Final Project Proposal

Project Title: Cross-Species Translation
Team Members: Ian Gifford, Deborah Sinishaw, Sara Yang
GitHub Repository: MAIS202_Final_Project
Assigned TPM: Emma

1. Introduction and Project Overview

Animal communication is a complex system that humans have yet to fully understand. While some species exhibit structured vocalization patterns, interpreting their meanings remains a challenge. This project aims to bridge that gap by developing a **Cross-Species Translation Model**, which will analyze animal vocalizations and generate human-readable interpretations.

2. Choice of Dataset

I. Dataset Source: https://github.com/earthspecies/library/tree/main
II. Why This Dataset:
- Large (mostly labeled) dataset of multiple animals (dogs, orcas, macaques, giant otters, etc.)
- One notable one is the dog dataset of 693 recordings of dog barks in 3 specific contexts

3. Methodology and Approach

I. Data Preprocessing
   A. Normalize & Noise Reduction: For raw data, we need to normalize all recordings to a consistent format and then filter out background noise. The Earth Species dataset already does this, but if we find other datasets, we will need to do this.
   B. Feature Extraction
      1. Spectrograms: Visual representation of frequencies over time.
      2. Chroma Features: Represent pitch classes (like musical notes). Useful if animal calls have tonal variations (e.g., bird songs, whale songs).
      3. Mel-Frequency Cepstral Coefficients (MFCCs): Extracts patterns of vocalization (e.g., barking intensity).
   C. Split training and testing data
II. Machine learning model

A. Given labeled animal vocalizations (labled by the context/rough English translation), our goal is to be able to have the model listen and classify other animal sounds.
B. CNN (Convolution Neural Network)
    1. Used to identify patterns in spatial data (spectrograms).
C. RNN (Recurrent Neural Network)
    1. Used to identify patterns in sequential data (audio file, sentences).
D. CNN seems more applicable with what we are working with. RNN is used more for machine translation and language classification. Since animal sounds do not have transcriptions or any known structure, it would be better to use a CNN to find patterns in short audiofiles represented by spectrograms.

III. Evaluation Metric
A. Classification
-Confusion matrix: Gives a breakdown of which species/contexts are correctly and incorrectly categorized.
-Accuracy: We will use this if our dataset has a similar amount of data in each class (balanced dataset).
-Precision: We will use this in addition to accuracy, especially if the dataset is unbalanced.
B. Translation
-BLEU score: Standard for evaluating machine translation (e.g., English-to-French translation). It compares the predicted text to a human reference.
-Perplexity: Measures if the model produces understandable and accurate translations.

4. **Application**

I. Overview: Our cross-species translation application will enable users to detect the animal species from audio input and interpret the vocalization's meaning in human-readable English.
II. Input: Record live audio using a microphone or upload an audio file
III. Output: The type of animal and context. This will probably be in an app or website.