

The background features several stylized icons related to audio and technology. In the top center is a blue turntable with red highlights. On the left, there's a red and blue headphones icon. On the right, there's a red and blue speaker icon. At the bottom left is a blue boombox with red speakers. At the bottom right is a white smartphone with a red speaker icon. Small white dots of varying sizes are scattered across the dark gray background.

TEAM - 36

REAL-TIME AUDIO MODERATION

Words Matter. We Keep Them in Check

INTRODUCTION

A REAL-TIME AUDIO MODERATION SYSTEM DESIGNED TO DETECT AND FLAG NSFW CONTENT USING ADVANCED AUDIO ANALYSIS AND TRANSCRIPTION TECHNOLOGIES.

Overview:

- A local AI system for *audio content moderation*.
- Ensures *privacy and security* by running directly on the user's device.
- Detects *profanity, offensive speech, and harmful content* before user access.

MOTIVATION

- Online moderation tools often require internet access, posing privacy risks.
- Existing moderation is centralized, limiting user control.
- A local AI-based approach improves security, speed, and flexibility.

OBJECTIVES

- DEVELOP AN “ON-DEVICE AI MODERATION” SYSTEM.
- DETECT PROFANITY, HATE SPEECH, AND OFFENSIVE CONTENT.
- ENSURE REAL-TIME AND OFFLINE FUNCTIONALITY.
- PROVIDE A USER-FRIENDLY INTERFACE FOR CONTROL & CUSTOMIZATION.

FEATURES

AUDIOMOD - WEB VERSION

- AUTOMATIC SPEECH-TO-TEXT TRANSCRIPTION USING WHISPER
- PROFANITY DETECTION USING BETTER-PROFANITY
- WEB INTERFACE FOR FILE UPLOAD AND RESULTS VIEWING
- NGROK INTEGRATION FOR PUBLIC URL

AUDIO MOD - MACOS & WINDOWS

- REAL-TIME MONITORING WATCHES FOLDERS FOR NEW AUDIO FILES AND PROCESSES THEM INSTANTLY.
- SPEECH TRANSCRIPTION USES THE WHISPER MODEL TO CONVERT AUDIO INTO TEXT LOCALLY.
- PROFANITY TAGGING DETECTS OFFENSIVE LANGUAGE AND ASSIGNS A CLEAN, MILD, OR EXPLICIT TAG.
- BACKGROUND EXECUTION RUNS SILENTLY IN THE BACKGROUND WITH A CUSTOM APP ICON ON BOTH PLATFORMS.

Progress So Far...

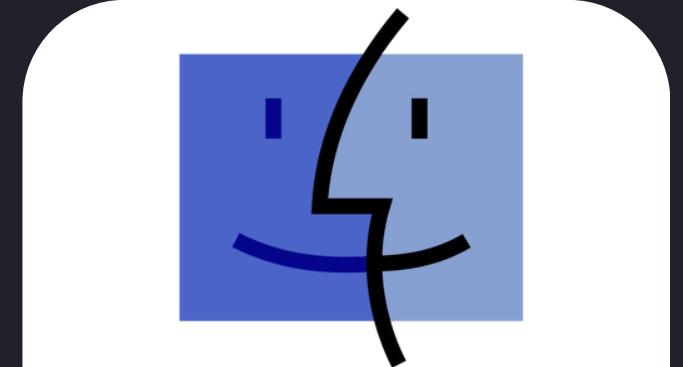
- We made a model that can efficiently run and can perform tasks efficiently.
- An App for Mac, Windows and Web browser version
- An app that runs in background and has an icon available
- The app is quite light and fast with storage space of 425 Mb
- Of which majorly taken by the lightweight Whisper model.
- These apps still have underlying issues which will be patched later especially with permissions in OS versions



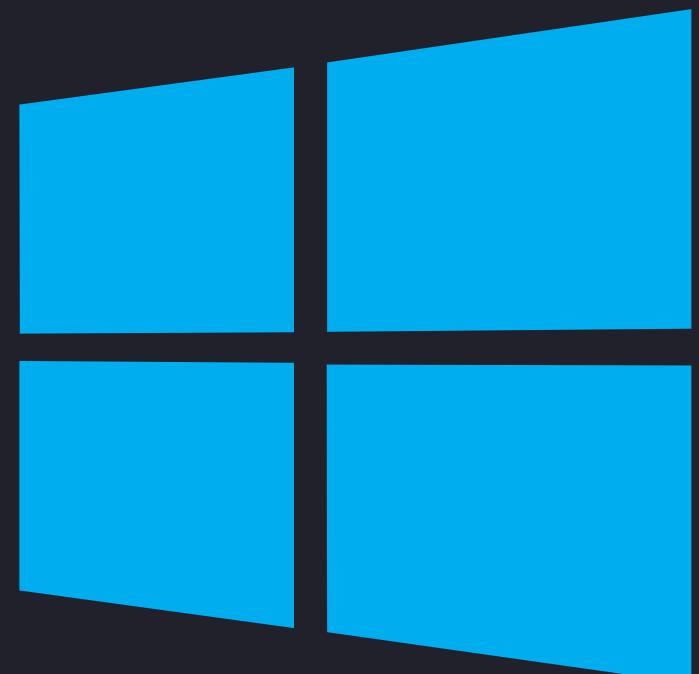
All versions:



Cloud



Mac



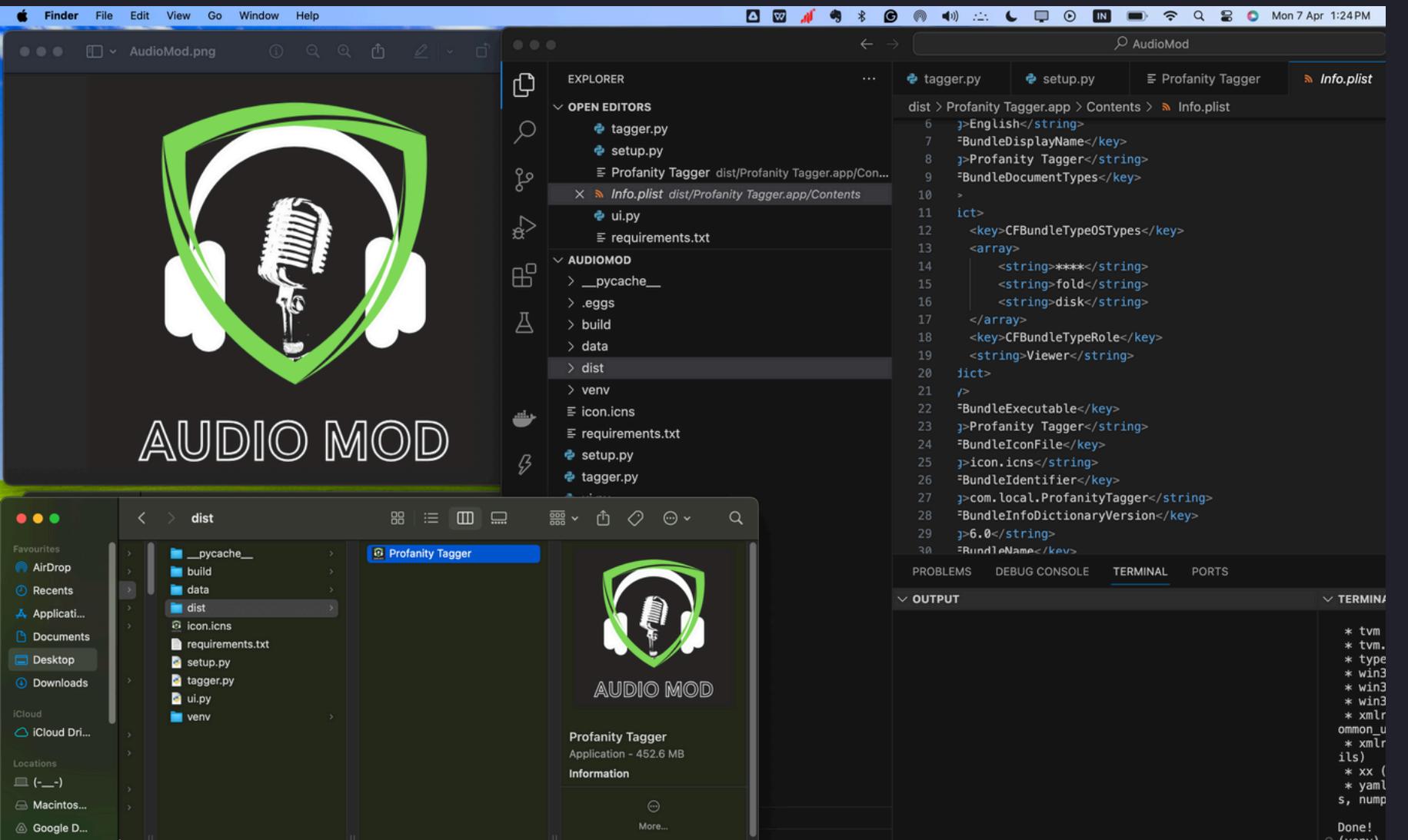
Windows

FUTURE GOALS

- Improve real-time audio moderation for live speech detection in accuracy and performance
- Optimize UI for better performance on local devices and deal with permission issues
- Expand detection to multi-lingual support
- Providing keyword options to user



Results OS version:



Watching: /Users/rohanrao/Desktop/WatchAudio
2025-04-07 13:28:26.079 python[67994:3831861] +[IMKClient subclass]: chose IMKClient_Legacy
2025-04-07 13:28:26.079 python[67994:3831861] +[IMKInputSession subclass]: chose IMKInputSession_Legacy

Results Web version:

Profanity Tagger

No file chosen

| File | Tag | Transcript |
|----------------|---------|--|
| sample-0.mp3 | CLEAN | Perhaps he made up to the party afterwards, and took her, and his heart at end continued. It was not Fieryce set it over her. Edmund |
| fuck-36089.mp3 | PROFANE | Fuck! |

```
Watching folder: /content/data/WatchAudio
* Web UI: https://a42b-35-185-249-25.ngrok-free.app
* Serving Flask app 'ui_web'
* Debug mode: off
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
* Running on all addresses (0.0.0.0)
* Running on http://127.0.0.1:5002
* Running on http://172.28.0.12:5002
Press CTRL+C to quit
127.0.0.1 - - [07/Apr/2025 05:16:34] "GET / HTTP/1.1" 200 -
127.0.0.1 - - [07/Apr/2025 05:16:35] "GET /favicon.ico HTTP/1.1" 404 -
127.0.0.1 - - [07/Apr/2025 05:18:40] "POST /upload HTTP/1.1" 200 -
127.0.0.1 - - [07/Apr/2025 05:19:03] "GET / HTTP/1.1" 200 -
127.0.0.1 - - [07/Apr/2025 05:19:08] "POST /upload HTTP/1.1" 200 -
127.0.0.1 - - [07/Apr/2025 05:19:44] "GET / HTTP/1.1" 200 -
```

Conclusion



This project provides a local, secure, and efficient moderation solution.

Testing with various AI models has shown promising results and a basic model that can run in the background with a basic UI and icons

Future improvements will enhance speed, accuracy, and user control and smoother UI with a pop up upload feature.