Emotion-Driven Audio-Visual Experience

Reality Architects

Aditya Vikram Singh

Ankit Gundewar

Shishir Kallapur

Yifan Tang

{singh.adityav, gundewar.a, kallapur.shi, tang.yifa}@northeastern.edu

Problem Statement

The Issue

- Current multimedia experiences are static and do not adapt to a user's emotions in real-time.
- Lack of seamless integration of AI-generated audio and visual content based on user mood and input.

Why it Matters

- Enhancing human-computer interaction through AI-driven emotional experiences.
- Potential applications in entertainment, therapy, meditation, and personalized media.

{singh.adityav, gundewar.a, kallapur.shi, tang.yifa}@northeastern.edu

Goal of the Project

To develop an AI system that generates personalized, immersive audio-visual experiences by understanding the emotional state and intent of the user based on the input provided.

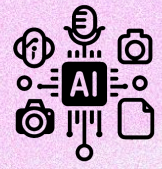
Key Features:

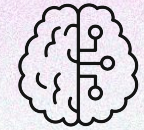
- Generating music and visuals to align with user's mood and intent.
- Real-time interaction and enhancements with the environment via webcam and hand gestures.
- Iterative feedback loop to tune experience.

{singh.adityav, gundewar.a, kallapur.shi, tang.yifa}@northeastern.edu

What makes it interesting?

- Interactivity: Users influence the generated content based on mood and intent.
- Multi-modal AI: Combining music generation, generative art, and real-time input processing.
- Human-AI Collaboration: Users guide AI instead of just consuming predefined media.
- User Personalization: System can be tuned to learn a specific user's likes and dislikes





{singh.adityav, gundewar.a, kallapur.shi, tang.yifa}@northeastern.edu

Related Work

- Multimodal emotional cues enhance perception and processing efficiency [1].
- The link between paintings and music in emotional expression has been studied through AI-driven music generation [2].
- AI-driven text-to-video generation enhances music visualization by mapping textual descriptions to dynamic visuals [3].

[1]A. B. M. Gerdes, M. J. Wieser, and G. W. Alpers, “Emotional pictures and sounds: A review of multimodal interactions of emotion cues in multiple domains,” Frontiers in Psychology, vol. 5, p. 1351, Nov. 2014.[Online]. Available: Frontiers

[2]T. Hisariya, H. Zhang, and J. Liang, "Bridging Paintings and Music – Exploring Emotion Based Music Generation Through Paintings," arXiv preprint arXiv:2409.07827, Sep. 12, 2024. [Online]. Available: arxiv

[3]V. Liu, T. Long, N. Raw, and L. Chilton, "Generative Disco: Text-to-Video Generation for Music Visualization," arXiv:2304.08551, Sep. 28, 2023. [Online]. Available: arxiv

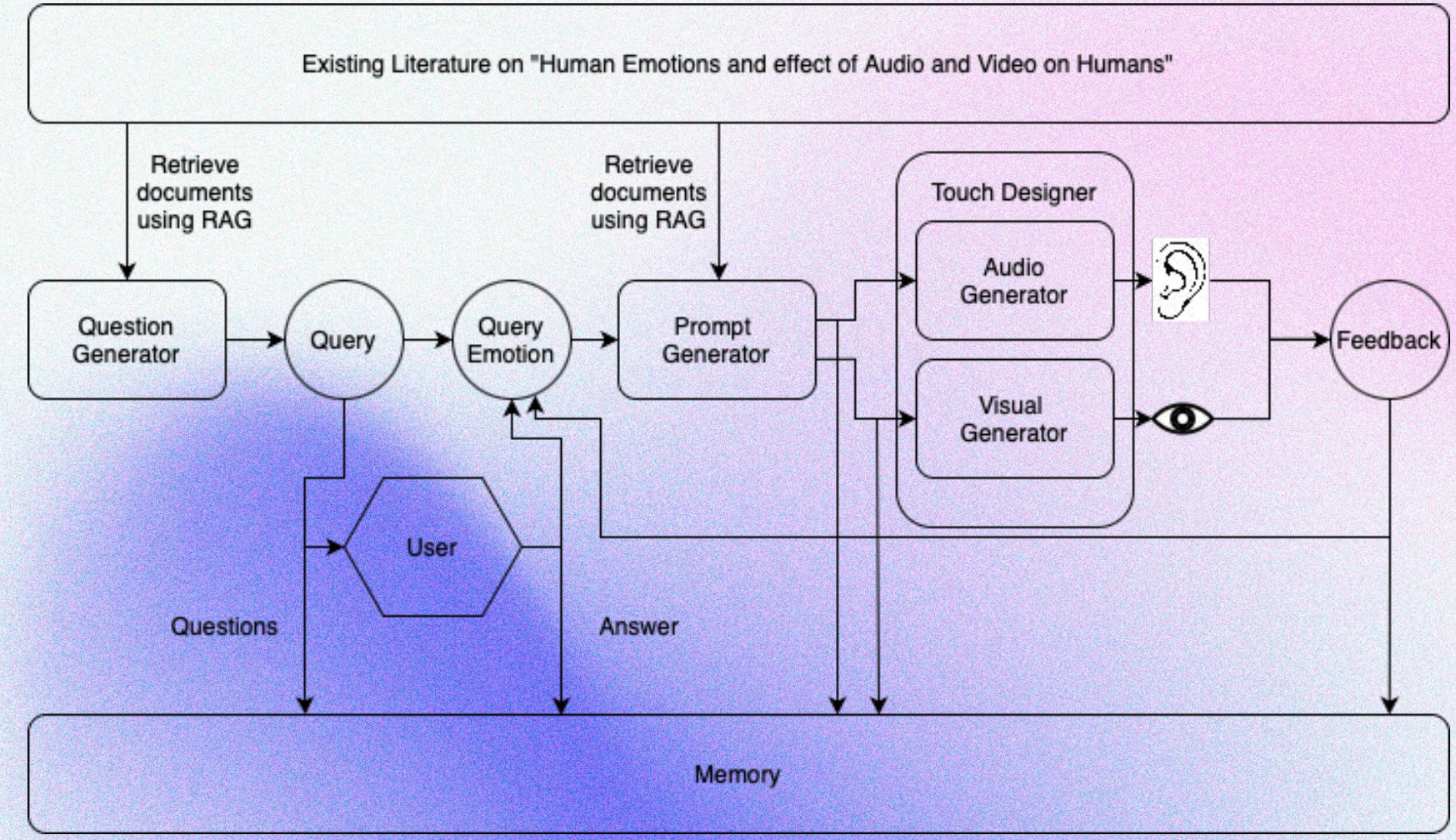
{singh.adityav, gundewar.a, kallapur.shi, tang.yifa}@northeastern.edu

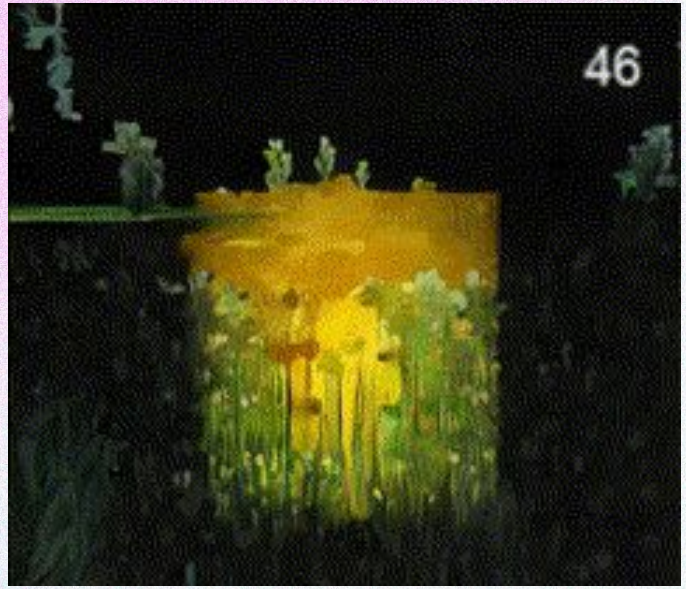
Hypothesis

- Main Hypothesis:
  - AI-generated personalized audio-visual experience enhances user engagement and emotional impact compared to static media.
- Sub-Hypotheses:
  - Emotion-driven generative content can lead to deeper user immersion.
  - Interactive elements can increase perceived control and satisfaction.

{singh.adityav, gundewar.a, kallapur.shi, tang.yifa}@northeastern.edu

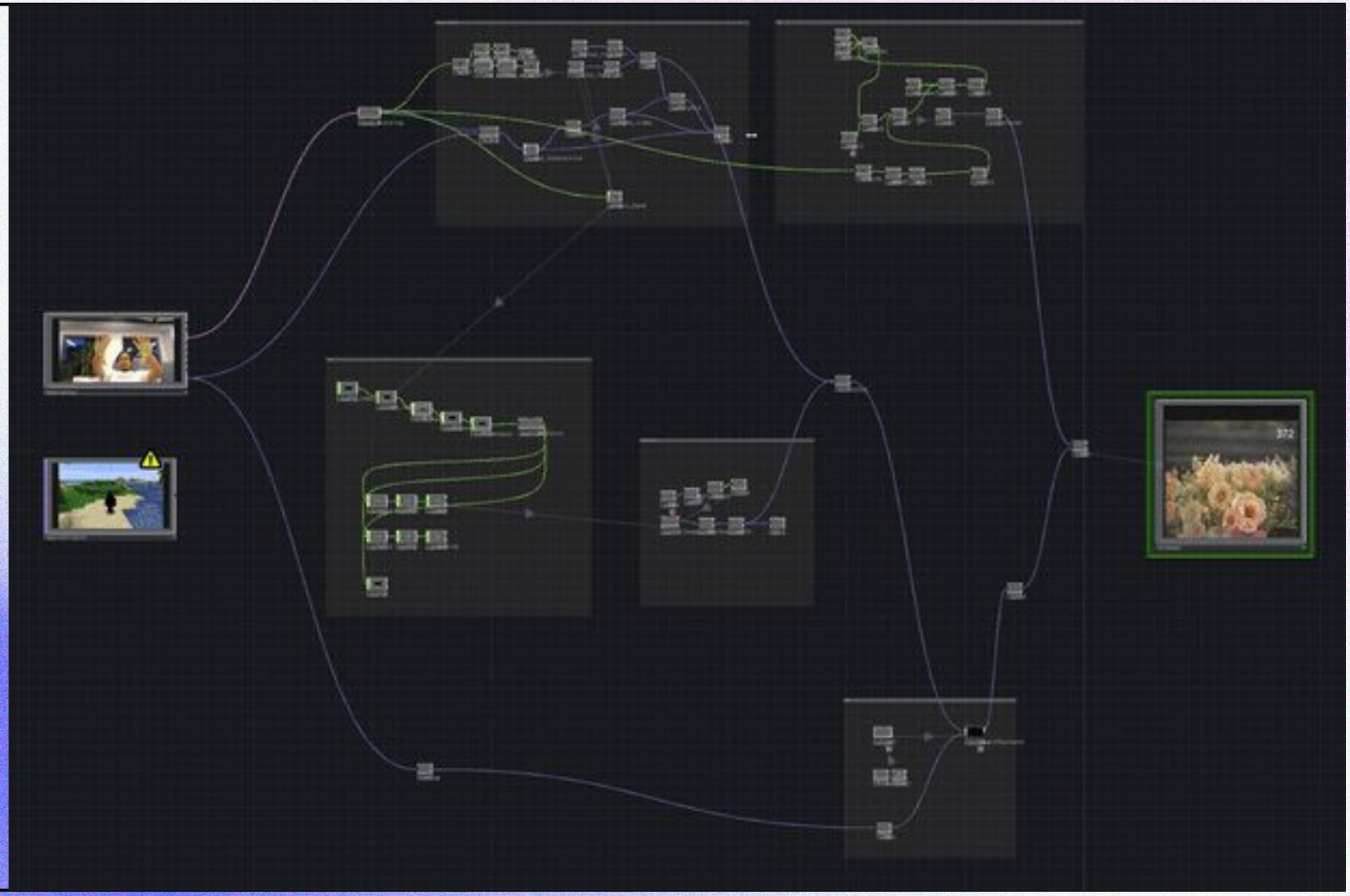
Design





{singh.adityav, gundewar.a, kallapur.shi, tang.yifa}@northeastern.edu

TouchDesigner Pipeline



{singh.adityav, gundewar.a, kallapur.shi, tang.yifa}@northeastern.edu

Evaluation

1. Survey for users after using the system:
    a. Users first experienced a start-to-end + feedback demo of the project.
    b. Given a short survey to answer about their experience.
    c. 5 point Likert Scale method used to present results.

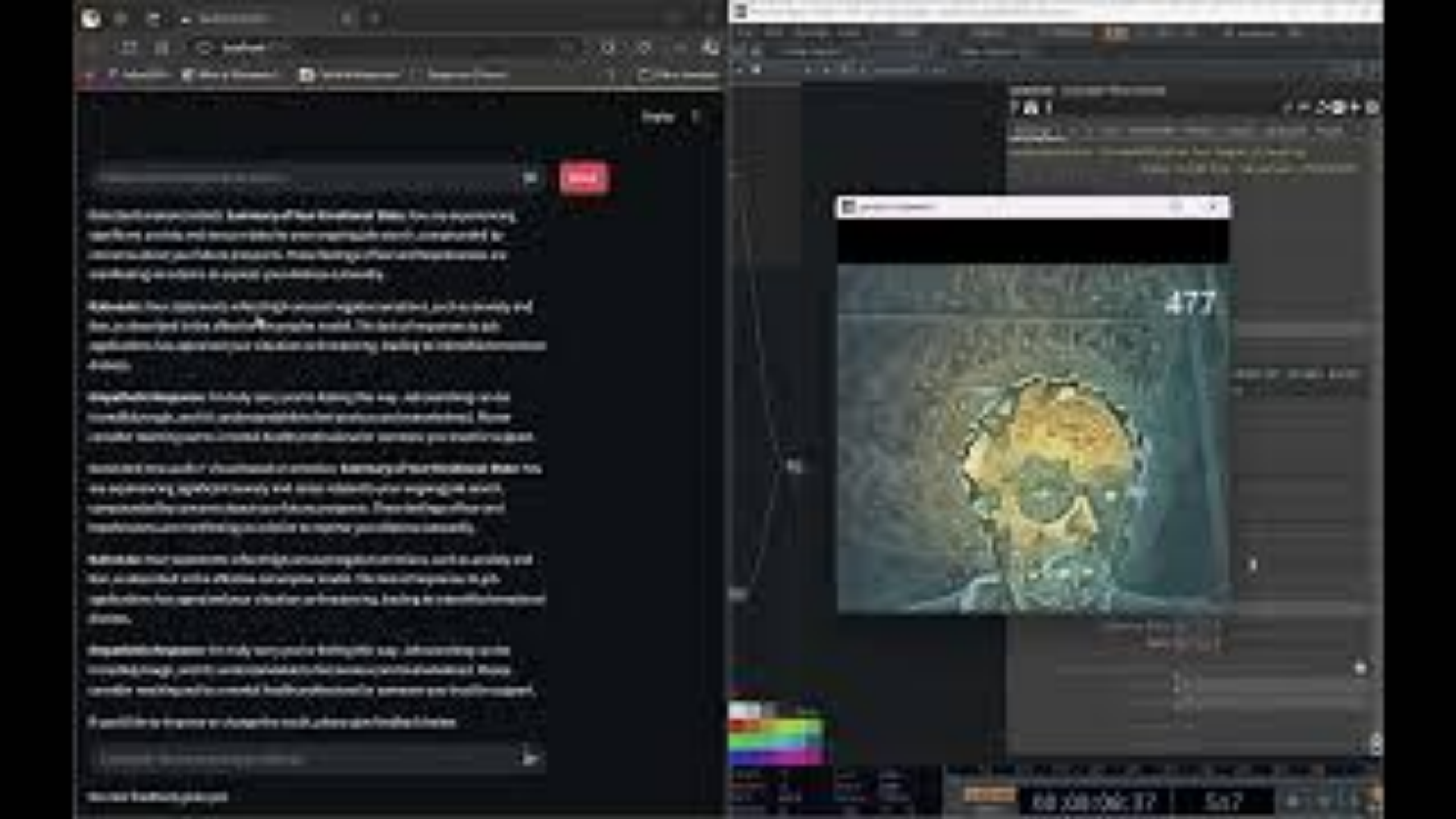
2. Quantitative metric to check how much a user has interacted (proof of concept)

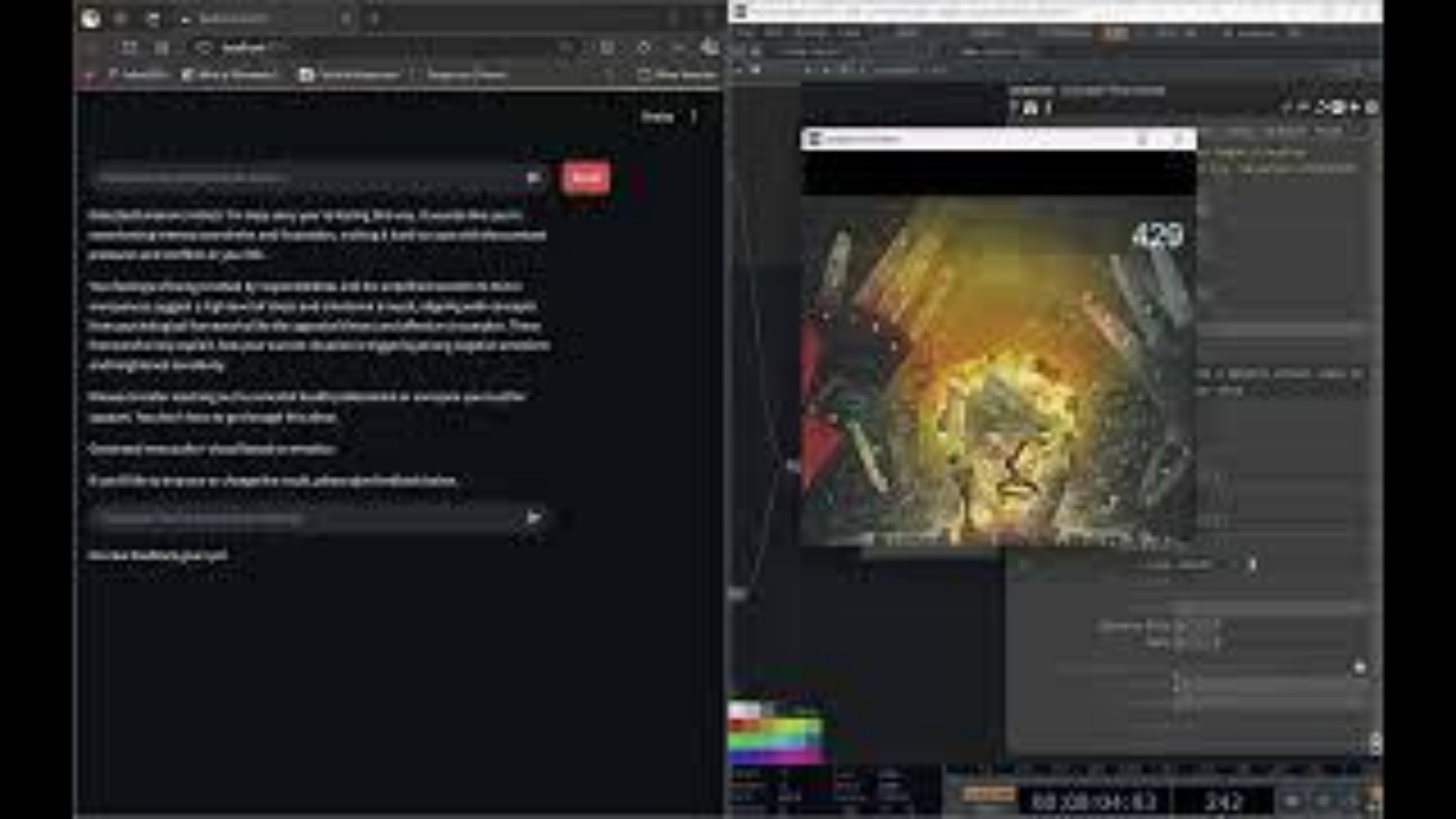
{singh.adityav, gundewar.a, kallapur.shi, tang.yifa}@northeastern.edu

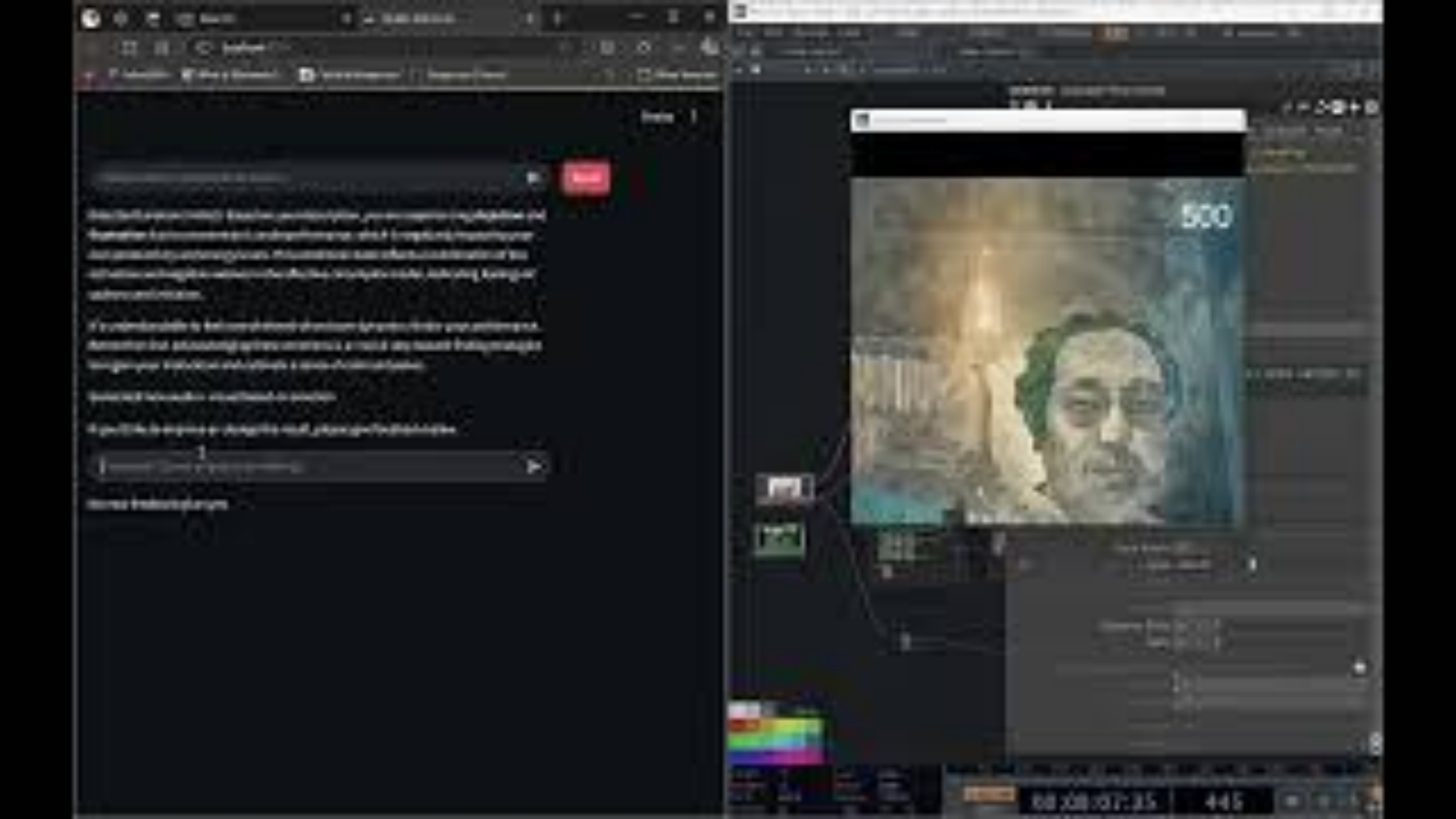
Challenges

- We could not find a good API based model for music generation. Setting up the MusicGen from Meta was more challenging than expected.
- We wanted to create a UI for the user to interact with. Using Streamlit as the UI was challenging.
- Setting up the pipelines in TouchDesigner took longer than expected.

{singh.adityav, gundewar.a, kallapur.shi, tang.yifa}@northeastern.edu


477


429


500

Potential Application of the system

System

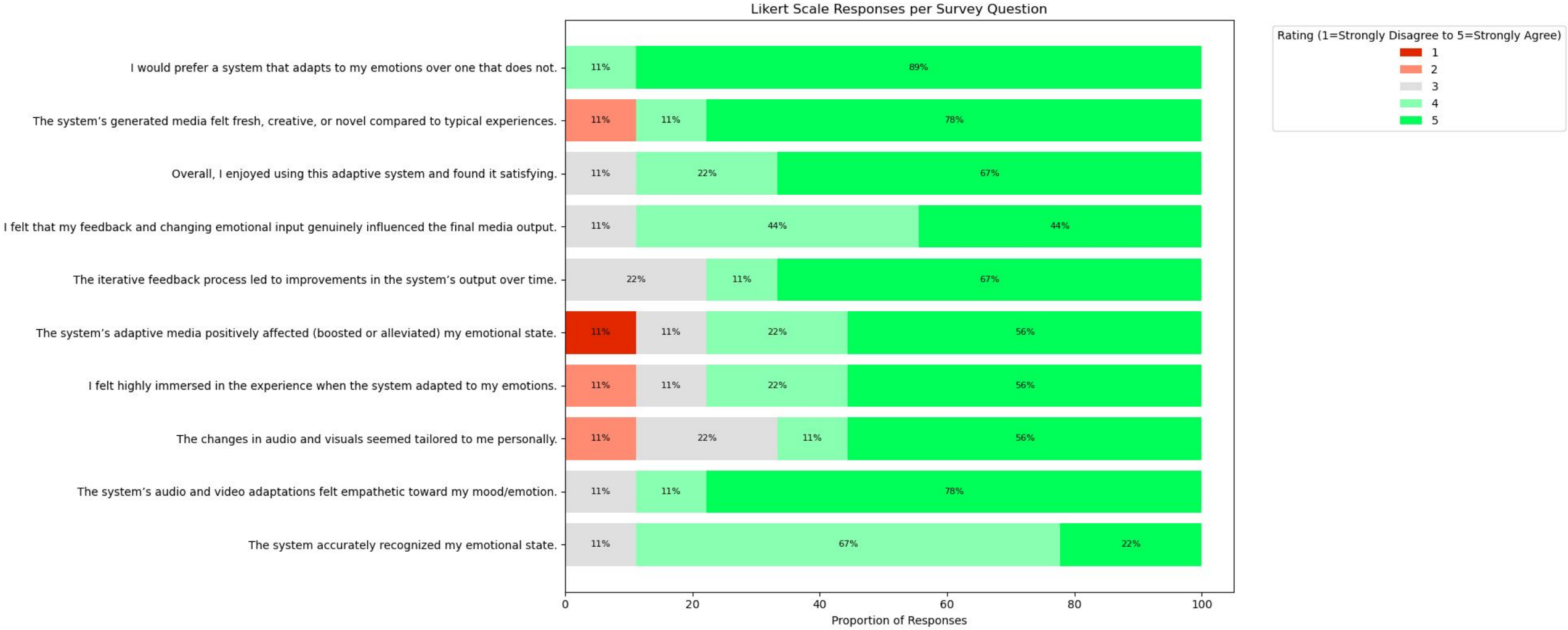
Emotion: Angry
Intent: Venting

Emotion: Happy, content
Intent: Relax, Reflect

Emotion: Dejection, Frustration
Intent: Feel hopeful

{singh.adityav, gundewar.a, kallapur.shi, tang.yifa}@northeastern.edu

Survey Results



{singh.adityav, gundewar.a, kallapur.shi, tang.yifa}@northeastern.edu

Thank You!

{singh.adityav, gundewar.a, kallapur.shi, tang.yifa}@northeastern.edu