Social Sensor PRD v1

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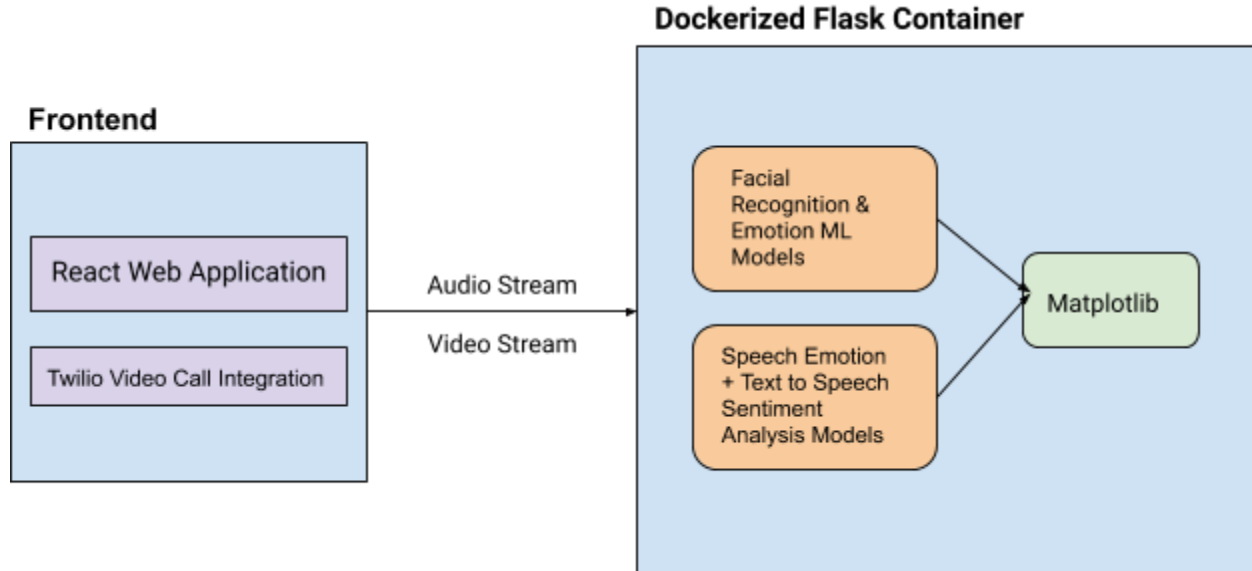
- Detect emotions of each party to aid the autistic individual in identifying social cues
- Provide appropriate suggestions to the autistic individual on how to respond adequately
- Use both visual and auditory input to create a more accurate analysis of emotions and tone

Social Sensor will be built around Twilio's existing visual and auditory API and will provide users with information regarding the social cues of the conversation happening in real-time. 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.

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

Architecture Diagram

Dockerized Flask Container

Frontend

React Web Application

Twilio Video Call Integration

Audio Stream

Video Stream

Facial Recognition & Emotion ML Models

Speech Emotion + Text to Speech Sentiment Analysis Models

Matplotlib

User Stories

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

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

Create Room
User Story: As a new user, I can create a Twilio 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.

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.

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

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.

User Interface

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

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

Toggle 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

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

User Interaction and Design
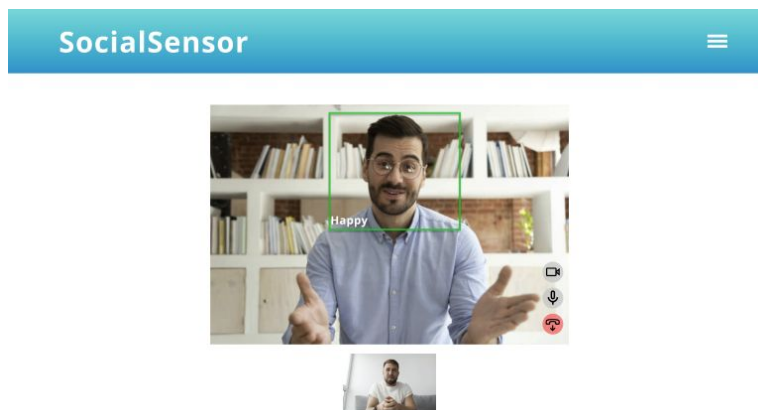
Appendices

Technologies
- **Frontend:** React
- **Backend:** Python, TensorFlow, DeepSpeech
- **APIs:** Twilio, Keras, GCP Speech-To-Text