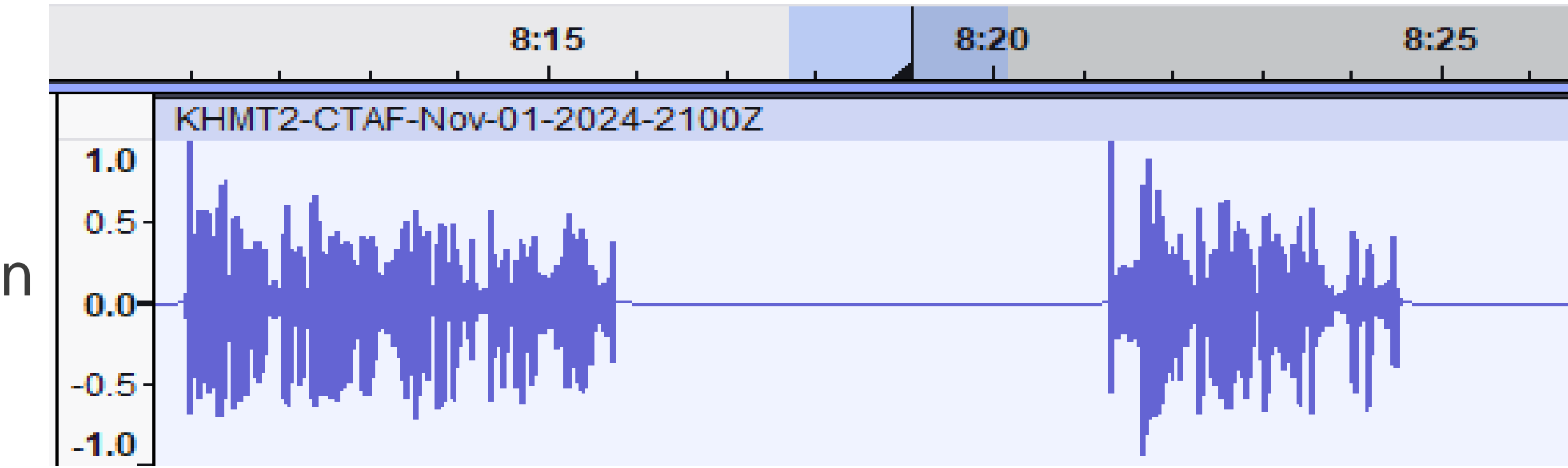


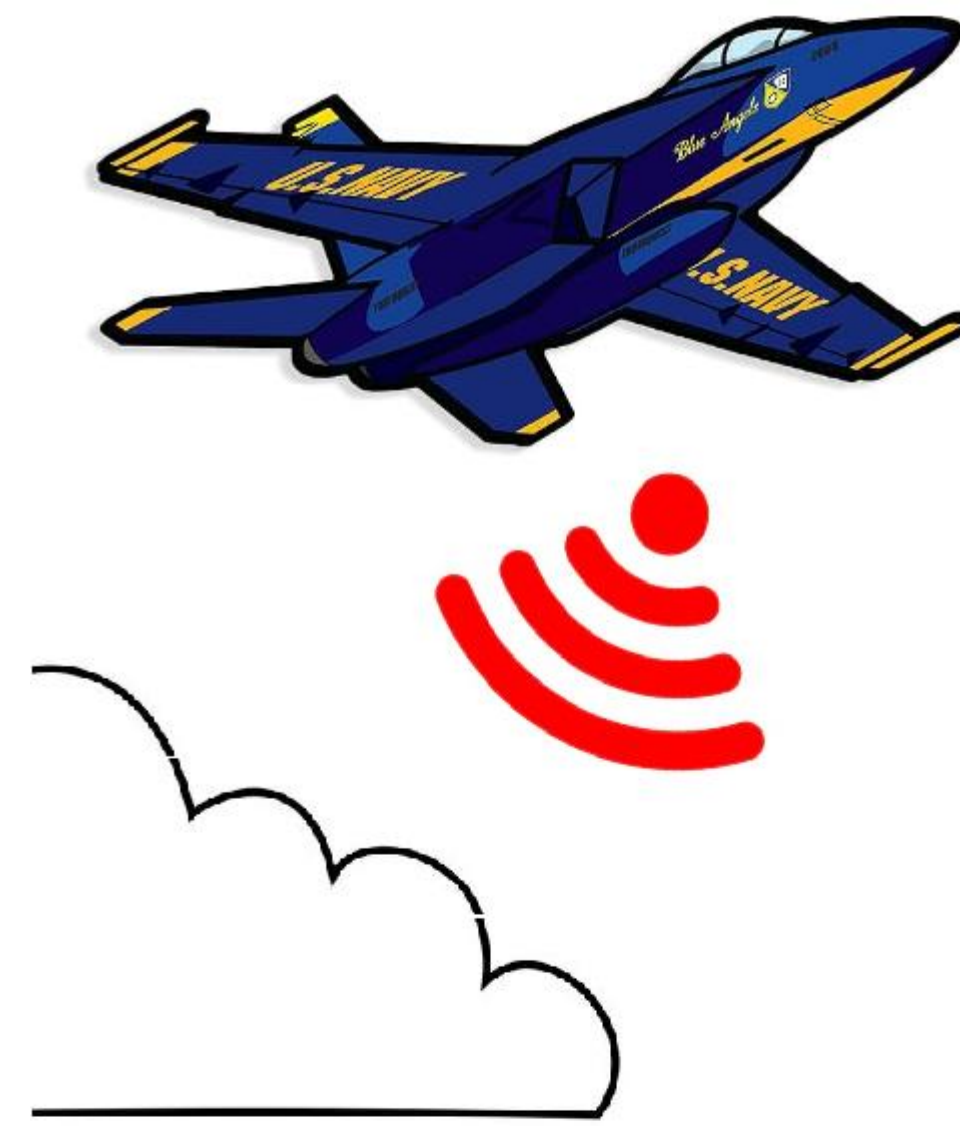
# AI Speech to Text for Military Communications

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

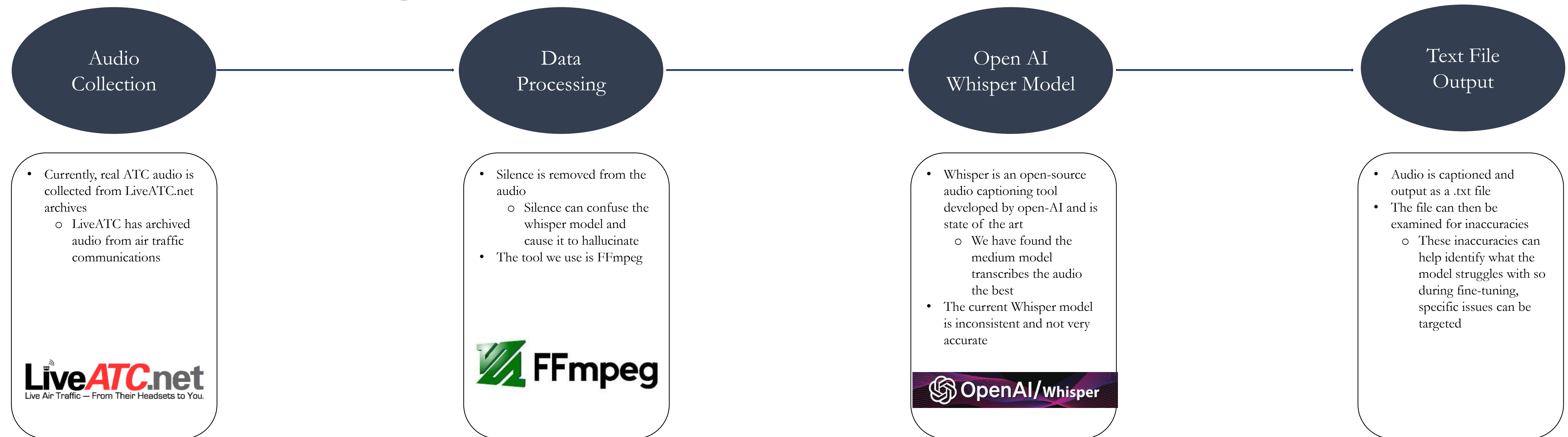


## Problem statement

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. Our project addresses this challenge by fine-tuning OpenAI's Whisper model, a robust speech-to-text system, for non-civilian environments. Initial testing with real air traffic communications shows the Whisper model performs inconsistently, struggling with proper nouns, and omitting sentences under high noise conditions. Our fine-tuning aims to overcome these issues, improving transcription accuracy in noisy, mission-critical settings.



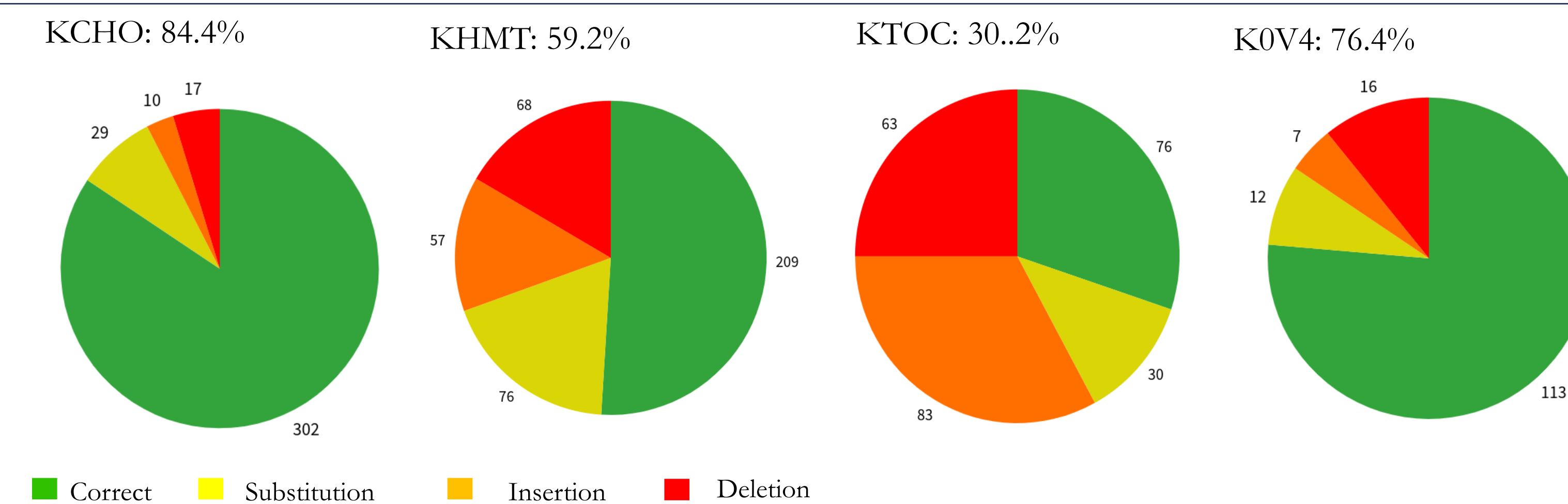
## Current Design Flow



## Test Results



To establish a baseline for transcription accuracy, we manually transcribed ATC audio and compared the results to outputs from the Whisper model. The dataset comprises four audio files of varying clarity, sourced from LiveATC.net archives. These recordings represent diverse real-world scenarios from multiple pilots, aircraft, and radio equipment. 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 the 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. Additionally, the model may learn to leverage contextual audio signatures to aid in aircraft identification and improve transcription accuracy.

## Fine-Tune Targets

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

## Ideal Design Flow

