BPD BODYCAM TIMESTAMPS PROJECT

Sai Krishna Sashank Madipally

Krishna Adithya V

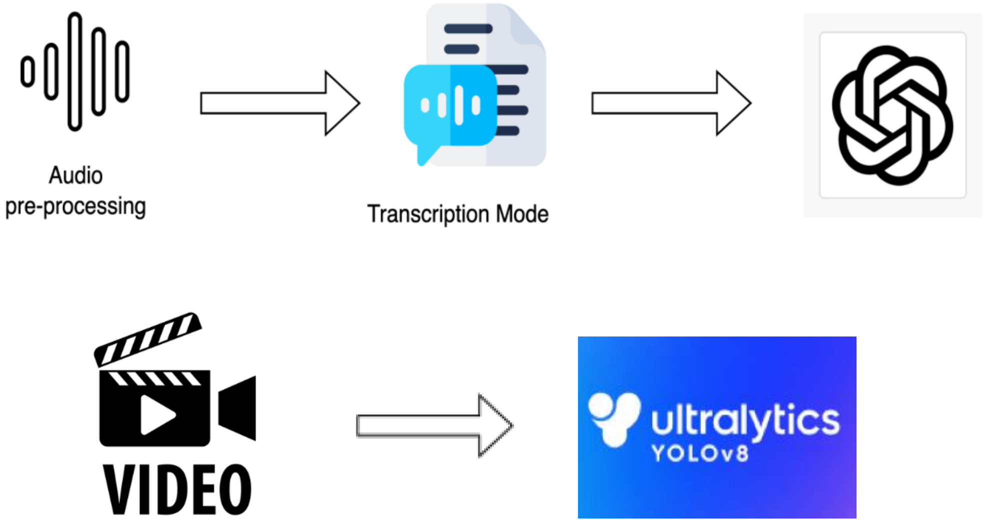
Aakash Bhatnagar

Problem Statement

- In this project, we analyzed police body camera footage, a very time consuming and labor intensive task if performed manually.
- Our client (Carmen Guhn-Knight representing the Law Offices of Howard Friedman) needed a more efficient way to identify key moments in long videos where officers make aggressive comments, complain about lack of planning, or fail to direct protesters.
- The desired outcomes were:
  - An ASR model that can transcribe the audio in all videos to text transcripts.
  - Models that can analyze the text to detect/timestamp three different incidents:
    - Instances of the police discussing the lack of a proper plan
    - Instances where the police failed to offer directions
    - Instances where the police directs unnecessarily aggressive and offensive comments towards the protestors
    - Instances where the police forcefully used batons

PROPOSED SOLUTION

Pipeline



Audio
pre-processing

Transcription Mode

Step 1 – Audio Transcription: OpenAI Whisper
Step 2 – LLMs: OpenAI GPT
Step 3 – Object Detection: Ultralytics YOLOv8

YOLOv8

- YOLOv8, or You Only Look Once version 8, is an Object Detection model.
- To fine-tune YOLOv8, one typically starts with a pre-trained model on a large dataset like COCO and then continues training on a smaller dataset that is more relevant to the target application.
- For YOLOv8 fine-tuning, we annotated 65 images using Roboflow. 60 images were used for training and the remaining 5 images for validation.
- Objects Being Detected: Police, Protestors, and Batons
- Final Metrics:
  - **Box Loss**: 0.5848
  - **Class Loss**: 0.5517
  - **mAP50**: 0.823

Whisper

- Version Experimented
  - Tiny
  - Base
  - Small
  - Medium
  - Large
- Findings:
  - Large performed best but is very time-consuming and needs a GPU
  - Medium seems a reasonable model that is time and GPU-efficient
  - Tiny has a bias towards word **“back” → “black people.”**

GPT-4 - Prompt

You are an AI system specialized in **detecting planning issues,** critiquing plans, and **analyzing conversations between police officers** regarding how to disperse. Additionally, identify any instances suggesting **1st Amendment violations**, criticizing the **lack of a plan**, and **aggressive comments**.

Give response only in the json format, for example: {"1": "What should we do now? I don't have a clue.", "2": "What the fuck is this", "3": "Beat the fuck out of them"}

"There can be multiple instances, find out all of them. If you do not find anything, just return **{"None": "None"}"**

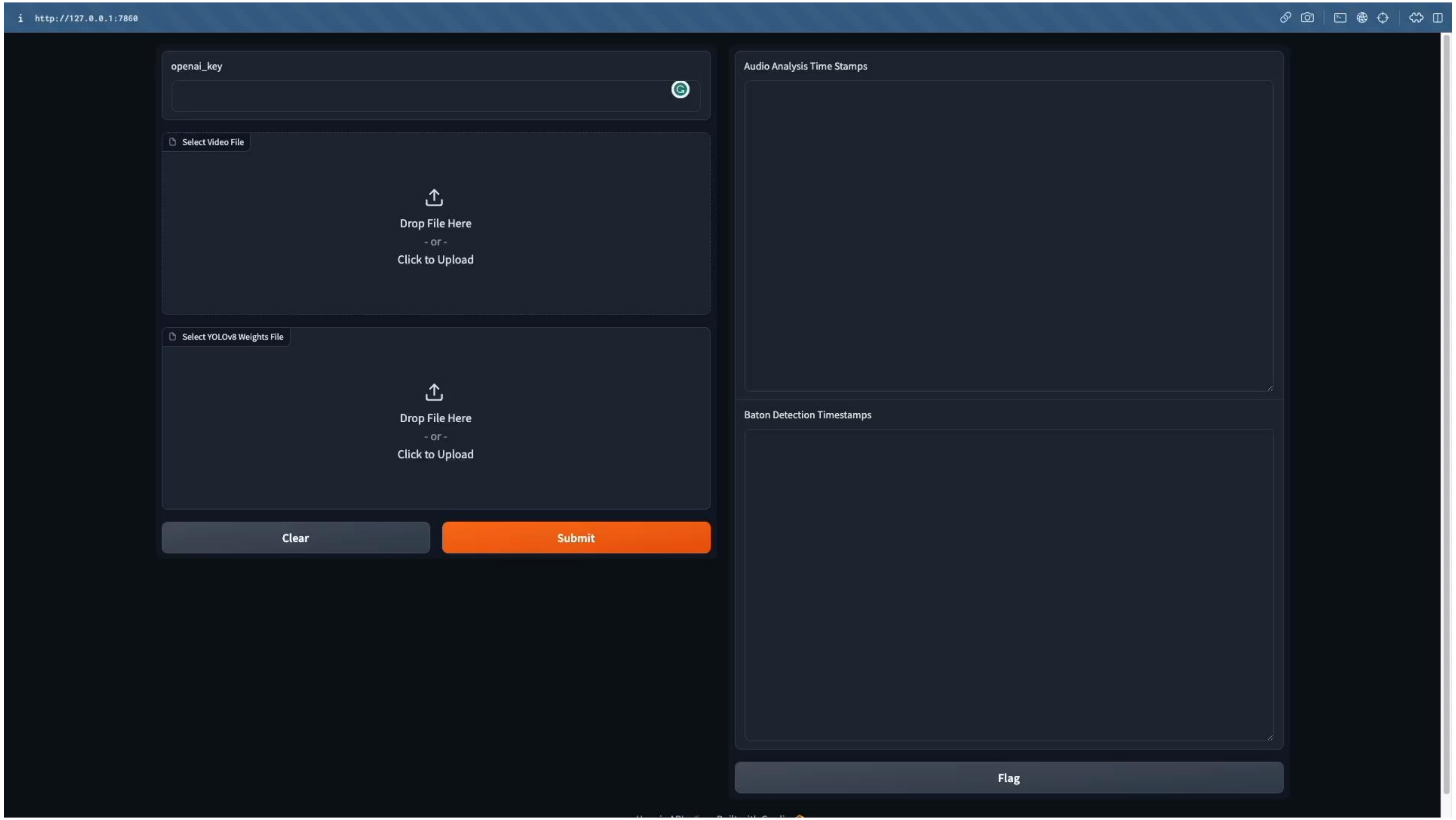
GPT-4

- Versions Experimented:
  - GPT-3.5-Turbo
  - GPT-4-1106
  - GPT-4
  - GPT-4-preview
- Findings:
  - Very structured **json** output for every iteration
  - Solved the randomness problem by setting the **seed**
  - The model was able to capture some very **subtle instances** that can be helpful to the case
  - We were also able to find if there was no violation in a transcript mitigating a lot of **false positives**
  - Models **unable** to automatically identify between the protestors and the police officers in transcripts

Deployment and Hosting

- Docker: Publicly available at **aakash0017/ml-nlgma-body-cam/**
- Deployment platform:
  - Huggingface Spaces (Tiny Whisper model)
  - AWS (Full deployment if credits)

Demo Video



- Runtime:
  - Whisper Large: 34 min 48 seconds (CPU)
  - GPT-4: 2 min
  - YOLO: 2 min 5 seconds

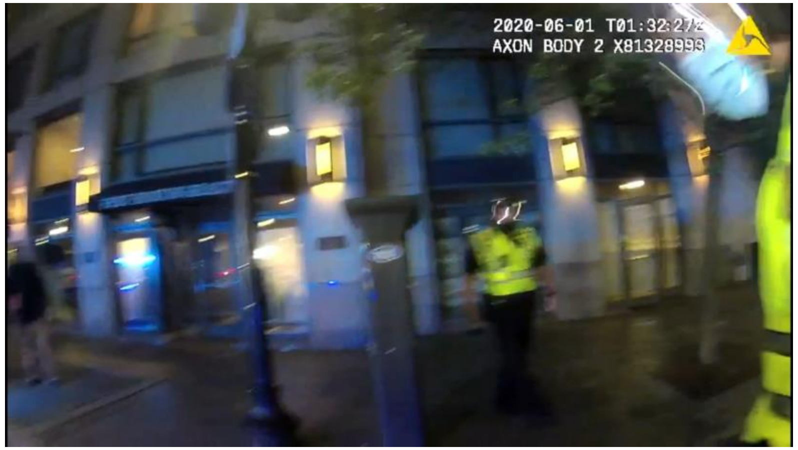
- Video Length: 9 min

Results

- Let's hit him with the fucking hose. Start Time: 00:35:32 End Time: 00:35:33 – Offensive Police
- What the fuck did I do? Start Time: 01:34:10 End Time: 01:34:13 – Confused Protestor
- Y'all ain't got nobody to fuck with. Start Time: 00:40:38 End Time: 00:40:40
- What the fuck are you doing? Start Time: 00:54:09 End Time: 00:54:12
- What's your fucking name? Start Time: 01:23:34 End Time: 01:23:36 – Rude Police
- This is fucking nuts. Start Time: 01:11:31 End Time: 01:11:33 – Police Confused

Clipped Videos

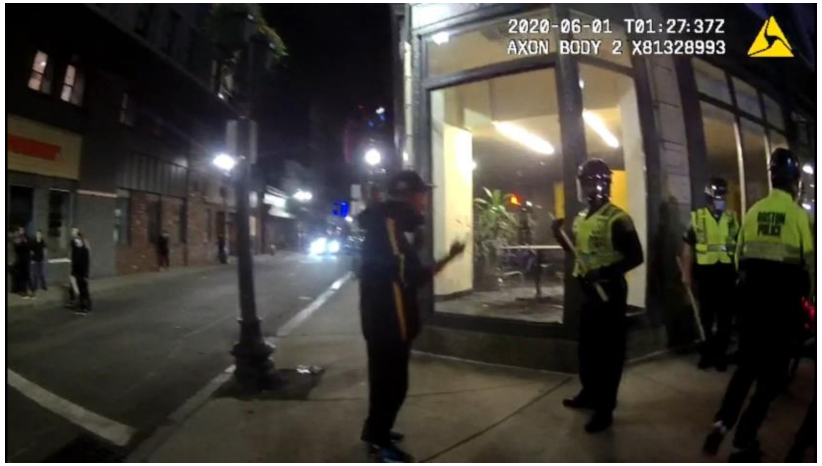
Protestor: I'm Fu**ng Scares



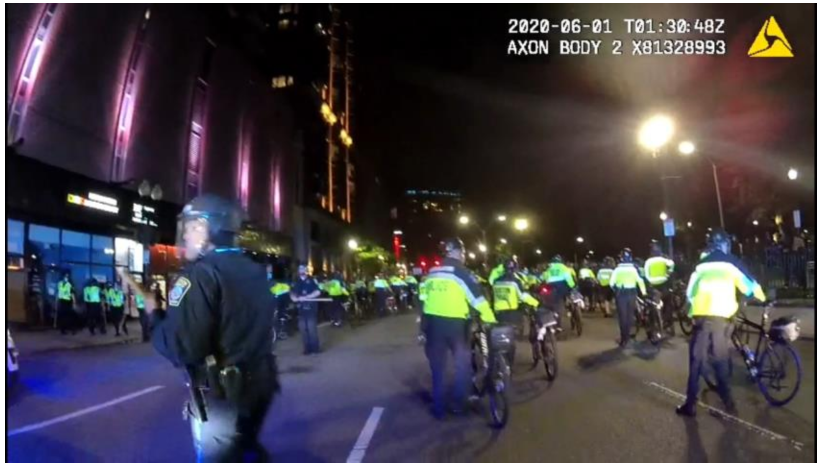
Police: He's a Fu**ng idiot

Protestor: Can you f**ng help me?

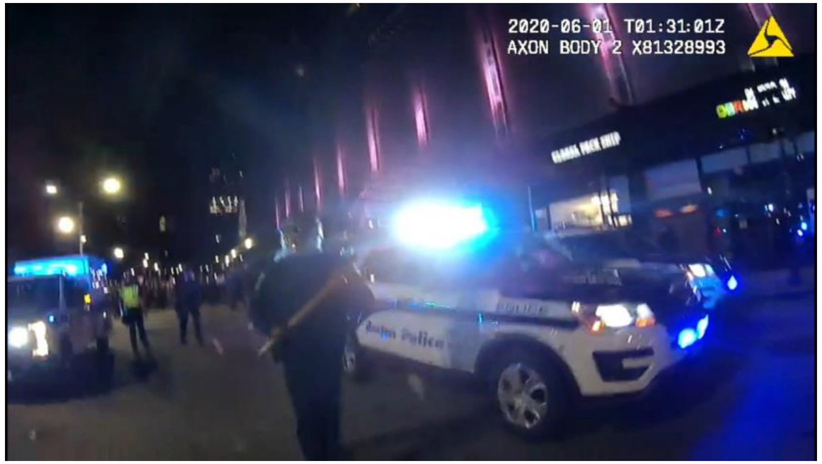
Findings



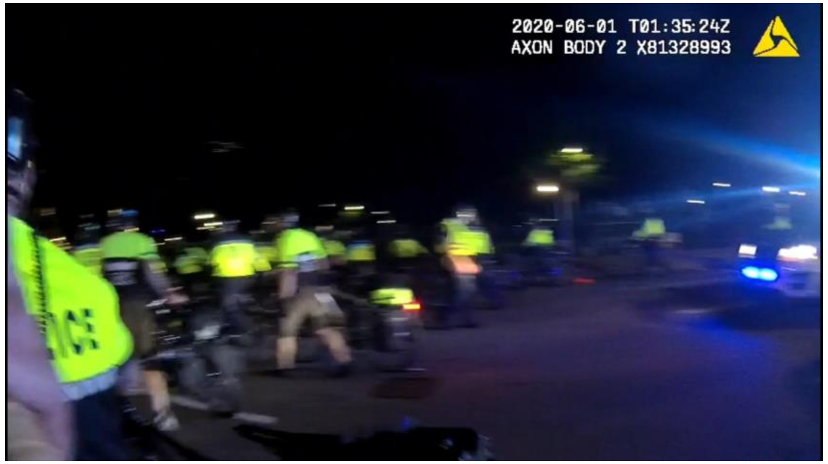
00:00:44 to 00:00:48



00:03:55 to 00:03:59



00:04:09 to 00:04:12



00:08:31 to 00:08:35

Future Work

- Implement Instant Whisper
- Show video clips in the UI
- Train more object detection models to catch violence
- Correlate object detection model with transcripts.

Thank You