Appendix

A. Category-level Texts produced by LLMs

We provide more examples of texts produced for different prompts, including both visual and habitat. For prompting about visual attributes we set ‘System Role’ as “You are a helpful assistant who can identify any `[domain]` based on its photo.” For prompting about habitat attributes we set ‘System Role’ as “You are a helpful assistant who knows about the habitat of any `[domain]`.”

A.1. GPT4 texts used in our method

Prompt
What characteristics can be used to differentiate a *White Spruce*,
a type of plant with scientific name Picea glauca from other organisms based on just a photo?
Produce an exhaustive list of all attributes or any information that can be used to identify the plant uniquely.
Texts should be of the form “Plant with <description>”.
Ensure to structure your response as a list of single sentences.
Output
- Plant with short, blunt, and four-sided needles that are blue-green to green in color.
- Plant with needles that are spirally arranged on the branches.
- Plant with a conical shape, often with a straight trunk.
- Plant with bark that is thin, scaly, and grayish-brown in color.
- Plant with cones that are slender, cylindrical, and hang down from the branches.
- Plant with cones that are light brown and have thin, flexible scales.
- Plant with branches that are densely covered with needles, giving it a full appearance.
- Plant with a height that can reach up to 60 feet or more.
- Plant with a habitat often found in cold, northern climates, such as the boreal forests of Canada and Alaska.
- Plant with a preference for well-drained, sandy or gravelly soils.
- Plant with a growth pattern that is slow to moderate.
- Plant with a crown that is narrow and pointed.
- Plant with lower branches that often droop downwards.
- Plant with a root system that is shallow and wide-spreading.
- Plant with a tolerance for cold temperatures and harsh winter conditions.
- Plant with a tendency to grow in pure stands or mixed with other conifer species.
- Plant with a lifespan that can exceed 200 years.
- Plant with a resistance to pests and diseases, making it a hardy species.
- Plant with a usage in landscaping for windbreaks, privacy screens, or as a specimen tree.
- Plant with a significance in the timber industry for its wood, which is used for pulp and construction.

Prompt
Where can we find a *White Spruce*, a type of plant with scientific name Picea glauca?
Produce a list of all habitat and geographic information that can be used to identify the plant.
Each text produced should start with “Plant which”.
Ensure to structure your response as a list of single sentences.
Output
- Plant which thrives in the cold, boreal forests of North America.
- Plant which is native to the northern parts of North America, including Alaska and Canada.
- Plant which can be found growing in the high altitudes of the Rocky Mountains.
- Plant which is commonly found in the taiga and the boreal forests of North America.
- Plant which can survive in a variety of soils, including sandy, loamy, and clay soils.
- Plant which can be found in the northeastern and northwestern parts of the United States.
- Plant which is often found in cool, moist locations, such as near rivers and lakes.
- Plant which can tolerate harsh winter conditions and is often found in areas with heavy snowfall.
- Plant which is a dominant species in the North American boreal forest, also known as the taiga.
- Plant which can be found in the northern parts of the United States, including Maine, Wisconsin, and Minnesota.

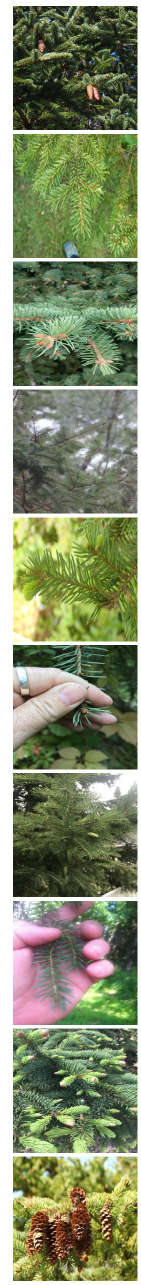


Figure 4. Example of texts produced for a category “White Spruce” of the iNaturalist dataset using GPT4-0613, example images on the right.

Prompt
What characteristics can be used to differentiate a *Common Tern* from other birds based on just a photo?
Produce an exhaustive list of all attributes or any information that can be used to identify the bird uniquely.
Texts should be of the form “a bird with <characteristic feature>”.
Ensure to structure your response as a list of single sentences.

Output
- A bird with a slender, pointed bill that is usually orange with a black tip.
- A bird with a black cap on its head that extends to the eyes.
- A bird with a deeply forked tail, which gives it a distinctive ’swallow-like’ appearance.
- A bird with long, narrow wings that are grey on top and lighter underneath.
- A bird with short, red legs.
- A bird with a white underbelly and chest.
- A bird with a size of approximately 31-35 cm in length and a wingspan of 70-80 cm.
- A bird with a distinctive hovering flight pattern before diving to catch fish.
- A bird with a sharp, piercing call that sounds like ’kee-arr’.
- A bird that is often seen in coastal areas, particularly during the breeding season.
- A bird that is often seen in large flocks, particularly during migration.
- A bird that is commonly found in the northern parts of North America, Europe, and Asia during the breeding season.
- A bird that migrates long distances, often to the southern parts of South America and Africa during the non-breeding season.
- A bird that is often seen diving into the water to catch small fish and invertebrates.
- A bird with a sleek, streamlined body that is well adapted for its aquatic lifestyle.
- A bird that is often seen perched on buoys, posts, and other structures near water.
- A bird that is known for its aggressive behavior towards intruders, including humans, during the breeding season.
- A bird that lays its eggs on the ground in a shallow scrape, often in a colony with other terns.
- A bird with juveniles that have a scaly pattern on their wings and back.

Prompt
Where can we find a *Common Tern*?
Produce a list of habitat and geographic location information that can be used to identify the bird.
Each text produced should start with “a bird”.
Ensure to structure your response as a list of single sentences.

Output
- A bird which can be found in coastal areas and inland water bodies across North America, Europe, and Asia during the breeding season.
- A bird which migrates to warmer coastal areas, offshore islands, and open seas in the tropics and Southern Hemisphere for the winter.
- A bird which nests in colonies on the ground in areas with loose sand, gravel, shell, or cobble pebbles typically less than 350 feet from the water.
- A bird which can be seen in estuaries, salt marshes, and sandy or rocky shorelines during migration and winter.
- A bird which can also be found in man-made habitats such as dredge spoil islands and gravel rooftops.
- A bird which is often seen hovering over water before diving down to catch a fish.
- A bird which is known for its long-distance migrations, with some individuals traveling from the Arctic to the Antarctic and back each year.
- A bird which is easily identifiable by its slender body, long pointed wings, forked tail, and black cap on its head.

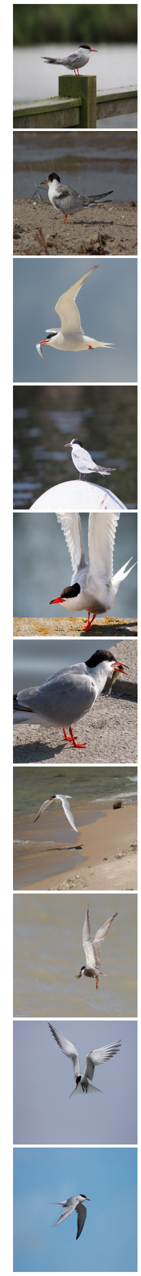


Figure 5. Example of texts produced for a category “Common Tern” of the CUB dataset using GPT4-0613, example images on the right.

A.2. LLaMA texts used during evaluation

Prompt
What characteristics can be used to differentiate a *Common Tern* from other birds based on just a photo?
Produce an exhaustive list of all attributes or any information that can be used to identify the bird uniquely.
Texts should be of the form “a bird with <characteristic feature>”.
Ensure to structure your response as a list of single sentences.
Output
- A bird with a white forehead and a distinctive black stripe running across its eyes.
- A bird with a pale yellow bill that is longer than its head.
- A bird with a dark grey back and wings, and a light grey underside.
- A bird with a distinctive black stripe running along its tail feathers.
- A bird with a distinctive black band on its outermost tail feathers.
- A bird with a pale yellow patch on its forehead that is shaped like a crescent moon.
- A bird with a distinctive black stripe running along its neck.
- A bird with a dark grey crown and a light grey nape.
- A bird with a pale yellow patch on its upperwing that is shaped like a crescent moon.
- A bird with a distinctive black stripe running along its upperwing.
- A bird with a dark grey patch on its upperwing that is shaped like a triangle.
- A bird with a pale yellow patch on its lowerwing that is shaped like a crescent moon.
- A bird with a distinctive black stripe running along its lowerwing.
- A bird with a dark grey patch on its lowerwing that is shaped like a triangle.
- A bird with a distinctive black stripe running along its legs and feet.
- A bird with a pale yellow patch on its upperwing that is surrounded by a dark grey border.
- A bird with a distinctive black stripe running along its upperwing that is surrounded by a pale yellow border.
- A bird with a pale yellow patch on its lowerwing that is surrounded by a dark grey border.
- A bird with a distinctive black stripe running along its lowerwing that is surrounded by a pale yellow border.

Prompt
Where can we find a *Common Tern*?
Produce a list of habitat and geographic location information that can be used to identify the bird.
Each text produced should start with “a bird”.
Ensure to structure your response as a list of single sentences.
Output
- A bird of freshwater lakes and rivers, the Common Tern can also be found in inland areas with large bodies of water, such as reservoirs, ponds, and wetlands. In North America, they can be spotted in the Great Lakes region and in the central United States. In Europe, they can be found in the British Isles, France, and Germany. In Asia, they can be seen in China, Korea, and Japan. In Australia, they can be found in the southeastern region.
- A bird of the Arctic tundra, the Common Tern can be found in the northernmost parts of Europe, Asia, and North America, including Alaska and Canada. In these regions, they can be seen nesting on rocky cliffs and in coastal areas.

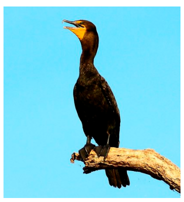
Table 6. Example of texts produced for category “Common Tern” of the CUB dataset using LLaMA2-7B.

B. LLMs for Fine-grained Image Captioning

We present captions generated using InstructBLIP, MiniGPT4 and recently released OpenAI's GPT4 Vision API. We specifically ask for concise single sentences in less than 70 words such that can be used to train CLIP using image-level supervision. This is in consideration with the fact that CLIP has a token length of 77. We use the following prompt:

Describe this image in fine-grained detail with as much information as possible in a single sentence less than 70 words.

Below we show examples of texts generated.

	MiniGPT4	InstructBLIP	GPT4
	This image shows a small bird perched on a tree branch. The bird has a brown and gray body with a white underside and a long, pointed tail. Its beak is black and its eyes are a bright yellow. The bird's wings are folded close to its body and its tail feathers are spread out to the sides.	The image depicts a small bird perched on a branch of a tree in the forest. The bird has a brown body with a reddish-brown head and beak, suggesting that it is a songbird. The bird is sitting on a branch near the top of the tree, which indicates that it is likely to be a passerine bird.	The image shows a diminutive bird with olive-brown plumage, subtle wing bars, and a delicate beak, perched on a bare twig, set against a diffused green backdrop suggesting a natural, leafy habitat.
	This image shows a large black bird perched on a branch with its beak open, as if it is singing. The bird has a yellow beak and black feathers, with a white underside. The bird's wings are spread out, and its tail feathers are spread out as well. The bird's body is elongated and slender, with a long neck and legs.	The image features a black bird sitting on a branch with its beak open. The bird is perched on a tree branch, and it appears to be enjoying the scenery around it. There are several birds in the background of the image, including a red-winged blackbird that can be seen near the top left corner of the image.	A black cormorant with a slender, hooked orange beak and glassy eyes is perched confidently on a gnarled, leafless tree branch, silhouetted against a clear, azure sky.
	This image shows a small bird perched on a rocky outcropping, with its beak open as if it is calling out. The bird has a black and white body with a red beak and legs. The background is a blurry green and brown landscape with some trees visible in the distance. The overall mood of the image is peaceful and serene.	The image features a black and white bird perched on a rock in the foreground. The bird is sitting on the edge of the rock, with its beak resting against the side of the rock. There is also a small amount of greenery surrounding the bird, suggesting that it is likely to be a wild bird.	A puffin-like seabird, exhibiting a contrasting plumage of black upperparts and white underparts, with a distinctive red bill, rests on the mossy and lichen-spattered edge of a rocky cliff, against a blurred backdrop of lush greenery, indicative of a coastal or island habitat.
	This image shows a small bird perched on a tree stump in a forest. The bird has brown feathers with a white underside and a black beak. Its eyes are bright and alert, and its wings are spread out as if it is ready to take flight. The forest is lush and green, with leaves and branches visible in the background.	The image depicts a small bird perched on a fallen tree trunk in the forest. The bird is standing upright, with its wings spread out and its beak slightly open. There are several leaves scattered around the area, which may indicate that the bird is enjoying the natural environment.	In the image, a brown-and-white thrush with distinctive spotting on its breast stands alert on the curled-over, weathered end of a fallen log, surrounded by a natural understory of brown leaf litter and sporadic green plants, with its sharp eye and beak visible, indicating a forest floor habitat.

MiniGPT4 and InstructBLIP are not able to produce a single long sentence, rather a set of sentences. We show all those sentences in the figure such that context length is less than 77. **GPT4 is much more accurate** than MiniGPT4 and InstructBLIP, but still makes some errors such as in the color of the bird in the first and last image. It identifies the color of the upperparts of the first bird - an Acadian Flycatcher - as olive brown whereas it is olive green. In the last bird it identifies the bird as brown-and-white whereas it is brown and yellow. It should be noted that our **category level descriptions capture the correct colors**. The descriptions are also very coarse descriptions of few body parts.

C. Training Hyperparameters

We list the training hyperparameters for all datasets. The temperature parameter is learned using learning rate (lr) 1E-2 and weight decay (wd) 1E-6 for all. The momentum parameter is set to 0.98 for all.

	CUB	FGVC Aircraft	Stanford Cars	Flowers102	NABirds	iNat	EuroSAT	Food101	ImageNet	CalTech101	DTD	Oxford Pets	Sun397	UCF101
lr proj	6E-07	4E-07	3E-07	7E-07	8E-07	1E-07	2E-06	1E-07	2E-06	1E-07	5E-07	8E-07	1E-06	8E-07
lr main	1E-07	1E-07	1E-07	1E-07	1E-07	5E-08	5E-07	5E-08	5E-07	5E-08	2E-07	2E-07	5E-07	2E-07
wd proj	1E-01	1E-06	1E-02	1E-02	1E-06	1E-03	1E-06	1E-06	1E-06	1E-03	1E-06	1E-04	1E-06	1E-06
wd main	1E-01	1E-06	1E-02	1E-03	1E-06	1E-03	1E-06	1E-06	1E-06	1E-03	1E-06	1E-04	1E-06	1E-06
temperature init	1.3	1.8	1.6	2	1	1	1.8	2	1.8	2	2	2	2	2

D. Texts and Per-task MAP on NeWT

Task	Negative Text	Positive Text	CLIP	Ours
inat_non_species_birds_near_signs	a photo of a bird which is not on a road sign	a photo of a bird which is on a road sign	95.75	97.94
inat_non_species_diseased_zebra_finch	a photo of a zebra finch bird which is healthy	a photo of a zebra finch bird which is diseased	61.10	61.20
inat_non_species_intersex_mallards	a photo of a mallard bird which is not intersex	a photo of a mallard bird which is intersex	48.12	47.72
inat_non_species_mating_aligator_lizard	a photo of a aligator lizard mating	a photo of a aligator lizard not mating	34.66	34.69
inat_non_species_mating_danaus_plexippus	a photo of danaus plexippus which are mating	a photo of danaus plexippus not mating	64.20	69.40
inat_non_species_mating_toxomerus_marginatus	a photo of a toxomerus marginatus mating	a photo of a toxomerus marginatus not mating	37.84	38.22
inat_non_species_white_american_robin	a photo of a white robin bird	a photo of an american robin bird	36.94	36.72
inat_observed_Allegheny_Mountain_Dusky_Salamander_vs_Dusky_Salamander	a photo of a Allegheny Mountain Dusky Salamander	a photo of a Dusky Salamander	46.30	50.10
inat_observed_Belize_Crocodile_vs_American_Crocodile	a photo of a Belize Crocodile	a photo of an American Crocodile	62.22	64.30
inat_observed_California_Sea_Lion_vs_Steller_Sea_Lion	a photo of a california sea lion	a photo of a stellar sea lion	58.22	63.28
inat_observed_Common_Grass_Yellow_vs_Three-spotted_Grass_Yellow	a photo of a Common Grass Yellow	a photo of a Three-spotted Grass Yellow	47.78	48.03
inat_observed_Eastern_Oyster_vs_Pacific_Oyster	a photo of a Eastern Oyster	a photo of a Pacific Oyster	63.34	72.70
inat_observed_Flea_Jumper_vs_Asiatic_Wall_Jumping_Spider	a photo of a Flea Jumper	a photo of a Asiatic Wall Jumping Spider	50.00	51.10
inat_observed_Jelly_Ear_vs_Ear_fungus	a photo of a Jelly Ear	a photo of a Ear fungus	58.16	59.22
inat_observed_Northern_Cinnabar_Polypore_vs_Cinnabar_Bracket	a photo of a Northern Cinnabar Polypore	a photo of a Cinnabar Bracket	49.53	52.72
inat_observed_Rough_Green_Snake_vs_Smooth_Greensnake	a photo of a Rough Green Snake	a photo of a Smooth Greensnake	52.97	57.34
inat_observed_Southern_Black_Widow_vs_Western_Black_Widow	a photo of a Southern Black Widow	a photo of a Western Black Widow	48.44	47.72
inat_observed_Western_Grey_Kangaroo_vs_Eastern_Grey_Kangaroo	a photo of a Western Grey Kangaroo	a photo of a Eastern Grey Kangaroo	60.70	62.56
inat_observed_southern_cattail_vs_lesser_reedmace	a photo of a southern cattail	a photo of a lesser reedmace	52.75	52.60
inat_unobserved_amanita_flavorubens_v_amanita_xanthocephala	a photo of a amanita flavorubens	a photo of a amanita xanthocephala	52.72	47.80
inat_unobserved_armillaria_luteobubalina_v_armillaria_novae-zelandiae	a photo of a Armillaria luteobubalina	a photo of a Armillaria novae-zelandiae	35.20	35.72
inat_unobserved_chloris_verticillata_v_chloris_cucullata	a photo of a chloris verticillata	a photo of a chloris cucullata	39.68	41.56
inat_unobserved_cladonia_squamosa_v_cladonia_portentosa	a photo of a Cladonia squamosa	a photo of a Cladonia portentosa	52.20	61.22
inat_unobserved_cuphea_aequipetala_v_cuphea_hyssopifolia	a photo of a Cuphea aequipetala	a photo of a Cuphea hyssopifolia	43.90	49.97
inat_unobserved_pinus_clausa_v_pinus_mugo	a photo of a Pinus clausa	a photo of a Pinus mugo	71.00	66.06
inat_unobserved_podarcis_virescens_v_podarcis_guadarramae	a photo of a Podarcis virescens	a photo of a Podarcis guadarramae	47.29	46.80
inat_unobserved_turdus_torquatus_v_turdus_atrogularis	a photo of a Turdus torquatus	a photo of a Turdus atrogularis	70.30	79.80
ml_age_black_bellied_plover	a photo of a black bellied plover bird which is not an adult	a photo of a black bellied plover bird which is an adult	45.96	41.12
ml_age_coopers_hawk	a photo of a cooper's hawk bird which is not an adult	a photo of a cooper's hawk bird which is an adult	53.47	68.90
ml_age_sanderling	a photo of a sanderling which is not an adult	a photo of a sanderling which is an adult	54.06	53.84
ml_bio_is_at_flower	a photo of a bird which is not at a flower	a photo of a bird which is at a flower	94.30	94.70
ml_bio_raptor_utility_pole	a photo of a raptor bird which is not on a utility pole	a photo of a raptor bird which is on a utility pole	92.50	94.05
ml_photo_rating_12_vs_45_v2	a photo with bad perceptual quality	a photo with good perceptual quality	83.90	80.44
ml_photo_rating_12_vs_45_v3	a photo with bad perceptual quality	a photo with good perceptual quality	82.30	78.94
ml_tag_back_of_camera	a photo not showing the back of a camera	a photo showing the back of a camera	32.30	34.06
ml_tag_copulation	a photo without copulation	a photo with copulation	51.50	49.10
ml_tag_egg	a photo without an egg	a photo with an egg	76.00	77.30
ml_tag_foraging_waterfowl	a photo of a waterfowl not foraging	a photo of a waterfowl foraging	58.03	65.00
ml_tag_in_hand	a photo of a bird which is not in hand	a photo of a bird which is in hand	96.20	97.00
ml_tag_nest	a photo without a nest	a photo with a nest	72.50	75.30
ml_tag_watermark	a photo without a watermark	a photo with a watermark	63.10	67.90
nabirds_species_classification_amecro_comrav	a photo of an American Crow	a photo of a Common Raven	70.25	71.06
nabirds_species_classification_bargol_comgol	a photo of a Barrow's Goldeneye	a photo of a Common Goldeneye	50.30	49.50
nabirds_species_classification_brwhaw_reshaw	a photo of a Broad-winged Hawk bird	a photo of a Red-shouldered Hawk bird	91.06	91.70
nabirds_species_classification_casvir_plsvir	a photo of a Cassin's Vireo	a photo of a Plumbeous Vireo	82.75	83.50
nabirds_species_classification_coohaw_shshaw	a photo of a Cooper's Hawk	a photo of a Sharp-shinned Hawk	66.00	66.90
nabirds_species_classification_easmea_wesmea	a photo of a eastern meadowlark	a photo of a western meadowlark	72.40	72.25
nabirds_species_classification_semsan_wessan	a photo of a Semipalmated Sandpiper	a photo of a Western Sandpiper	62.03	63.90
nabirds_species_classification_sursco_whwsco2	a photo of a Surf Scoter	a photo of a White-winged Scoter	49.97	51.72
nabirds_species_classification_truswa_tunswa	a photo of a trumpeter Swan	a photo of a tundra Swan	70.10	70.50
Average			**60.25**	**61.90**

E. Alternate Training Strategies : Details

As mentioned in the main paper, our first strategy involves masking texts using ground truth visibility annotations, which are available for the CUB dataset. We manually create a dictionary of mapping of parts to words that might be used to describe the part in the GPT generated texts. For example, `["leg", "foot", "feet"]` can be used to describe the legs of the bird in the GPT generated texts. Based on this, we mask those sentences for a given image which describe parts for which the gt annotations indicate that the part is not visible. Although this strategy offers improvement (Tab. 7 row 2), obtaining these human labelled captions is resource intensive, which is why training using LLM generated texts is useful.

In the second strategy, we assume that CLIP's similarity scores are good indication of visibility. The first way is to pair texts above a threshold of similarity for each image (Tab. 7 row 3). We take this threshold as 0.5 probability after taking softmax over texts. The second way is max pooling at instance level (Tab. 7 row 4) over image instances and text instances involves generating similarity scores for a small batch of images and texts of the same category to find the pair that has the max similarity and using only that pair to backprop.

For FixMatch (Tab. 7 row 5) we generate psuedo soft labels using weaker augmentations and thresholding them with a probability value of 0.5. We finally use the loss we described in § 3.2 combined with the cross entropy loss between logits and pseudo labels (with a ratio of 1:0.33). For knowledge distillation (Tab. 7 row 6), we use the § 3.2 loss combined with the KL divergence (with a ratio 1:4) between the logits of a teacher network scaled by a temperature term (of value 3) and the student logits. For both these methods we initialize every network with pre-trained CLIP.

Method	Accuracy
Ours - CLIP^{FT} + A	54.23
Ground Truth Visibility Masks	**54.47**
Pairing texts with similarity above a treshold	53.99
Max-pooling at image and text instance level	54.10
FixMatch	54.19
Knowledge distillation	54.38

Table 7. **Performance of alternate training strategies trained and tested on CUB dataset.** We use the same train and test class splits as described in the main paper, so as to **evaluate on unseen classes**. Using GT visibility annotations to mask texts offers improvement. Knowledge distillation offers very little improvement. All others have worse accuracy.

Training Dataset	CUB	FGVC Aircraft	Stanford Cars	Flowers102	NABirds	iNat		EuroSAT	Food101	ImageNet	CalTech101	DTD	Oxford Pets	Sun397	UCF101
Testing Dataset					CUB	CUB	Flowers								
B/32 CLIP	50.54	29.27	69.72	71.78	50.54	50.54	71.78	68.89	89.94	62.61	93.87	58.45	96.64	73.03	71.12
B/32 CLIP + A	50.71	30.35	69.47	75.37	50.71	50.71	75.37	72.07	90.06	65.11	94.54	59.90	96.53	75.97	75.71
B/32 CLIP^{FT} + A	**53.34**	**36.41**	**71.63**	**77.05**	**55.29**	**54.58**	**77.05**	**78.56**	**93.71**	**65.98**	**95.75**	**62.20**	**96.88**	**77.75**	**75.99**
B/16 CLIP	51.91	36.47	74.94	77.05	51.91	51.91	77.05	64.05	92.49	67.41	93.89	60.26	97.04	75.49	77.45
B/16 CLIP + A	53.58	36.47	73.83	80.84	53.58	53.58	80.84	71.51	93.72	69.74	94.87	64.13	96.81	78.88	80.47
B/16 CLIP-A-Self	-	33.00	72.90	75.30	-	-	75.30	70.50	91.20	68.30	95.90	62.30	97.00	76.80	76.40
B/16 CLIP^{FT} + A	**55.63**	**40.75**	**75.78**	**81.26**	**56.76**	**56.77**	**81.20**	**81.82**	**95.08**	**71.87**	**96.03**	**65.21**	**97.21**	**80.32**	**80.69**

Table 8. **Results on 14 datasets.** We show performance on ViT B/32 and B/16 with 1:1 train/test class split, obtaining superior accuracy across all settings.

Method	Accuracy
CLIP	71.18
CLIP + A	72.54
CLIP^{FT} + A	**74.80**
CLIP-A-Self	71.30

Table 9. **Performance on CUB with 3:1 train/test split.** Following CLIP-A-Self CUB split we show that we outperform significantly.

	CUB	FGVC Aircraft	Stanford Cars
Multiply Probs	53.10	36.35	71.58
Average Probs	53.34	36.41	71.63

Table 10. **Averaging vs Multiplying Probabilities for aggregating.** Averaging works better than multiplying per-text probabilities for each class for final classification accuracy.

F. Improvement over CLIP-A-Self [19]

CLIP-A-self only tunes a adapter network and lags in performance compared to ours (CLIP^{FT} + A). This again points to the need for fine-tuning CLIP encoders to recognize fine-grained attributes. Another reason why CLIP-A-self suffers is because they query GPT to produce descriptive texts for a fixed set of pre-defined attributes. It should also be noted that CLIP-A-self performs worse than CLIP + A in most cases, which shows that even though it does not tune the CLIP encoders, it still overfits to the training classes. We provide results on the all 14 datasets in Tab. 8. For B/16 we compare with CLIP-A-Self in a 16-shot setting for 1:1 train/test split. In Tab. 9 we also compare with 3:1 train/test split for CUB dataset.

G. Correctness of LLM generated texts

We select 4-6 classes from each of CUB, Stanford Cars and FGVC Aircraft test classes and use all the sentences (about 20 per class) produced by GPT for manual evaluation of correctness. For each sentence participants mark whether an attribute is correct or incorrect. Participants mark some sentences as unsure ($\sim$ 2-4 %). Correctness is determined through various sources across the web such as Wikipedia and All About Birds, by searching the characteristics of the category on Google. Among the sentences marked correct or incorrect over all classes, 96% (CUB), 90% (FGVC Aircraft) and 96% (Stanford Cars) are marked as correct. In FGVC Aircraft where correctness is lower compared to the other two, the incorrect sentences mainly contain measurements in metric units which are usually not easily identifiable visually. For example, the sentence "An aircraft with a range of approximately 3,600 kilometers." is marked as incorrect as Wikipedia states the number as 3100 kilometers later reduced to 2770 kilometers. We list the sentences for an example class of each dataset with their marked correctness in the following tables.

Category Name	Sentences generated using GPT4	Correctness
Downy Woodpecker	A bird with a small size, typically measuring between 5.5 to 6.7 inches in length.	1
	A bird with a black and white color pattern.	1
	A bird with a white underbelly and lower parts.	1
	A bird with black wings that have white spots.	1
	A bird with a black tail that has white outer feathers.	1
	A bird with a white back.	1
	A bird with a black head and neck.	0
	A bird with a white stripe above and below the eyes.	1
	A bird with a small, pointed beak that is perfect for pecking at wood.	1
	A bird with a red patch at the back of the head, but only in males.	1
	A bird with a black bill that is shorter than its head.	1
	A bird with a straight, chisel-like bill.	1
	A bird with a large head compared to its body.	1
	A bird with a short, stiff tail that provides support against tree trunks.	not sure
	A bird with a fluttering flight pattern that alternates between flapping and folding its wings.	1
	A bird that is often found in deciduous forests, orchards, parks, and suburban areas.	1
	A bird that is commonly seen alone or in pairs.	1
	A bird that is often seen on tree trunks or branches, especially those of deciduous trees.	1
	A bird that is frequently seen feeding on insects, seeds, and berries.	1
	A bird that is native to North America, particularly the United States and Canada.	1
	A bird that is often seen pecking at tree bark in a vertical position.	1
	A bird that is known for its drumming sound on tree trunks, which is a form of communication.	1
	A bird that is often seen in the lower parts of trees or shrubs, unlike other woodpeckers.	1
	A bird that is smaller and more delicate than the similar-looking Hairy Woodpecker.	1
	A bird that is known for its ability to adapt to human-altered habitats, such as orchards and residential areas.	1

Table 11. Correctness study of the category “Downy Woodpecker” of the CalTech-UCSD Birds dataset

Category Name	Sentences generated using GPT4	Correctness
Spitfire	An aircraft with an elliptical wing shape.	1
	An aircraft with a single propeller at the front.	1
	An aircraft with a long, pointed nose.	1
	An aircraft with a large, bubble-shaped cockpit canopy.	1
	An aircraft with a tail wheel landing gear.	1
	An aircraft with two exhaust stacks on each side of the engine cowling.	1
	An aircraft with a distinctive, rounded vertical stabilizer.	1
	An aircraft with a Rolls-Royce Merlin or Griffon engine, identifiable by the specific arrangement of exhausts.	1
	An aircraft with a thin, streamlined fuselage.	1
	An aircraft with a relatively small horizontal stabilizer compared to the size of the wings.	1
	An aircraft with a four-bladed propeller, especially in later models.	1
	An aircraft with a radiator under each wing in a characteristic rectangular shape.	1
	An aircraft with a retractable landing gear.	1
	An aircraft with a relatively short wingspan compared to its length.	0
	An aircraft with a distinctive ’kink’ in the leading edge of the wing, near the wingtip.	1
	An aircraft with a large spinner covering the hub of the propeller.	1
	An aircraft with a relatively high-set cockpit, giving the pilot a good view.	1
	An aircraft with a narrow track undercarriage.	1
	An aircraft with a single pilot seat.	1
	An aircraft with a distinctive, rounded tailplane.	1

Table 12. Correctness study of the category “Spitfire” of the FGVC Aircraft dataset

Category Name	Sentences generated using GPT4	Correctness
Rolls-Royce Ghost Sedan 2012	A car with a large, rectangular front grille with vertical slats.	1
	A car with a Spirit of Ecstasy hood ornament.	1
	A car with a long, sleek, and luxurious body design.	1
	A car with a high beltline and a low roofline.	1
	A car with a pair of round LED headlights on each side of the grille.	0
	A car with a three-box design, meaning separate compartments for the engine, passenger, and cargo.	1
	A car with a large, prominent Rolls-Royce logo on the center of the wheel hubs.	1
	A car with a pair of exhaust pipes located at the corners of the rear bumper.	1
	A car with a rear-hinged back door, also known as 'suicide doors'.	1
	A car with a large, flat hood and a short front overhang.	1
	A car with a long rear overhang, giving it a classic limousine look.	1
	A car with a two-tone paint job, often with the hood, roof, and trunk in a different color than the rest of the body.	1
	A car with a large, luxurious, and spacious interior visible through the windows.	1
	A car with a large, flat trunk lid.	1
	A car with a distinctive Rolls-Royce clock on the dashboard visible through the windshield.	1
	A car with a large, round fuel cap on the right rear side.	0
	A car with a distinctive Rolls-Royce treadplate on the door sill.	1
	A car with a large, rectangular rear window.	1
	A car with a small, triangular window at the rear of the side windows.	1
	A car with a distinctive Rolls-Royce umbrella stored in the rear door.	1

Table 13. Correctness study of the category "Rolls-Royce Ghost Sedan 2012" of the Stanford Cars dataset