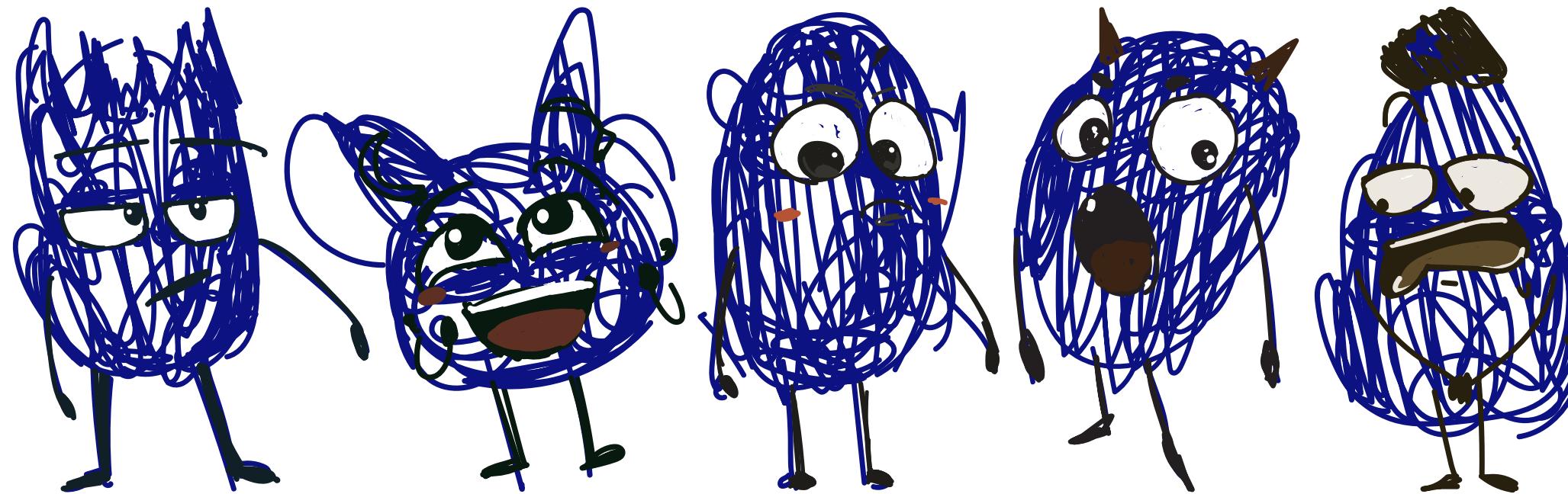
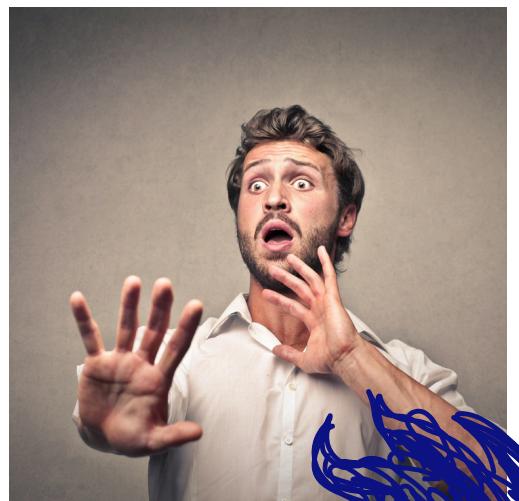
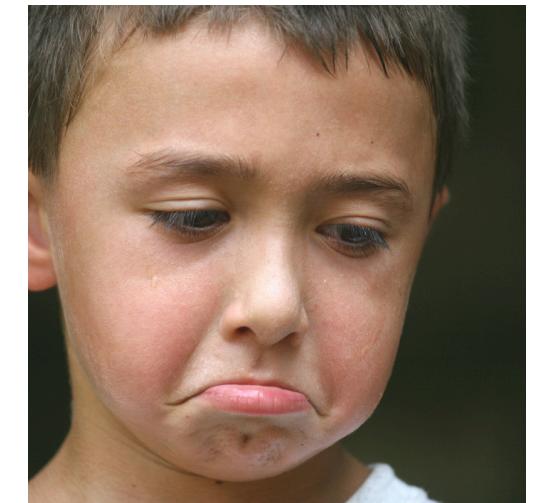


EMOTION-BASED SONG RECOMMENDER



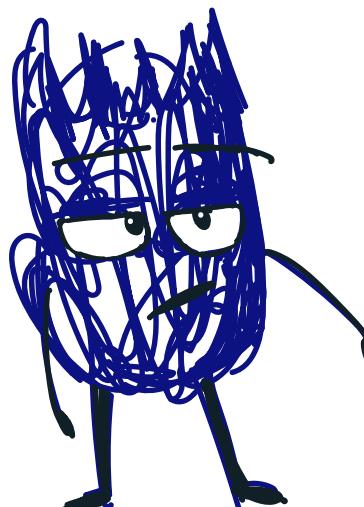
BACKGROUND

This project builds an intelligent, real-time song recommendation system that analyzes facial expressions to detect emotions and suggest mood-appropriate music. By combining webcam-based emotion recognition with Spotify's music database, it delivers a personalized and seamless listening experience.



OBJECTIVES

- **To create a system that detects human emotions using facial expression recognition and recommends songs that correspond to the detected mood.**
- **To implement real-time face detection and emotion classification using OpenCV and You Only Look Once (YOLO).**
- **Provides a real-time user interface that displays the webcam feed alongside detected faces, bounding boxes, and emotion labels. Based on these detected emotions, the system generates personalized song recommendations and presents the recommended tracks directly through the interface.**
- **Songs are sourced from Spotify via Spotipy which is the Python library for the Spotify Web API.**



SCOPES

Real-Time Face Detection and Emotion Classification

- **OpenCV** → capture video frames from a webcam
- **trained YOLO11n** → detect faces and draw bounding boxes in real time
- **trained YOLO11n-cls** → Classify facial emotions (8 classes)

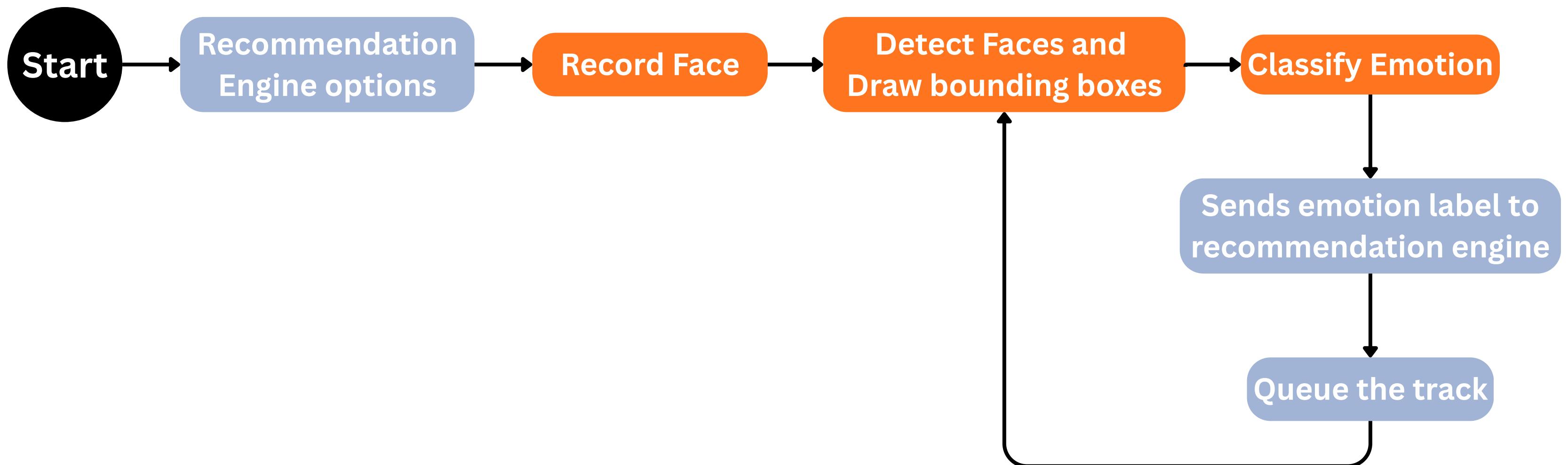
Song Recommendation System

- Build Preference metrics from the user's selected playlist
- Map detected emotions to corresponding song moods or genres.
- Retrieve suitable tracks using the Spotify Web API and automatically queue a track for the user.

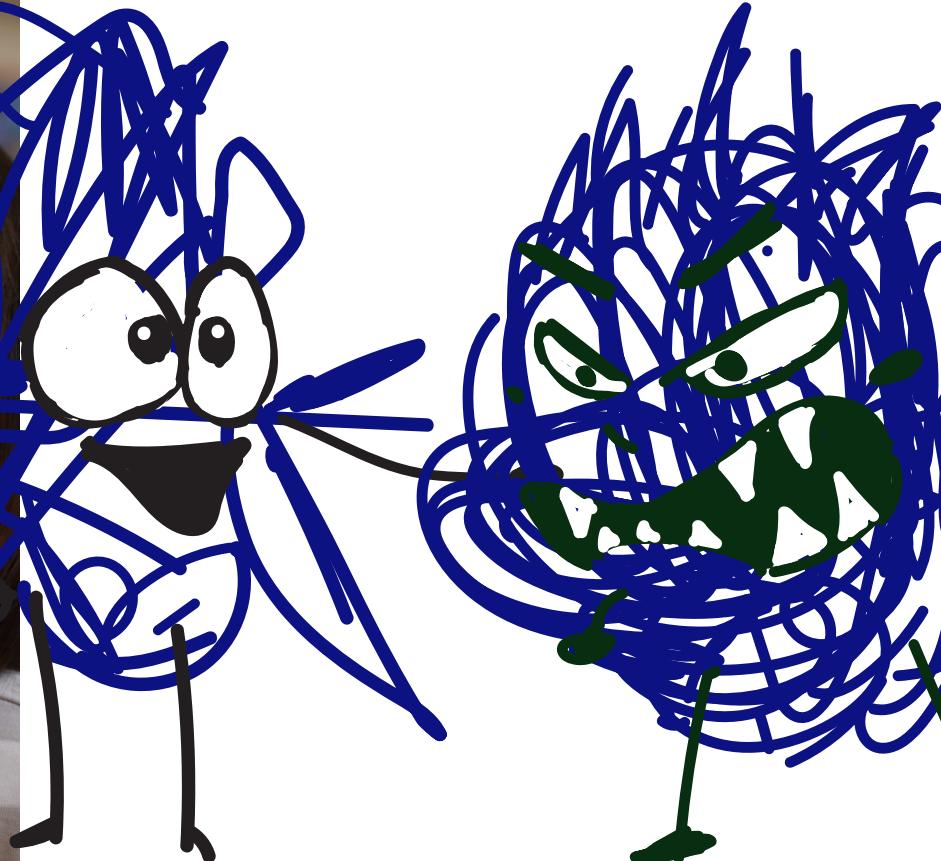
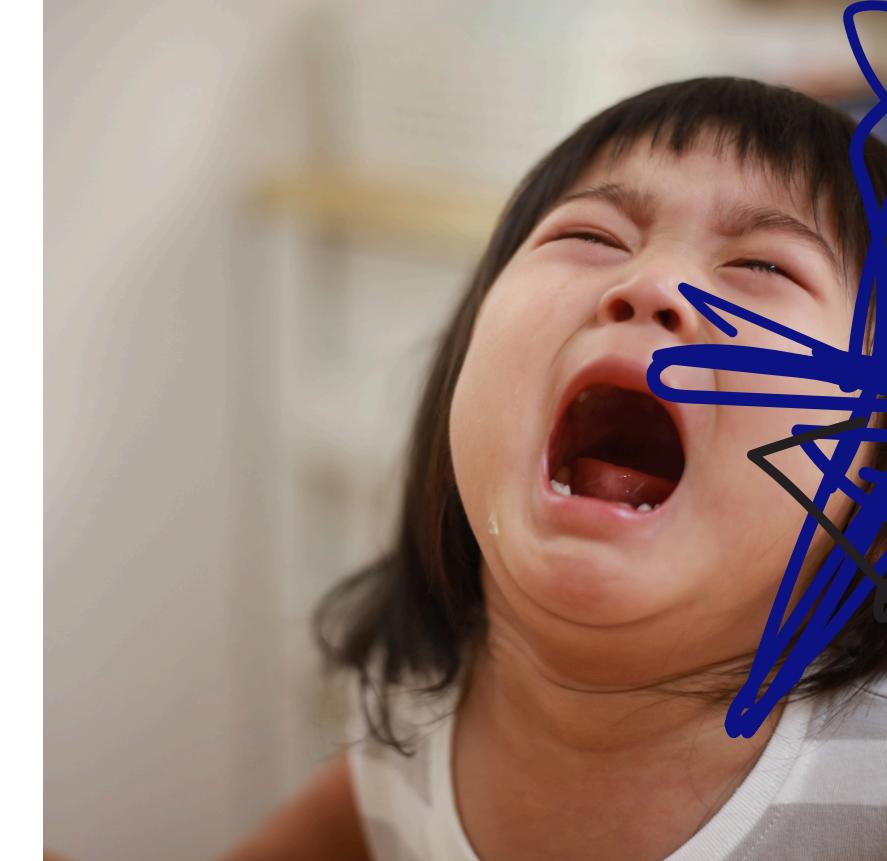


OVERVIEW

- Application flow



METHODOLOGY



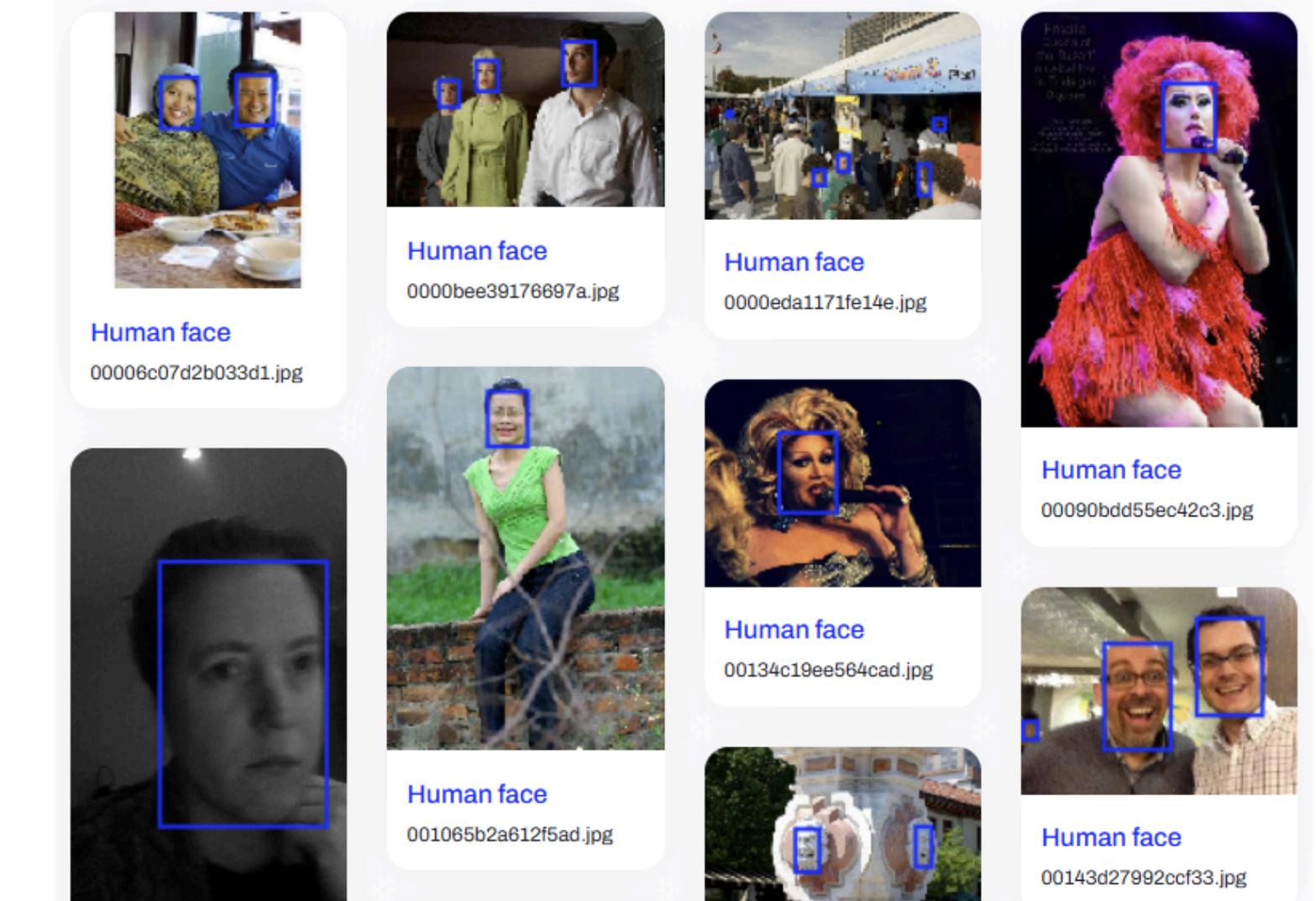
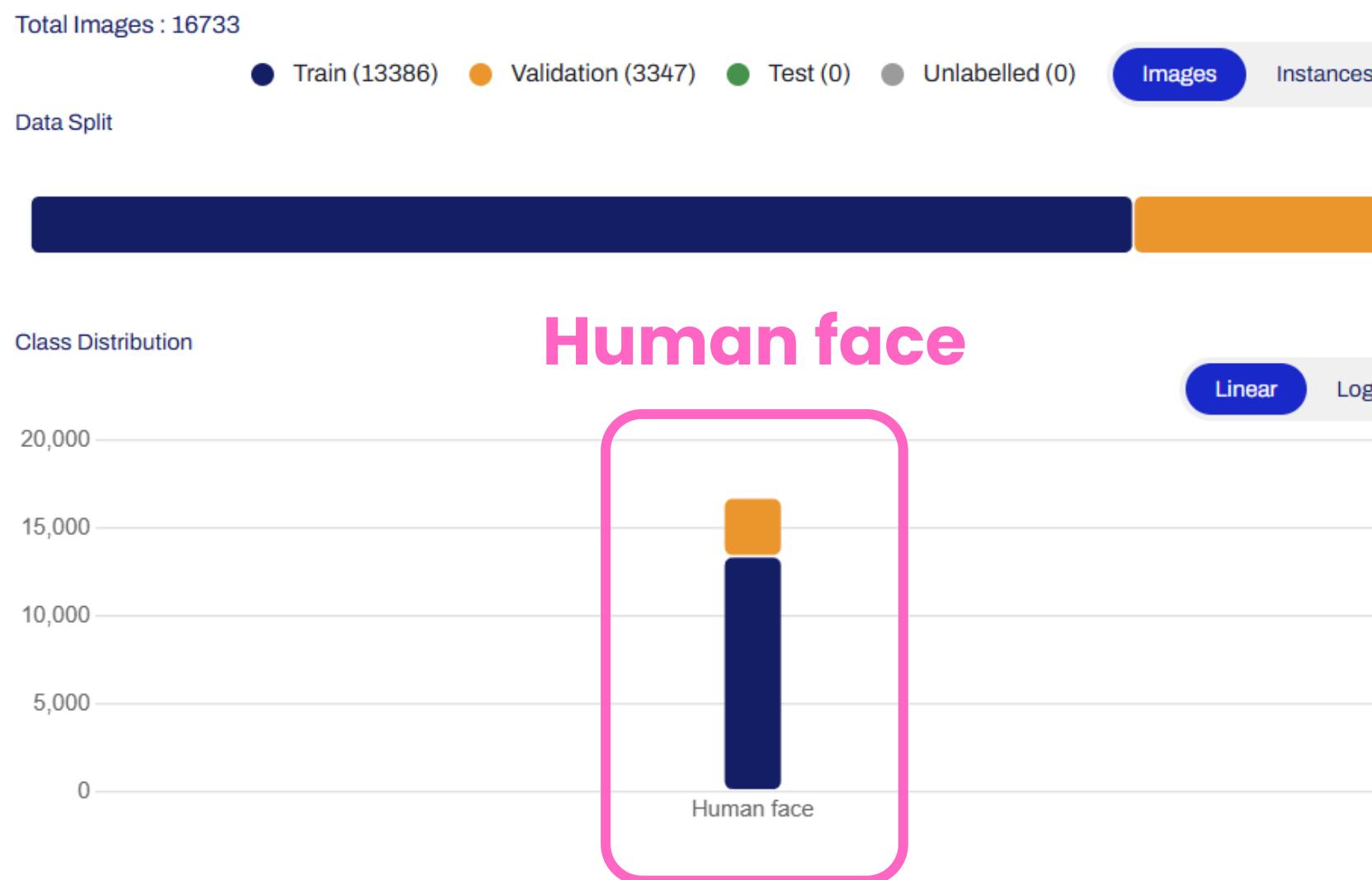
FACE-DETECTION-DATASET

Fares Elmenshawii

- **Size:** ~16.7k high-quality images
- **Purpose:** Specifically created for face detection training
- **Source:** Scrapped from Google Open Images using the OIDv4 toolkit
- **Annotations:**
 - **Original scale labels:** Pixel-based bounding box coordinates
 - **YOLO format labels:** Normalized bounding box coordinates
- **Quality:** Images selected for clarity and relevance; all contain detectable faces
- **Use Case:** Well-suited for training deep learning models focused on face localization and detection
- **Coverage:** Only face-relevant images preserved; non-related images excluded for dataset purity

<https://www.kaggle.com/datasets/fareselmenshawii/face-detection-dataset>

FACE DETECTION MODEL



Face-Detection-Dataset

FACE DETECTION MODEL

Dataset: Face-Detection-Dataset

Size: 16,733 images in total

- **Train set:** 13,386 images
- **Validation set:** 3,347 images

Architecture: YOLOv1n

CONFIGURATIONS:

Pretrained	Yes
Epochs	20
Image Size	640
Patience	100
Cache	None
Device	GPU
Batch Size	10

FACE DETECTION RESULT

Path to Ultralytics HUB: <https://hub.ultralytics.com/models/YaCSN3TYdCmXVLnAFtZW>

```
Validating /content/runs/detect/train/weights/best.pt...
Ultralytics 8.3.232 🚀 Python-3.12.12 torch-2.9.0+cu126 CUDA:0 (NVIDIA L4, 22693MiB)
YOLOv1n summary (fused): 100 layers, 2,582,347 parameters, 0 gradients, 6.3 GFLOPs
      Class      Images   Instances      Box(P      R      mAP50  mAP50-95): 100% ----- 168/168 11.6it/s 14.5s
                  all        3347      10299      0.891      0.799      0.875      0.584
Speed: 0.1ms preprocess, 0.8ms inference, 0.0ms loss, 0.8ms postprocess per image
Results saved to /content/runs/detect/train
Ultralytics HUB: Syncing final model...
: 100% ----- 5.2MB 4.0MB/s 1.3s
Ultralytics HUB: Done ✅
Ultralytics HUB: View model at https://hub.ultralytics.com/models/YaCSN3TYdCmXVLnAFtZW 🚀
```

FACE DETECTION RESULT



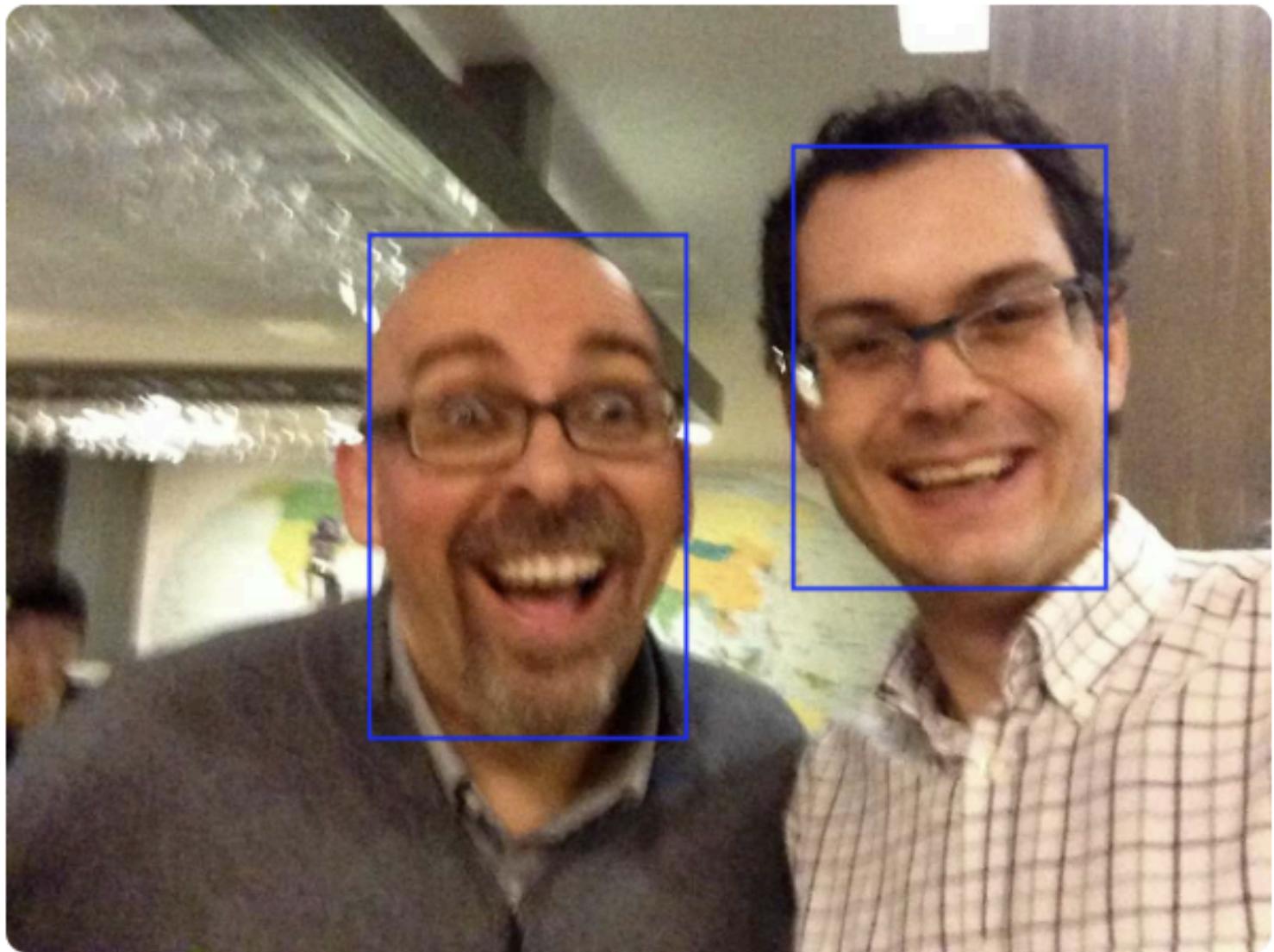
- **Mean Average Precision 50 (mAP50(B)):** 0.875
- **Mean Average Precision 50-95 (mAP50-95(B)):** 0.584
- **Precision (B):** 0.891
- **Recall (B):** 0.799

FACE DETECTION RESULT

Test

Preview your model

[Image](#) [Camera](#)



Settings

[Image Size](#)

320px [640px](#)

[Confidence Threshold](#)

[0.25](#)

[IoU Threshold](#)

[0.45](#)

Inference results

[Human face](#)

[90.2%](#)

[Human face](#)

[88.7%](#)

FERPLUS DATASET

Arnab Kumar Roy

- **Task:** Facial Emotion Recognition with 8 emotion classes : Angry, Contempt, Disgust, Fear, Happy, Neutral, Sad, Surprise
- **Dataset Structure:**
 - **Train:** 66,379 images
 - **Validation:** 8,341 images
 - **Test:** 3,579 images
- **Balance:** Dataset is now balanced using multiple data augmentation techniques
- **Content:** Contains labeled facial images with clear emotion categories
- **Use Case:** Suitable for training, validating, and testing deep learning models for emotion classification

<https://www.kaggle.com/datasets/arnabkumarroy02/ferplus>

EMOTION CLASSIFIER MODEL

Total Images : 78293

● Train (66379) ● Validation (8341) ● Test (3573) ● Unlabelled (0)

Images Instances

Data Split



Class Distribution

Linear Log

15,000

10,000

5,000

0

angry

contempt

disgust

fear

happy

neutral

sad

surprise

DATASET: FERPLUS

EMOTION CLASSIFIER MODEL

Dataset: FERPlus

Size: 78,293 images in total

- **Train set: 66,379 images**
- **Validation set: 8,341 images**
- **Test set: 3,573 images**

Architecture: YOLO11n-cls

8 classes: Angry, Contempt, Disgust, Fear, Happy, Neutral, Sad, Surprise

CONFIGURATIONS:

Pretrained	Yes
Epochs	20
Image Size	640
Patience	100
Cache	None
Device	GPU
Batch Size	10

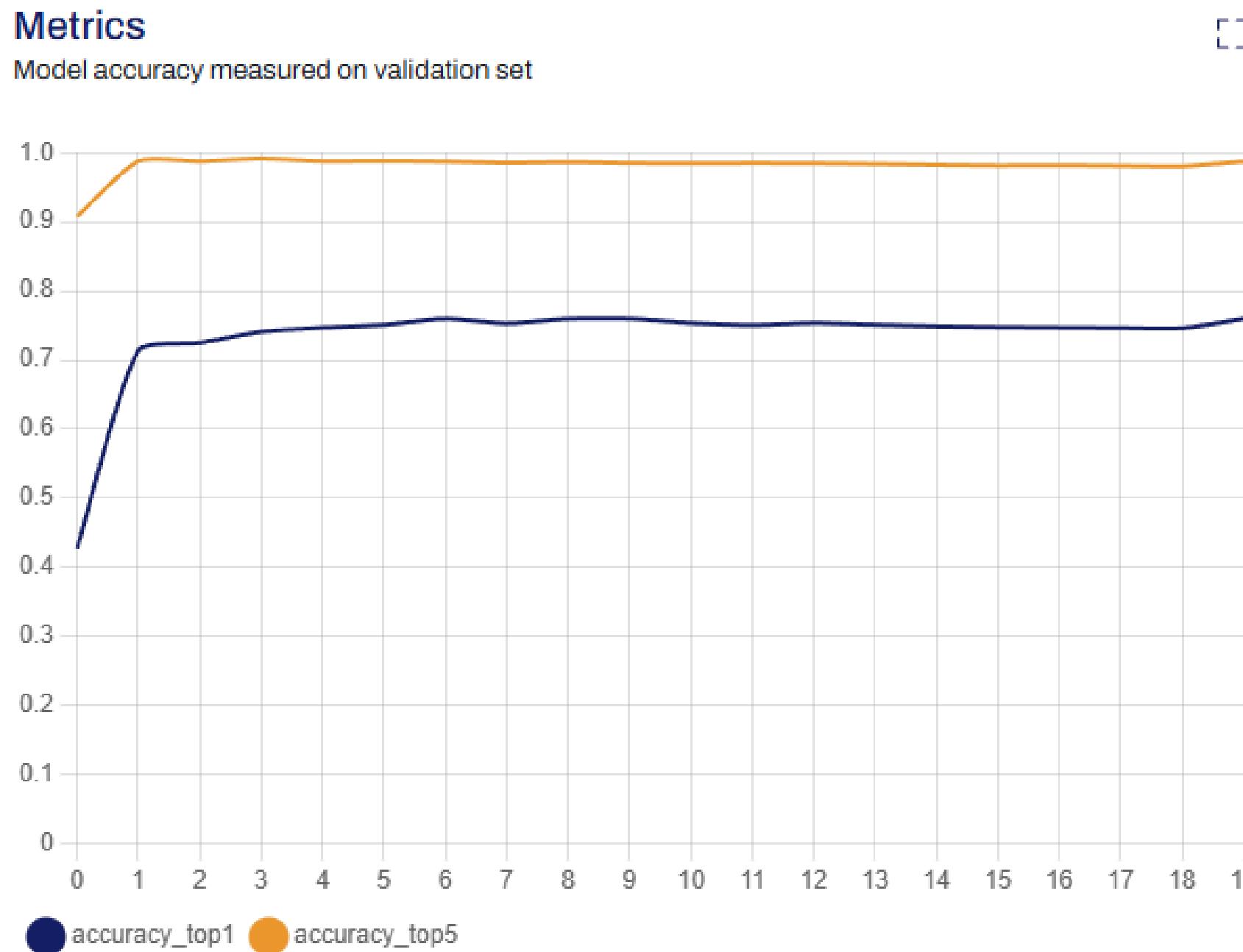
EMOTION CLASSIFIER RESULTS

PATH TO ULTRALYTICS HUB: [HTTPS://HUB.ULTRALYTICS.COM/MODELS/6KXLXRSMOYFROEPTU0BP](https://hub.ultralytics.com/models/6kxLxRsM0Yfr0EPtU0bP)

```
20 epochs completed in 2.017 hours.
Optimizer stripped from /content/runs/classify/train/weights/last.pt, 3.2MB
Optimizer stripped from /content/runs/classify/train/weights/best.pt, 3.2MB

Validating /content/runs/classify/train/weights/best.pt...
Ultralytics 8.3.233 🚀 Python-3.12.12 torch-2.9.0+cu126 CUDA:0 (NVIDIA L4, 22693MiB)
YOLOv1n-cls summary (fused): 47 layers, 1,536,272 parameters, 0 gradients, 3.2 GFLOPs
WARNING ⚠️ Skipping /content/datasets/ferplus.zip unzip as destination directory /content/datasets/ferplus is not empty.
train: /content/datasets/ferplus/train... found 66379 images in 8 classes ✅
val: /content/datasets/ferplus/validation... found 8341 images in 8 classes ✅
test: /content/datasets/ferplus/test... found 3573 images in 8 classes ✅
      classes  top1_acc  top5_acc: 100% ━━━━━━━━ 418/418 20.2it/s 20.7s
      all        0.76      0.987
Speed: 0.5ms preprocess, 0.6ms inference, 0.0ms loss, 0.0ms postprocess per image
Results saved to /content/runs/classify/train
Ultralytics HUB: Syncing final model...
: 100% ━━━━━━━━ 3.1MB 2.7MB/s 1.1s
Ultralytics HUB: Done ✅
Ultralytics HUB: View model at https://hub.ultralytics.com/models/6kxLxRsM0Yfr0EPtU0bP 🚀
```

EMOTION CLASSIFIER RESULTS



OVERALL MODEL ACCURACY: 0.76

EMOTION CLASSIFIER RESULTS

Test

Preview your model

[Image](#) [Camera](#)



Settings

Image Size

320px [640px](#)

Confidence Threshold

0.25

IoU Threshold

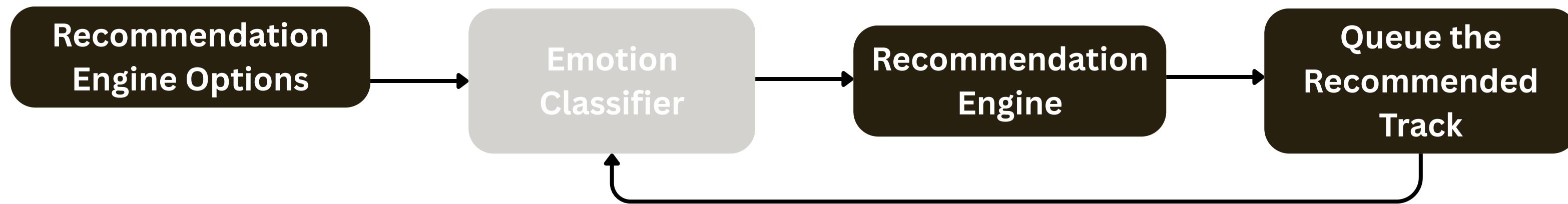
0.45

Inference results

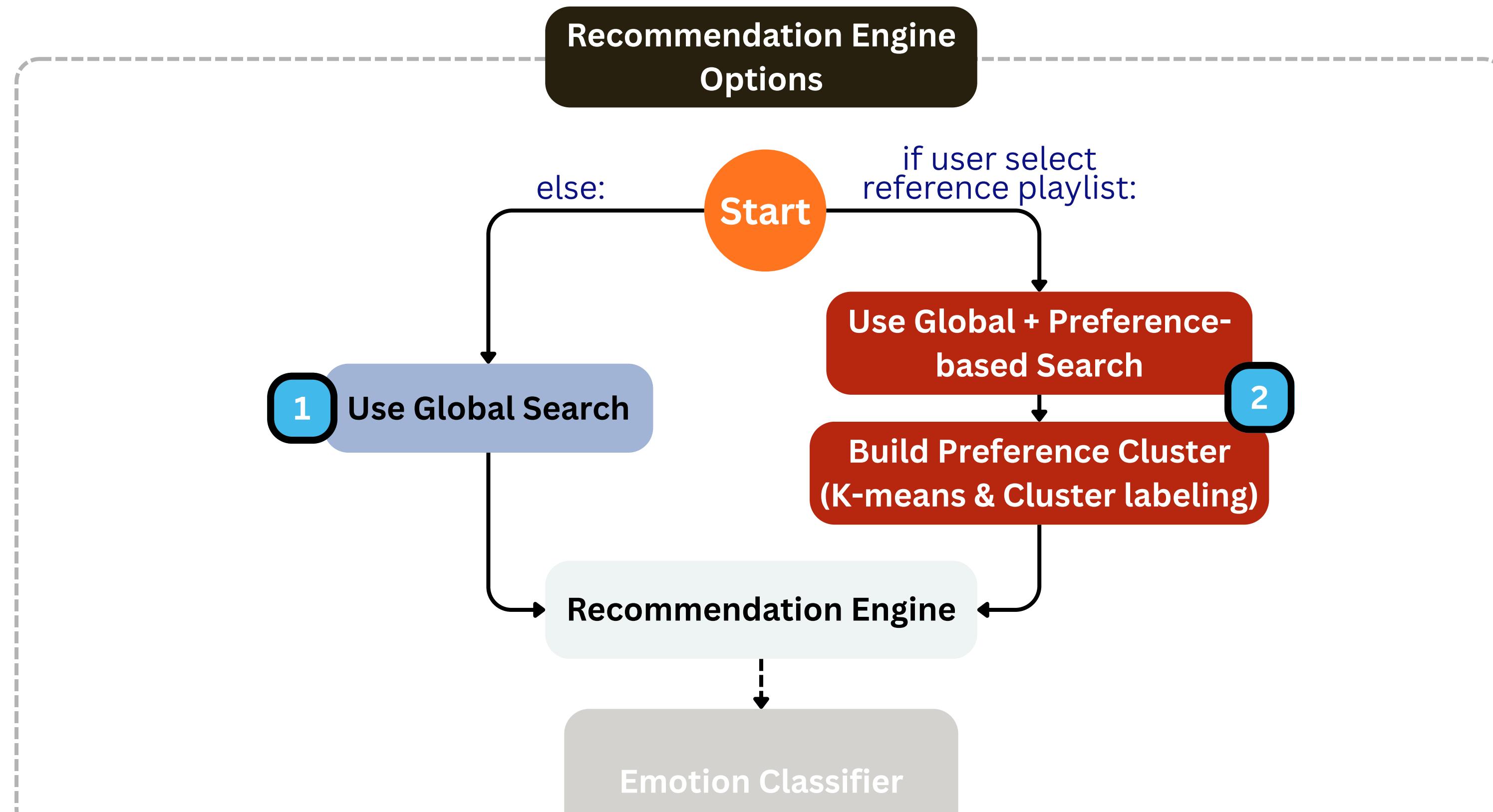
happy

44.1%

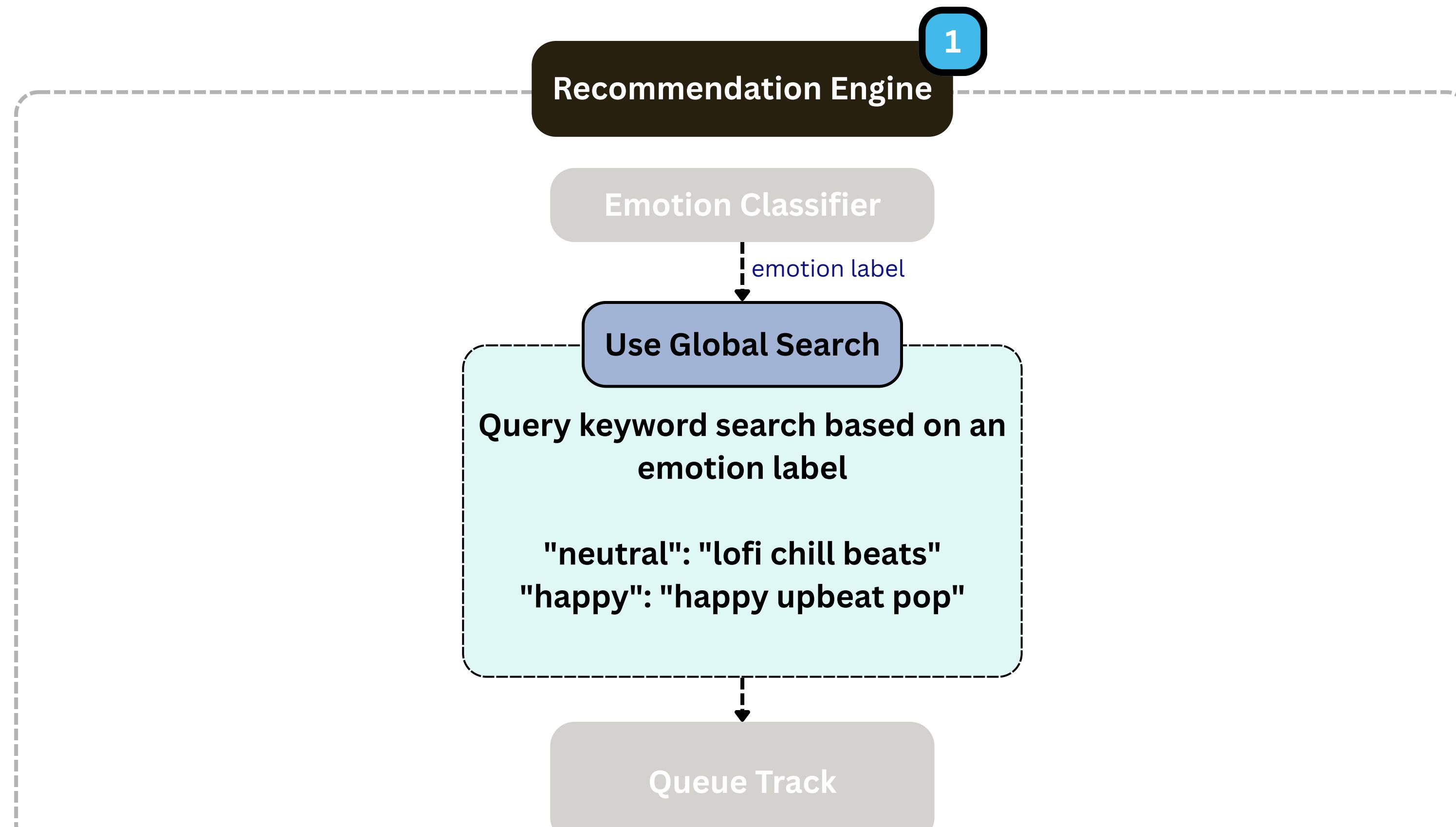
SONGS RECOMMENDATION



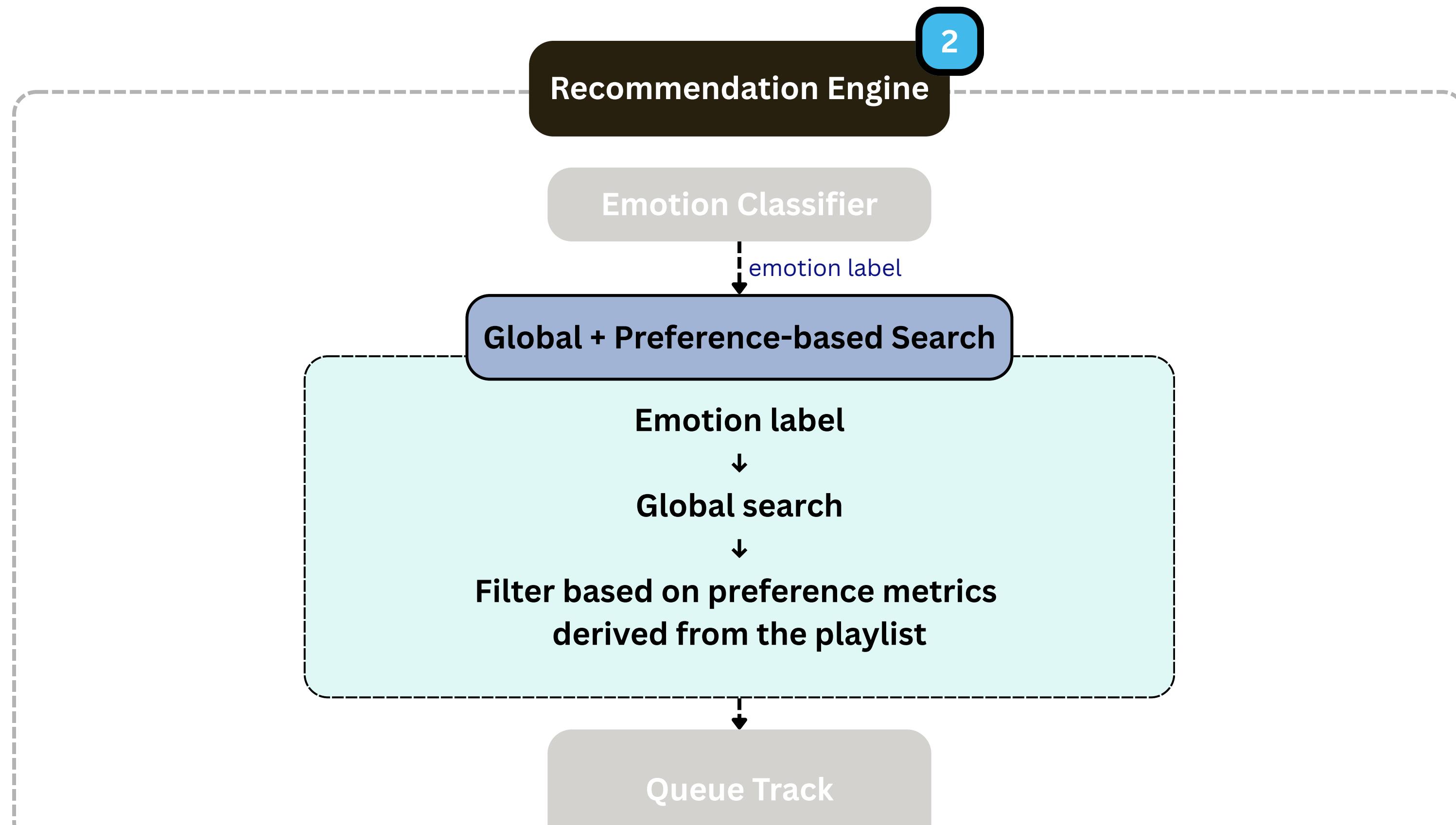
SONGS RECOMMENDATION



SONGS RECOMMENDATION



SONGS RECOMMENDATION



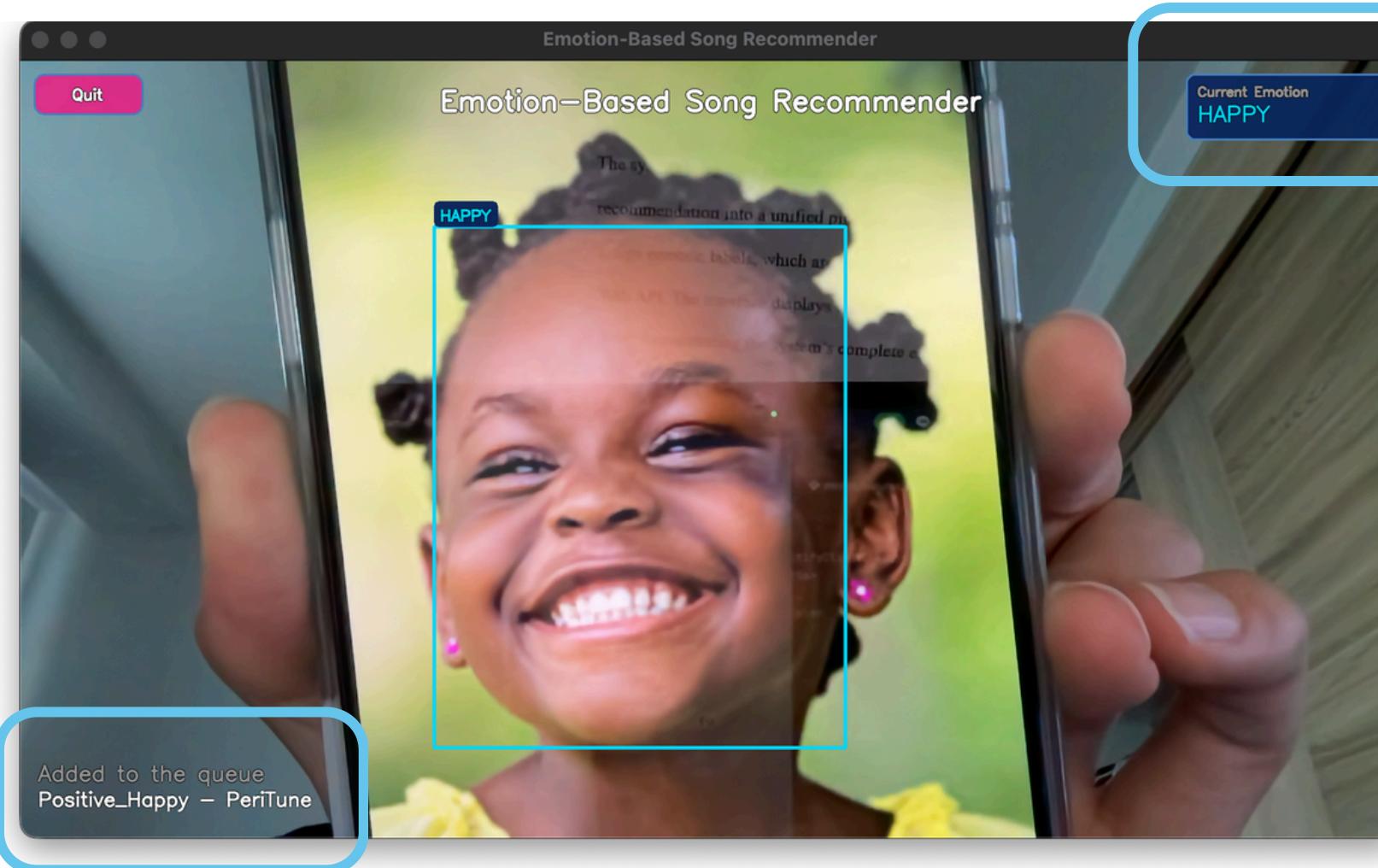
RESULT

****Disclaimer****

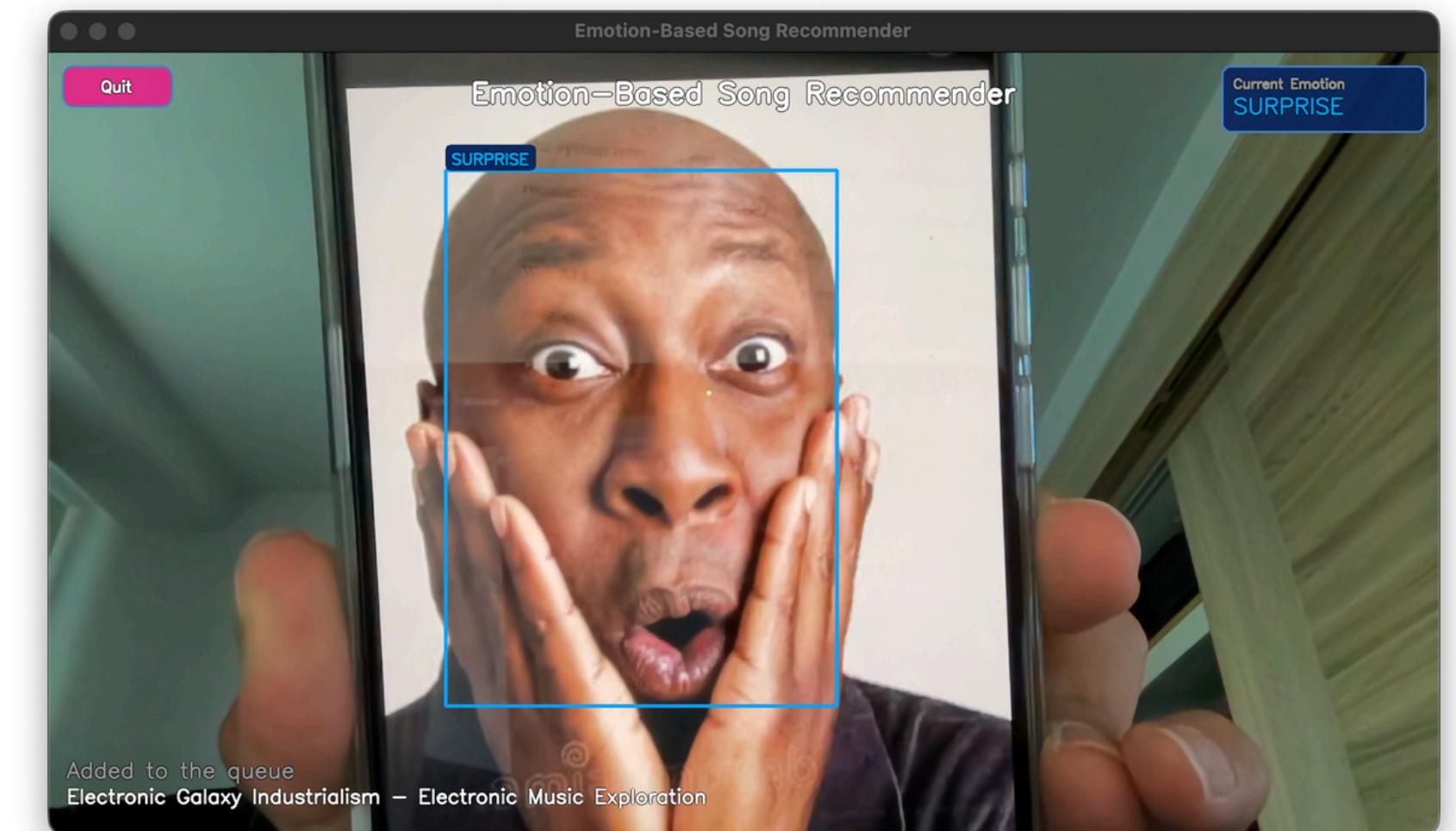
We are not great at acting so we used the photos of unknown person to test the model

RESULTS

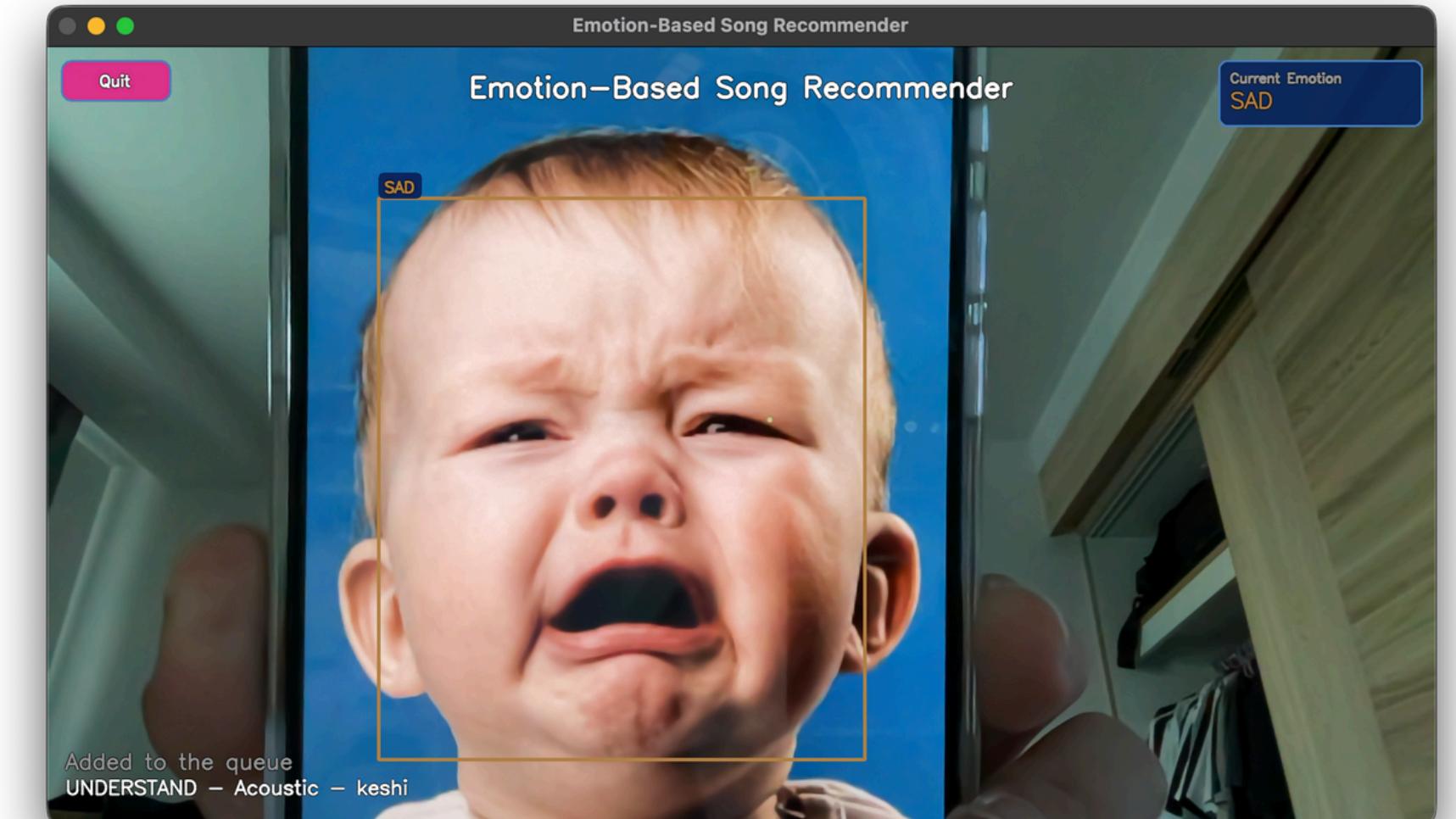
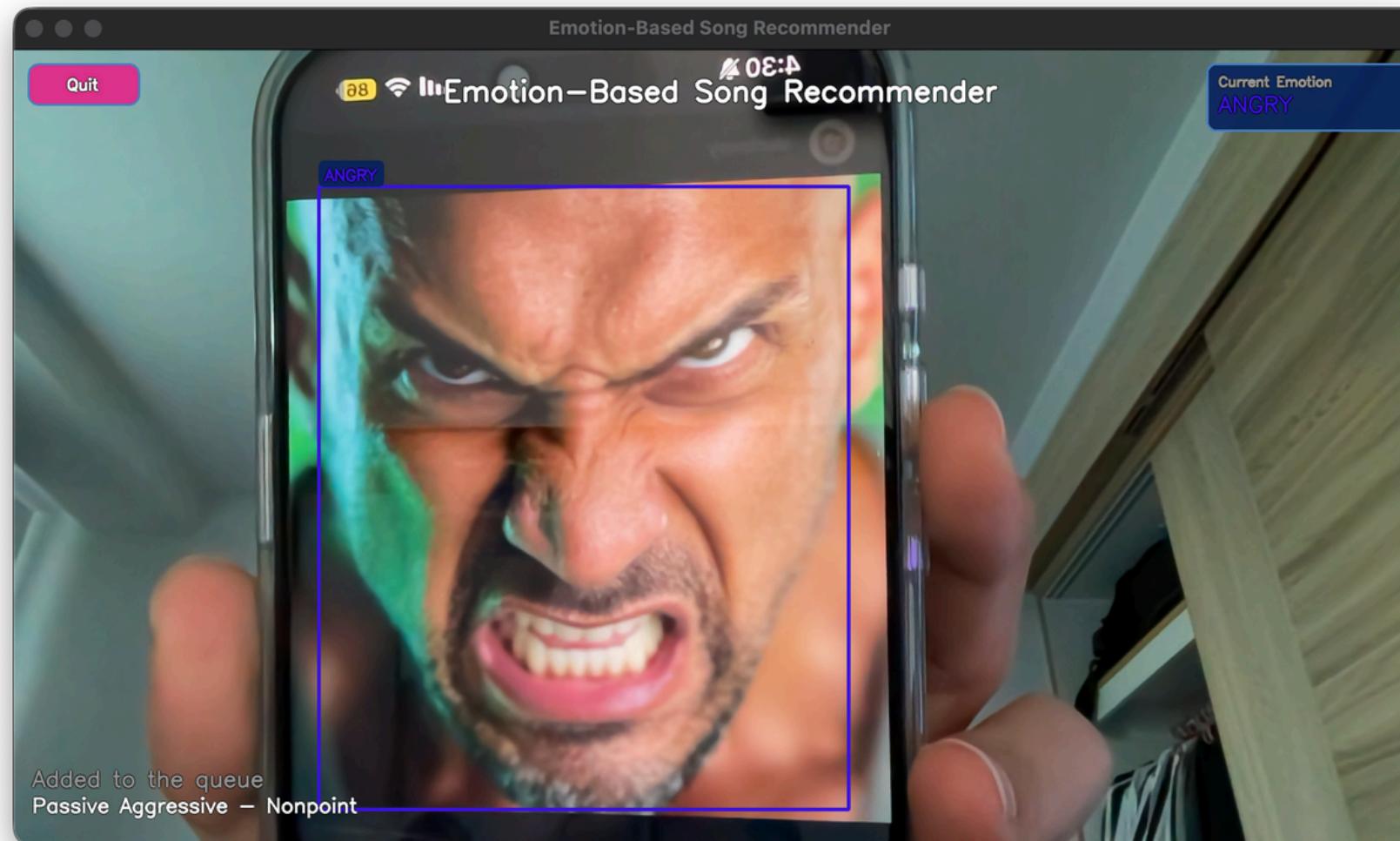
Emotion Label



Emotion Label

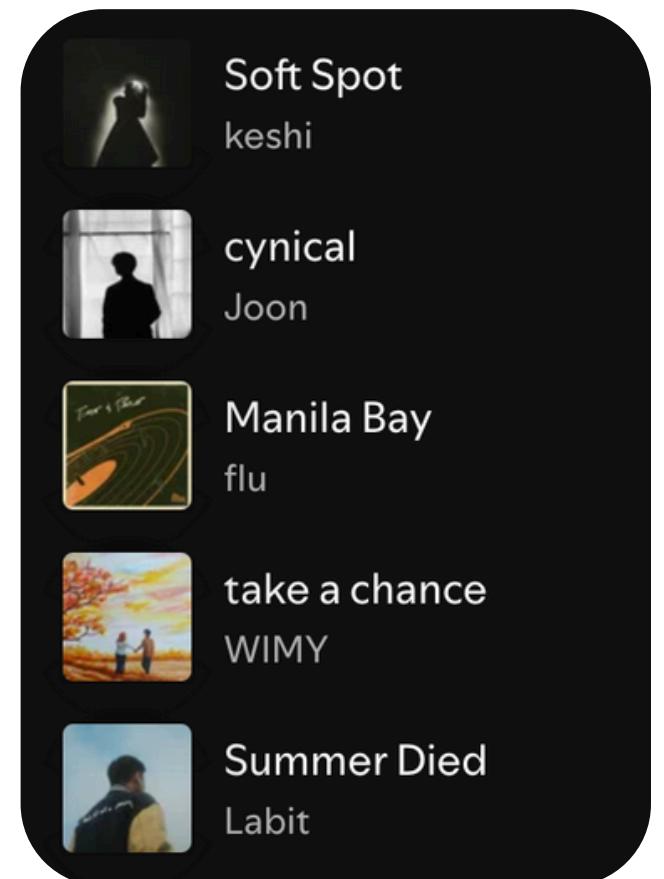


RESULTS

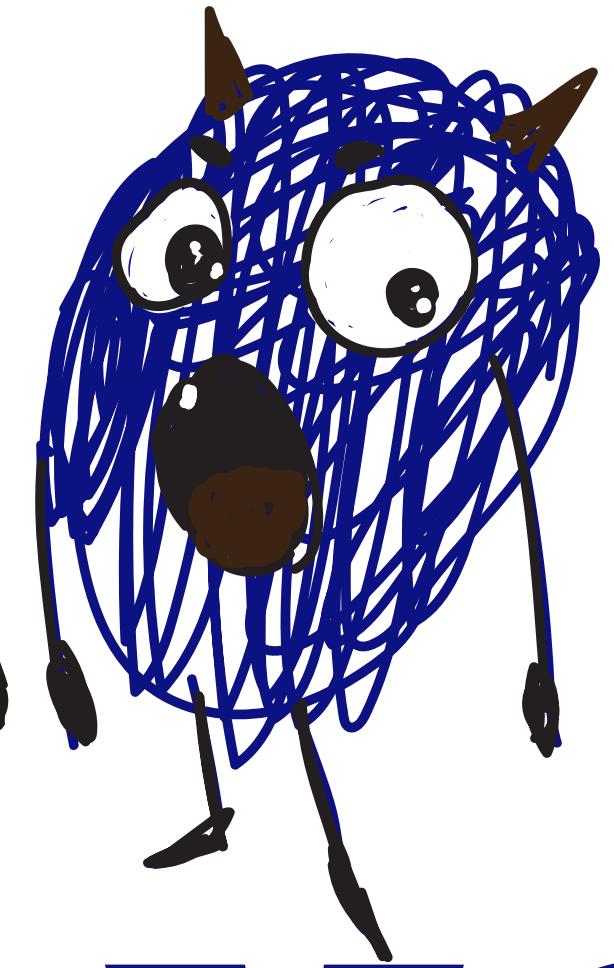
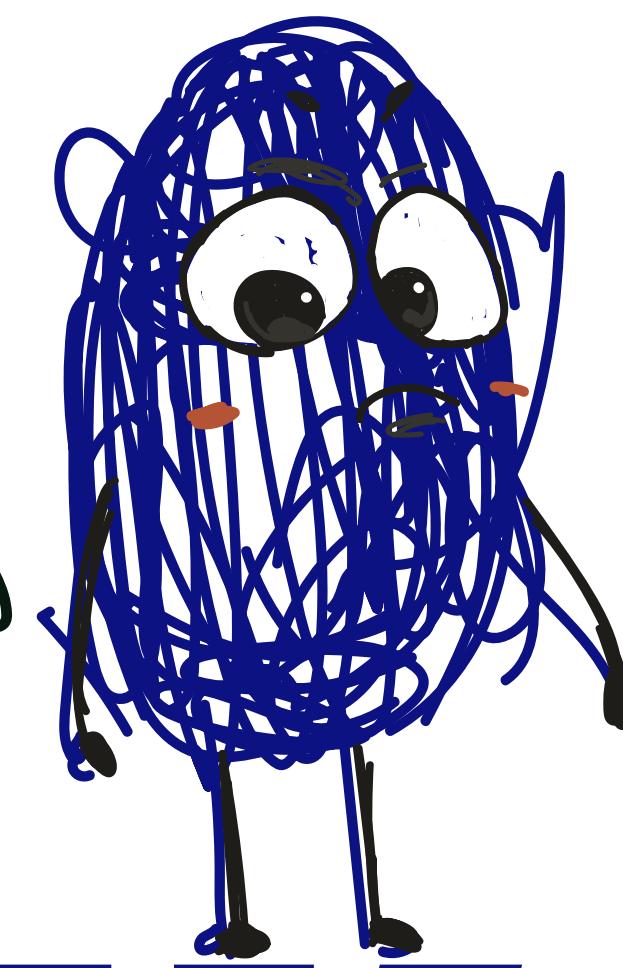


LIMITATIONS

- Emotion detection accuracy depends on lighting/camera quality
- Requires Spotify Premium
- The recommendation model relies only on numeric acoustic features (valence, energy, tempo, etc.), not lyrics or mood of the songs
- As Spotify removed the API for audio features, so the clustering evaluation rely on an older reference file instead.



THANK YOU



U

DEMO

