 2020.

[14] Gabriela Montejo-Kovacevich, Chris Jiggins, and Ian Warren. Cambridge butterfly wing collection batch 2, May 2019.

Question: Each biological species has a unique scientific name composed of two parts: the first for the genus and the second for the species within that genus.
What is the scientific name of the fish in the image?

Options: A) *Lepomis cyanellus*, B) *Notropis blennius*, C) *Esox lucius*, D) *Gambusia affinis.*

Write the answer after writing 'The answer is: '.

Answer:

The answer is: Esox lucius

Ground Truth: Esox lucius

Figure 15: Contextual Prompting. GPT-4o Correct prediction. Section M.2.

Question: Each biological species has a unique scientific name composed of two parts: the first for the genus and the second for the species within that genus.
What is the scientific name of the fish in the image?

Options: A) *Phenacobius mirabilis*, B) *Notropis atherinoides*, C) *Lepomis humilis*, D) *Lepomis macrochirus.*

Write the answer after writing 'The answer is: '.

Answer:

The answer is: Lepomis humilis.

Ground Truth: Lepomis macrochirus

Figure 16: Contextual Prompting. GPT-4o Incorrect prediction. Section M.2.

[15] Chris Jiggins, Gabriela Montejo-Kovacevich, Ian Warren, and Eva Wiltshire. Cambridge butterfly wing collection batch 3, May 2019.

[16] Gabriela Montejo-Kovacevich, Chris Jiggins, and Ian Warren. Cambridge butterfly wing collection batch 4, May 2019.

[17] Gabriela Montejo-Kovacevich, Chris Jiggins, Ian Warren, and Eva Wiltshire. Cambridge butterfly wing collection batch 5, May 2019.

[18] Ian Warren and Chris Jiggins. Miscellaneous Heliconius wing photographs (2001-2019) Part 1, February 2019.

[19] Ian Warren and Chris Jiggins. Miscellaneous Heliconius wing photographs (2001-2019) Part 3, February 2019.

[20] Gabriela Montejo-Kovacevich, Chris Jiggins, Ian Warren, and Eva Wiltshire. Cambridge butterfly wing collection batch 6, May 2019.

[21] Chris Jiggins and Ian Warren. Cambridge butterfly wing collection - Chris Jiggins 2001/2 broods batch 1, January 2019.

[22] Chris Jiggins and Ian Warren. Cambridge butterfly wing collection - Chris Jiggins 2001/2 broods batch 2, January 2019.

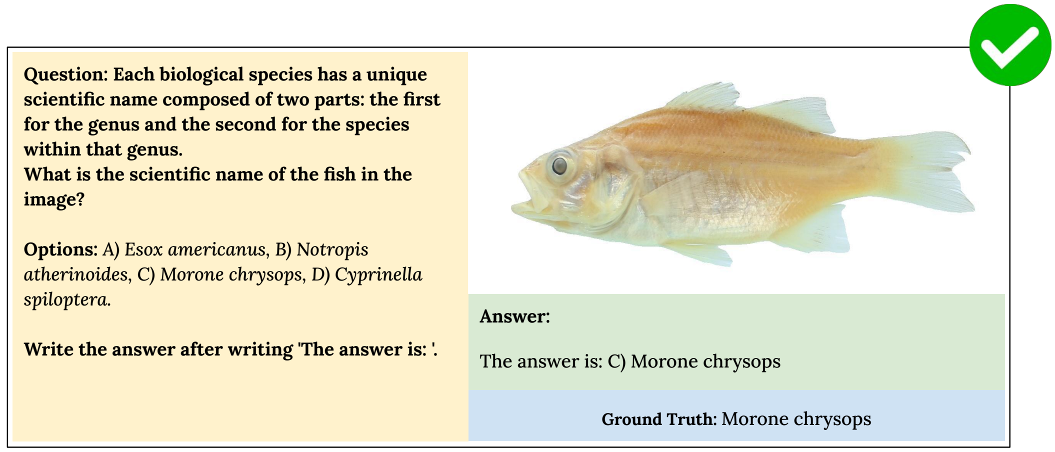


Figure 17: Contextual Prompting. LLaVa-13B Correct prediction. Section M.2.

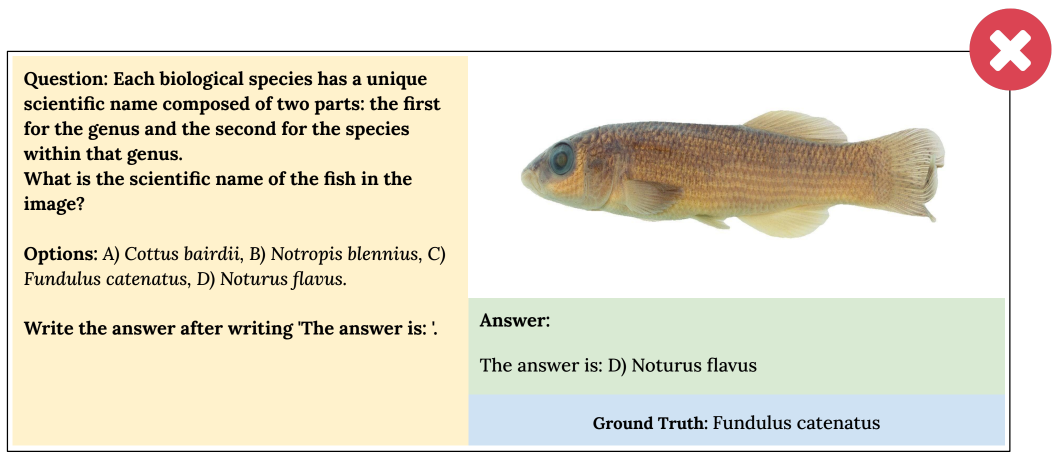


Figure 18: Contextual Prompting. LLaVa-13B Incorrect prediction. Section M.2.

[23] Joana I. Meier, Patricio Salazar, Gabriela Montejo-Kovacevich, Ian Warren, and Chris Jggins. Cambridge butterfly wing collection - Patricio Salazar PhD wild specimens batch 3, October 2020.

[24] Gabriela Montejo-Kovacevich, Chris Jiggins, and Ian Warren. Cambridge butterfly wing collection batch 1- version 2, May 2019.

[25] Gabriela Montejo-Kovacevich, Chris Jiggins, Ian Warren, Camilo Salazar, Marianne Elias, Imogen Gavins, Eva Wiltshire, Stephen Montgomery, and Owen McMillan. Cambridge and collaborators butterfly wing collection batch 10, May 2019.

[26] Patricio Salazar, Gabriela Montejo-Kovacevich, Ian Warren, and Chris Jiggins. Cambridge butterfly wing collection - Patricio Salazar PhD wild and bred specimens batch 1, December 2018.

[27] Gabriela Montejo-Kovacevich, Chris Jiggins, Ian Warren, and Eva Wiltshire. Cambridge butterfly wing collection batch 7, May 2019.

[28] Patricio Salazar, Gabriela Montejo-Kovacevich, Ian Warren, and Chris Jiggins. Cambridge butterfly wing collection - Patricio Salazar PhD wild and bred specimens batch 2, January 2019.

Dense Caption: The image showcases a small fish with a slender, elongated body shape. The body is covered in yellowish-brown hues, with a gradient that lightens towards the ventral side and deepens dorsally. The scales are distinctly visible, giving a textured appearance to the body. Starting from the head, the fish has a rounded snout and a relatively large, prominent eye with a dark pupil and a lighter surrounding area. The mouth is small and positioned slightly downturned. The dorsal fin is placed towards the rear of the body, exhibiting a rather elongated and low profile with delicate, translucent rays. The anal fin mirrors the dorsal fin but is placed closer to the rear ventral side, displaying similar morphological traits. The pectoral fins are modest in size, almost triangular, and located near the gill covers. The pelvic fins are also small and positioned ventrally, slightly behind the pectoral ones. The caudal fin is fan-shaped and extends symmetrically, marked by an array of thin rays. This fin appears semi-transparent with slight yellowish tinges that match the overall body coloration. Other distinctive features include a lateral line that is faintly visible and runs along the midline of the body. The skin reveals an intricate network of fine lines, which might suggest sensory functions in identifying environmental changes. Given these detailed anatomical descriptions, the fish could potentially be identified as a member of the Poeciliidae family, possibly a species of Gambusia or Heterandria.

Question: Use the above dense-caption and the image to answer the following question. What is the scientific name of the fish in the image?
Options: A) *Notropis volucellus*, B) *Notropis percobromus*, C) *Gambusia affinis*, D) *Lepomis gibbosus*.

Therefore, the answer is:

Answer:

The detailed description and the image closely match the morphological characteristics of the fish species Gambusia affinis, which belongs to the Poeciliidae family. Given the options:
A) Notropis volucellus
B) Notropis percobromus
C) Gambusia affinis
D) Lepomis gibbosus

The scientific name of the fish in the image is:
C) Gambusia affinis

Ground Truth: Gambusia affinis

Figure 19: Dense Captions in Prompts. GPT-4o Correct prediction. Section M.3.

Dense Caption: The fish displayed has an elongated, cylindrical body adorned with a pattern of small, well-defined scales that are golden-brown on the dorsal side and a lighter, creamy hue on the ventral side. The head is slightly pointed with a terminal mouth, featuring a prominent, spherical eye located on either side. The dorsal fin, located slightly more than halfway down the back, is relatively small and rounded. The caudal fin at the tail is forked, matching the body's light coloration and adorned with fine, subtle spots.
Pectoral fins are positioned just behind the gill covers and are almost translucent, slightly broader at the base and tapering to a point. The pelvic fins are minute and situated abdominally, while the anal fin mirrors the size and shape of the dorsal fin but is placed closer to the caudal peduncle. A closer inspection of the body reveals faint horizontal striping along the lateral line, a key feature in taxonomic differentiation. The absence of specialized pigment patterns, such as vibrant or contrasting spots, suggests a species adapted for camouflage in its environment. Distinctive anatomical features such as the body coloration pattern, shape, and position of fins, presence of lateral line markings, and general morphology are crucial for identifying the scientific classification of this fish species.

Question: Use the above dense-caption and the image to answer the following question. What is the scientific name of the fish in the image?
Options: A) *Notropis dorsalis*, B) *Lepomis megalotis*, C) *Notropis stramineus*, D) *Gambusia affinis*.

Therefore, the answer is:

Answer:

Based on the dense-caption provided and the image, the description matches the characteristics of the fish known as **Notropis stramineus**.

Therefore, the answer is: **C) Notropis stramineus**.

Ground Truth: Gambusia affinis

Figure 20: Dense Captions in Prompts. GPT-4o Incorrect prediction. Section M.3.

[29] Erika Pinheiro de Castro, Christopher Jiggins, Karina Lucas da Silva-Brand0̆0e3o, Andre Victor Lucci Freitas, Marcio Zikan Cardoso, Eva Van Der Heijden, Joana Meier, and Ian Warren. Brazilian Butterflies Collected December 2020 to January 2021, February 2022.

[30] Gabriela Montejo-Kovacevich, Chris Jiggins, Ian Warren, and Eva Wiltshire. Cambridge butterfly wing collection batch 8, May 2019.

[31] Gabriela Montejo-Kovacevich, Chris Jiggins, Ian Warren, Eva Wiltshire, and Imogen Gavins. Cambridge butterfly wing collection batch 9, May 2019.

Dense Caption: The image features a fish with a long, slender body, a pointed head, and a row of small, round fins along its length. The fins are located towards the back of the fish, and their positioning is consistent with the body shape. The fish has a yellowish-brown color, with a pattern of darker stripes running along its body. The fish's overall appearance is typical of a long, slender fish species, such as a snake fish or a similar species from the family Channidae. The presence of the fins and the body shape, along with the color pattern, can help identify the specific species within this family.

Question: Use the above dense-caption and the image to answer the following question. What is the scientific name of the fish in the image?
Options: A) *Esox americanus*, B) *Perca flavescens*, C) *Esox lucius*, D) *Fundulus catenatus*.

Therefore, the answer is:

Answer:

A) Esox americanus

Ground Truth: Esox americanus

Figure 21: Dense Captions in Prompts. LLaVa-7B Correct prediction. Section M.3.

Dense Caption: The image features a small, long, and slender fish with a yellow body and a black stripe running along its length. The fish has pointed head and a slightly curved tail. The fins are small and located towards the back of the fish, with a pair of dorsal fins and a pair of ventral fins. The fish's body is covered in a pattern of small, dark spots, which are a characteristic feature of the species. The fish's overall appearance suggests that it belongs to the family Cyprinidae, which includes species such as the common carp and the goldfish.

Question: Use the above dense-caption and the image to answer the following question. What is the scientific name of the fish in the image?
Options: A) *Cottus bairdii*, B) *Notropis atherinoides*, C) *Gambusia affinis*, D) *Notropis buccatus*.

Therefore, the answer is:

Answer:

A) Cottus bairdii

Ground Truth: Gambusia affinis

Figure 22: Dense Captions in Prompts. LLaVa-7B Incorrect prediction. Section M.3.

[32] Gabriela Montejo-Kovacevich, Eva van der Heijden, and Chris Jiggins. Cambridge butterfly collection - GMK Broods Ikiam 2018, November 2020.

[33] Gabriela Montejo-Kovacevich, Quentin Paynter, and Amin Ghane. Heliconius erato cyrbia, Cook Islands (New Zealand) 2016, 2019, 2021, September 2021.

[34] Ian Warren and Chris Jiggins. Miscellaneous Heliconius wing photographs (2001-2019) Part 2, February 2019.

[35] Camilo Salazar, Gabriela Montejo-Kovacevich, Chris Jiggins, Ian Warren, and Imogen Gavins. Camilo Salazar and Cambridge butterfly wing collection batch 1, May 2019.

[36] Anniina Mattila, Chris Jiggins, and Ian Warren. University of Helsinki butterfly collection - Anniina Mattila bred specimens, February 2019.

[37] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011.

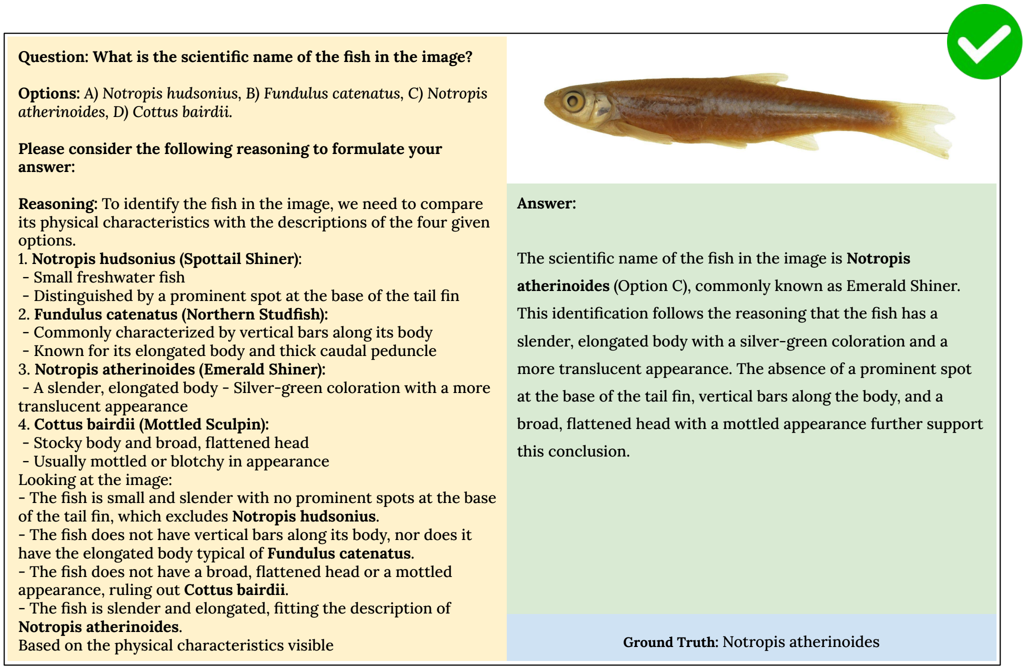


Figure 23: Chain-Of-Thought Prompting. GPT-4o Correct prediction. Section M.4.

[38] Elijah Cole, Kimberly Wilber, Grant Van Horn, Xuan Yang, Marco Fornoni, Pietro Perona, Serge Belongie, Andrew Howard, and Oisin Mac Aodha. On label granularity and object localization. In *European Conference on Computer Vision*. Springer, 2022.

[39] Richard C Edmunds, Baofeng Su, James P Balhoff, B Frank Eames, Wasila M Dahdul, Hilmar Lapp, John G Lundberg, Todd J Vision, Rex A Dunham, Paula M Mabee, et al. Phenoscape: identifying candidate genes for evolutionary phenotypes. *Molecular biology and evolution*, 33(1):13–24, 2015.

[40] Boris Sekachev, Nikita Manovich, Maxim Zhiltsov, Andrey Zhavoronkov, Dmitry Kalinin, Ben Hoff, TOsmanov, Dmitry Kruchinin, Artyom Zankevich, DmitriySidnev, Maksim Markelov, Johannes222, Mathis Chenuet, a andre, telenachos, Aleksandr Melnikov, Jijoong Kim, Liron Ilouz, Nikita Glazov, Priya4607, Rush Tehrani, Seungwon Jeong, Vladimir Skubriev, Sebastian Yonekura, vugia truong, zliang7, lizhming, and Tritin Truong. opencv/cvat: v1.1.0, August 2020.

[41] R OpenAI. Gpt-4 technical report. *arXiv*, pages 2303–08774, 2023.

[42] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. *arXiv preprint arXiv:2304.08485*, 2023.

[43] Wei-Lin Chiang, Zhuohan Li, Zi Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E Gonzalez, et al. Vicuna: An open-source chatbot impressing gpt-4 with 90%* chatgpt quality. *See https://vicuna. lmsys. org (accessed 14 April 2023)*, 2023.

[44] Weihan Wang, Qingsong Lv, Wenmeng Yu, Wenyi Hong, Ji Qi, Yan Wang, Junhui Ji, Zhuoyi Yang, Lei Zhao, Xixuan Song, et al. Cogvlm: Visual expert for pretrained language models. *arXiv preprint arXiv:2311.03079*, 2023.

[45] Deyao Zhu, Jun Chen, Xiaoqian Shen, Xiang Li, and Mohamed Elhoseiny. Minigpt-4: Enhancing vision-language understanding with advanced large language models. *arXiv preprint arXiv:2304.10592*, 2023.

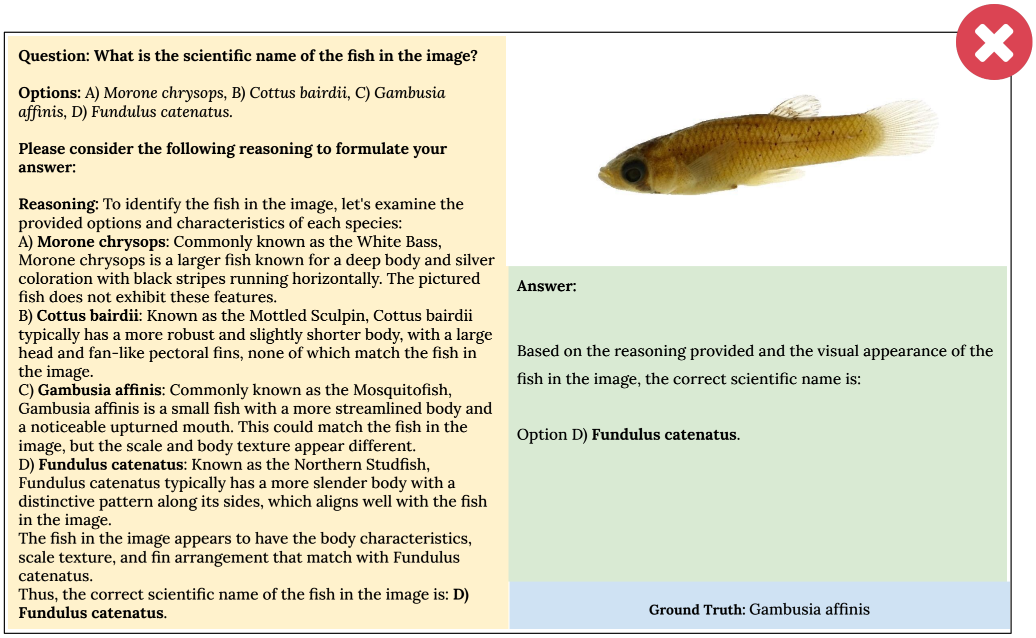


Figure 24: Chain-Of-Thought Prompting. GPT-4o Incorrect prediction. Section M.4.

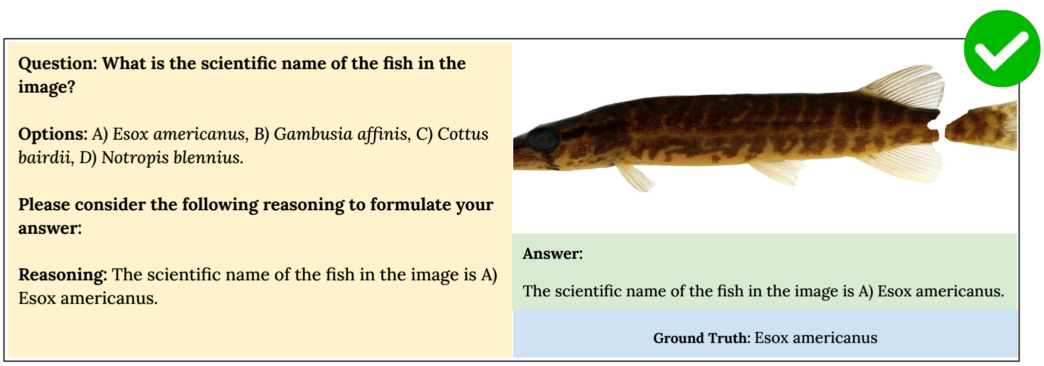


Figure 25: Chain-Of-Thought Prompting. LLaVa-13B Correct prediction. Section M.4.

[46] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. *arXiv preprint arXiv:2301.12597*, 2023.

[47] Wenliang Dai, Junnan Li, Dongxu Li, Anthony Meng Huat Tiong, Junqi Zhao, Weisheng Wang, Boyang Li, Pascale Fung, and Steven Hoi. Instructblip: Towards general-purpose vision-language models with instruction tuning, 2023.

[48] Samuel Stevens, Jiaman Wu, Matthew J Thompson, Elizabeth G Campolongo, Chan Hee Song, David Edward Carlyn, Li Dong, Wasila M Dahdul, Charles Stewart, Tanya Berger-Wolf, et al. Bioclip: A vision foundation model for the tree of life. *arXiv preprint arXiv:2311.18803*, 2023.

[49] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.

Question: What is the scientific name of the fish in the image?

Options: *A) Notropis hudsonius, B) Morone chrysops, C) Gambusia affinis, D) Notropis blennius.*

Please consider the following reasoning to formulate your answer:

Reasoning: The scientific name of the fish in the image is B) Morone chrysops.

Answer:

B) Morone chrysops

Ground Truth: Gambusia affinis

Figure 26: Chain-Of-Thought Prompting. LLaVa-13B Incorrect prediction. Section M.4.

Question: What is the scientific name of the fish in the image?

Options: A) *Notropis blennius*, B) *Notropis dorsalis*, C) *Cottus bairdii*, D) *Phenacobius mirabilis.*

Chosen Answer: Notropis dorsalis.

Please provide: 1) Whether the chosen answer is correct (True/False). 2) The correct answer .

Answer:

1) The chosen answer is correct (True/False): False
2) The correct answer: **Notropis blennius**"

Ground Truth: Notropis blennius

Figure 27: GPT-4o correct prediction. Section N.1.

Question: What is the scientific name of the fish in the image?

Options: A) *Lepomis macrochirus*, B) *Perca flavescens*, C) *Notropis volucellus*, D) *Notropis blennius.*

Chosen Answer: Notropis volucellus.

Please provide: 1) Whether the chosen answer is correct (True/False). 2) The correct answer .

Answer:

The correct answer is Notropis volucellus. The fish in the image is a long, slender fish with a yellowish color. It is swimming in a body of water, possibly a river or a lake.

Ground Truth: Notropis blennius

Figure 28: LLaVa-13B incorrect prediction. Section N.1.

Question: What is the scientific name of the bird in the image?

Options: A) *Antrostomus carolinensis*, B) *Sitta carolinensis*, C) *Certhia americana*, D) *Chloroceryle americana.*

Chosen Answer: Sitta carolinensis

Please provide: 1) Whether the chosen answer is correct (True/False). 2) The correct answer .

Ground Truth: Chloroceryle americana

Answer:

1) The chosen answer is correct: False.
2) The correct answer: D) Chloroceryle americana.
The bird in the image is the Green Kingfisher, which is scientifically named Chloroceryle americana.

Figure 29: GPT-4o Correct prediction. Section N.1.

Question: What is the scientific name of the bird in the image?

Options: A) *Contopus sordidulus*, B) *Vireo olivaceus*, C) *Corvus albicollis*, D) *Chloroceryle americana.*

Chosen Answer: Contopus sordidulus

Please provide: 1) Whether the chosen answer is correct (True/False). 2) The correct answer .

Ground Truth: Chloroceryle americana

Answer:

True. The scientific name of the bird in the image is Contopus sordidulus.

Figure 30: LLaVa-13B Incorrect prediction. Section N.1.

Question: What is the scientific name of the butterfly in the image?

Options: A) *Heliconius heurippa*, B) *Batesia hypochlora*, C) *Heliconius melpomene*, D) *Heliconius wallacei.*

Chosen Answer: Heliconius heurippa

Please provide: 1) Whether the chosen answer is correct (True/False). 2) The correct answer .

Ground Truth: Batesia hypochlora

Answer:

1) The chosen answer is incorrect (False).
2) The correct answer is B) Batesia hypochlora.

Figure 31: GPT-4o Correct prediction. Section N.1.

Question: What is the scientific name of the butterfly in the image?

Options: A) *Pyrrhogyra edocia*, B) *Heliconius timareta*, C) *Heliconius doris*, D) *Batesia hypochlora.*

Chosen Answer: Heliconius doris

Please provide: 1) Whether the chosen answer is correct (True/False). 2) The correct answer .

Ground Truth: Batesia hypochlora

Answer:

1. False
2. The correct answer is B) Heliconius timareta.

Figure 32: LLaVa-13B Incorrect prediction. Section N.1.

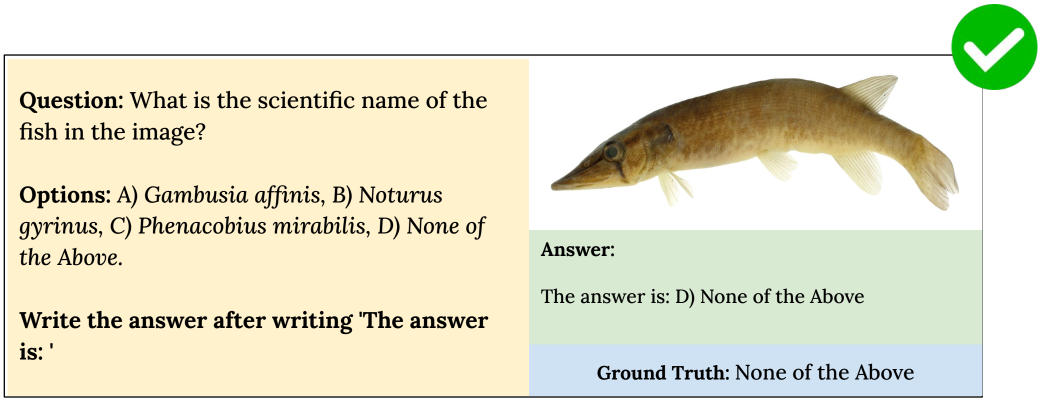


Figure 33: GPT-4o Correct prediction. Actual species name is Esox Americanus. Section N.2.

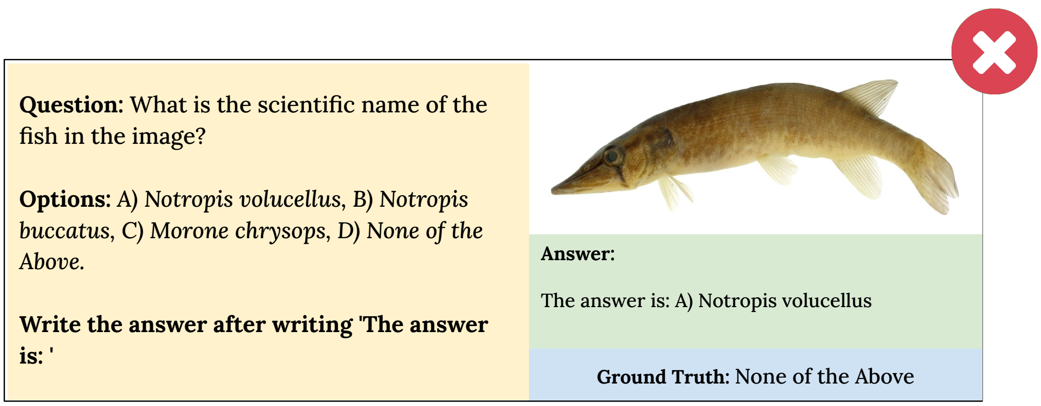


Figure 34: LLaVa-13B Incorrect prediction. Actual species name is Esox Americanus. Section N.2.

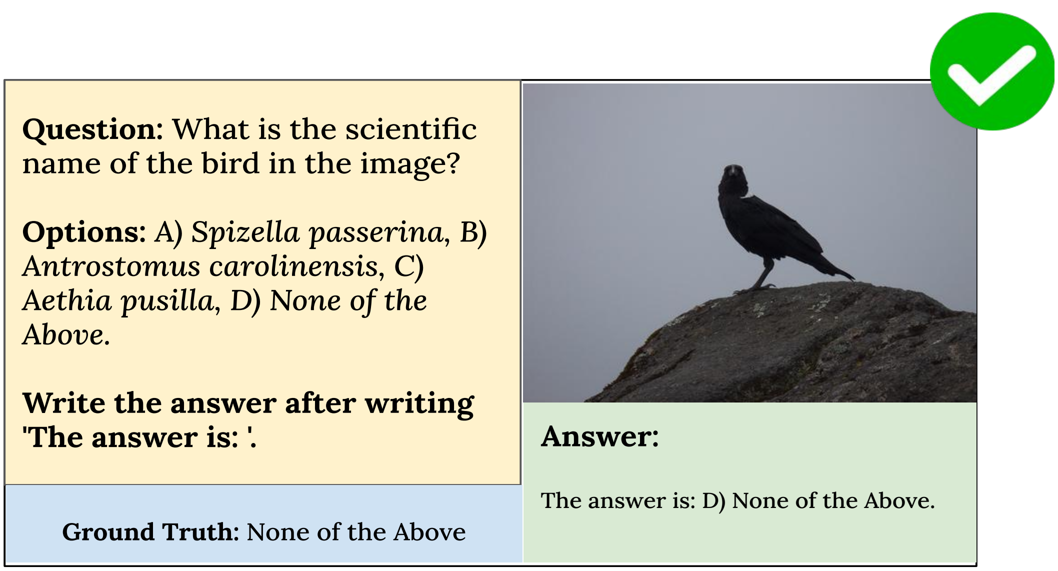


Figure 35: GPT-4o Correct prediction. Actual species name is Corvus Albicollis. Section N.2.

Question: What is the scientific name of the bird in the image?

Options: A) *Cardellina pusilla*, B) *Thryothorus ludovicianus*, C) *Passer domesticus*, D) *None of the Above.*

Write the answer after writing 'The answer is: '.

Ground Truth: None of the Above

Answer:

C

Figure 36: Blip-Flan-XL Incorrect prediction. Actual species name is Corvus Albicollis. Section N.2.

Question: What is the scientific name of the butterfly in the image?

Options: A) *Temenis laothoe*, B) *Heliconius doris*, C) *Heliconius heurippa*, D) *None of the Above.*

Write the answer after writing 'The answer is: '.

Ground Truth: None of the Above

Answer:

The answer is: A) Temenis laothoe

Figure 37: GPT-4o Incorrect prediction. Actual species name is *Batesia Hypochlora*. Section N.2.

Question: What is the scientific name of the butterfly in the image?

Options: A) *Heliconius erato*, B) *Heliconius melpomene*, C) *Heliconius eleuchia*, D) *None of the Above.*

Write the answer after writing 'The answer is: '.

Ground Truth: None of the Above

Answer:

D

Figure 38: Blip-Flan-XL Correct prediction. Actual species name is *Batesia Hypochlora*. Section N.2.

Question: What is the scientific name of the fish in the image?

Options: A) *Cyprinella spiloptera*, B) *Phenacobius mirabilis*, C) *Notropis hudsonius*, D) *Notropis buccatus*.

Write the answer after writing 'The answer is: '.

Answer:

The answer is: B) Phenacobius mirabilis

Ground Truth: Phenacobius mirabilis

Figure 39: No Prompting. GPT-4o Correct prediction. Section M.1.

Question: What is the scientific name of the fish in the image?

Options: A) *Esox lucius*, B) *Perca flavescens*, C) *Notropis hudsonius*, D) *Phenacobius mirabilis*.

Write the answer after writing 'The answer is: '.

Answer:

The answer is: C) Notropis hudsonius

Ground Truth: Phenacobius mirabilis

Figure 40: No Prompting. GPT-4o Incorrect prediction. Section M.1.

Question: What is the scientific name of the fish in the image?

Options: A) *Notropis hudsonius*, B) *Notropis blennius*, C) *Noturus flavus*

Write the answer after writing 'The answer is: '.

Answer:

The answer is: Notropis hudsonius

Ground Truth: Notropis hudsonius

Figure 41: No Prompting. COG-VLM Correct prediction. Section M.1.

Question: What is the scientific name of the fish in the image?

Options: A) *Notropis buccatus*, B) *Lepomis humilis*, C) *Notropis dorsalis*, D) *Esox lucius*.

Write the answer after writing 'The answer is: '.

Answer:

The scientific name of the fish in the image is Notropis buccatus.

Ground Truth: Lepomis humilis

Figure 42: No Prompting. COG-VLM Incorrect prediction. Section M.1.

**Question: Each biological species has a unique scientific name composed of two parts: the first for the genus and the second for the species within that genus.
What is the scientific name of the fish in the image?**

Options: A) *Lepomis cyanellus*, B) *Notropis blennius*, C) *Esox lucius*, D) *Gambusia affinis*.

Write the answer after writing 'The answer is: '.

Answer:

The answer is: Esox lucius

Ground Truth: Esox lucius

Figure 43: Contextual Prompting. GPT-4o Correct prediction. Section M.2.

**Question: Each biological species has a unique scientific name composed of two parts: the first for the genus and the second for the species within that genus.
What is the scientific name of the fish in the image?**

Options: A) *Phenacobius mirabilis*, B) *Notropis atherinoides*, C) *Lepomis humilis*, D) *Lepomis macrochirus*.

Write the answer after writing 'The answer is: '.

Answer:

The answer is: Lepomis humilis.

Ground Truth: Lepomis macrochirus

Figure 44: Contextual Prompting. GPT-4o Incorrect prediction. Section M.2.

Question: Each biological species has a unique scientific name composed of two parts: the first for the genus and the second for the species within that genus.
What is the scientific name of the fish in the image?

Options: *A) Esox americanus, B) Notropis atherinoides, C) Morone chrysops, D) Cyprinella spiloptera.*

Write the answer after writing 'The answer is: '.

Answer:

The answer is: C) Morone chrysops

Ground Truth: Morone chrysops

Figure 45: Contextual Prompting. LLaVa-13B Correct prediction. Section M.2.

Question: Each biological species has a unique scientific name composed of two parts: the first for the genus and the second for the species within that genus.
What is the scientific name of the fish in the image?

Options: *A) Cottus bairdii, B) Notropis blennius, C) Fundulus catenatus, D) Noturus flavus.*

Write the answer after writing 'The answer is: '.

Answer:

The answer is: D) Noturus flavus

Ground Truth: Fundulus catenatus

Figure 46: Contextual Prompting. LLaVa-13B Incorrect prediction. Section M.2.

Dense Caption: The image showcases a small fish with a slender, elongated body shape. The body is covered in yellowish-brown hues, with a gradient that lightens towards the ventral side and deepens dorsally. The scales are distinctly visible, giving a textured appearance to the body. Starting from the head, the fish has a rounded snout and a relatively large, prominent eye with a dark pupil and a lighter surrounding area. The mouth is small and positioned slightly downturned. The dorsal fin is placed towards the rear of the body, exhibiting a rather elongated and low profile with delicate, translucent rays. The anal fin mirrors the dorsal fin but is placed closer to the rear ventral side, displaying similar morphological traits. The pectoral fins are modest in size, almost triangular, and located near the gill covers. The pelvic fins are also small and positioned ventrally, slightly behind the pectoral ones. The caudal fin is fan-shaped and extends symmetrically, marked by an array of thin rays. This fin appears semi-transparent with slight yellowish tinges that match the overall body coloration. Other distinctive features include a lateral line that is faintly visible and runs along the midline of the body. The skin reveals an intricate network of fine lines, which might suggest sensory functions in identifying environmental changes. Given these detailed anatomical descriptions, the fish could potentially be identified as a member of the Poeciliidae family, possibly a species of Gambusia or Heterandria.

Question: Use the above dense-caption and the image to answer the following question. What is the scientific name of the fish in the image?
Options: A) *Notropis volucellus*, B) *Notropis percobromus*, C) *Gambusia affinis*, D) *Lepomis gibbosus*.

Therefore, the answer is:

Answer:

The detailed description and the image closely match the morphological characteristics of the fish species Gambusia affinis, which belongs to the Poeciliidae family.
Given the options:
A) Notropis volucellus
B) Notropis percobromus
C) Gambusia affinis
D) Lepomis gibbosus

The scientific name of the fish in the image is:
C) Gambusia affinis

Ground Truth: Gambusia affinis

Figure 47: Dense Captions in Prompts. GPT-4o Correct prediction. Section M.3.

Dense Caption: The fish displayed has an elongated, cylindrical body adorned with a pattern of small, well-defined scales that are golden-brown on the dorsal side and a lighter, creamy hue on the ventral side. The head is slightly pointed with a terminal mouth, featuring a prominent, spherical eye located on either side. The dorsal fin, located slightly more than halfway down the back, is relatively small and rounded. The caudal fin at the tail is forked, matching the body's light coloration and adorned with fine, subtle spots.
Pectoral fins are positioned just behind the gill covers and are almost translucent, slightly broader at the base and tapering to a point. The pelvic fins are minute and situated abdominally, while the anal fin mirrors the size and shape of the dorsal fin but is placed closer to the caudal peduncle. A closer inspection of the body reveals faint horizontal striping along the lateral line, a key feature in taxonomic differentiation. The absence of specialized pigment patterns, such as vibrant or contrasting spots, suggests a species adapted for camouflage in its environment. Distinctive anatomical features such as the body coloration pattern, shape, and position of fins, presence of lateral line markings, and general morphology are crucial for identifying the scientific classification of this fish species.

Question: Use the above dense-caption and the image to answer the following question. What is the scientific name of the fish in the image?
Options: A) *Notropis dorsalis*, B) *Lepomis megalotis*, C) *Notropis stramineus*, D) *Gambusia affinis*.

Therefore, the answer is:

Answer:

Based on the dense-caption provided and the image, the description matches the characteristics of the fish known as **Notropis stramineus**.

Therefore, the answer is: **C) Notropis stramineus**.

Ground Truth: Gambusia affinis

Figure 48: Dense Captions in Prompts. GPT-4o Incorrect prediction. Section M.3.

Dense Caption: The image features a fish with a long, slender body, a pointed head, and a row of small, round fins along its length. The fins are located towards the back of the fish, and their positioning is consistent with the body shape. The fish has a yellowish-brown color, with a pattern of darker stripes running along its body. The fish's overall appearance is typical of a long, slender fish species, such as a snake fish or a similar species from the family Channidae. The presence of the fins and the body shape, along with the color pattern, can help identify the specific species within this family.

Question: Use the above dense-caption and the image to answer the following question. What is the scientific name of the fish in the image?
Options: A) *Esox americanus*, B) *Perca flavescens*, C) *Esox lucius*, D) *Fundulus catenatus*.

Therefore, the answer is:

Answer:

A) Esox americanus

Ground Truth: Esox americanus

Figure 49: Dense Captions in Prompts. LLaVa-7B Correct prediction. Section M.3.

Dense Caption: The image features a small, long, and slender fish with a yellow body and a black stripe running along its length. The fish has a pointed head and a slightly curved tail. The fins are small and located towards the back of the fish, with a pair of dorsal fins and a pair of ventral fins. The fish's body is covered in a pattern of small, dark spots, which are a characteristic feature of the species. The fish's overall appearance suggests that it belongs to the family Cyprinidae, which includes species such as the common carp and the goldfish.

Question: Use the above dense-caption and the image to answer the following question. What is the scientific name of the fish in the image?
Options: A) *Cottus bairdii*, B) *Notropis atherinoides*, C) *Gambusia affinis*, D) *Notropis buccatus*.

Therefore, the answer is:

Answer:

A) Cottus bairdii

Ground Truth: Gambusia affinis

Figure 50: Dense Captions in Prompts. LLaVa-7B Incorrect prediction. Section M.3.

Question: What is the scientific name of the fish in the image?

Options: A) *Notropis hudsonius*, B) *Fundulus catenatus*, C) *Notropis atherinoides*, D) *Cottus bairdii*.

Please consider the following reasoning to formulate your answer:

Reasoning: To identify the fish in the image, we need to compare its physical characteristics with the descriptions of the four given options.
1. **Notropis hudsonius (Spottail Shiner):**
 - Small freshwater fish
 - Distinguished by a prominent spot at the base of the tail fin
2. **Fundulus catenatus (Northern Studfish):**
 - Commonly characterized by vertical bars along its body
 - Known for its elongated body and thick caudal peduncle
3. **Notropis atherinoides (Emerald Shiner):**
 - A slender, elongated body - Silver-green coloration with a more translucent appearance
4. **Cottus bairdii (Mottled Sculpin):**
 - Stocky body and broad, flattened head
 - Usually mottled or blotchy in appearance
Looking at the image:
- The fish is small and slender with no prominent spots at the base of the tail fin, which excludes **Notropis hudsonius**.
- The fish does not have vertical bars along its body, nor does it have the elongated body typical of **Fundulus catenatus**.
- The fish does not have a broad, flattened head or a mottled appearance, ruling out **Cottus bairdii**.
- The fish is slender and elongated, fitting the description of **Notropis atherinoides**.
Based on the physical characteristics visible

Answer:

The scientific name of the fish in the image is **Notropis atherinoides** (Option C), commonly known as Emerald Shiner. This identification follows the reasoning that the fish has a slender, elongated body with a silver-green coloration and a more translucent appearance. The absence of a prominent spot at the base of the tail fin, vertical bars along the body, and a broad, flattened head with a mottled appearance further support this conclusion.

Ground Truth: Notropis atherinoides

Figure 51: Chain-Of-Thought Prompting. GPT-4o Correct prediction. Section M.4.

Question: What is the scientific name of the fish in the image?

Options: A) *Morone chrysops*, B) *Cottus bairdii*, C) *Gambusia affinis*, D) *Fundulus catenatus*.

Please consider the following reasoning to formulate your answer:

Reasoning: To identify the fish in the image, let's examine the provided options and characteristics of each species:
A) **Morone chrysops**: Commonly known as the White Bass, Morone chrysops is a larger fish known for a deep body and silver coloration with black stripes running horizontally. The pictured fish does not exhibit these features.
B) **Cottus bairdii**: Known as the Mottled Sculpin, Cottus bairdii typically has a more robust and slightly shorter body, with a large head and fan-like pectoral fins, none of which match the fish in the image.
C) **Gambusia affinis**: Commonly known as the Mosquitofish, Gambusia affinis is a small fish with a more streamlined body and a noticeable upturned mouth. This could match the fish in the image, but the scale and body texture appear different.
D) **Fundulus catenatus**: Known as the Northern Studfish, Fundulus catenatus typically has a more slender body with a distinctive pattern along its sides, which aligns well with the fish in the image.
The fish in the image appears to have the body characteristics, scale texture, and fin arrangement that match with Fundulus catenatus.
Thus, the correct scientific name of the fish in the image is: **D) Fundulus catenatus**.

Answer:

Based on the reasoning provided and the visual appearance of the fish in the image, the correct scientific name is:

Option D) **Fundulus catenatus**.

Ground Truth: Gambusia affinis

Figure 52: Chain-Of-Thought Prompting. GPT-4o Incorrect prediction. Section M.4.

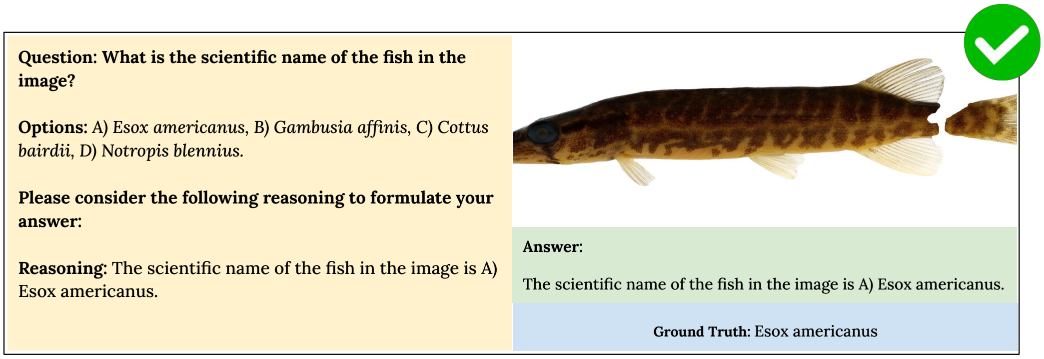


Figure 53: Chain-Of-Thought Prompting. LLaVa-13B Correct prediction. Section M.4.

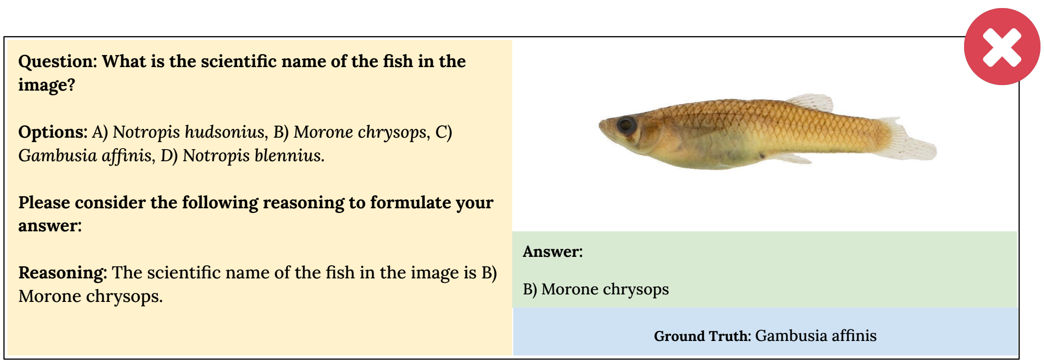


Figure 54: Chain-Of-Thought Prompting. LLaVa-13B Incorrect prediction. Section M.4.

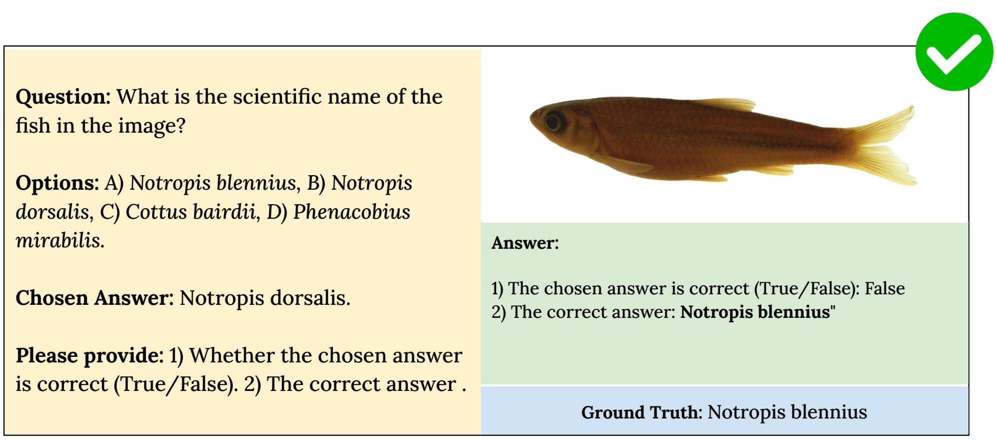


Figure 55: GPT-4o correct prediction. Section N.1.

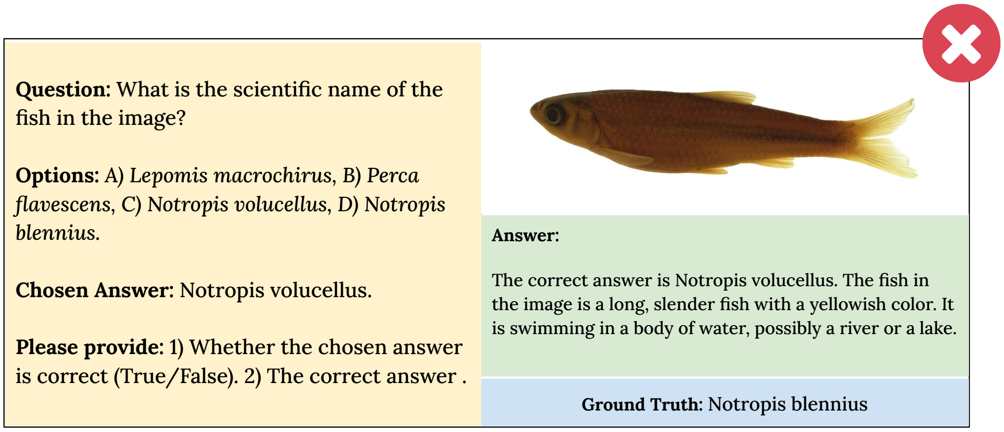


Figure 56: LLaVa-13B incorrect prediction. Section N.1.

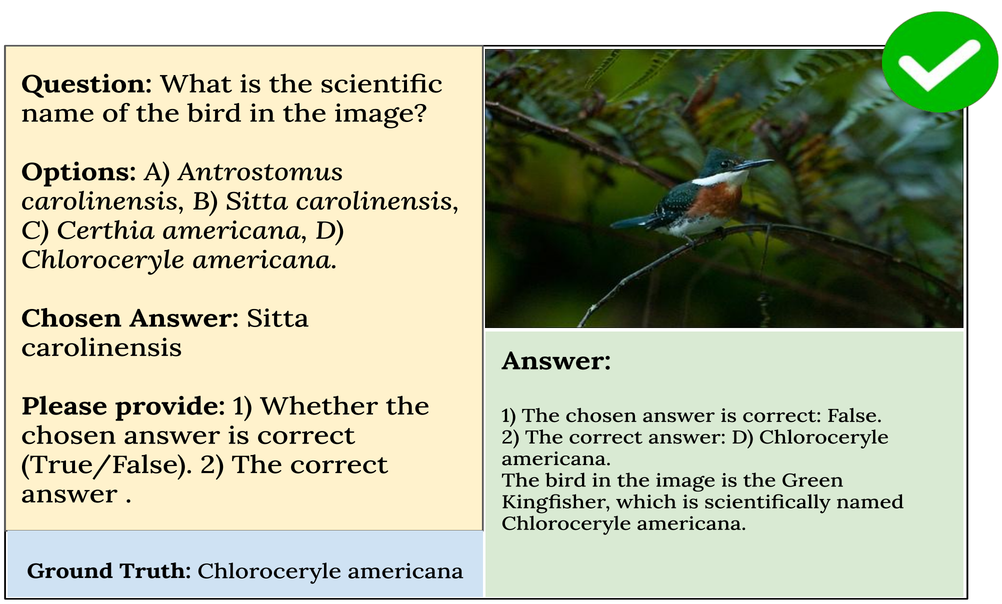


Figure 57: GPT-4o Correct prediction. Section N.1.

Question: What is the scientific name of the bird in the image?

Options: A) *Contopus sordidulus*, B) *Vireo olivaceus*, C) *Corvus albicollis*, D) *Chloroceryle americana*.

Chosen Answer: Contopus sordidulus

Please provide: 1) Whether the chosen answer is correct (True/False). 2) The correct answer .

Ground Truth: Chloroceryle americana

Answer:

True. The scientific name of the bird in the image is Contopus sordidulus.

Figure 58: LLaVa-13B Incorrect prediction. Section N.1.

Question: What is the scientific name of the butterfly in the image?

Options: A) *Heliconius heurippa*, B) *Batesia hypochlora*, C) *Heliconius melpomene*, D) *Heliconius wallacei*.

Chosen Answer: Heliconius heurippa

Please provide: 1) Whether the chosen answer is correct (True/False). 2) The correct answer .

Ground Truth: Batesia hypochlora

Answer:

1) The chosen answer is incorrect (False).
2) The correct answer is B) Batesia hypochlora.

Figure 59: GPT-4o Correct prediction. Section N.1.

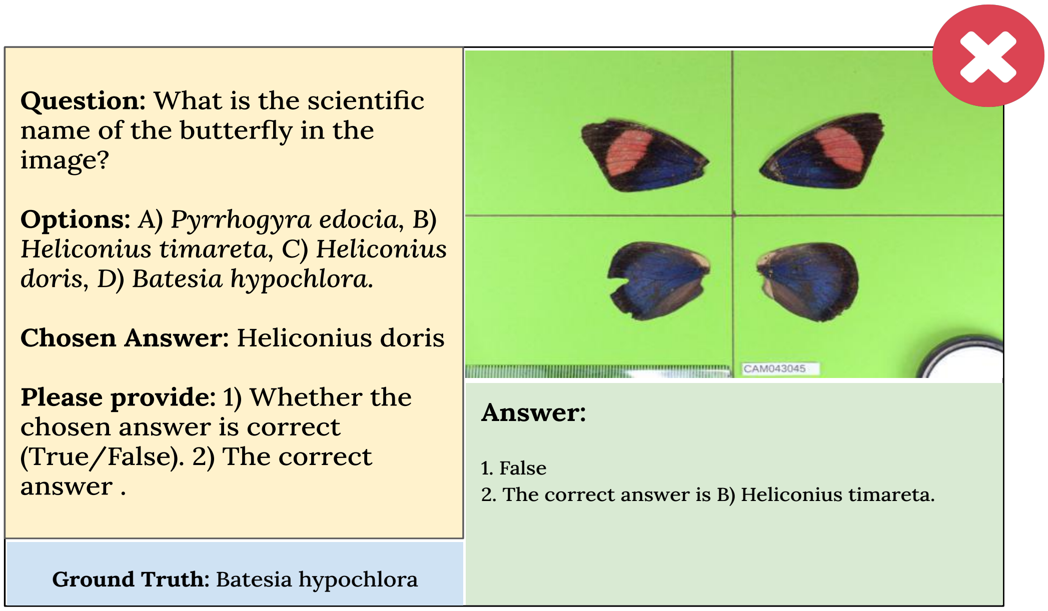


Figure 60: LLaVa-13B Incorrect prediction. Section N.1.

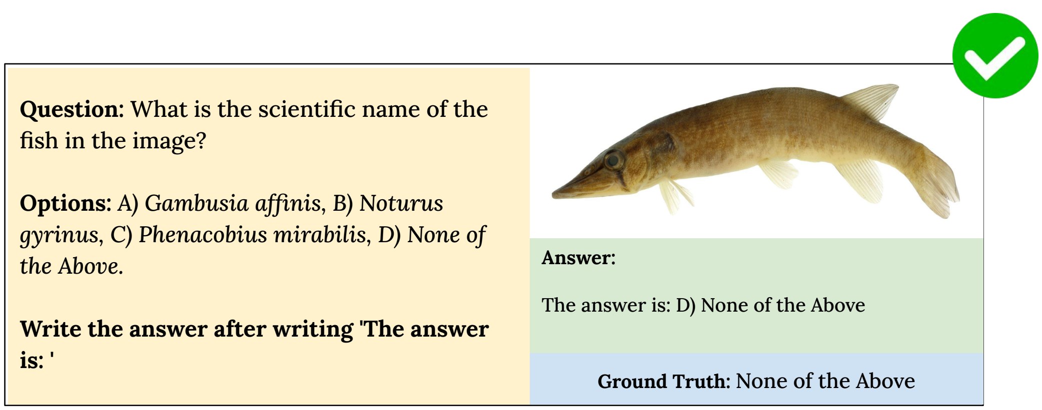


Figure 61: GPT-4o Correct prediction. Actual species name is Esox Americanus. Section N.2.

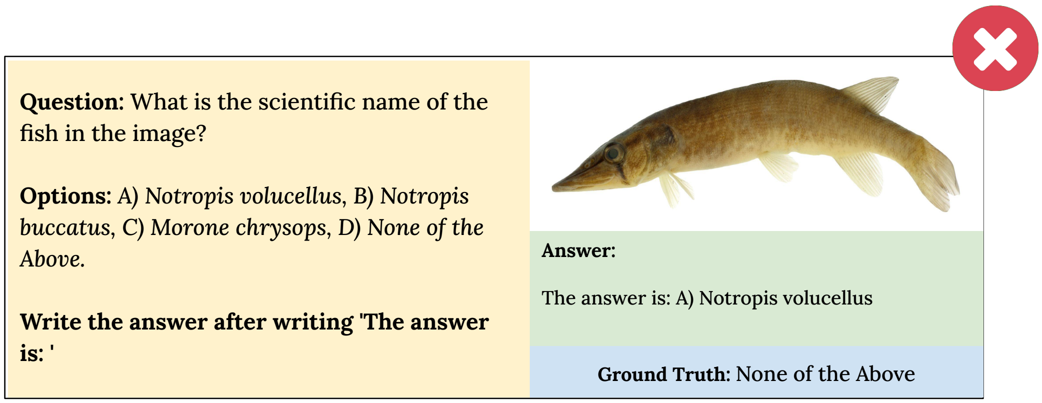


Figure 62: LLaVa-13B Incorrect prediction. Actual species name is Esox Americanus. Section N.2.

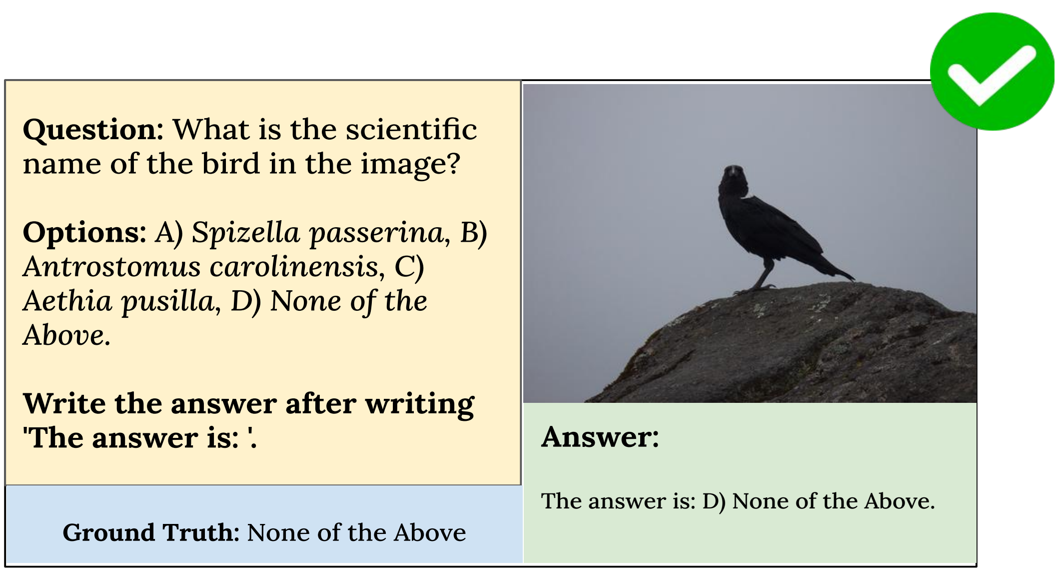


Figure 63: GPT-4o Correct prediction. Actual species name is Corvus Albicollis. Section N.2.

Question: What is the scientific name of the bird in the image?

Options: A) *Cardellina pusilla*, B) *Thryothorus ludovicianus*, C) *Passer domesticus*, D) None of the Above.

Write the answer after writing 'The answer is: '.

Ground Truth: None of the Above

Answer:

C

Figure 64: Blip-Flan-XL Incorrect prediction. Actual species name is Corvus Albicollis. Section N.2.

Question: What is the scientific name of the butterfly in the image?

Options: A) *Temenis laothoe*, B) *Heliconius doris*, C) *Heliconius heurippa*, D) None of the Above.

Write the answer after writing 'The answer is: '.

Ground Truth: None of the Above

Answer:

The answer is: A) Temenis laothoe

Figure 65: GPT-4o Incorrect prediction. Actual species name is *Batesia Hypochlora*. Section N.2.

Question: What is the scientific name of the butterfly in the image?

Options: A) *Heliconius erato*, B) *Heliconius melpomene*, C) *Heliconius eleuchia*, D) *None of the Above.*

Write the answer after writing 'The answer is: '.

Ground Truth: None of the Above

Answer:

D

Figure 66: Blip-Flan-XL Correct prediction. Actual species name is *Batesia Hypochlora*. Section N.2.