

TEAM CTRL + V

# SNAP SOUND

GEORGE WANG, ANANYA PRAKASH, YASH SINGH, BELLA QU  
TA: PREETHAM MANAPURI



# PROJECT INTRO

Snap Sound is an innovative solution for finding music that correlates to an image's mood. The user's Spotify data will be used to create a personalized playlist for them that matches the mood of the input image.

It has various applications including:

- Increased accessibility to visual content
- Bringing life to still images
- Educational tool



# PROJECT OVERVIEW

Part 1 (MVP): Facial Emotion -> Playlist



A screenshot of a mobile application interface. At the top, there's a camera view showing a woman smiling. A blue dashed square box highlights her face. Below the camera view, the title "Feeling Happy" is displayed in large white letters. Underneath the title, it says "A highly personalized playlist created with Snap Sound." and "ananya • 73 songs, about 3 hr 30 min". There's a green play button icon. The main part of the screen shows a list of songs with columns for #, Title, Album, and Duration. The first few songs are:

| # | Title                         | Album      | Duration |
|---|-------------------------------|------------|----------|
| 1 | Talk To Me                    | Talk To Me | 3:18     |
| 2 | I think about it all the time | BRAT       | 2:15     |
| 3 | Atlantis                      | Fantasea   | 2:29     |

Part 2: General Mood -> Playlist



A screenshot of a mobile application interface. At the top, there's a camera view showing a tiger roaring. Below the camera view, the title "Feeling Angry" is displayed in large white letters. Underneath the title, it says "A highly personalized playlist created with Snap Sound." and "ananya • 76 songs, about 4 hr 15 min". There's a green play button icon. The main part of the screen shows a list of songs with columns for #, Title, Album, and Duration. The first few songs are:

| # | Title            | Album         | Duration |
|---|------------------|---------------|----------|
| 1 | Marry The Night  | Born This Way | 4:24     |
| 2 | Break Your Heart | Rokstarr      | 3:06     |

# PIPELINE



01

Take an input image

02

If FaceNet model detects a face, classify the emotion of the face using Visual Transformer. Else capture the general mood using a CNN

03

03

Use the user's Spotify data to create personalized playlist that fits the classification

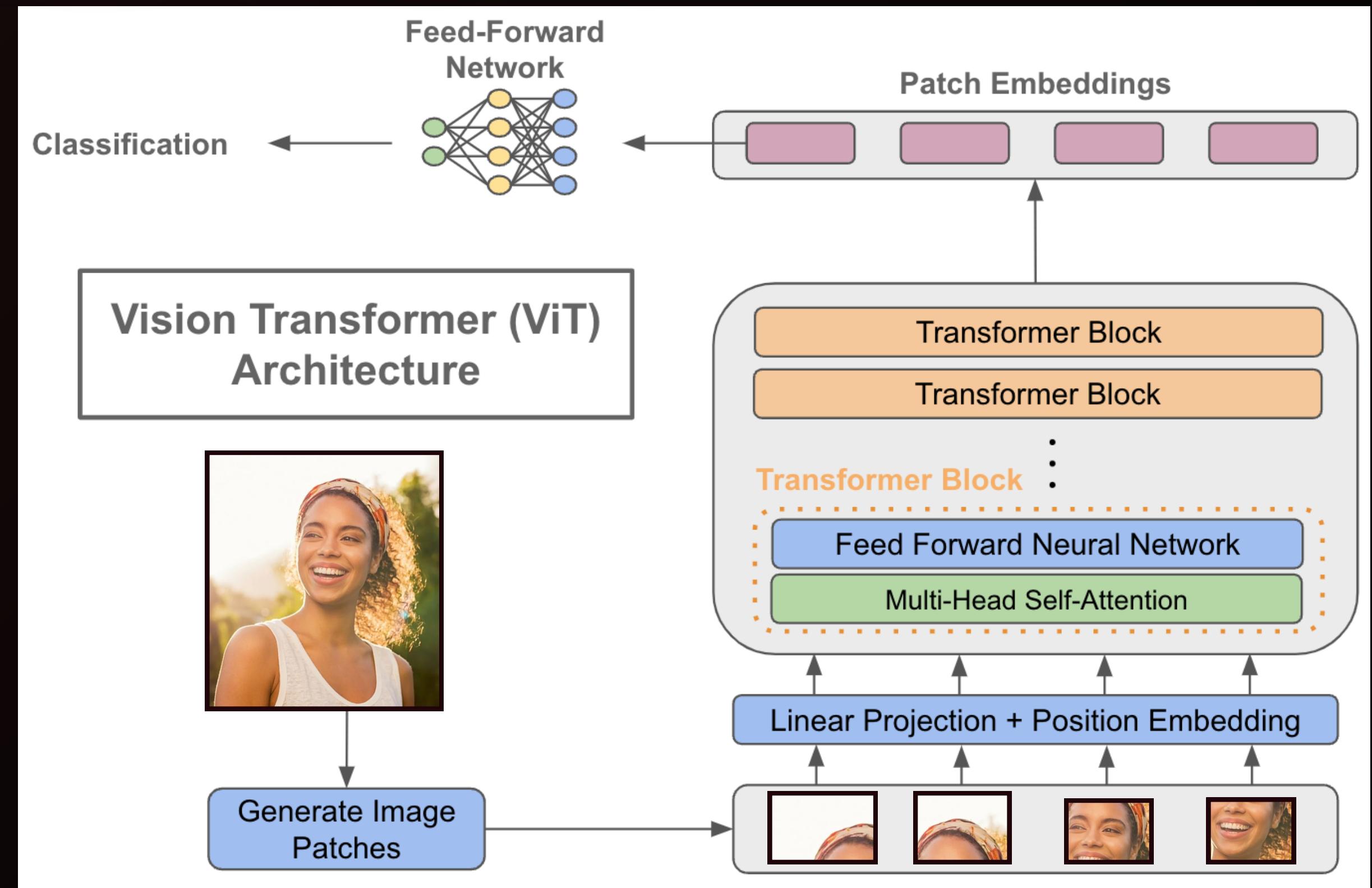
04

Output the playlist, and allow user to download

# EMOTION DETECTION (PT. 1)

The input image is broken into smaller pieces, and each piece is analyzed by the transformer to figure out the big picture: the mood.

Using a fine-tuned version of Google ViT (Vision Transformer): **87%** accuracy

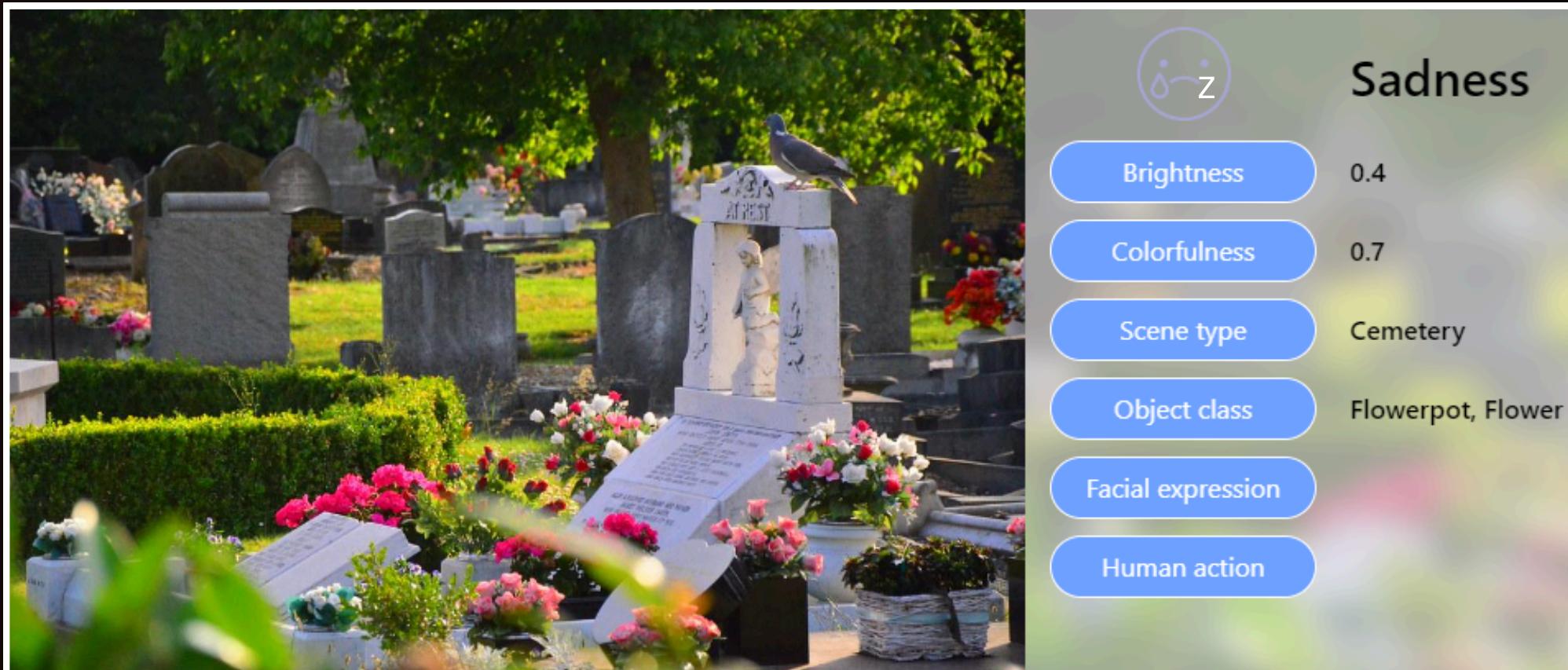


# MOOD DETECTION DATASET (PT. 2)

01

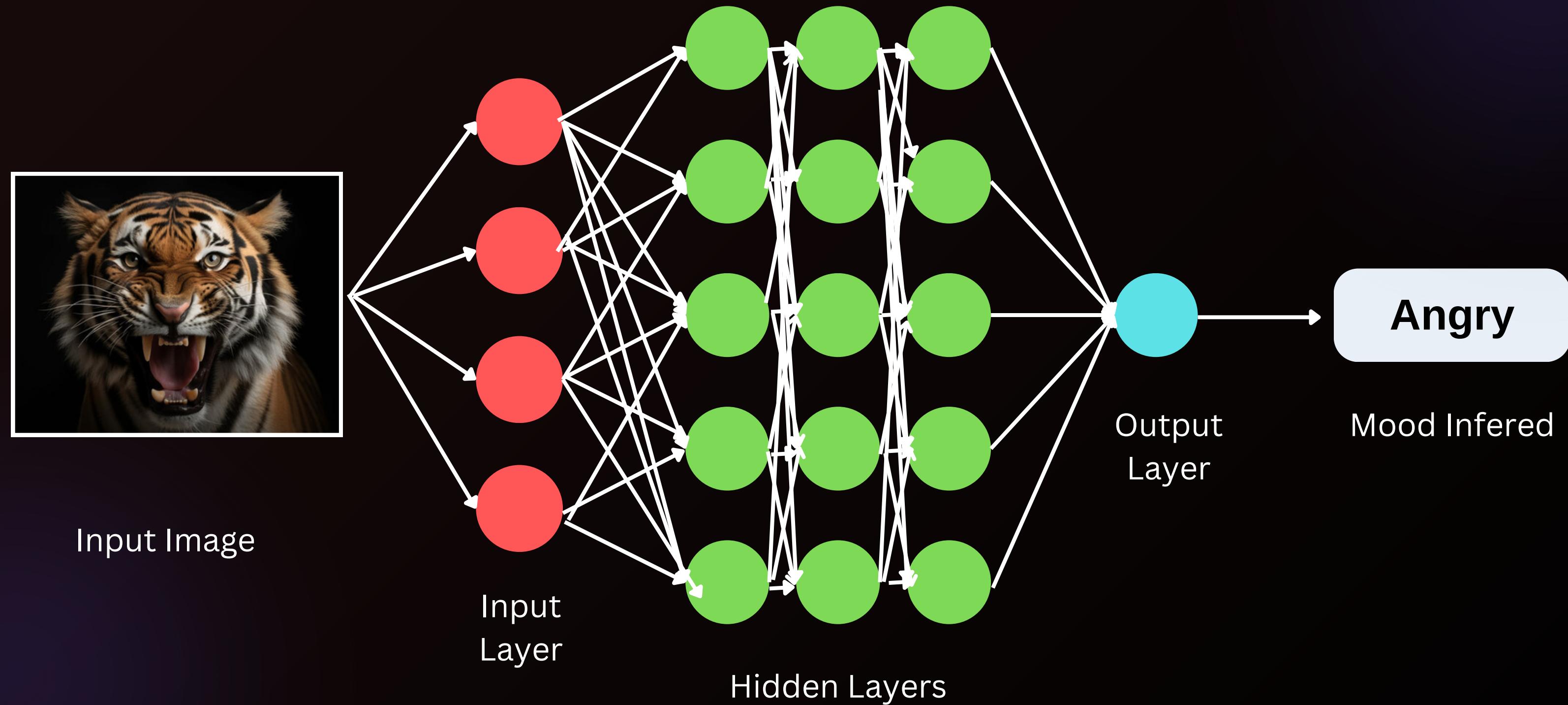
Collected data from EmoSet - 118K images and 8 mood labels

Example:



|                | Amusement | Awe  | Contentment | Excitement | Anger | Disgust | Fear | Sadness |
|----------------|-----------|------|-------------|------------|-------|---------|------|---------|
| Cemetery       | 0.00      | 0.00 | 0.00        | 0.00       | 0.00  | 0.00    | 0.00 | 0.99    |
| Landfill       | 0.00      | 0.00 | 0.00        | 0.00       | 0.00  | 0.94    | 0.00 | 0.05    |
| Corridor       | 0.00      | 0.00 | 0.00        | 0.01       | 0.00  | 0.02    | 0.68 | 0.29    |
| Boxing ring    | 0.00      | 0.00 | 0.00        | 0.03       | 0.97  | 0.00    | 0.00 | 0.00    |
| Stage          | 0.06      | 0.00 | 0.00        | 0.91       | 0.00  | 0.00    | 0.01 | 0.02    |
| Lawn           | 0.18      | 0.07 | 0.43        | 0.15       | 0.08  | 0.01    | 0.02 | 0.06    |
| Mountain snowy | 0.00      | 0.92 | 0.02        | 0.06       | 0.00  | 0.00    | 0.00 | 0.00    |
| Carousel       | 0.97      | 0.00 | 0.00        | 0.03       | 0.00  | 0.00    | 0.00 | 0.00    |

# MOOD DETECTION MODEL (PT. 2)



We finetuned the ResNet-50 (pre-trained on ImageNet) CNN using data from EmoSet, achieved 67% accuracy

# CHOOSING SPOTIFY SONGS

01

Depending on the mood, parameters are selected for track happiness, energy, tempo, danceability, acousticness, and valence

02

The user's top and recommended songs are shuffled and parsed through based on the parameters

03

A playlist with a specified number of songs is outputted

E.g. parameters for "happy":

Happiness: 0.7 - 1.0

Tempo: 0.6 - 1.0

Energy: 0.6 - 1.0

Danceability: 0.7 - 1.0

Acousticness: 0.0 - 0.3

Valence: 0.7 - 1.0



# CHALLENGES WE FACED

## Spotify API

- Initially used Spotify -> wasn't compatible with our project
- We switched to the normal API

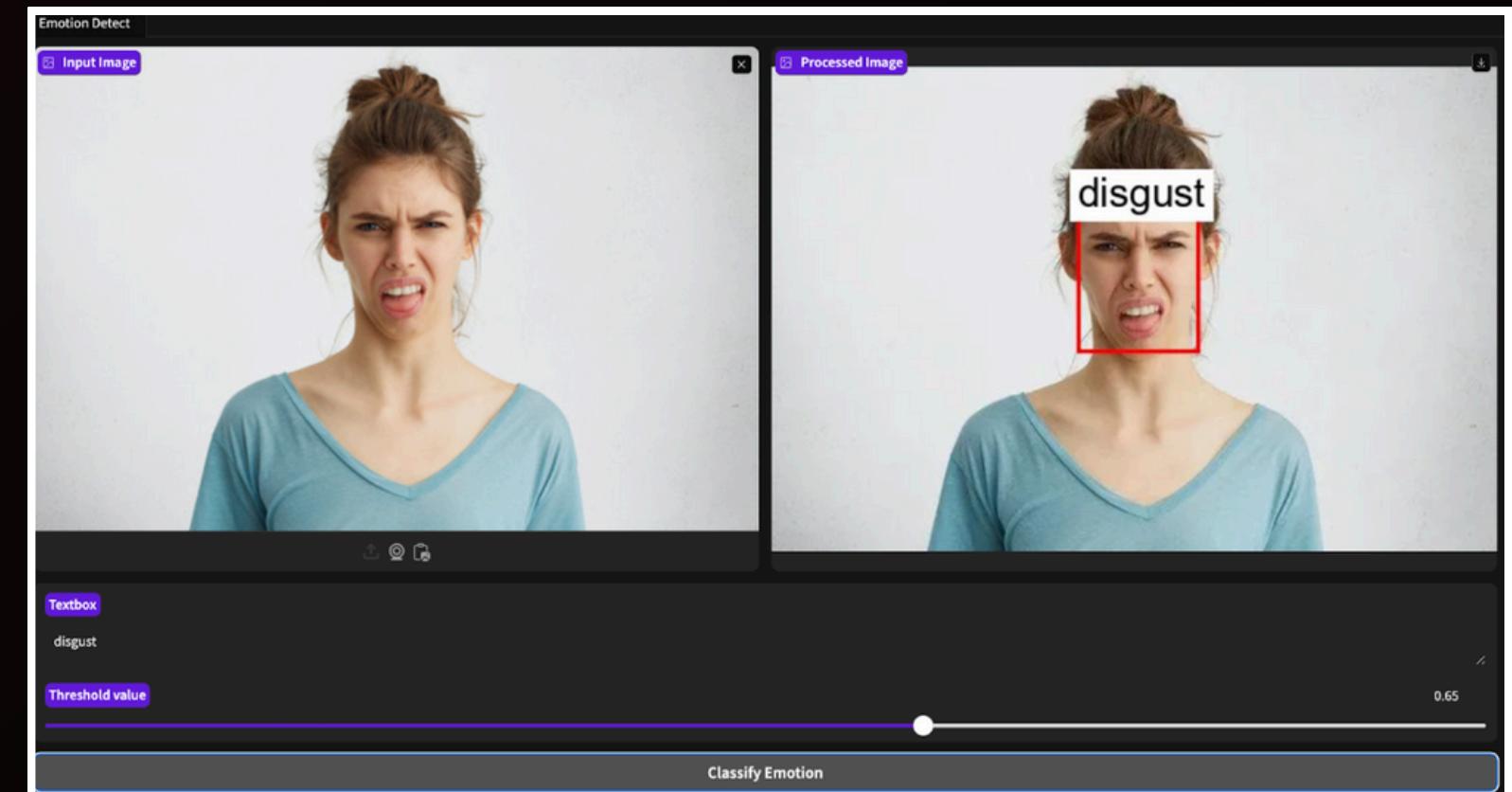
## Data

- ResNet-50 overfits in fine-tuning process
- Data set hadn't been used publicly before
- Other datasets weren't good quality or had too much bias (e.g. B&W images only)



# OUR LEARNING

- We implemented learning from through our time in Cog\*Works
  - Facial Recognition: week 2 - vision module
  - CNNs for mood detection from an image
- During this project specifically:
  - Spotify API
  - Flask for backend work
  - Built an interactive interface
  - Fine tune models



# DEMO

Allow Spotify to connect to:  
**snapSound**

George  
[Not you?](#)

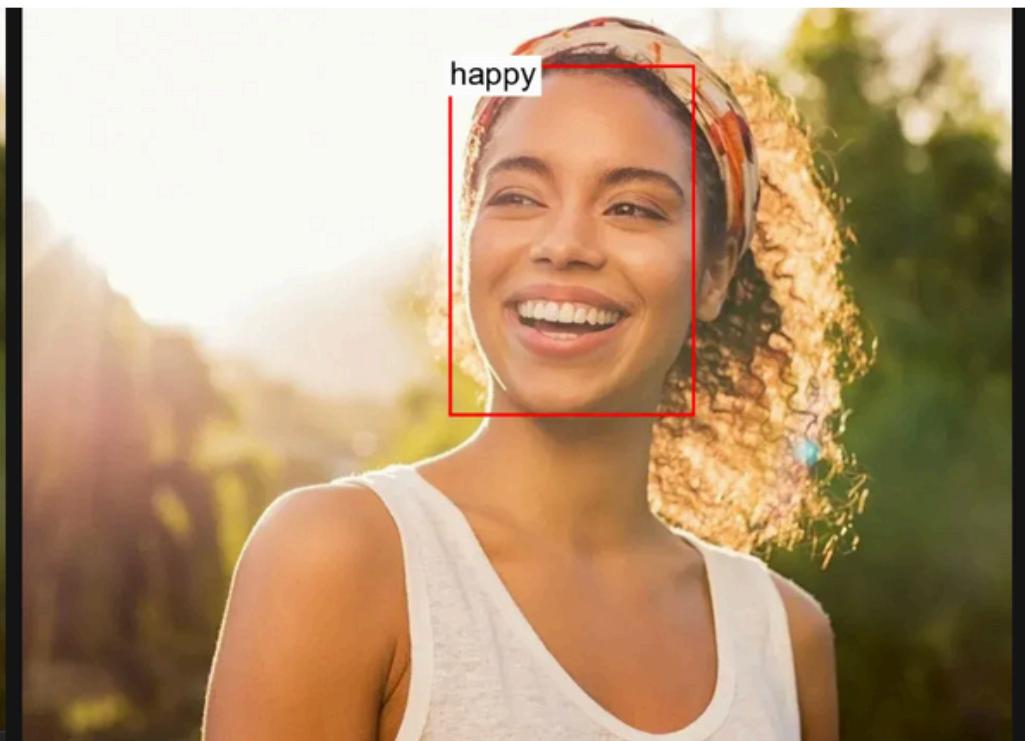
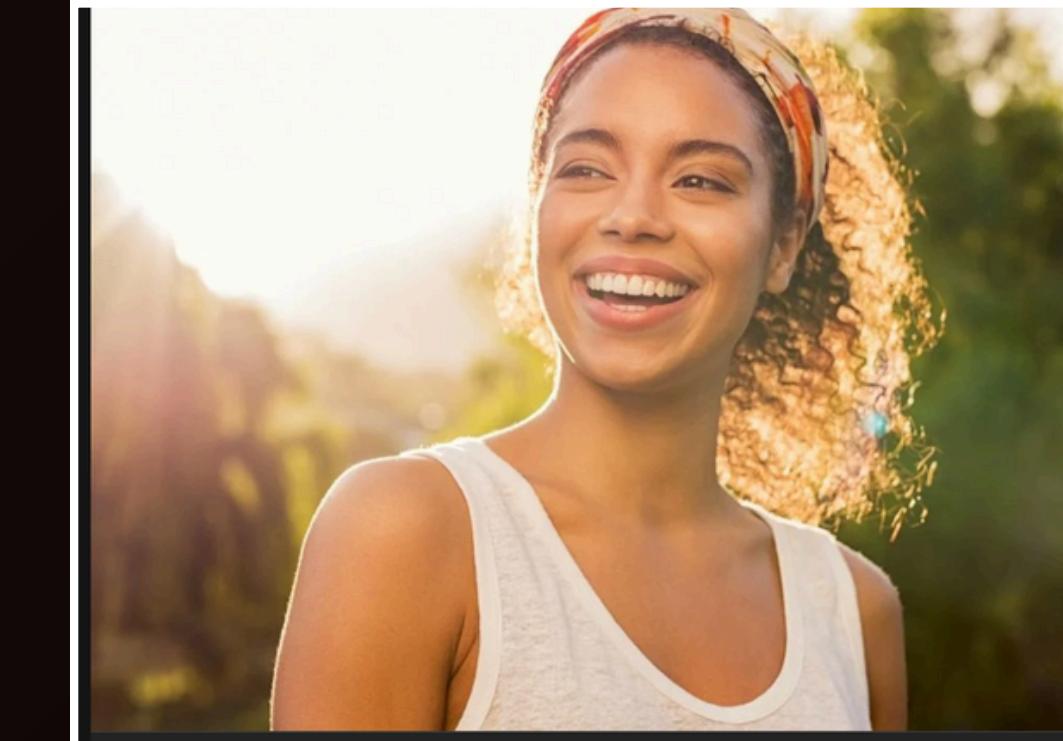
You agree that snapSound will be able to:

- View your Spotify account data
  - Your email
  - Your Spotify subscription, account country, and explicit content filter settings.
  - Your name, username, profile picture, Spotify followers, and public playlists.

**Agree**

**Cancel**

You can remove this access at any time in your account settings.  
For more information about how snapSound can use your personal data, please see snapSound's [privacy policy](#).



**Detected Emotion**  
happy

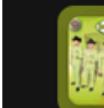
**Threshold value**  
0.45

**Classify Emotion and Get Playlist**

**Song List**

1. La Chona by Los Tucanes De Tijuana
2. When They Fight, They Fight by Generational
3. Entre Beso Y Beso by La Arrolladora Banda El Limón De Rene Camacho
4. Should I Stay or Should I Go - Remastered by The Clash
5. La ingrata by Café Tacvba
6. Date Night (Same Time) [feat. Chris Brown] by Kirko Bangz
7. Walking On A Dream by Empire Of The Sun
8. Tiburon - English Version by Proyecto Uno
9. La Mordidita (feat. Yotuel) by Ricky Martin
10. Eternamente Bella by Alejandra Guzman

La Chona by Los Tucanes De Tijuana



La Chona  
Los Tucanes De Tijuana

03:19

When They Fight, They Fight by Generational



When They Fight, They Fight  
Generational

03:20

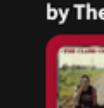
Entre Beso Y Beso by La Arrolladora Banda El Limón De Rene Camacho



Entre Beso Y Beso  
La Arrolladora Banda El Limón De Ren

03:02

Should I Stay or Should I Go - Remastered by The Clash



Should I Stay or Should I Go  
The Clash

03:08

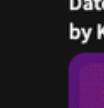
La ingrata by Café Tacvba



La ingrata  
Café Tacvba

03:32

Date Night (Same Time) [feat. Chris Brown] by Kirko Bangz



Date Night (Same Time) [feat.  
Chris Brown]  
Kirko Bangz, Chris Brown

03:46

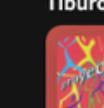
Walking On A Dream by Empire Of The Sun



Walking On A Dream  
Empire Of The Sun

03:18

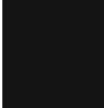
Tiburon - English Version by Proyecto Uno



Tiburon - English Version  
Proyecto Uno

05:02

La Mordidita (feat. Yotuel) by Ricky Martin



La Mordidita (feat.  
Yotuel)  
Ricky Martin, Yotuel

03:31

Eternamente Bella by Alejandra Guzman

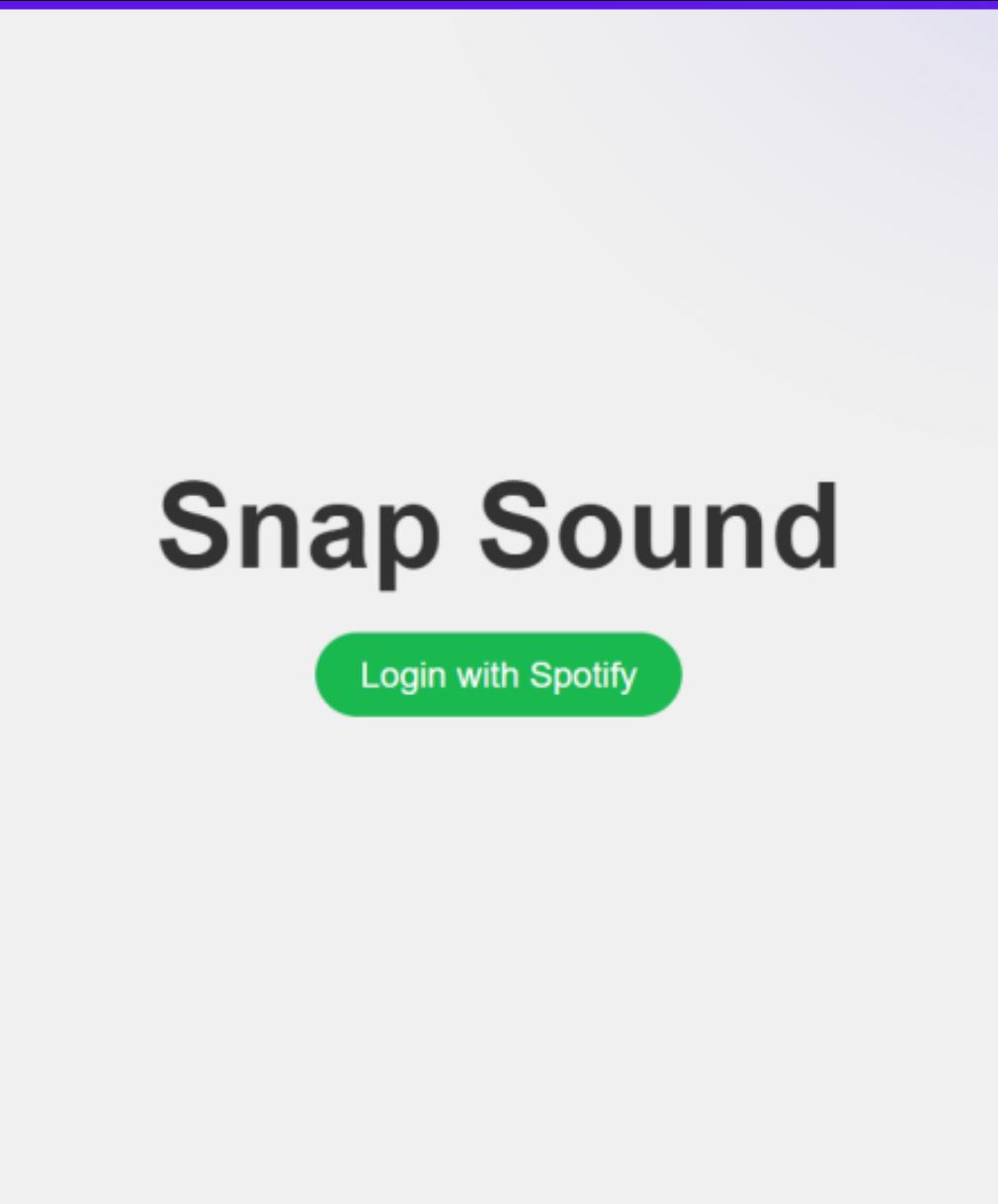


Eternamente Bella  
Alejandra Guzman

03:26

# FUTURE IMPROVEMENTS

- Improved UI
- Make the program more cohesive
- Fine tune our model further
  - Expand to event detection (e.g. weddings, Halloween, etc.)
- An expanded set of moods/emotions detected



Example of future UI



# THANK YOU!

TEAM CTRL + V