

# Social Sensor PRD v2

## About the Team

**Team Name:** StorkeAI

**Project Name:** Social Sensor

**Company:** Invoca

### Members:

Krishna Nanda (Team Lead)

Renee Labutay (Scribe)

Alex Rasla

Raymond Deng

Regina Weinreb

## Background

---

### The Problem

Autism Spectrum Disorder (ASD) refers to a broad range of conditions that impacts the nervous system and impairs an individual's ability to communicate and interact. About 1 in 54 children have been identified with ASD.

With a deficit in social cognition and communication, individuals on the autism spectrum often struggle with sensing emotions, understanding the perspectives of others, and using their interpersonal skills to appropriately respond in different types of situations. The lack of communicative intuition limits them from having meaningful conversations and creating strong relationships with others.

### Motivation

Our incentive for this project is to help people with autism better understand social cues and allow them create more meaningful relationships with others and help them gain more intuition by sensing emotion. It will enable them to express themselves more eloquently and recognize others' sentiments. Furthermore, the technology to analyze audio and visual sentiment in real time can be applied to a variety of different industries to help people and businesses analyze customer tones and react accordingly.

## **Existing Solutions**

The majority of existing solutions focus on teaching social cues rather than detecting them within conversations and are mainly targeted to children affected by ASD. Additionally, sentiment analysis on written text is common, but to our knowledge, there are no tools that incorporate both audio and visual tools.

One product is Otsimo, a mobile Application full of educational games for children with special needs. It is aimed to help people with non-verbal autism, ADHD, late speakers, muscle weakness, and other health conditions improve their articulation rather than analyzing cues. However, Otsimo is limited to the education industry, thus it cannot be used outside of disabled students to learn. Additionally, the IBM Tone Analyzer is a tool that uses linguistic analysis to detect emotional and language tones in written text. It helps businesses understand and improve customer conversations through analysis at the document and sentence level. Unfortunately, it is not open source, so users would have to pay to use this product

Additionally, there are a handful of ML models for facial emotions, but their accuracy levels vary. For instance, the Facial Emotion Classifier is a deep CNN model trained from FER-2013 dataset has the best accuracy of 66-71%, which is still quite low. The Speech Emotion Recognizer recognizes emotions through speech, but also has low accuracy as well.

As of now, there are no existing models that recognize body language, backchanneling, advanced speech analysis (ex: long pauses between words might be a social cue), or are multi-modal (both speech and video).

## **Our Goals**

Our goal is to improve the social experience for individuals with ASD by developing SocialSensor, a video conferencing platform that will:

- Detect emotions of each party to aid the autistic individual in identifying social cues
- Provide appropriate suggestions to the autistic individual on how to respond adequately
- Use both visual and auditory input to create a more accurate analysis of emotions and tones

Social Sensor is built around a P2P decentralized design, which allows clients to connect directly without using a middle man. When using SocialSensor, users will have the ability to meet with each other online, through video conferencing software. During the conversation, SocialSensor will utilize ML models to analyze one's word choice, tonality, and facial expressions, to pick up and display social cues and emotions of the other party that the autistic individual may have missed. We hope to design a minimalist yet effective UI that will not distract the user from the conversation, but will help them have a successful interaction with another person. Our current UI vision is to have the real time emotion and sentiment displayed near the face of each person in the video conversation.

## **Innovation**

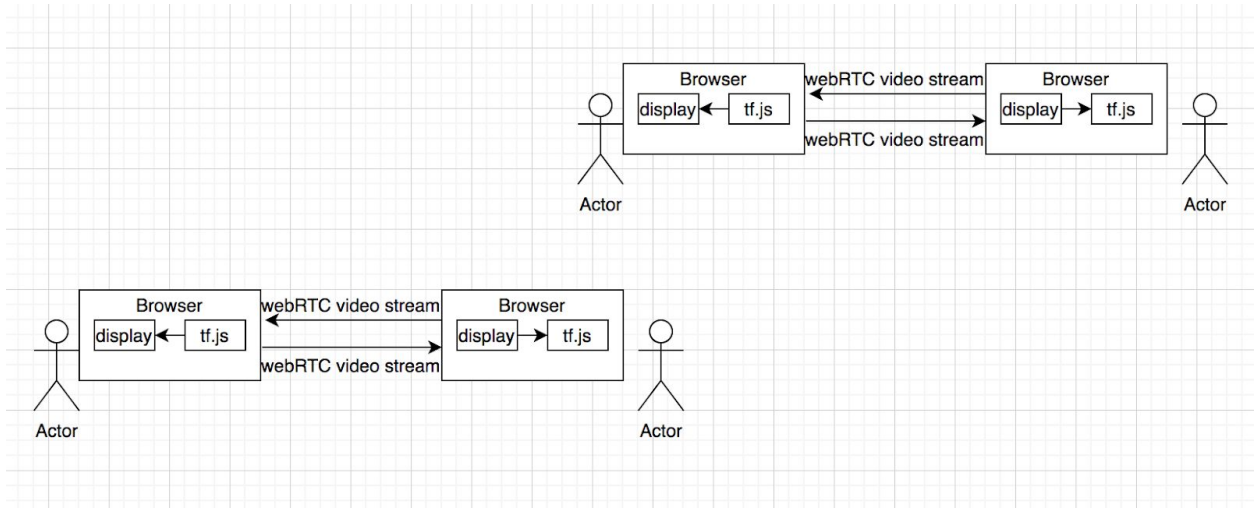
For this project, we pivoted from using Twilio Programmable Video Calling to Peer-to-Peer video calling. The reason for this is that the P2P design is incredibly powerful and scalable compared to the Twilio Client-Server model. Additionally since we are running our sentiment analysis in the browser using Tensorflow.js, running inference on a video stream takes very few lines of code. Compared to our code base when we were implementing our platform with Twilio, we have seen a reduction in both lines of code and number of files by 250 times and 400 times respectively. Our application is now more lightweight and decentralized. Additionally, we chose this approach to address issues of latency with running machine learning models on real-time video streams. Since we are doing this all in the browser, without having to send data to and from a server, we are able to provide sentiment analysis in real time with, and a more valuable user experience. The notion of real time sentiment analysis during a video call is something we at SocialSensor have yet to see be developed. Most other platforms who try to achieve a similar effect result by using screenshots of the person speaking within some sort of interval and analyzing the sentiment of these various images.

## **Assumptions**

- One individual using this platform has ASD and/or has difficulties detecting social cues.
- The individuals using the platform does not have any visual or auditory disabilities

# System Architecture Overview

## High Level Diagram



## UX/UI Design

As you can see in the image below, the user will be able to see the sentiment of the other user during the video call.

The buttons on the left will enable the user to...

- Turn their camera on/off
- Turn their audio on/pff



# Requirements

---

## User Stories

### 1: Video Conferencing

**User Story: As a user, I can see and hear my conversation partner on a call, so that I can communicate effectively.**

- **Acceptance Criteria:** Pass manual test for stable video and audio connection.
- **Scenario 1:** Call is successfully received
  - Given two users are in a call session,
  - When the user looks at the region of the screen that should have the other user's feed
  - The user will see and hear the other user's feed
- **Scenario 2:** Call is not successfully received
  - Given two users are in a call session,
  - When the user looks at the region of the screen that should have the other user's feed
  - The user will see an error message in that region
- **Github issue:** <https://github.com/CS189A-SocialSensor/SocialSensor/issues/2>

### 2: Create Room

**User Story: As a new user, I can create a room, so that I can initiate a call.**

- **Acceptance Criteria:** On clicking on the 'Create Room' button, new room is created and user redirected to new room.
- **Scenario 1:** User has an account.
  - Modal will appear that asks if the user has an account.
  - Clicking on the button will launch the login page.
  - User will log in.
  - User will enter a room.
- **Scenario 2:** User does not have an account.
  - Modal popup will ask the user if he/she has an account.
  - Selecting no will launch a "Create Account" page.
  - User will enter name, email, DOB, and password.
  - User will enter a room.

### **3: Inviting User to Existing Call**

**User Story: As a user, I can share my Twilio room link with my conversation partner, so that they can join my call.**

- **Acceptance Criteria:** A room will have a corresponding join link that will allow another user to join the call.
- **Scenario 1:** The link is valid & user #1 sends the link to user #2
  - User #2 will click on the link.
  - User #2 will enter the room.
  - User #2 will see User #1 in the video room.
- **Scenario 2:** The link is invalid.
  - User #2 will see an error message on the screen.
  - The error message will include a button that allows them to initiate a call instead.
- **Scenario 3:** User #1 sends the link to multiple people.
  - An error message will pop up stating that the room has reached maximum capacity.

### **4: Joining Call**

**User Story: As a user, I can join a video call by clicking on an invitation link.**

- **Acceptance Criteria:** Entering a certain room link URL should redirect users to desired room.
- **Scenario 1:** Link is valid and user joins call
  - User clicks on link
  - The meeting room is loaded with all the users currently in the meeting
  - User can show video and audio to other users in the meeting room
- **Scenario 2:** Link is invalid
  - Webpage is loaded signaling the link is invalid or expired
  - User can choose to go to their login page/dashboard

### **5: Receiving Real Time Sentiment Analysis**

**User Story: As a user with ASD, I will receive real time sentiment analysis during the call, so that I can understand how the other person is feeling.**

- **Acceptance Criteria:** The emotion analysis model projects a box around the face of the speaker continuously for the duration of the call, with a textual label for each emotion.
- **Scenario 1:** User is currently in a call.
  - Given that the user wants sentiment analysis enabled

- The emotional facial model will project a box around the face of the speaker color coded, along with a textual label to describe the sentiment associated with the speaker's facial expression.
- **Scenario 2:** User is currently in a call.
  - Given that the user wants sentiment analysis disabled
  - The user will be able to communicate with the video conferencing platform, similar to any off the shelf video conferencing platform with no sentiment analysis performed on the speaker.

## **6: Logging into User Account**

**User Story: As a user, I can log into my account, so I can host my own personal meeting and keep track of my upcoming meetings.**

- **Acceptance Criteria:** After entering a username and password successfully, the user is redirected to their personal dashboard.
- **Scenario 1:** Username and password match
  - User dashboard shows up showing upcoming scheduled meetings
  - User can browse upcoming meeting details
- **Scenario 2:** Username and password do not match
  - Error message signaled on login page indicating invalid username/password
  - User can enter username and password again
- **Github issue:** <https://github.com/CS189A-SocialSensor/SocialSensor/issues/5>

## **7: Creating a User Account**

**User Story: As a user, I can sign up & create my own account, so I can host my own personal meeting and keep track of my upcoming meetings.**

- **Acceptance Criteria:** Through Auth0, the user will be able to create an account by entering valid credentials.
- **Scenario 1:** User will enter a valid email address & password manually.
  - Once Auth0 confirms that the email address is valid and the password meets the criteria, then an account will be created.
- **Scenario 2:** User will sign up using Google.
  - User can click on the "Continue with Google" option, and an account will be created.
- **Github Commit:**

<https://github.com/CS189A-SocialSensor/socialsensor/commit/801dfdfc79bd8975f996e0de65351c090ed0482a>

## **8: Turning the Sentiment Analysis On & Off**

**User Story: As a user, I can toggle on and off the social sensor depending on if I need it or not, so that if I am a neurotypical, it appears as a normal call.**

- **Acceptance Criteria:** Selecting options from a menu should allow clients to pick 'on', 'off' for model analysis.
- **Scenario 1:** User toggles button on
  - Given two users are in a call
  - And the video can be seen
  - The user's interface will display model analysis of their partner's current emotions
- **Scenario 2:** User toggles button off
  - Given two users are in a call
  - And the video can be seen
  - The user's interface will not display model analysis of their partner's current emotions
- **Github Issue:** <https://github.com/CS189A-SocialSensor/SocialSensor/issues/4>

## **9: Receiving Notifications**

**User Story: As a user, I can receive notifications so I can be informed about the current sentiment of the call and about my upcoming meetings.**

- **Acceptance Criteria:** When the user receives a call, they should be informed of it via notification on their dashboard.
- **Scenario 1:** User receives a call and is not currently in a call.
  - Given the user is on the dashboard.
  - Given the user wants to be notified.
  - The user's dashboard will prompt them with a notification of the incoming call.
- **Scenario 2:** User receives a call and is currently in a call
  - Given the user does not want to be disturbed.
  - The user's interface will quietly put the notification of the missed call in their inbox.
- **Scenario 3:** User receives a call and is currently in a call
  - Given the user does not mind being disturbed.
  - The user's interface will notify the user with a popup of the incoming call.
- **Github Issue:** <https://github.com/CS189A-SocialSensor/SocialSensor/issues/3>

## **10: Toggling Analysis between users**

**User Story: As a user, I can change which user's sentiment is being analyzed on a group call by clicking on their window.**



- **Acceptance Criteria:** Clicking on the user window should show up a menu on the UI with the option to toggle user sentiment on/off.
- **Scenario 1:** User turns off model analysis of another user in a call.
  - Given two users are in a call
  - And the video can be seen
  - Toggling the menu option on the user's menu should disable model analysis on specified user
- **Scenario 2:** User turns on model analysis of another user in a call.
  - Given two users are in a call
  - And the video can be seen
  - Toggling the menu option on the user's menu should enable model analysis on specified user

### **11: Toggle Text**

**User Story: As a user, I can toggle on and off between having a textual representation of my conversation partner's emotions displayed, so that I can understand to which color it corresponds.**

- **Acceptance Criteria:** Clicking on the user window should show up a menu on the UI with the option to toggle text labels on/off, and text should show or hide as a result of the button.
- **Scenario 1:** User toggles button on
  - Given two users are in a call
  - And the video can be seen
  - The user's interface will display a textual representation of their partner's current emotions
- **Scenario 2:** User toggles button off
  - Given two users are in a call
  - And the video can be seen
  - The user's interface will not display a textual representation of their partner's current emotions
- **Github issue:** <https://github.com/CS189A-SocialSensor/SocialSensor/issues/1>

### **12: Blurring out the Background**

**User Story: As a user, I can choose to blur out the background of the other participant so that I won't be bothered by any distractions.**

- **Acceptance Criteria:** By toggling this option, the background of the other participant will be blurred out.
- **Scenario 1:** User toggles button on.
  - Given two users are in a call

- And the video can be seen
- The background of the other participant will be blurred out.
- **Scenario 2:** User toggles button off.
  - Given two users are in a call
  - And the video can be seen
  - The background of the other participant will not be blurred out.
- **Github Issue:**  
<https://github.com/CS189A-SocialSensor/sentiment-analysis-prototype/issues/6>

### **13: Black and White UI**

**User Story: As a user, I can choose to view the video if the other participant is in black and white so that I can focus and am not overstimulated by bright colors.**

- **Acceptance Criteria:** By toggling this option, the video of the other participant will be displayed in black and white.
- **Scenario 1:** User toggles button on.
  - Given two users are in a call
  - And the video can be seen
  - The video of the other participant will be in black and white.
- **Scenario 2:** User toggles button off.
  - Given two users are in a call
  - And the video can be seen
  - The video of the other participant will not be in black and white.
- **Github Issue:**  
<https://github.com/CS189A-SocialSensor/sentiment-analysis-prototype/issues/10>

### **14: Reports**

**User Story: As a user, I can view a report of the call so that I can review the sentiments the other participant expressed, and learn from the experience.**

- **Acceptance Criteria:** After the call ends, a button should appear, and when clicked, a detailed report of the sentiments that occurred in the call will be displayed.
- **Scenario 1:** User clicks Show Report button.
  - Given the call has ended
  - The user's interface will display a sentiment report of the call.
- **Scenario 2:** User does not click the Show Report button.
  - Given the call has ended
  - The user's interface will display the same screen.

## **15: Turning Camera On & Off**

**User Story: As a user, I can turn my camera on & off during the video call.**

- **Acceptance Criteria:** Create a button with a camera logo that appears during the video call. When the user presses the button, their camera will be turned off.
- **Scenario 1:** User presses the button to turn off his/her camera.
  - The video feed from the user's camera will no longer be accessible.
  - The other participant will not be able to see the other user.
- **Scenario 2:** User presses the button to turn on his/her camera.
  - The program will be able to access the video feed from the device's camera.
  - The other participant will be able to see the user.

## **16: Mute Button**

**User Story: As a user, I can turn my audio on and off.**

- **Acceptance Criteria:** Create a button with a microphone logo that appears during video call. When the user presses the button, their microphone will be turned off.
- **Scenario 1:** User presses the button to turn off their microphone.
  - Given two user are in a call
  - And audio is connected
  - The user's microphone will be turned off and the other participant will not be able to hear them.
- **Scenario 2:** User presses the button to turn on their microphone.
  - Given two user are in a call
  - And audio is connected
  - The user's microphone will be turned on and the other participant will be able to hear them.

## **17: WebRTC Stream**

**User Story: As an http request, I can fetch the WebRTC video stream from the other actor**

- **Acceptance Criteria:** Ensure the other actor's video streams are successfully displayed on the user's UI
- **Scenario 1:** The WebRTC stream is successfully fetched
  - The client calls the fetch method to retrieve the stream
  - The client receives the WebRTC stream and can be displayed on the client's UI
- **Scenario 2:** The WebRTC stream is unsuccessfully fetched
  - The client calls the fetch method

- The WebRTC stream is not sent/delivered to the client
- The client doesn't see the other user's video stream in their UI
- **Github Issue:**  
<https://github.com/CS189A-SocialSensor/sentiment-analysis-prototype/issues/9>

## **18: Tensorflow.js**

**User Story:** As an http request, I can use tensorflow.js to do sentiment analysis on the WebRTC video stream

- **Acceptance Criteria:** Tensorflow.js analyzes the videostream and does sentiment analysis
- **Scenario 1:** Tensorflow.js successfully analyzes the sentiment of the WebRTC videostream
  - The WebRTC stream is sent to tensorflow.js to do sentiment analysis
  - Tensorflow.js sends the sentiment analysis back to the client
  - The sentiment analysis is displayed on the client's UI
- **Scenario 2:** Tensorflow.js unsuccessfully analyzes the sentiment of the WebRTC videostream
  - The WebRTC stream is sent to tensorflow.js to do sentiment analysis
  - Tensorflow.js fails to do analysis/send the sentiment analysis back to the client
  - The sentiment analysis is not displayed on the UI, and the previous sentiment stays displayed until the next frame is sent
- **Github Issue:**  
<https://github.com/CS189A-SocialSensor/sentiment-analysis-prototype/issues/11>

## **19: Retrieve Peer ID**

**User Story:** As a client I can initially fetch the address to connect to my peer from the Twilio server.

- **Acceptance Criteria:** The connection is established.
- **Scenario 1:** The identifier was not successfully received.
  - The client and their peer attempt to make a call.
  - The client makes a fetch attempt to retrieve the identifier for their peer from the server.
  - The fetch attempt is unsuccessful.
- **Scenario 2:** The identifier was successfully received.
  - The client and their peer attempt to make a call.

- The client makes a fetch attempt to retrieve the identifier for their peer from the server.
- The fetch attempt is successful.
- **Github Issue:**  
<https://github.com/CS189A-SocialSensor/sentiment-analysis-prototype/issues/7>

## **20: Retrieve ML Models Client Side**

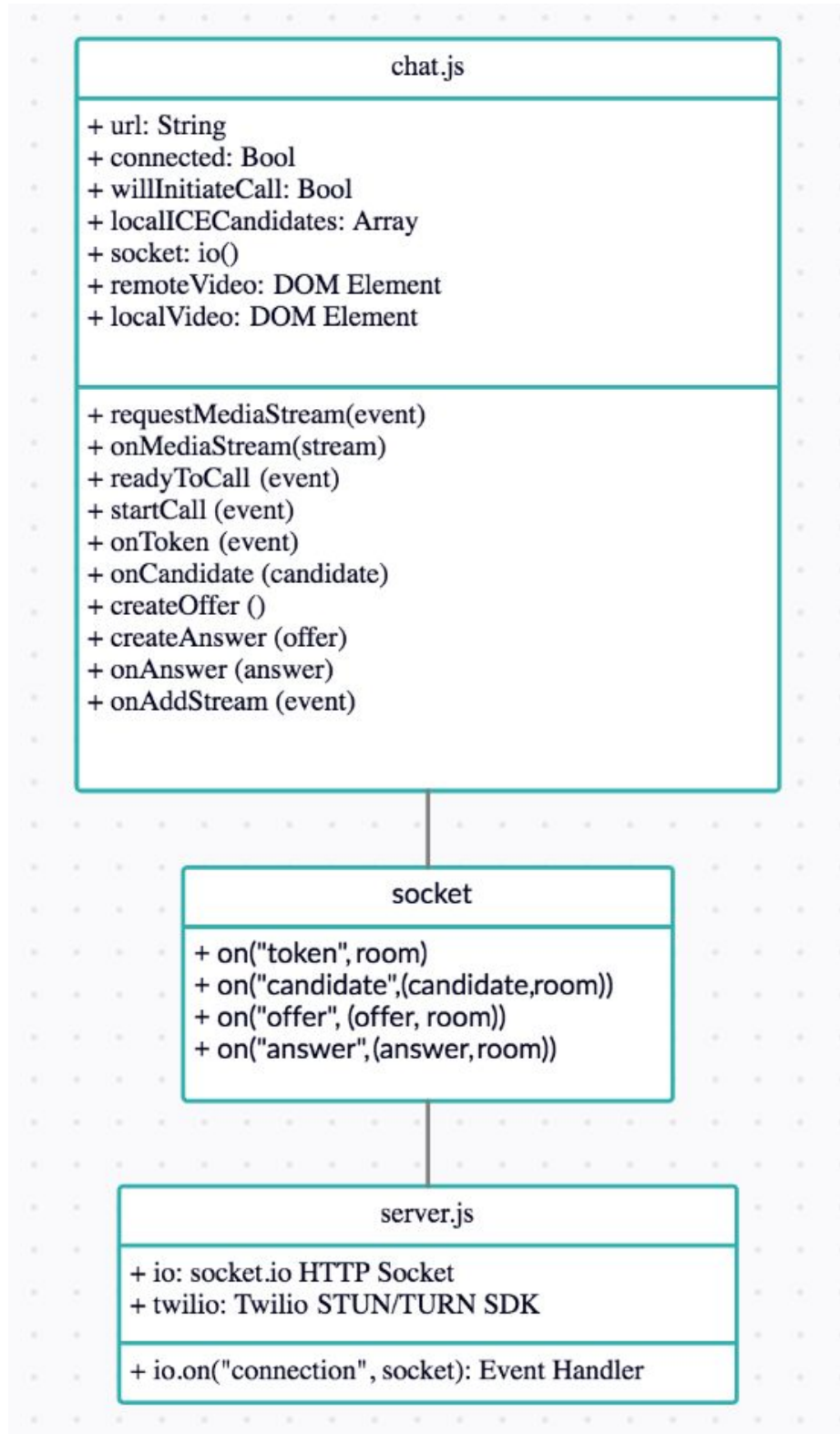
**User Story: As a client I can initially fetch the ML models from the Twilio server.**

- **Acceptance Criteria:** The client successfully retrieves the ML models from the server.
- **Scenario 1:** The ML models are not retrieved from the server.
  - The client attempts to retrieve the ML models from the server using a fetch attempt.
  - This fetch attempt fails.
- **Scenario 2:** The ML models are retrieved from the server.
  - The client attempts to retrieve the ML models from the server using a fetch attempt.
  - This fetch success.
- **Github Issue:**  
<https://github.com/CS189A-SocialSensor/sentiment-analysis-prototype/issues/8>

# System Models

---

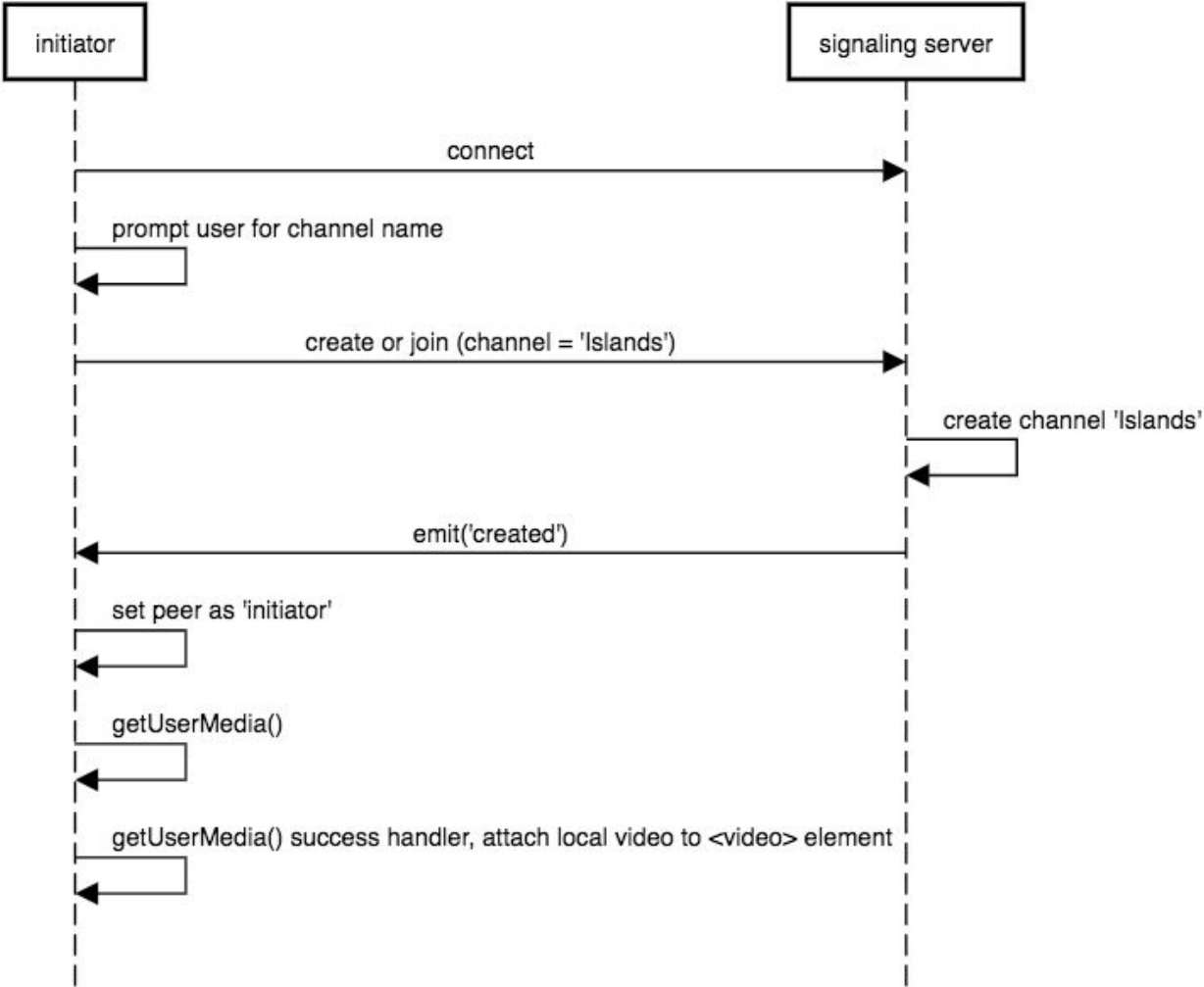
## UML Diagram



## Sequence Diagrams - User Interactions

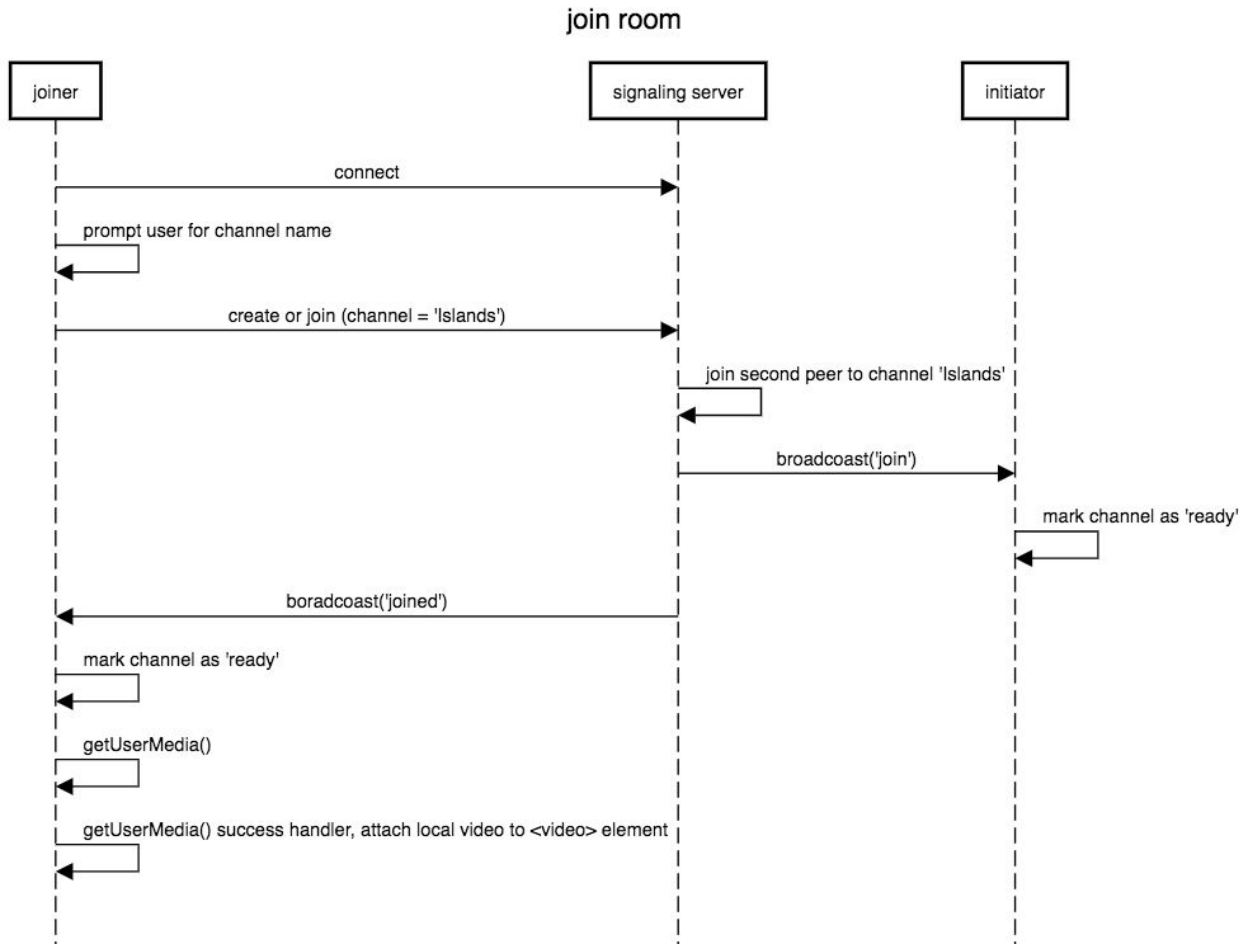
# Create Room (webRTC)

## create room



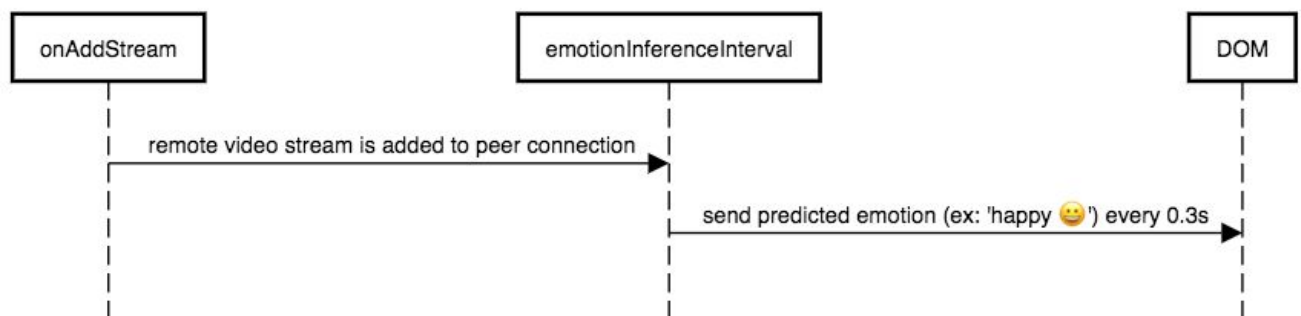
# Join Room (webRTC)





## Emotion Recognition (tensorflow.js)

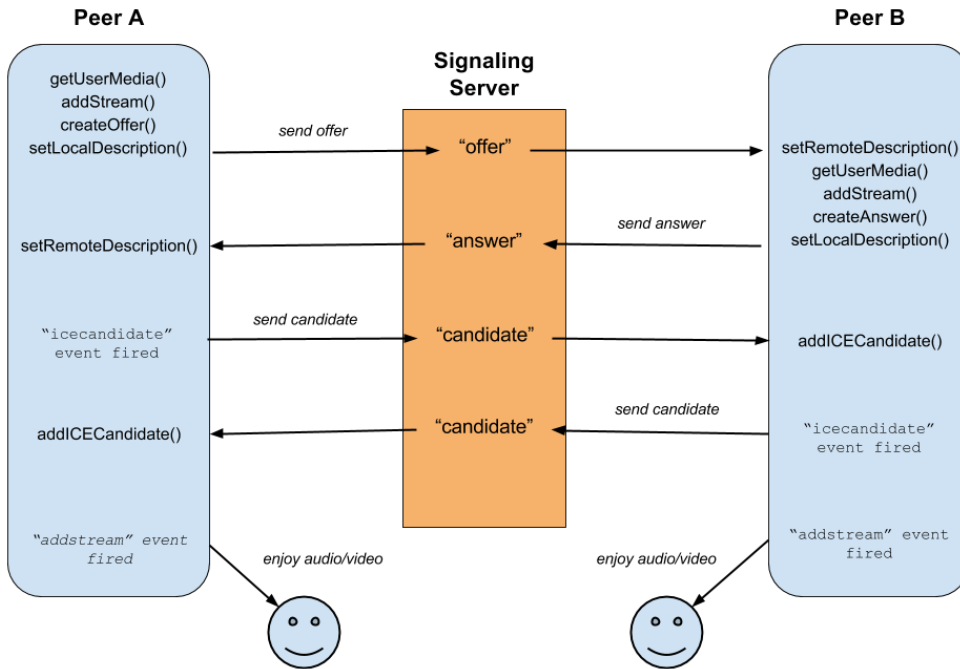
### tensorflow.js emotion recognition



## Sequence Diagrams - Function Interactions

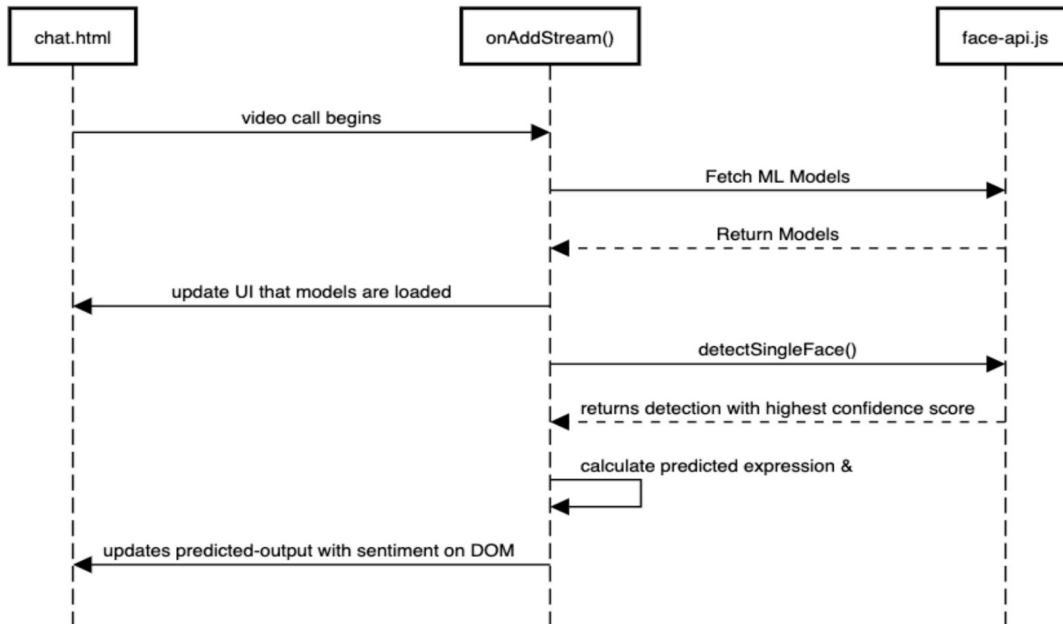
## User Connection (webRTC)

- Implemented through series of functions rather than with classes

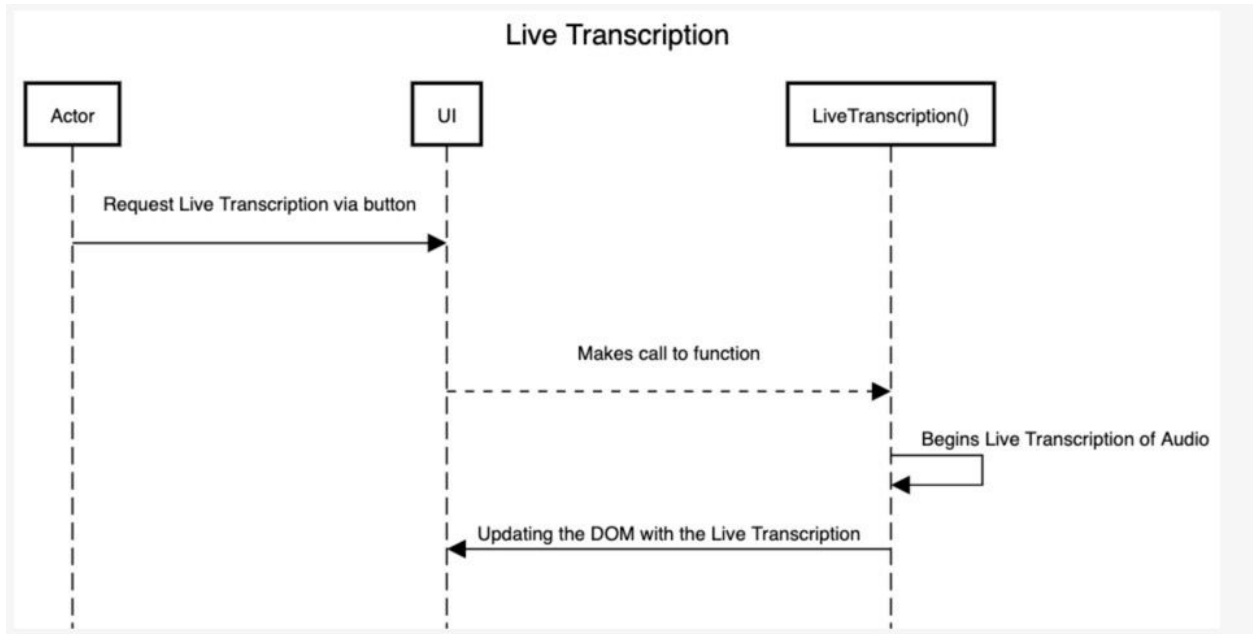


## Running the ML Inference (webRTC)

Running the ML Inference with `onAddStream()`



## Live Transcription



## Appendices

---

### Technologies

**Front End:** HTML/CSS/JS, React

**Back End:** node.js, Express

**Machine Learning:** tensorflow.js, face-api.js

**Video Call:** webRTC, Twilio STUN/TURN