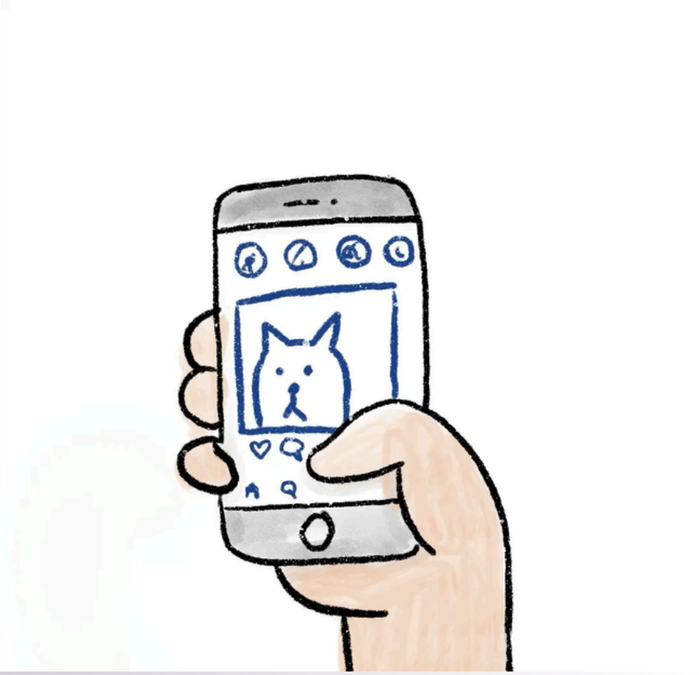
Emotion detection and personalized behaviour prediction



Students:
Medeea-Elena Geanovu
Octavian-Mihail Gheorghe
Mihai

Coordinator:
PhD Dr. Kuderna-Iulian Benţa

Babeş Bolyai Universtiy
Cluj-Napoca, Romania

Faculty Of Mathematics and
Computer Science

STEP 1 | Find Domain and Existing Research

OBJECTIVE

Develop an app capable of detecting human emotions from real-time video stream, in response to a stimulus (a topic based video) and analyze its effects on the viewer's affective state.

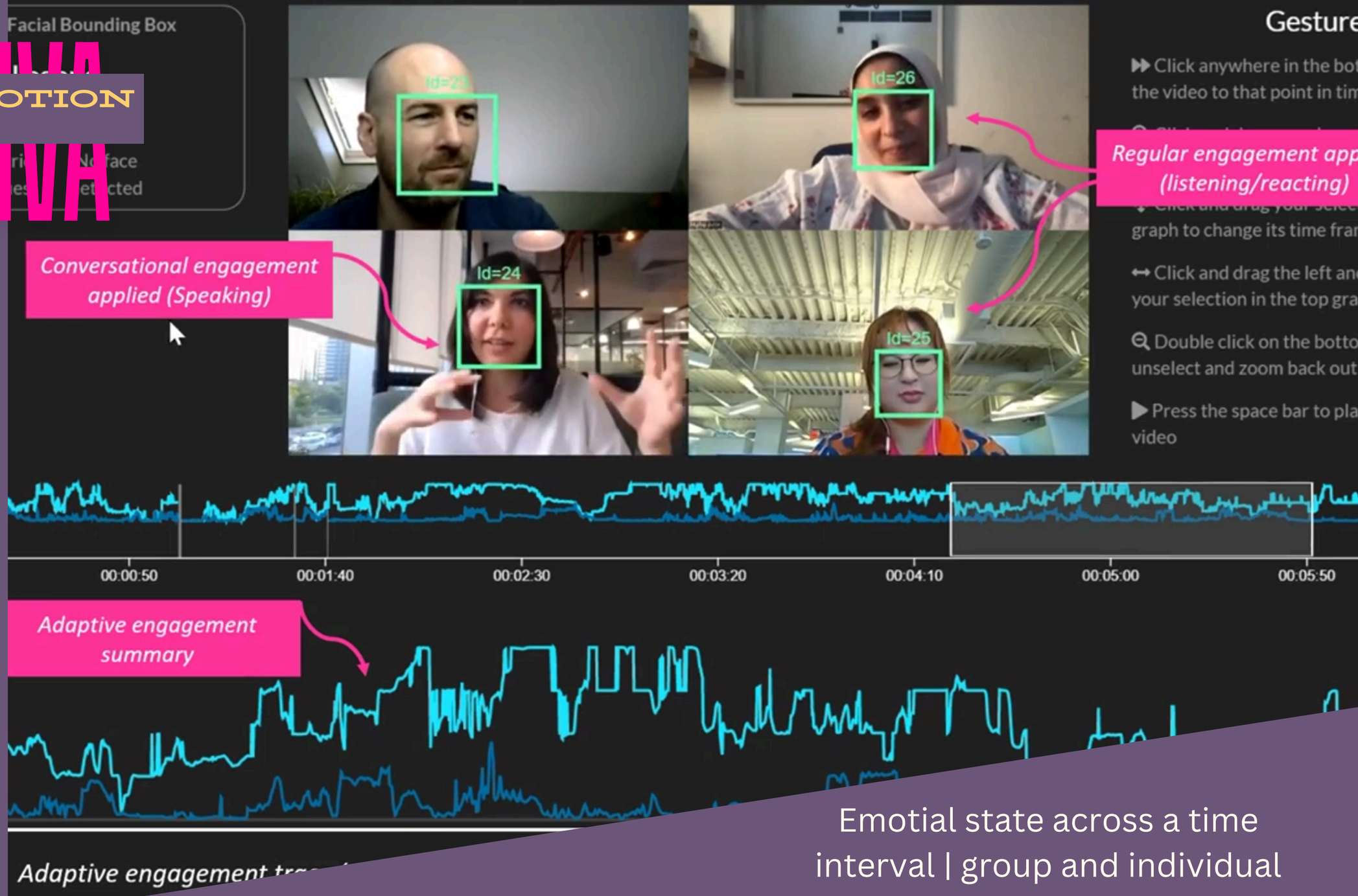
REASON

While existing projects focus on recognition of the affective state, they often overlook the specific influence of video content on the emotional experience.

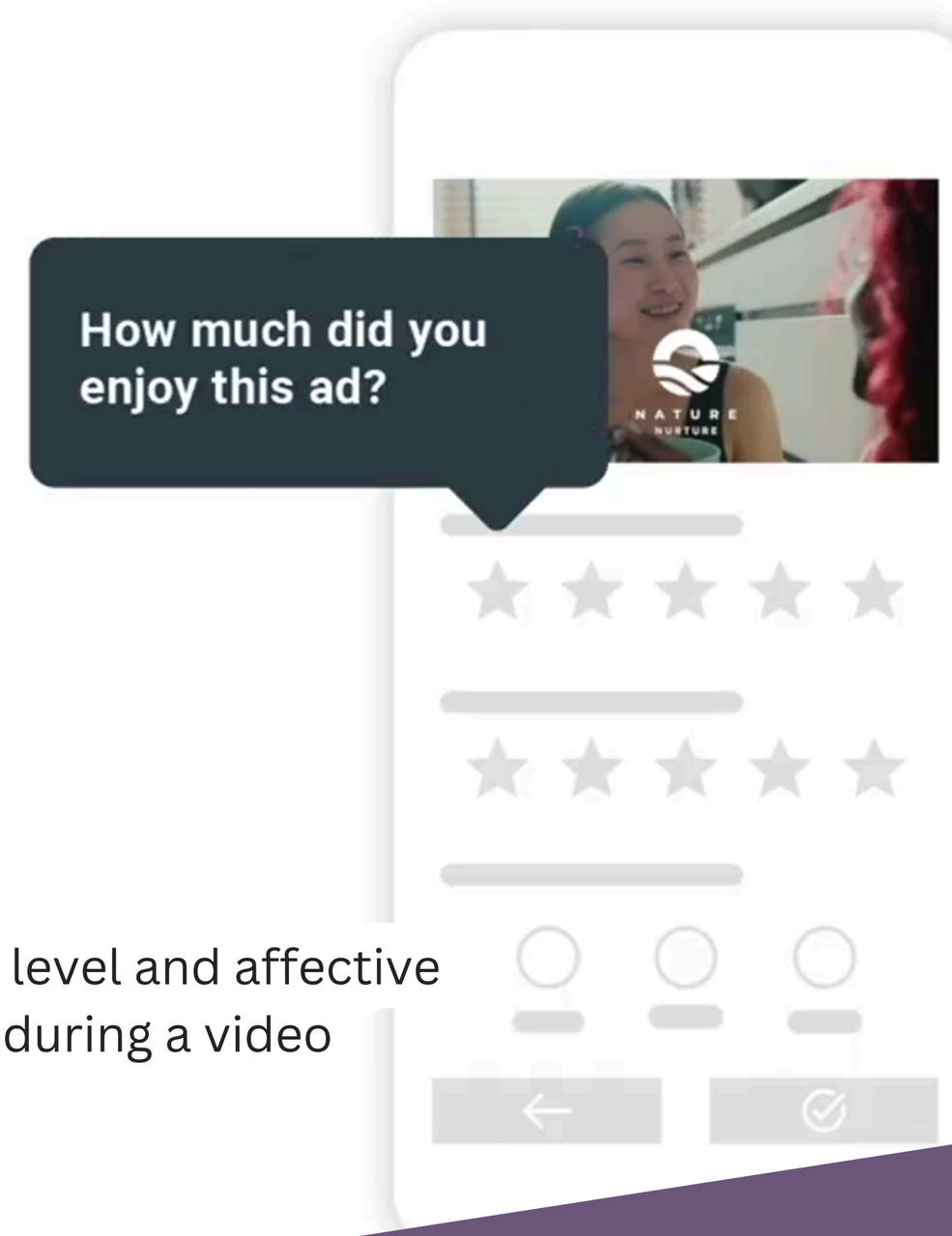
AFFECTIVA

EMOTION AI

- CNN
- 9 states, 30 classifiers
- Affectiva DB



Emotial state across a time interval | group and individual

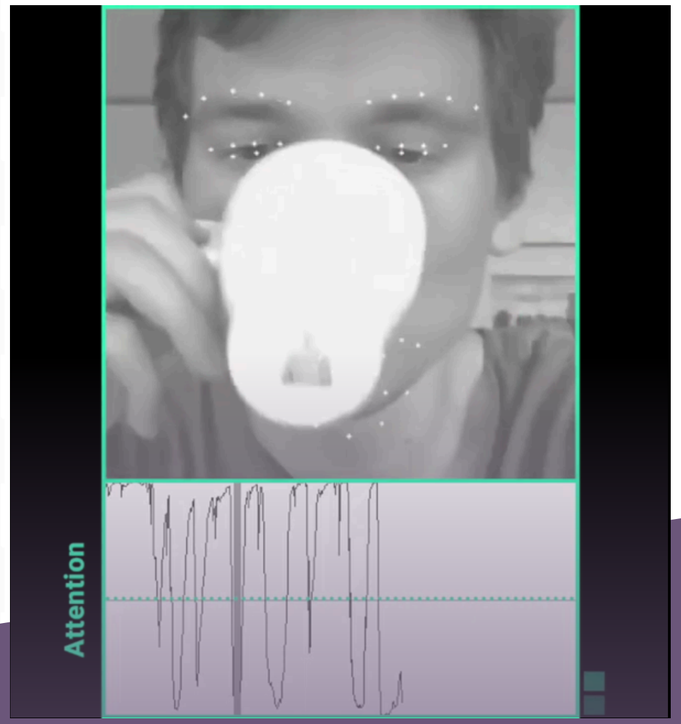


Attention level and affective state during a video

REALEYES

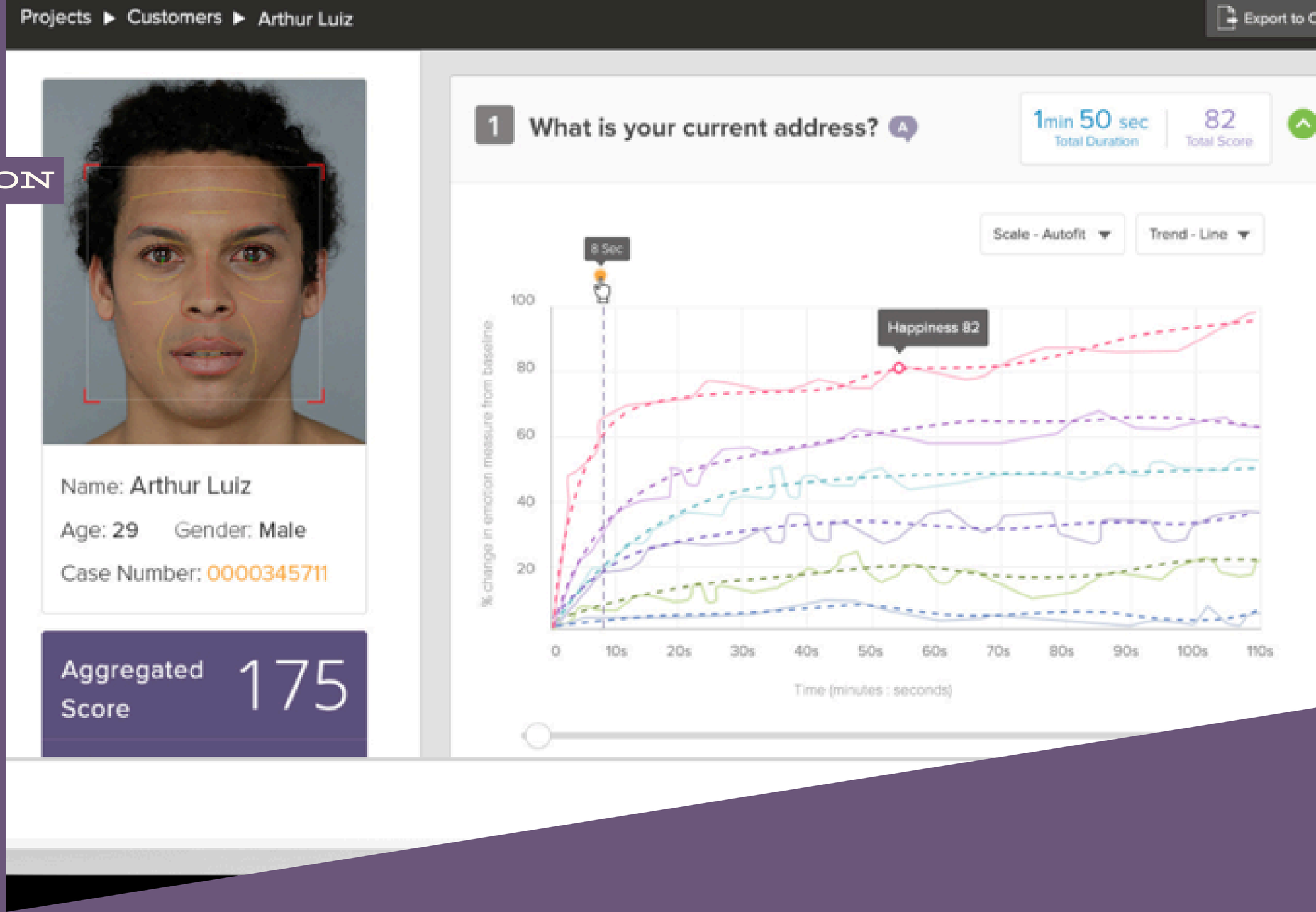
ATTENTION & EMOTION

- CNN
- ‘In the wild’ datasets
- Personal baseline



NVISO EMOTION

- CNN
- 7 states, 100+ landmarks
- Non-deterministic approach



STEP 2 | Research Question

MAIN QUESTION

How do different types of video content affect viewers' emotions, mental states, and decision-making processes?

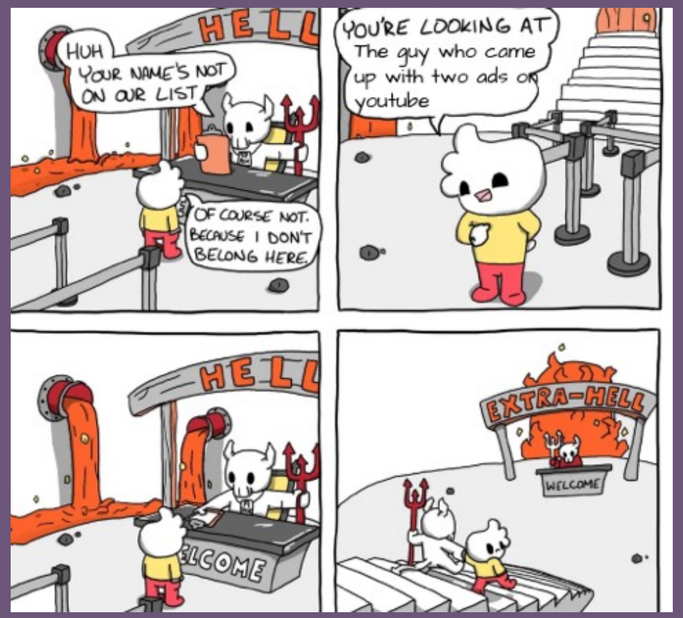
Can sentiment analysis be employed while viewing videos to predict lasting effects in viewer state and decisions?

AAAAND?

STEP 3 | Hypothesis

Video content related to specific topics (e.g. charitable , product advertisements, personal stories, commentaries) elicits distinct emotional responses that can be captured through real-time analysis. These emotional responses can be used to monitor viewers' mental states which in turn, can help prevent through their report influence on immediate decisions, like mood shifts, or longer-term choices, such emotional state, behavior or opinion changes.

Doomscrolling
Rage-Bait
Short attention and emotion?



STEP 4 | Testing and Analysis

Test the system with participants, recording and analyzing their emotional changes in real-time as they watch various types of content directly on their front-facing phone and desktop camera.

REAL-TIME CAPTURE

INSIGHT GENERATION

Identify correlations between content type, emotional response, and any reported changes in decision-making.

1. Face Detection and Tracking

MediaPipe for real-time face tracking and OpenCV for pre-processing and feedback. IP Webcam to stream the video content to the server.

VGG Image Annotator (VIA) for segmenting the video to be presented to the viewer into organized units - the topics corresponding to the entities and actions.

References:
Introduction to MediaPipe [1]
VGG Image Annotator (VIA) [2]
Lightweight real-time hand segmentation leveraging MediaPipe landmark detection [3]
Facial Landmarks Detection System with OpenCV MediaPipe and Python using Optical Flow (Active) Approach [4]
Real-Time Head Pose Estimation FaceMesh with MediaPipe and OpenCV: A Comprehensive Guide [5]

2. High-Frame-Rate Processing

Gradientfaces as a real-time pre-processing step to enhance frame consistency.

Principal Component Analysis (PCA) for dimensionality reduction before merging features.

References:
Edge detection examples using gradients [1]
PCA Based Human Face Recognition with Improved Method for Distorted Images due to Facial Makeup [2]
Unsupervised Method for Face Photo - Sketch Synthesis and Recognition [3]

3. Feature extraction

Landmark Detection with MediaPipe. FACS will be employed to track facial features in detail (Action Units)

Vision for the future: 3D Facial Modeling for improved accuracy across different angles.

References:
Detecting Facial Landmarks on 3D Models Based on Geometric Properties – A Review of Algorithms, Enhancements, Additions and Open-Source Implementations [1]
3D Shape-based Face Recognition Using Automatically Registered Facial Surfaces [2]
Facial Action Coding System (FACS) – A Visual Guidebook [3]

4. Emotion Classification

Convolutional Neural Networks (CNNs) trained on labeled emotion datasets (e.g. happiness, sadness, anger) - ResNet50, MobileNetV2, LightCNN-9/25. K Nearest Neighbour for supervised learning.

More sparkle: Bayesian Networks to handle ambiguities in emotional classification.

References:
Facial Expressions Recognition system using Bayesian Inference [1]
Dynamic adaptive threshold based learning for noisy annotations robust facial expression recognition [2] [2.a]
Affectiva model [3] Which CNNs and Training Settings to Choose for Action Unit Detection? A Study Based on a Large-Scale Dataset [3.1]
Automatic Detection of Sentimentality from Facial Expressions [3.2]
Combining Gradientfaces, principal component analysis, and Fisher linear discriminant for face recognition [4]
Computational Creative Advertisements [5]

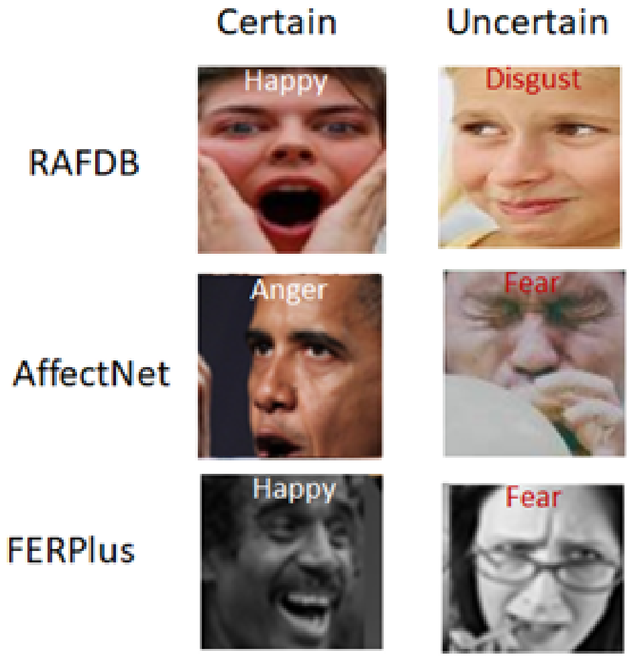
5.

User Emotion Patterns and Trends Analysis

K-Means for unsupervised learning to identify user-specific emotional patterns and biases.

Hope for improvement: The result is then presented to the user in a natural language manner in a dashboard.

DATASET



AffectNet seems most doable

Add new images to custom dataset

AM-FED for action units

SCHEMA

OVERVIEW

Gradientfaces and Principal Component Analysis (PCA) for High-Frame-Rate Processing.

Landmark Detection and CNN will be implemented to identify and classify critical facial points.

Unsupervised learning for personal trends and adjusting classification results.

Bayesian Networks for distinguishing between ambiguous emotional states.

Perhaps? NLP algorithms tocategorize and find more relations between video topics based on keywords in transcripts.

Thank You