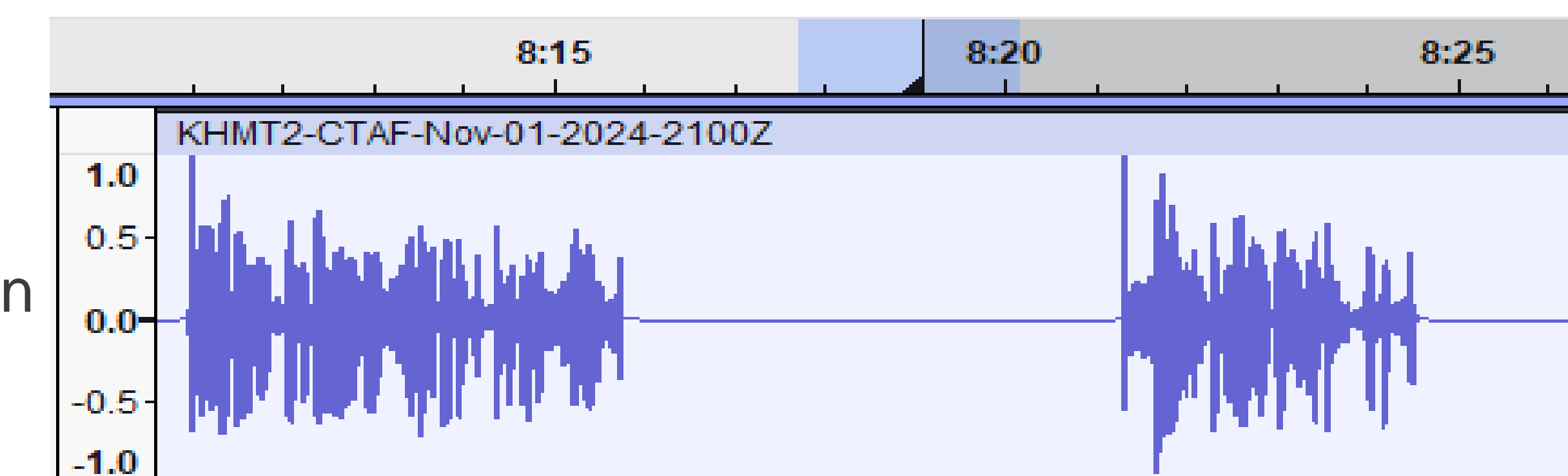


AI Speech to Text for Military Communications

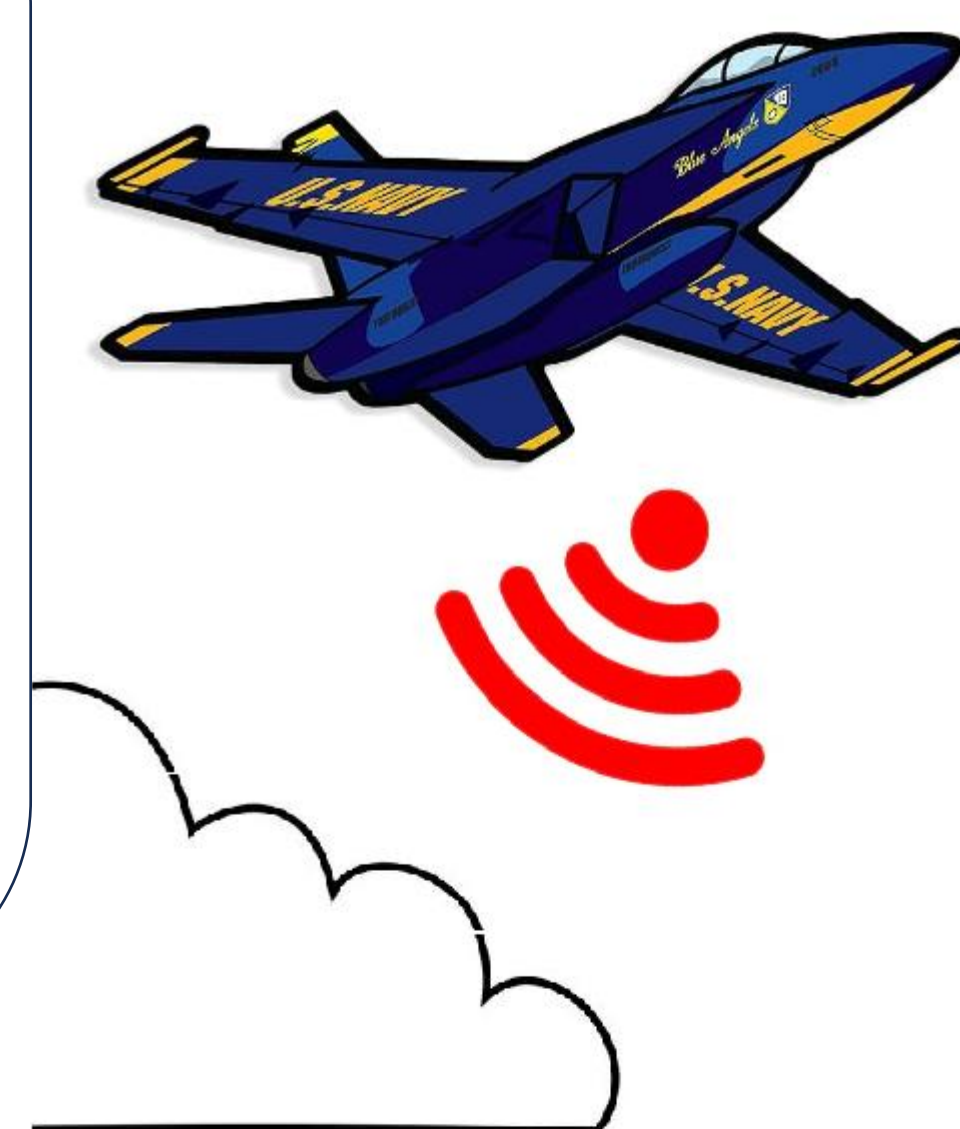
Team members: Nathan Devore, Nate Eldering, Connor Kohout, Allen Lee | **Faculty adviser:** Tamer Nadeem, Ph.D. | **Sponsor:** DoD | **Mentor:** Clinton



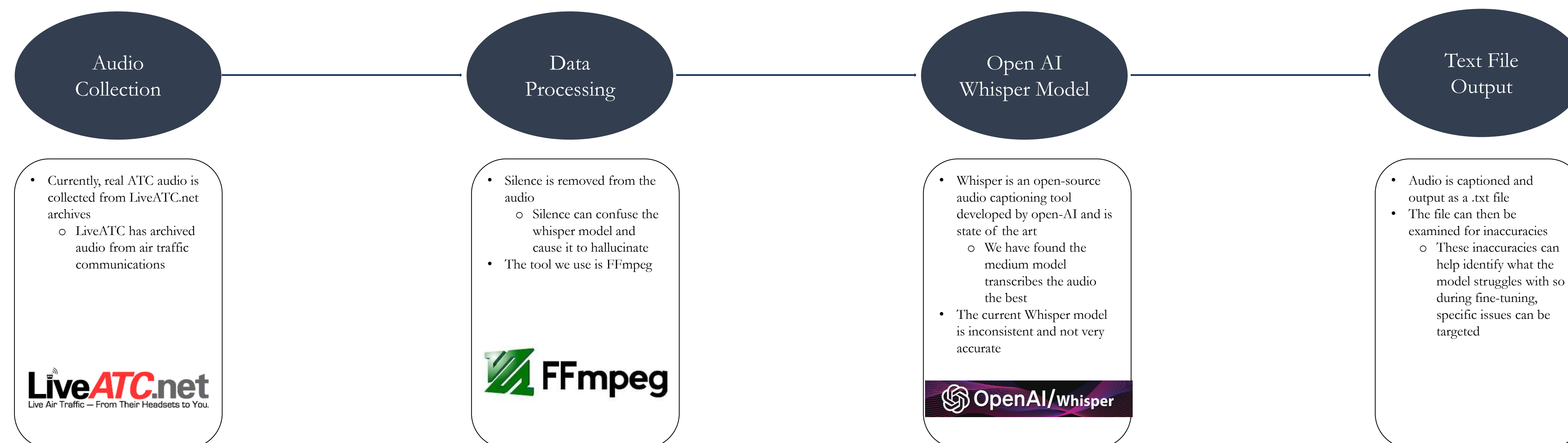
Problem Statement

Military operations occur in noisy environments and involve specific terminology that current speech-to-text models cannot handle effectively. Air communications, in particular, are hindered by wind, radio static, and engine noise, degrading transcription quality.

Initial testing with real air traffic communications shows the Whisper model performs inconsistently in these environments. Our project addresses this challenge by fine-tuning OpenAI's Whisper model, a robust speech-to-text system, to improve transcription accuracy in noisy, mission-critical settings.



Current Design Flow



What is Whisper?

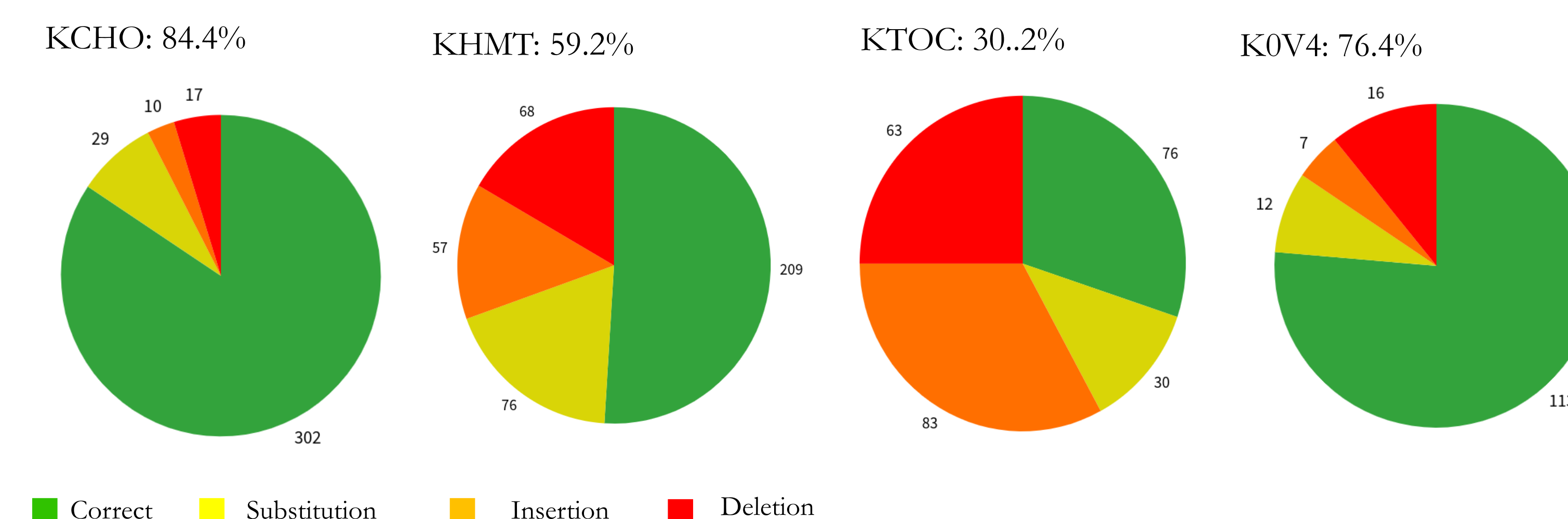
Whisper is an open-source, general-purpose speech recognition model developed by OpenAI, introduced in the December 2022 paper "Robust Speech Recognition via Large-Scale Weak Supervision." Trained on a diverse dataset of multilingual and multi-accented audio, Whisper offers robust performance across various scenarios, excelling at tasks like transcription, translation, and language identification.

It comes in six model sizes, four of which are English-only, each offering a tradeoff between speed and accuracy. This flexibility allows users to select models suited to specific needs, from real-time, lightweight processing to high-accuracy transcription. Whisper's versatility and robust performance make it a powerful tool for a range of speech and language applications.



Test Results

To establish a baseline for transcription accuracy, we manually transcribed ATC audio and compared the results to outputs from the Whisper model. Because manually labeling data is so time consuming, we are currently researching better methods. The dataset comprises four audio files of varying clarity, sourced from LiveATC.net archives. The audio was collected from the airports KCHO (Charlottesville Albemarle Airport, Virginia), KHMT (Hemet-Ryan Airport, California), K0V4 (Brookneal-Campbell County Airport, Virginia), and KTOC (Toccoa Airport, Georgia). This provides a robust foundation for evaluating the Whisper model.



Fine-Tuning

Whisper's baseline performance can be enhanced through targeted fine-tuning using aviation-specific data. By having pilots read scripted communications during flights and collecting audio through LiveATC archives, we ensure the training data matches real-world radio conditions. This precisely labeled dataset will help the model filter out cockpit and radio interference while improving speech detection in noisy environments. The fine-tuning process will also enhance recognition of proper nouns and the phonetic alphabet.

Fine-Tune Targets

In its current state, Whisper struggles with identifying proper nouns and the phonetic alphabet, significantly hurting its accuracy. For example, it was unable to recognize "Hemet" (the name of the airport), incorrectly labeling every instance as "have it." When we fine-tune the model, we will target proper noun recognition and the phonetic alphabet; this includes airport names, company names, aircraft manufacturers, and types.

Ideal Design Flow

