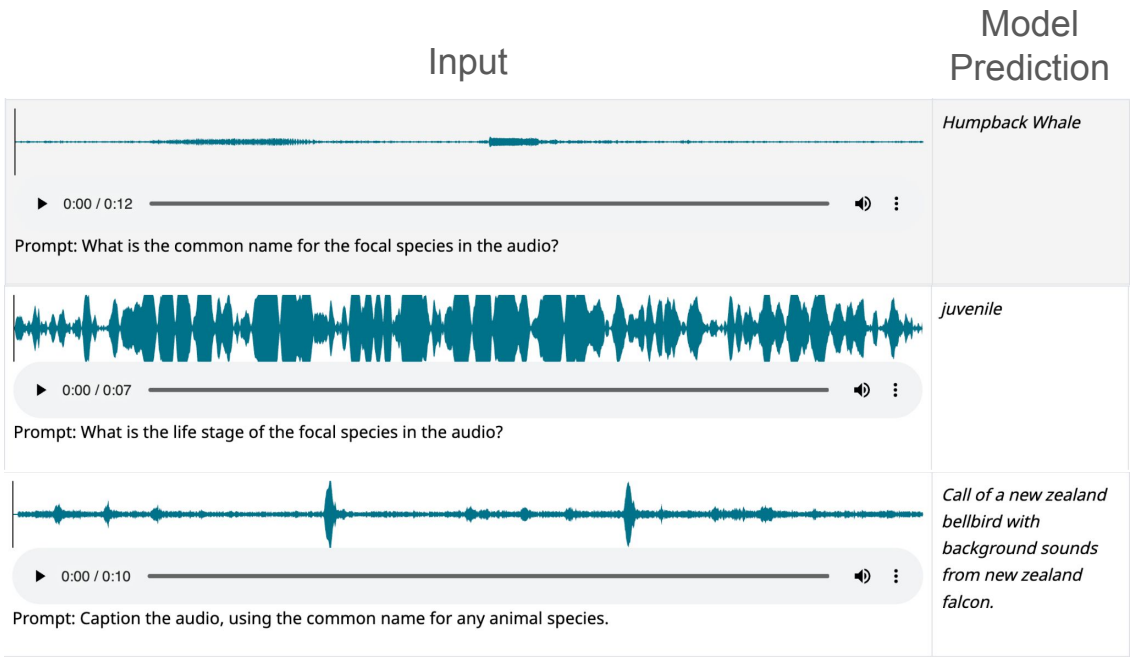
# Model Merging Improves Zero-Shot Generalization in Bioacoustic Foundation Models

D. Marincione*, D. Crisostomi, R. Dessì, E. Rodolà, **E. Rossi***

ESP Community Deep Dive, NeurIPS 2025 Workshop

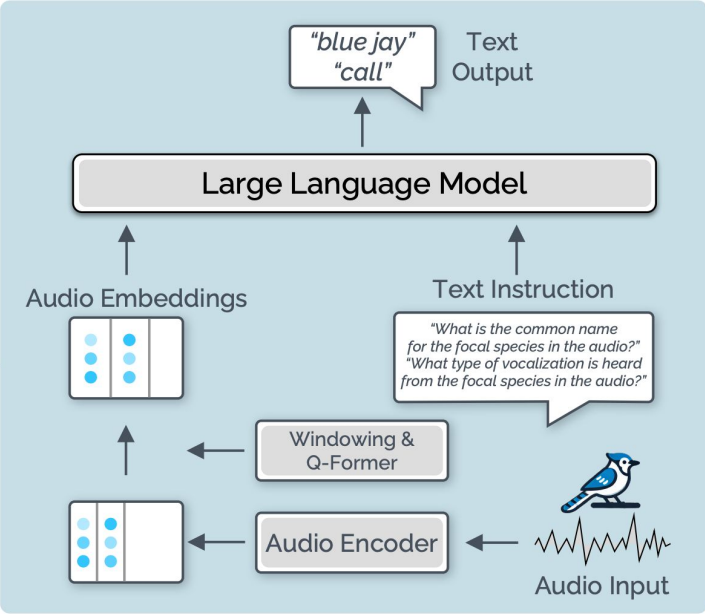
# Background

# NatureLM: the first bioacoustics audio-language model



# NatureLM: technical details

- Audio encoder
- LLM: Llama-3.1-8B-instruct (fine-tuned with LoRa)



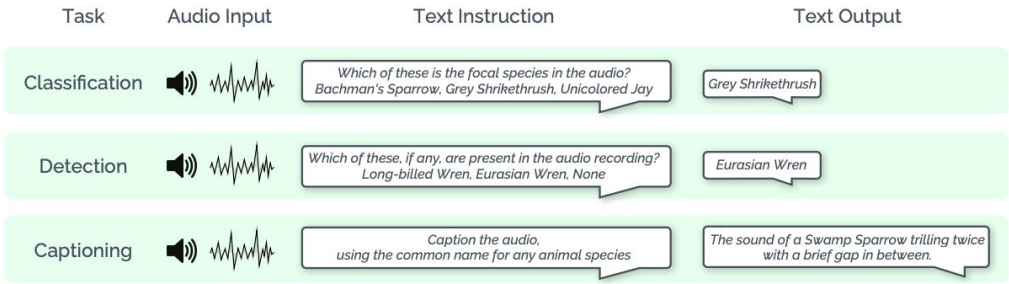




Figure 2: Examples of training instances

# Problem Analysis

# NatureLM is not robust to prompt variations

- Even small deviations from training prompts lead to large drops in accuracy

**Common Name Prompt**

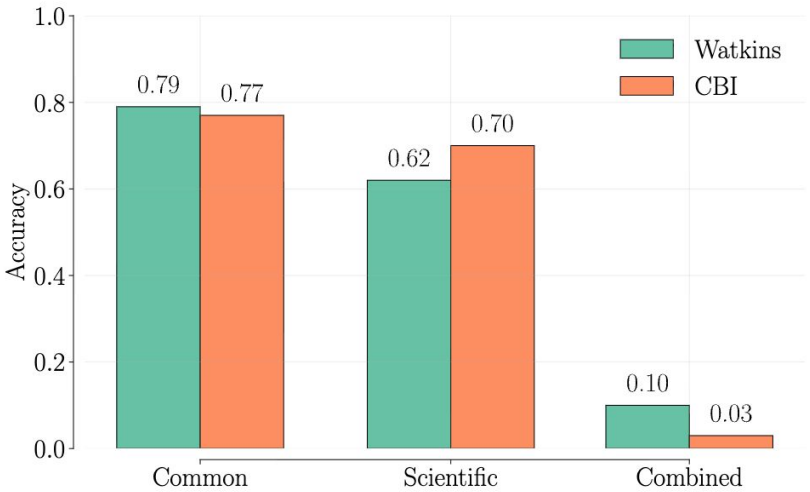
```
What is the common name for the focal species
in the audio?
```

**Scientific Name Prompt**

```
What is the scientific name for the focal
species in the audio?
```

**Combined Prompt**

```
Identify the focal species in the audio and
provide its scientific name, followed by a
colon and its common name.
```



# NatureLM is not able to follow instructions

**id 4186fc55-0ac2-4f44-9026-009f239fcf96**

GT: Stenella clymene: Clymene Dolphin
Common: Clymene Dolphin
Scientific: Stenella clymene
Combined: Stenella clymene: Clymene Dolphin

**id 4cbdd583-23c6-408d-aea5-3719dc6d9654**

GT: Lagenodelphis hosei: Frasers Dolphin
Common: Fraser's Dolphin
Scientific: Lagenodelphis hosei
Combined: Lagenodelphis hosei

**id 185c974a-d905-475c-8fc0-0cbab1e383b9**

GT: Eubalaena australis: Southern Right Whale
Common: Fin- Finback Whale
Scientific: Balaenoptera physalus
Combined: Balaenoptera physalus: 52 Hz Pulses

**id aecd4535-ebef-438d-811b-1ffb7be5c22e**

GT: Odobenus rosmarus: Walrus
Common: Walrus
Scientific: Odobenus rosmarus
Combined: Odobenus rosmarus - male courtship ...

- NatureLM-Audio has lost its instruction following capabilities in favor of task-specific ones acquired during fine-tuning

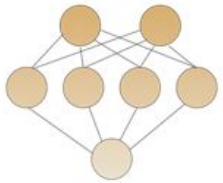
# Method and Results

# What do we need?

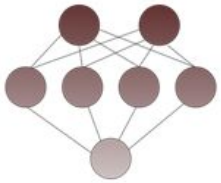
- Improving the instruction following capabilities, while retaining the bioacoustics expertise
- Avoiding the costly process of retraining the model

# Model Merging

Researcher: solving narrow problems

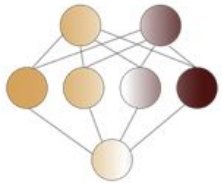


Navigator: navigating complex systems



MERGE

University Professor



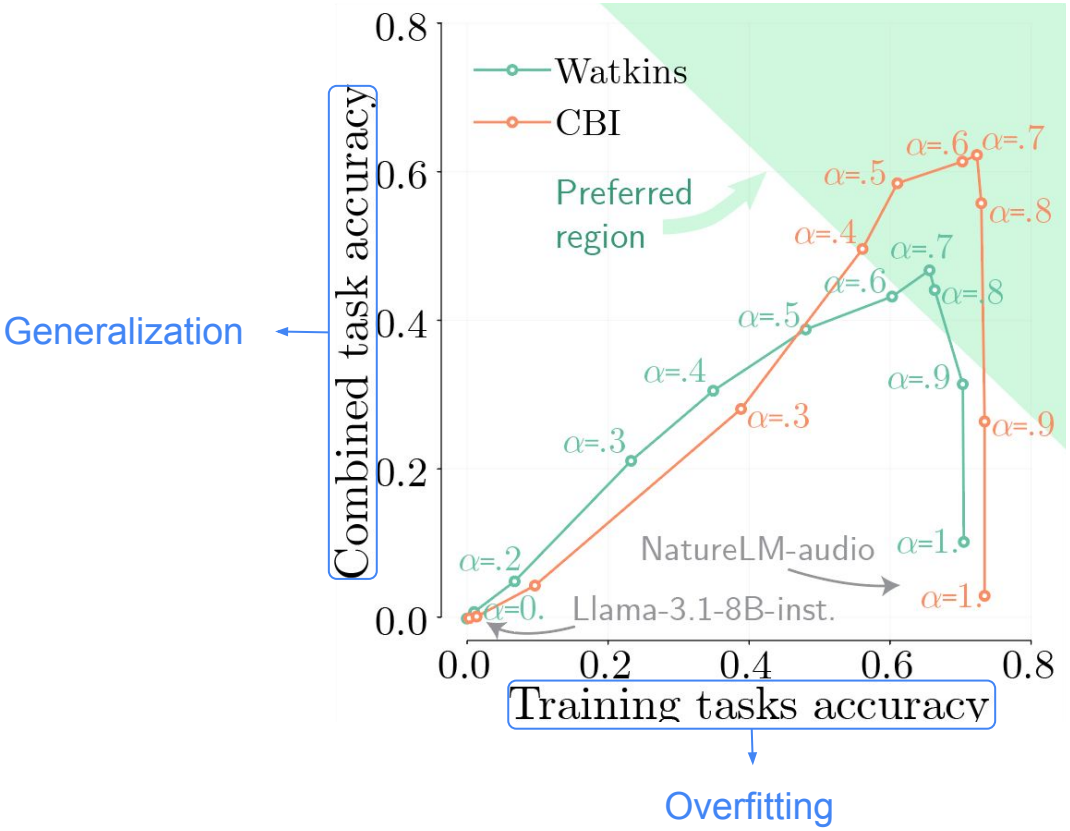
# Model Merging with the base model

- We merge the fine-tuned LM with its base model through linear interpolation

LoRa update

$$(1-\alpha)\,\mathbf{W}_{\text{base}} + \alpha\,\mathbf{W}_{\text{ft}} = (1-\alpha)\,\mathbf{W}_{\text{base}} + \alpha\,(\mathbf{W}_{\text{base}} + \mathbf{AB})$$
$$= \mathbf{W}_{\text{base}} - \cancel{\alpha\,\mathbf{W}_{\text{base}}} + \cancel{\alpha\,\mathbf{W}_{\text{base}}} + \alpha\,\mathbf{AB} = \mathbf{W}_{\text{base}} + \alpha\,\mathbf{AB}.$$

# Model Merging restores the lost capabilities

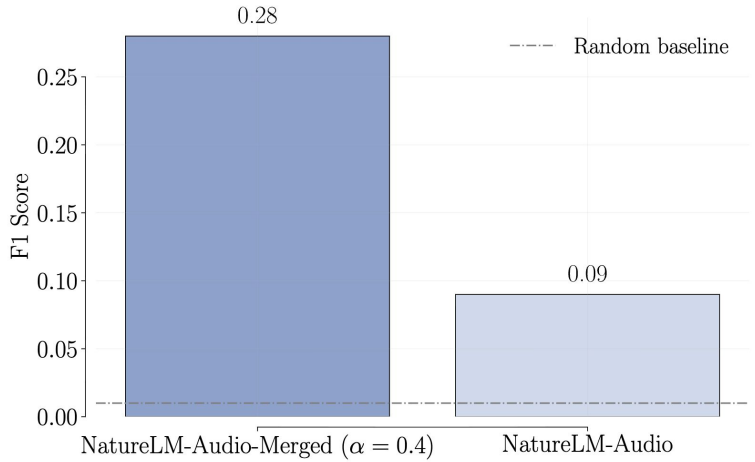


# New SOTA on Unseen Species Classification

- Zero-shot closed-set classification of species never seen during training

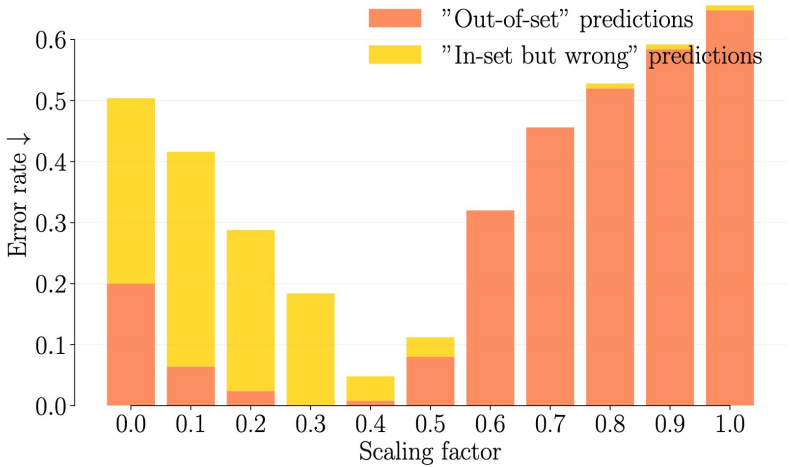
**Closed-Set Classification Prompt**

```
What is the common name for the focal species in the audio?  Output exactly one of:  {species_list}
```



# What is actually going on? Simplified Binary Setup

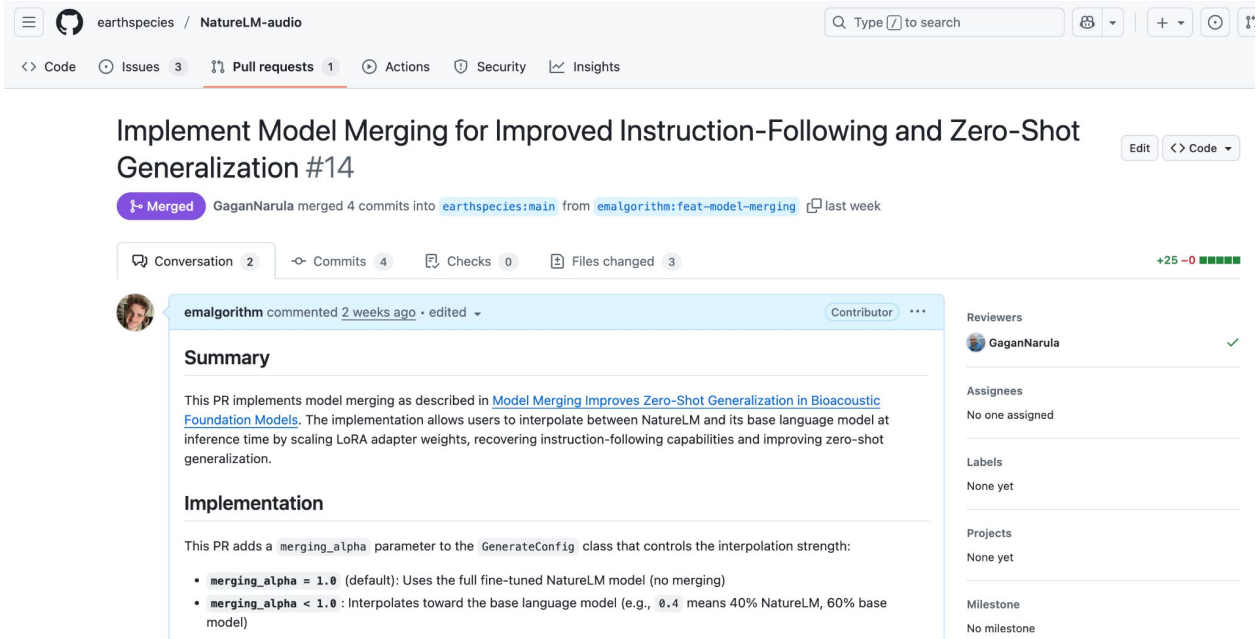
- NatureLM (α=1) often predicts labels outside the provided set
  - causes worse-than-random performance




- Reducing α mitigates this effect, with α=0.4 offering the best trade-off

# Already Merged into the Main Nature-LM Repo

- Original authors (ESP) found it super useful and it has already been integrated



# Thank you!