Technical Report

EngageAI

MSIS 522 Purple Cohort Team 1

Ephrem Tilahun, Megan Louie, Parham Hajzavar, Regina Geng, Zula Battulga

February 14, 2025

Table of Contents

1 | Team Information

1.1 Team members:

- Epherem Tilahun
- Megan Louie
- Parham Hajzavar
- Regina Geng
- Zula Battulga

1.2 Individual contributions:

- **Ideation** – Ephrem, Megan, Parham, Regina, Zula
- **Research** – Ephrem, Megan, Parham, Regina, Zula
- **Data Collection** – Megan, Zula, Parham
- **Feature Engineering** – Ephrem, Parham
- **Model Selection** – Ephrem, Megan, Parham, Regina, Zula
- **Hyperparameter Tuning** – Ephrem, Parham
- **Model Training & Evaluation** – Ephrem, Parham
- **App Development** – Ephrem, Parham
- **Testing** – Ephrem, Megan, Parham, Regina, Zula
- **Visualization & Reporting** – Ephrem, Megan, Parham, Regina, Zula
- **Presentation Preparation** – Ephrem, Megan, Parham, Regina, Zula

2 | Executive Summary

The rise of virtual learning has transformed education by providing unprecedented global access to knowledge. While this digital shift has enabled flexible learning and collaboration across borders, it presents significant challenges that will be explored in this report. Traditional engagement monitoring methods have proven insufficient for digital environments, leaving educators struggling to optimize virtual learning while maintaining student engagement.

EngageAI emerges as a pioneering solution to this challenge, introducing an AI-powered engagement analytics platform that transforms how educators measure and respond to student engagement. The system employs sophisticated facial recognition technology and deep learning models to analyze student reactions in real-time, detecting nuanced expressions of confusion, boredom, and attentiveness. This objective measurement system enables educators to move beyond the limitations of "teaching into the Zoom void" and make data-driven decisions about their instructional approaches.

Our technical implementation features a lightweight yet powerful web application architecture that prioritizes both performance and security. The frontend serves as an intuitive visualization layer, allowing instructors to upload recorded lecture videos and receive comprehensive engagement analytics dashboards within minutes. The backend, built on robust deep learning frameworks including MobileNetV2 and ResNet50, handles all computational-intensive tasks and sensitive data processing. To ensure data privacy and security, the system implements a self-contained architecture that automatically purges all analyzed videos and generated data after each session, maintaining strict compliance with FERPA regulations.

The development process leveraged a dataset of over 15,000 facial emotion images, though careful curation reduced this to 11,000 high-quality samples after removing corrupted or irrelevant data. Through sophisticated data augmentation techniques, we enhanced the model's resilience to variations in lighting, noise, and orientation, ensuring robust performance across diverse real-world scenarios. The system's architecture was specifically designed to scale, accommodating both individual instructor needs and institutional-level deployment.

EngageAI's potential extends far beyond higher education into K-12 classrooms, EdTech platforms, and corporate training environments. With the majority of instructors struggling to maintain student engagement in online learning, our technology addresses a critical market need across all digital education sectors. By equipping educators with real-time, actionable insights, ***EngageAI*** is transforming how we teach and learn—ultimately creating more engaging, personalized, and effective learning experiences for everyone.

3 | Business Problem

3.1 The New Age of Learning

As students and educators, we know far too well that we moved into online learning at full throttle. The COVID-19 pandemic forced educational institutions worldwide to pivot abruptly to virtual learning, reshaping the landscape of education in ways that continue to evolve. Five years later, platforms like Zoom, Teams, and virtual meetings have become ingrained in academic and professional settings, and it's safe to say they aren't going anywhere. However, this new era of digital learning has created significant challenges for professionals at the forefront of education.

Before the pandemic, online learning was often seen as an alternative or supplementary option, primarily utilized for distance learning programs. However, post-pandemic, virtual classrooms have become a core component of mainstream education, forcing institutions to rethink engagement, interaction, and instructional design. While the shift has brought undeniable benefits such as increased accessibility and flexibility, it has also introduced challenges in maintaining student engagement, fostering interaction, and ensuring effective knowledge retention.

Research has indicated that many students and instructors find virtual classrooms to be less interactive and less effective for learning (Walker and Koralesky). The lack of in-person cues, such as facial expressions, non-verbal gestures, and real-time feedback, has led to a significant drop in perceived engagement. With 18.1 million students in the U.S. engaging in online learning activities daily, a staggering 57% of those students do not pay attention in class (Peck). While remote learning is here to stay, it is clear that the education system must evolve beyond traditional teaching strategies to address these pressing challenges and ensure students receive a high-quality learning experience that mirrors or surpasses in-person instruction.

3.2 Market Gap & Pain Points

Despite the surge in online education, a fundamental gap remains in accurately measuring student engagement. Current engagement tracking tools rely heavily on manual observation, attendance records, and voluntary participation, none of which provide continuous, objective insights into student attentiveness. Instructors are left with limited visibility into student engagement levels, making it difficult to identify when learners are disengaged, confused, or unresponsive. Many institutions and universities still depend on outdated, non-automated methods that fail to scale across diverse student populations, resulting in inconsistent learning outcomes and reduced instructional effectiveness.

While online education offers unparalleled flexibility and accessibility, it also presents significant challenges for educators. Studies show that 74% of instructors report increased stress from online teaching (BMC Medical Education). The lack of real-time engagement analytics forces educators to guess whether students are absorbing content effectively, rather than relying on concrete data. Traditional tools such as chat box interactions and post-session surveys offer subjective, delayed feedback and fail to capture real-time indicators of engagement, such as facial cues of confusion or distraction. This issue is further exacerbated by the absence of immediate feedback loops, reducing instructor adaptability and limiting their ability to intervene proactively. Research indicates that 60% of instructors and 57% of students observed lower engagement levels after transitioning to online learning (Walker and Koralesky 7). Given that student engagement is a key predictor of academic success, this trend raises significant concerns about the long-term effectiveness of virtual education.

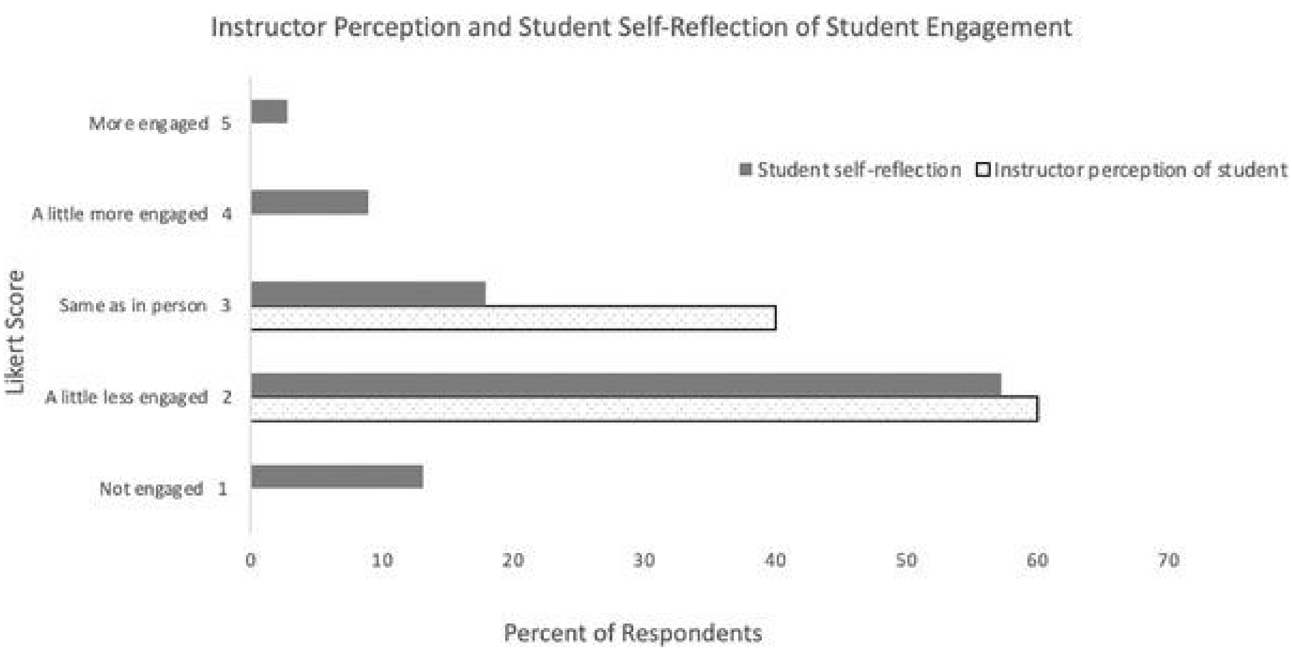


***Figure 1**. Instructor perception and student self-reflection of student engagement*

Furthermore, educators are burdened with increased workloads due to the demands of remote teaching, making it unrealistic for them to manually track engagement for every student. Institutions lack the ability to quantify engagement trends over time, making it difficult to optimize curriculum design and ensure that students remain actively involved in their learning journeys. The absence of real-time insights means that many students who struggle with remote learning remain undetected until it's too late, contributing to lower academic performance and higher dropout rates. Addressing these challenges requires a data-driven, scalable approach that empowers educators with real-time insights into student engagement, allowing them to adapt instruction dynamically, enhance learning experiences, and ultimately improve educational outcomes.

4 | Solution Approach & Design Process

4.1 Business Solution

The primary business problem ***EngageAI*** solves is the lack of an objective engagement measurement system for virtual and hybrid learning environments. The rigid dichotomy between instructors and online learning is evident, but ***EngageAI*** seeks to empower instructors to make data-driven decisions about their virtual teaching approaches. Instructors must be able to effectively measure engagement, identify at-risk students, and optimize lesson delivery based on real-time behavioral data.

One of the biggest gaps in outdated engagement tracking tools is their inability to provide continuous insights, unlike the superpowers of AI. ***EngageAI*** helps instructors proactively adjust their teaching strategies to prevent disengagement before it negatively impacts learning outcomes. It leverages AI-powered facial recognition and deep learning models to analyze student reactions and generate engagement analytics for instructors to reflect on post lectures. The system detects facial expressions indicative of confusion, boredom, and attentiveness, allowing educators to adjust their delivery styles accordingly. By offering automated reports and trend analysis, it allows institutions to track engagement at both the individual and cohort levels, ensuring that instructors have the insights needed to intervene and enhance student performance.

4.2 Design Process

When we set out to build ***EngageAI***, our primary objective was to create an intuitive and easy-to-understand product. This was central to the way we envisioned the solution even before we had written a single line of code. With such limited time for both ideation, design, and development, we understood that we needed to keep our focus narrow and the features intentional, and incorporated design principles from Lean in order to maintain tunnel vision during our sprint. It also led us to the conclusion that this product is not only valuable for higher education instructors, but for any instructor teaching online and interested in self-reflecting on their content and lecture delivery. Through AI, we are able to produce valuable insights for instructors who struggle to gauge their classrooms' engagement levels, an industry-wide challenge sometimes referred to as "teaching into the Zoom void." (Barba) Since most online lectures are already recorded for students who are absent, this helped us better understand how our AI solution can be most effective.

We designed our web app frontend to simply take a recorded lecture video as input, and to generate in under a minute a modern analytics dashboard with various engagement summaries and graphs, available for download to review offline later as needed. This dashboard serves as a guide for instructors to pinpoint peaks and dips in engagement levels, and better understand which parts of the lecture were successful and which might need improvement. We intentionally kept our frontend lightweight and only responsible for the visualization layer – pushing all the heavy lifting such as data processing, facial emotion analysis, and engagement scoring computation to the backend. This makes our system more secure, self-contained, and eliminates the need to expose our API. Recognizing the gravity of collecting facial biometric data, we decided that we were not going to save student info in production. After each session, the lecture video that is uploaded by an instructor and any analysis generated using the video is wiped from our system, leaving us safe from most cyber attacks.

4.3 Data & Methodology

Like all AI applications, the model is only ever going to be as good as the data it was trained on and validated against. We leveraged a free dataset found on Kaggle that consisted of over 15,000 images of various facial emotions such as anger, neutrality, happiness, and surprise. While cleaning this dataset, we had to drop at least 4,000 rows since they either contained corrupted images or nothing at all. From the data that was left, the images were mostly crops from larger images scraped from across the internet, like celebrities and stock images, many of which were exaggerated and not easily mappable to emotional emotions one would find in a Zoom lecture. We augmented these images to increase our dataset size, improve generalization, and boost our model's robustness to be resilient to variations in noise, lighting, and orientation changes.

Looking at the code, particularly dataset.py, we implemented a comprehensive data augmentation pipeline for training our emotion recognition model. Here's the detailed technical breakdown:

Data Augmentation Strategy: In dataset.py, we implemented multiple augmentation techniques specifically for training:

Spatial Augmentations: Initial resize to 256x256 followed by random crops to 224x224, creating position variance, Random horizontal flips with 50% probability, Random rotations up to 15 degrees, Random affine transformations:, Translations up to 10% in both directions, Scale variations between 90% and 110%, Color Space Augmentations:, Brightness adjustments up to ±20%, Contrast variations up to ±20%, Saturation changes up to ±20%, Hue shifts up to ±10%, Data Preprocessing Pipeline:

Normalization:
Used ImageNet statistics for normalization (mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225]), This helps with transfer learning since our ResNet-50 was pretrained on ImageNet
Video Preprocessing (prepare_video.py):

- Automatic orientation correction for portrait videos
- Standardization to landscape format
- Frame rate preservation during orientation fixes
- Progress tracking during video processing
- Efficient memory handling with frame-by-frame processing

Training/Validation Split Considerations:

- Separate transforms for training and validation
- Validation only uses resize and normalization without augmentations
- Uses ImageFolder structure, assuming emotion classes are organized in directories

The evaluation metrics show this pipeline was effective, achieving:

- Training accuracy: 87.92%
- Validation accuracy: 84.25% with minimal overfitting, suggesting our augmentation strategy successfully improved model generalization.

Processing Optimizations:

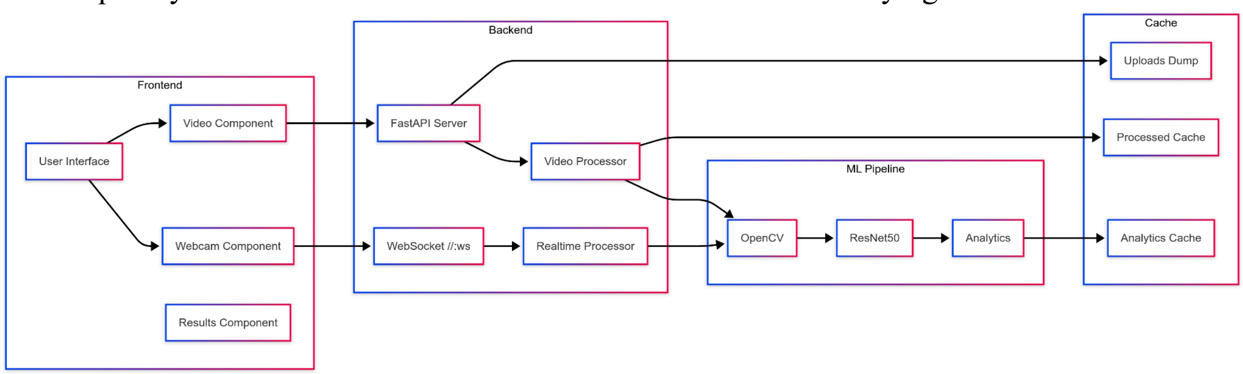
- Implemented efficient data loading with configurable num_workers
- Pin memory enabled for CUDA devices
- Batch size optimization based on available GPU memory
- Mixed precision training support for faster processing

4.4 Technical Implementation

The system is built with a multi-tier architecture:
Core Components from Source:

```
emotion_recognition/
├── backend/                # FastAPI services (processing and real-time)
├── frontend/               # React-based web interface
└── src/                    # ML pipeline components
    ├── inference.py        # Base inference engine
    ├── video_inference.py  # Video processing
    ├── realtime_inference.py # Real-time processing
    └── train.py            # Model training
```

The project evolved from a baseline photo inference system to a comprehensive emotion analysis platform. We began with fundamental scripts for different inference modes - photo, video, and real-time processing. The initial implementation focused on basic OpenCV-based face detection, but we quickly realized we needed more robust face detection for varying conditions.



The core ML architecture centers on our custom-trained ResNet-50 model for emotion classification. We adapted ResNet-50's convolutional layers and modified the final classification layers to output probabilities for five emotion states: Happy, Sad, Neutral, Angry, and Surprise.

The model weights were managed through Git LFS due to their size, and we developed a custom .pth weights converter to ensure compatibility across different PyTorch versions.

A significant upgrade came with the integration of Google's MediaPipe face detection, replacing our initial OpenCV Haar Cascade implementation. MediaPipe offered superior face detection accuracy and better handling of varying face angles and lighting conditions. We configured MediaPipe with a detection confidence threshold of 0.5 and implemented a padding system around detected faces to ensure we captured the full emotional context.

The backend service was built using FastAPI, implementing GPU acceleration for video processing. We optimized the video processing pipeline to handle 4 frames per second, striking a balance between real-time performance and accuracy. The service includes robust error handling and automatic GPU/CPU detection, falling back to CPU processing when GPU isn't available.

For real-time processing, we implemented a WebSocket-based system in the backend service, handling frame-by-frame analysis at 15 FPS. This required careful optimization of the frame processing pipeline, including JPEG compression for WebSocket transmission and efficient frame buffering to prevent memory leaks.

The frontend evolved from a basic video upload interface to a sophisticated React application. We implemented a custom webcam component that handles multiple camera sources, manages WebSocket connections for real-time processing, and provides immediate visual feedback with emotion annotations. The UI components were built with Material-UI, featuring a dark theme and responsive design.

We added an analytics dashboard that processes emotion detection results in real-time, generating engagement scores based on detected emotions. Each environment (classroom, seminar, product demo) has its own scoring algorithm, weighing emotions differently based on context. For example, in a classroom setting, sustained neutral expressions contribute positively to engagement scores, while in a product demo, higher weights are given to expressions of surprise and happiness.

The deployment architecture was containerized using Docker Compose, with separate containers for the frontend, backend processing service, and real-time inference service. This microservices approach allows for independent scaling of different components and simplified deployment across development and production environments.

Throughout development, we maintained strict testing protocols, with separate testing branches for model evaluation, WebSocket performance testing, and end-to-end system integration tests. Each feature branch (frontend, backend, docker, webcam, model) underwent thorough testing before being merged into the mainline.

The final system represents a full-stack implementation of emotion recognition, from low-level face detection and model inference to high-level analytics and user interface, all optimized for both real-time and batch-processing scenarios. *(Refer to the final image for a holistic system design in the appendix section)*

4.5 Results & Evaluation

From the training metrics shown, our emotion recognition model went through 50 epochs of training, with the screenshot showing epoch 8 reaching completion and epoch 9 at 60% progress.

```
Epoch 8/50: 100%|
[2025-02-03 19:25:38] Epoch 8/50:
[2025-02-03 19:25:38] Train Loss: 0.3151, Train Acc: 87.92%
[2025-02-03 19:25:38] Val Loss: 0.4419, Val Acc: 84.25%
[2025-02-03 19:25:38] New best model saved with validation accuracy: 84.25%
Epoch 9/50:  60%|
```

Training Performance:

- Train Loss: 0.3151
- Training Accuracy: 87.92%

Validation Performance:

- Validation Loss: 0.4419
- Validation Accuracy: 84.25%

The system logged this as a new best model based on the validation accuracy of 84.25%, triggering an automatic model checkpoint save. The gap between training accuracy (87.92%) and validation accuracy (84.25%) suggests a well-balanced model without significant overfitting, as the difference is only about 3.67%. This indicates our training strategy, likely including regularization techniques and our data augmentation pipeline, was effective in producing a robust model

Later on through just brute forcing repetitions, we got a model with 85.75% validation accuracy as the main model in epoch 12.

4.6 Limitations & Future Work

While advanced facial recognition technology can offer valuable real-time insights into student engagement, it is essential to recognize several key limitations to ensure a holistic understanding of learners' experiences. Relying exclusively on visual cues may overlook students who remain cognitively involved despite showing neutral expressions, or those who appear focused yet become distracted by non-visual factors such as background tabs or chat applications. Additionally, biases within the training data—ranging from underrepresented demographic groups to artificially posed expressions—can reduce model accuracy and fairness for certain student populations. Environmental and technical variations, including lighting inconsistencies, varying camera quality, and unstable bandwidth, further influence the system's ability to capture reliable data, particularly in large-scale real-time applications. Despite strict privacy protocols, continuous facial recognition can still raise ethical concerns for students who feel uneasy about being monitored, especially in regions governed by rigorous data protection regulations like GDPR or FERPA. Finally, although automated engagement metrics yield convenient insights, educators should avoid over-reliance on these scores alone; integrating qualitative feedback and a variety of pedagogical strategies remains critical for delivering a meaningful, student-centered learning experience.

EngageAI's technology is poised to evolve in tandem with broader advancements in AI, enabling widespread adoption across a range of educational and professional settings. While universities, K–12 programs, and EdTech platforms stand to benefit from integrated analytics, our substantial market potential also extends to corporate training. By empowering businesses to measure and enhance employee engagement in virtual sessions, we aim to boost knowledge retention and skill development.

Our core mission is to elevate the quality of online learning experiences for all. Recognizing the importance of regulatory and societal considerations, we are actively refining our model to respect privacy while delivering actionable insights. For instance, we plan to analyze facial meshes rather than storing full video feeds and to assign randomized IDs that do not reveal personal identities. These measures help ensure adherence to data protection standards, fostering trust and transparency among educators, learners, and organizations alike.

5 | Ethical & Legal Considerations

The development and deployment of ***EngageAI*** must prioritize ethical and legal considerations to ensure the system is both responsible and compliant with relevant regulations. First, privacy concerns are paramount, as the system relies on facial emotion recognition interaction analysis. It is essential to obtain explicit consent from students and educators before collecting and processing any data, ensuring compliance with data protection laws such as GDPR (General Data Protection Regulation) and FERPA (Family Educational Rights and Privacy Act). Additionally, the system must anonymize and securely store data to prevent misuse or unauthorized access. Ethical considerations also include addressing potential biases in the dataset, particularly in facial emotion recognition, to avoid reinforcing stereotypes or disadvantage certain demographic groups. Transparency in how the AI models operate and how engagement scores are calculated is crucial to maintain trust among users. Finally, the system should be designed to avoid over-surveillance, ensuring that it enhances the learning experience without creating a sense of constant monitoring or pressure for students. By adhering to these ethical and legal principles, ***EngageAI*** can foster a positive and inclusive educational environment

6 | Appendix

GitHub:

https://github.com/EphremTil17/emotion_recognition/tree/mainline?tab=readme-ov-file

Presentation Slide Deck:

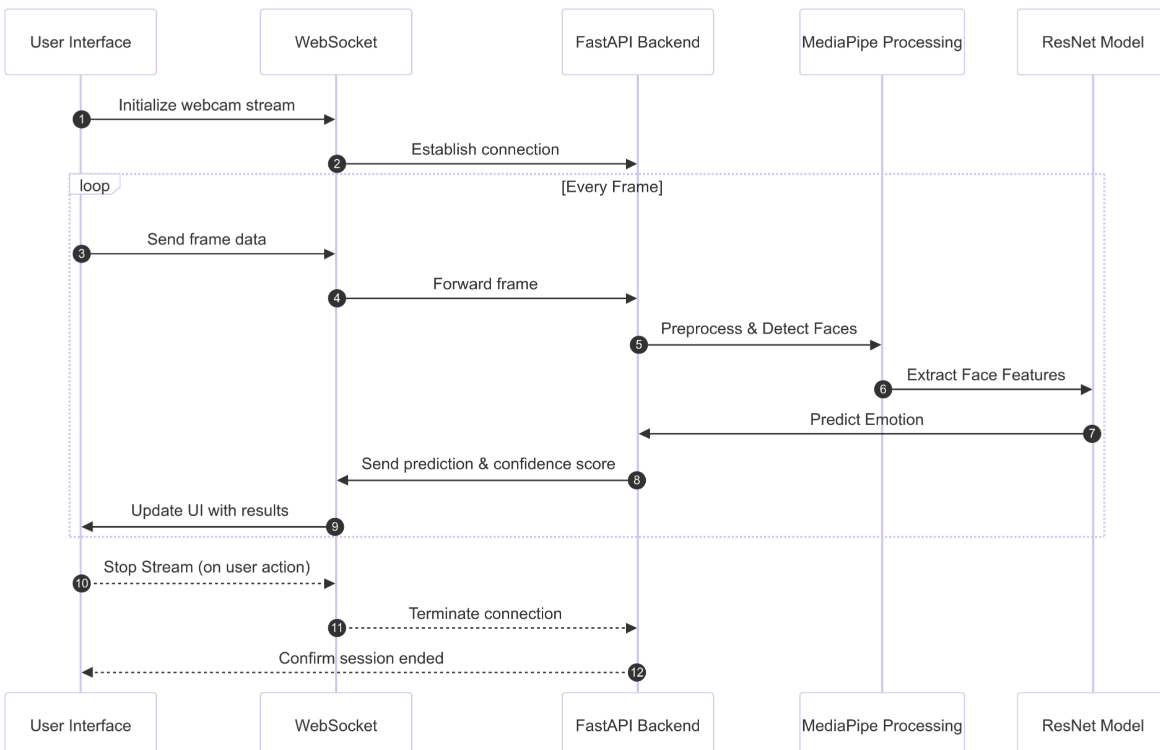
https://drive.google.com/file/d/1CFjqR4oW3WHpPwCUtHD8xPmMOXqrtJzn/view?usp=sharing

Dataset:

https://www.kaggle.com/datasets/sujaykapadnis/emotion-recognition-dataset/data

EngageAI:

https://engageai.ephremst.com/



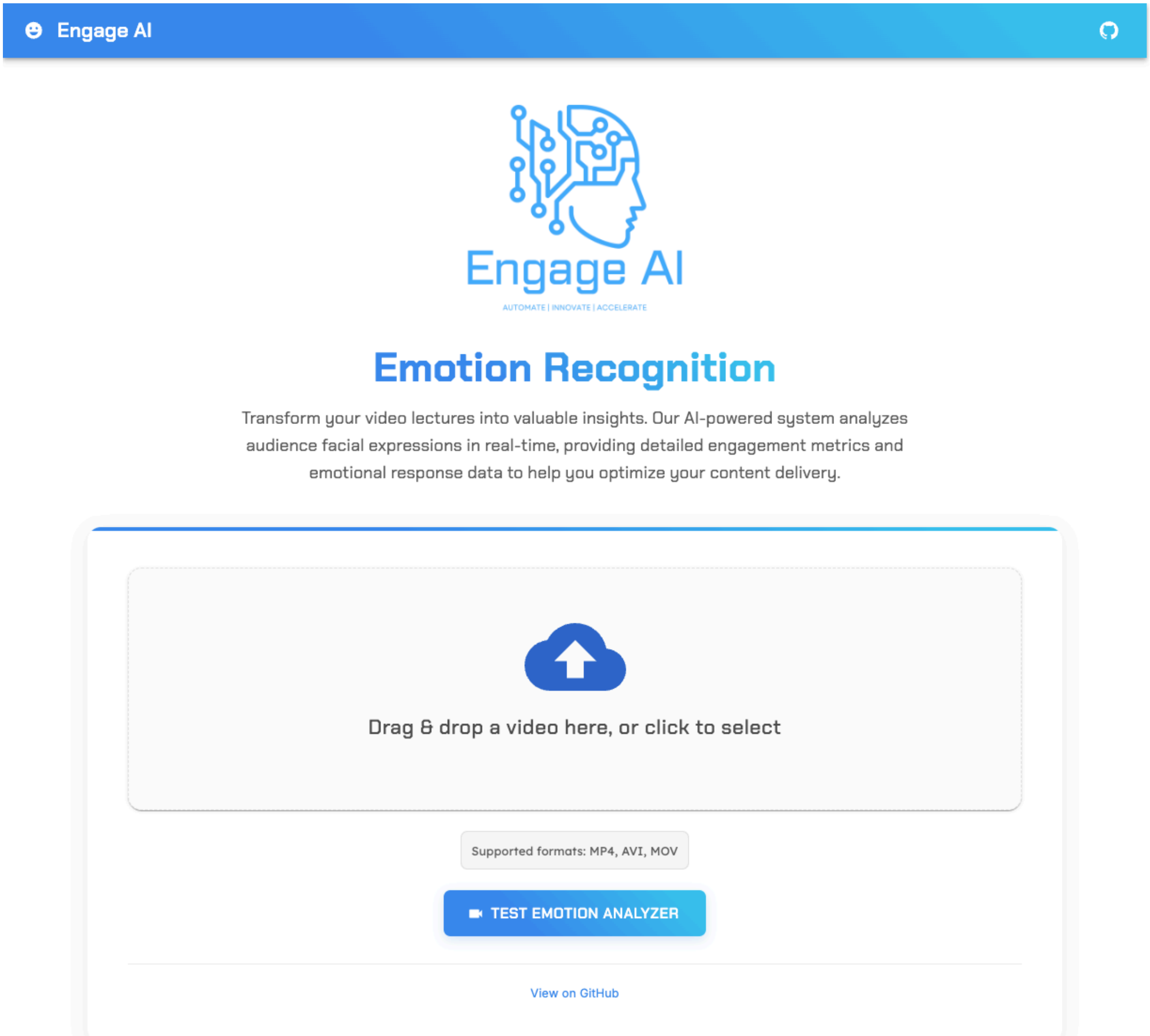


Figure 2*. *EngageAI *landing page (video upload)*

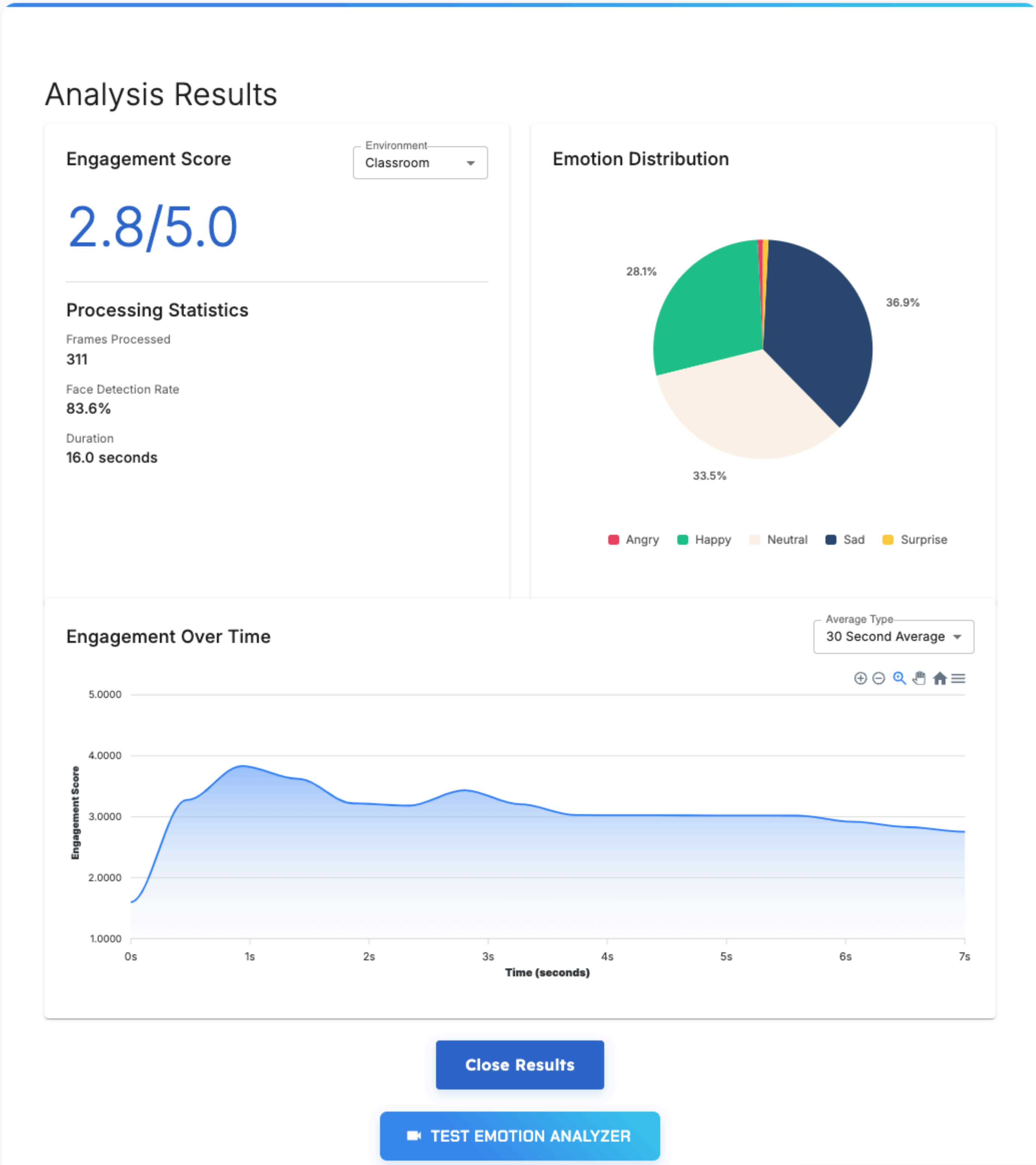


Figure 3. *Sample engagement analytics dashboard*

7 | Works Cited

Barba, Jacqueline. "'Teaching into the Zoom Void': Teachers struggle with students' lack of engagement." *Blueprint*, 28 January 2021, https://southblueprint.com/23416/features/teaching-into-the-zoom-void-teachers-struggle-with-students-lack-of-engagement/.

"The impact of online teaching on stress and burnout of academics during the transition to remote teaching from home - BMC Medical Education." *BMC Medical Education*, 20 June 2022, https://bmcmededuc.biomedcentral.com/articles/10.1186/s12909-022-03496-3. Accessed 12 February 2025.

Peck, Devlin. "Online Learning Statistics: The Ultimate List in 2025." *Devlin Peck*, 3 January 2025, https://www.devlinpeck.com/content/online-learning-statistics. Accessed 12 February 2025.

Walker, Kristen A., and Katherine E. Koralesky."Student and instructor perceptions of engagement after the rapid online transition of teaching due to COVID‐19." *PubMed Central*, 7 February 2021, https://pmc.ncbi.nlm.nih.gov/articles/PMC7995165/. Accessed 12 February 2025.

Verma, Nikita. "15 Stats and Case Studies About Student Disengagement." *Axon Park*, 24 March 2023, https://www.axonpark.com/15-stats-and-case-studies-about-student-disengagement/

8 | AI Integrity

We used AI tools, such as ChatGPT, to refine and improve the clarity of specific sections in our report, including the executive summary, business problem, and ethical & legal considerations. These tools helped streamline wording, enhance readability, and ensure coherence across different sections. However, all work remains entirely original, as the content, analysis, and critical thinking were our own. AI assistance was limited to refining language and structure rather than generating ideas or conducting research.