

# Vision Statement

## About the Team

**Team Name:** StorkeAI

**Project Name:** SocialSensor

**Company:** Invoca

**Team Lead:** Krishna Nanda

**Team Scribe:** Renee Labutay

**Members:** Alex Rasla, Krishna Nanda, Raymond Deng, Regina Weinreb, Renee Labutay

## Project Summary

Our goal for this project is to develop a telecommunications tool that will detect certain emotions and suggest appropriate responses that will ultimately improve the social experience for individuals with ASD.

## 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.

| Name of Product                 | Otsimo                                                                                                                                                                   | IBM Tone Analyzer                                                                                                                           |
|---------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------|
| Description                     | Mobile Application full of education games for children with special needs                                                                                               | Uses linguistic analysis to detect emotional and language tones in written text.                                                            |
| Strengths of Existing Solutions | Aimed at helping people with non-verbal autism, ADHD, late speakers, muscle weakness, and other health conditions improve their articulation rather than analyzing cues. | Performs analysis at both document & sentence levels, which is helpful to understanding and improving customer conversations.               |
| Gaps in Existing Solutions      | Limited to the education industry; technology cannot be used outside of helping disabled students learn.                                                                 | Only includes written text analysis and nothing about Not open source, so either users would have to pay to use the product to utilize this |

Additionally, there are a handful of ML models for facial emotions, but their accuracy levels vary:

- **Facial Emotion Classifier**
  - Deep CNN model trained from FER-2013 dataset has the best accuracy of 66-71%, which is still quite low
- **Speech Emotion Recognizer**
  - Poor accuracy for low numbers of emotions (i.e. happy, neutral, sad, etc.)
  - Very low accuracy when the number of emotions increased (i.e. fearful, angry, pleasant surprise, etc.)

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.

## **Potential Technologies**

Python, TensorFlow, Keras, GCP Speech-To-Text, Twilio, DeepSpeech, React

## **Milestones**

1. Research ML models, datasets, technologies, APIs, and other tools that will help us build our platform.
2. Develop a basic UI mockup that would display social cues of the other party.
3. Build a prototype that takes in facial expressions from a video stream and performs basic sentiment analysis using a ML model.
4. Create the web interface to our application that integrates Twilio video conferencing.
5. Get a real time data stream connected between our frontend and backend.
6. Begin doing analysis on the real-time data feed and sending the results back to the frontend.
7. Add account creation aspect (username + password)
8. Set up a database to store user information.
9. Expand on our output model and include suggested responses for the user.
10. Implement a tooltip for displaying suggested responses.
11. Perform user testing with the Autism Center to receive feedback on functionality and UI.
12. Expand on the sentiment analysis aspect (include body language, non-literal language, sarcasm, etc.).
13. Improve accuracy of the ML models (stretch goal).