

STAT8021 Project: Artificial Intelligence Video Interview Analyzer
(AVIA)
Group 6.2

Jason Chan Jin An Tong Si Man Yau Tsin Wai Chan Tsz Hin

2021-07-19

Contents

1 Chapter 1: Project Background, Problem, Scope of Study, Project Goals, and Objectives.	3
1.1 Project Background	3
1.2 Problem	3
1.3 Project Aims	3
1.4 Objectives	3
1.5 Scope of Study	3
2 Chapter 2: Literature Review	4
2.1 Facial Detection and Facial Expression Recognition	4
2.2 Eye Gaze Detection	5
2.3 Text Analytics	6
2.4 AVIA vs Similar Applications	7
3 Chapter 3: Research methods	7
3.1 Data Source	7
3.1.1 Video Analytics for Video Interview	7
3.1.2 Text Analytics for Writing Assessment	7
3.2 Data Preprocessing	8
3.3 Models Design	9
3.3.1 Video Analytics	9
3.3.2 Text Analytics	12
3.4 Data Visualization	12
4 Chapter 5: Discussions	13
4.1 Video Analytics	13
4.2 Text Analytics	14
4.3 Combined Overall	15
5 Chapter 6: Conclusion	15
6 Chapter 7: References	17
7 Chapter 8: Appendix	18
7.1 Results Summary	18
7.2 Program Documentation	20
7.3 Program Coding	22

1 Chapter 1: Project Background, Problem, Scope of Study, Project Goals, and Objectives.

1.1 Project Background

Currently, COVID-19 is affecting every country and city in the world and many companies decided to use online interviews to recruit new employees such as HireVue and Talview. Candidates' facial expressions, gestures, and behaviour throughout the interview are marked when answering sets of pre-set questions. They are also required to type in their answers for certain questions, which are marked by AI to give an overall score.

1.2 Problem

Applicants are now facing a new challenge to be interviewed by AI interviewers for being selected to be the right candidates. In order to get a higher score at the first-round AI interview, candidates need to get used to the new technologies and practice to express themselves to AI interviewers. However, the relevant resources are very limited for this new norm, candidates are struggling and find it difficult to get well-prepared.

1.3 Project Aims

To help the candidates, we decided to create AVIA (Artificial Intelligence Video Interview Analyzer) to provide practices to help candidates gain experience and enhance their performance to achieve a better overall score. We aim to capture candidates' facial expressions, eye contact and with the written answers that candidates type in, text analysis will be able to extract their keywords and see how close it is from the model answer. Combining these facial expressions, eye contact and text analysis, an overall score will be provided for reference.

1.4 Objectives

In order to provide an accurate overall score, our model is built with two important parts, video and text analysis. Firstly, our objective is to distinguish different types of facial expressions and eye contact accurately. Secondly, together with the text analysis, our objective is to check whether the typed-in responses are close to our model answer. By training the past available data, we wish to build a model which is able to identify facial expressions, eye contact and also the typed-in response from the candidates.

1.5 Scope of Study

When using this application, candidates can see statistics of his or her facial expressions and eye contact in the first part. Then, their type-in responses will be classified according to the wording they use to see how close it is from the model answer. An overall score will be generated based on their facial expression and type-in responses for the representation of their overall performance. **Therefore, the scope of study at includes both video analytics and text analytics.**

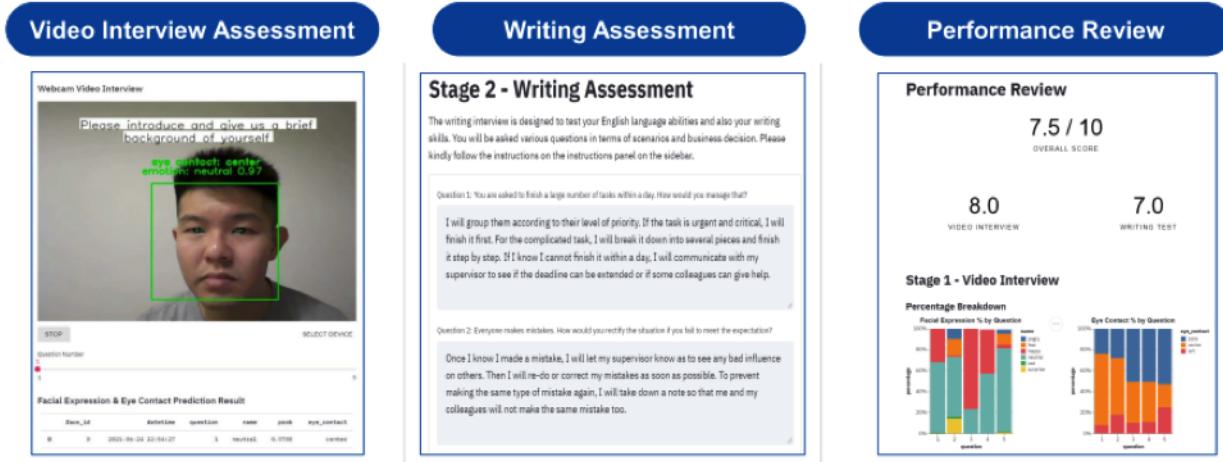


Figure 1: Scope and workflow of AVIA.

2 Chapter 2: Literature Review

2.1 Facial Detection and Facial Expression Recognition

In “Introduction to Computer Vision on Mobile Devices, Facial Detection and Recognition on Mobile Devices” by (Lui, 2015) states that facial detection is the key for analyzing facial expression. In “3D Approaches and Challenges in Facial Expression Recognition Algorithms”, (Nonis et al, 2019) states that Facial Expression Recognition is a computer-based technology that uses mathematical algorithms to analyze faces in images or video. There is great potential for facial expression application in the surveillance, security, and communication field. In our project, we are trying to use it in the career development field, enhancing the interview skill of the candidates so that they can have a larger chance to get into the second round of the interview.

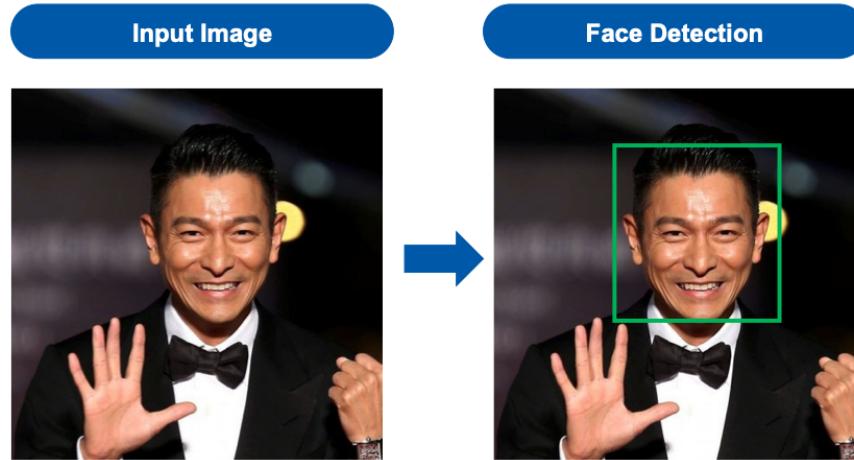


Figure 2: Face detection concept review.

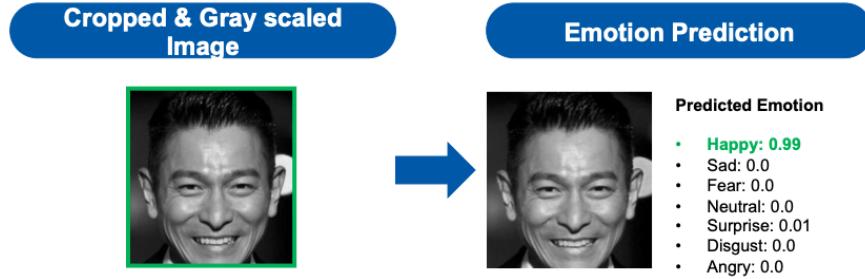


Figure 3: Facial expression recognition concept review.

(Ekman, 1971) defined a set of six emotions that are accepted as universal: anger, disgust, fear, happiness, sadness, and surprise. However, Martinez et al. state that “there are approximately 7000 different expressions that people frequently use in everyday life”. As we can see, emotions can be classified into two main groups: basic emotions and compounded emotions. When happiness and sadness are very distinctive from each other, some emotions share the same signals. For example, cheerfulness and optimism both expressed by rising lips. Those compound emotions are very difficult to be distinguished by people so are the same for AI.

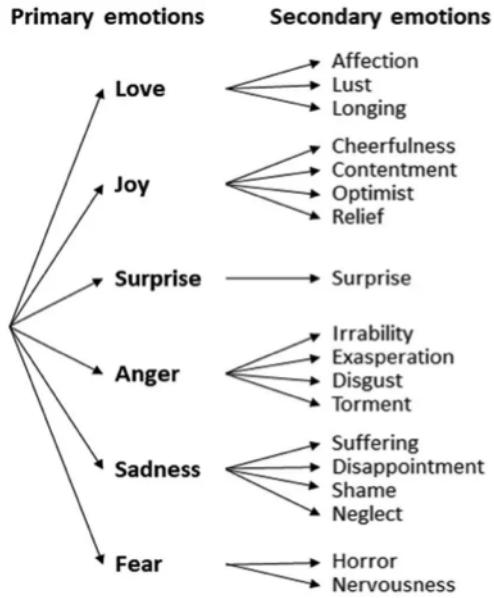


Figure 4: Universal primary emotions and secondary emotions.

2.2 Eye Gaze Detection

In our project, we also implemented eye contact detection since it is an essential part of body language in interviews. “Review of Eye Gaze Tracking Application” by (Bin Aziz, 2014) states that eye gaze tracking is the process of measuring where the eyes are looking and can be used to understand a person’s intentions. Eye contact is also an important part in the interview training.

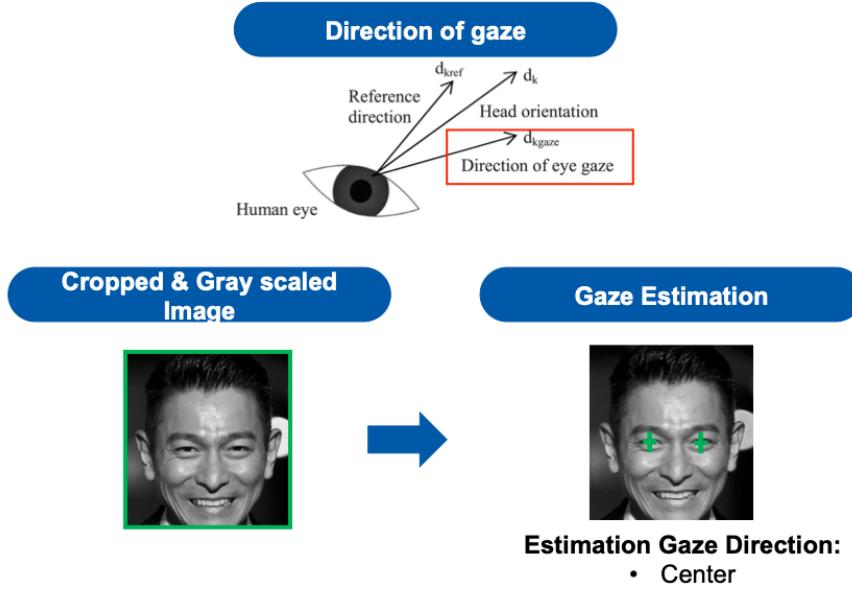


Figure 5: Eye gaze tracking concept review.

2.3 Text Analytics

For the second part the type-in responses, text analysis is required. “Text Mining for Big Data Analysis in Financial Sector”, (Bach et al, 2019) states that text mining extracts relevant words (N-grams) and relationships between them in order to categorize them. The text mining model is frequently used in the financial industry to relate certain financial news for positive or negative impact. In AVIA, we will use a similar method in our application. After the candidates typed in their responses, keyword extraction and sentiment analysis are performed. In order to reduce the noise of their responses, names, spaces, punctuations and stop-words will be removed. After getting the keywords, we are going to create an analysis to identify if it is close to the ideal response.

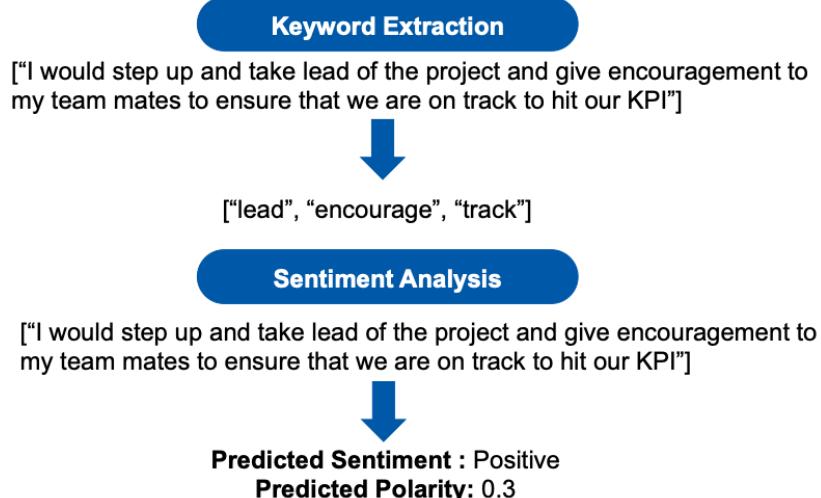


Figure 6: Text analytics examples that is used in AVIA.

With the combination of facial expressions, eye detection and text analysis, candidates can better understand

their performance and improve accordingly.

2.4 AVIA vs Similar Applications

For a similar application like HireVue, it is very easy to use and user-friendly from the interviewer view. It is easy to install and gives the flexibility to the interview to use anytime and anywhere. There is a large number of questions stored in the application and it will randomly pop some questions to the candidates. However, there are some negative comments on HireVue from the candidates that they dislike this application. Since the questions are too random, some of them are asked questions about boosting sales while they are applying for a clerk position. For another similar application Talview Hiring, it has a similar function like HireVue, which is used to screen the candidates. It is easy to use, very structured and standardized. Their customer service is good. However, some customers are complaining that there are still bugs after using it for 6 months and the customer support is terrible. Some complain that the user interface is quite confusing.

While creating our application AVIA, we are **standing from the candidates' point of view, which is very different from the existing applications such as HireVue and Talview. We aim to help candidates to get well-prepared for AI interviews.** We take notes for the strength of the similar applications, making sure it is easy to use with a very clear user interface. Since we let the candidates type in the questions they would like to practice with their facial expressions, we can avoid the mismatch of interview questions with the position the candidates are applying for. This makes us unique when comparing with the applications which are already selling in the market. This also helps us to reach out to the new market share, the candidates. Learning from the negative comments on the existing application, we will spend more time on testing to make sure all the bugs are fixed when launching our application. We will spend more time in customer support to make sure customers feel easy and comfortable when using our application. With our application, we help the candidates to practice and show confidence when having interviews with all AI interviewers, not only limited to HireVue and Talview.

3 Chapter 3: Research methods

3.1 Data Source

3.1.1 Video Analytics for Video Interview

We obtain data sources from Kaggle, which has a variety of large datasets for data analysis. Inside FER 2013 (Facial Expression Research) for grayscale images of faces, there are 32298 annotated examples with labels of seven emotions, where 0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral. They are split into a training set (28,709 examples) and test set (3,589 examples).

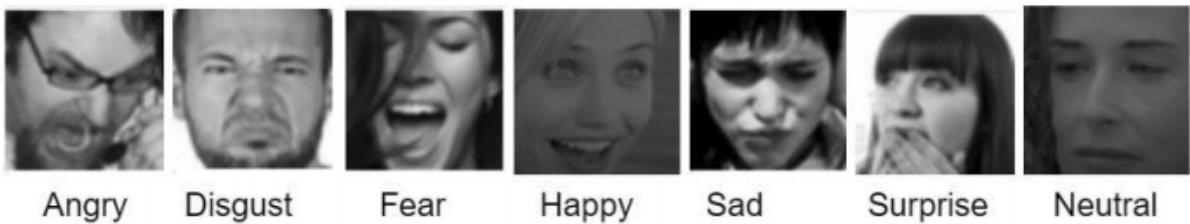


Figure 7: Sample image data from FER2013 dataset.

3.1.2 Text Analytics for Writing Assessment

We selected 5 common questions which will appear during the interview. We have come up with a model answer for each of the questions. We expect to see the positive keywords such as ‘communicate’, ‘helpful’, ‘division of labor’ when candidates type in their responses. Below are examples of the 3 of 5 questions and ideal answers.

Question 1: You are asked to finish a large number of tasks within a day. How would you manage that?

Model Answer 1: I will group them according to their level of priority. If the task is urgent and critical, I will finish it first. For the complicated task, I will break it down into several pieces and finish it step by step. If I know I cannot finish it within a day, I will communicate with my supervisor to see if the deadline can be extended or if some colleagues can give help.

Question 2: Everyone makes mistakes. How would you rectify the situation if you fail to meet the expectation?

Model Answer 2: Once I know I made a mistake, I will admit my mistakes and acknowledge my supervisor. Then I rectify my mistakes and assist others with the same mistake as soon as possible. To prevent making the same type of mistake again, I will have self reflection on what caused the mistake and have a discussion with my supervisor to find the source of the mistake and learn from it.

Question 3: When there is an expansion of your team and many new teammates join in. How would you adjust for this?

Model Answer 3: I will be friendly to them and give a helpful hand whenever they need it. Then, I will see if my supervisor has any training plan for the new teammates in which I can contribute my experience and knowledge. Clear division of labor and communication is needed since the team is getting larger.

3.2 Data Preprocessing

Since we test our model with a real-time webcam, we must crop our face image after face detection and convert it to grayscale since the data source we use to train our model is in greyscale.

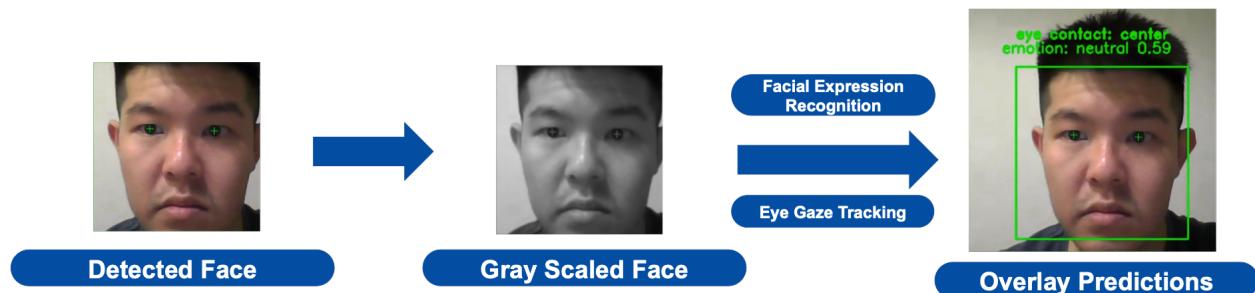


Figure 8: Cropped face and grayscaling before applying AI techniques.

For the text analysis part, we use tokenizing and lemmatization to normalize keywords before we can fit the model answers to create our model. We also use spaCy to determine if it is a noun or verb.

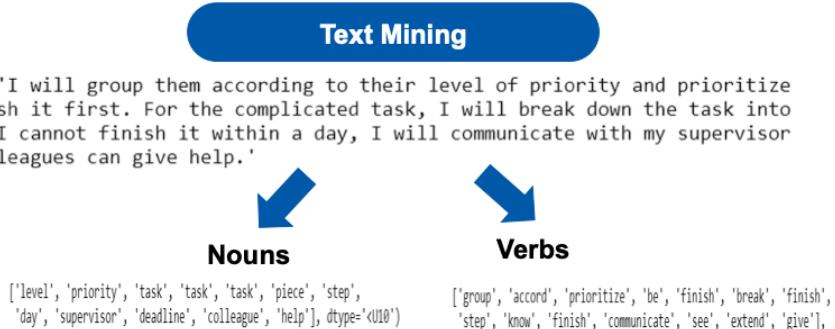


Figure 9: Tokenizing, lemmatizing, and then using spaCy to determine if it is verb or noun.

3.3 Models Design

3.3.1 Video Analytics

3.3.1.1 Face Detection - Haar Cascade

First, we need to build a face detection system before a facial expression recognition system. The face detection system can be constructed using Haar Cascade method or CNN based method. Haar Cascade method is faster but less accurate while the CNN method is more accurate but needs more resources. Our model proceeded with Haar Cascade with tuned parameters which is suitable for webcam frontal pose.

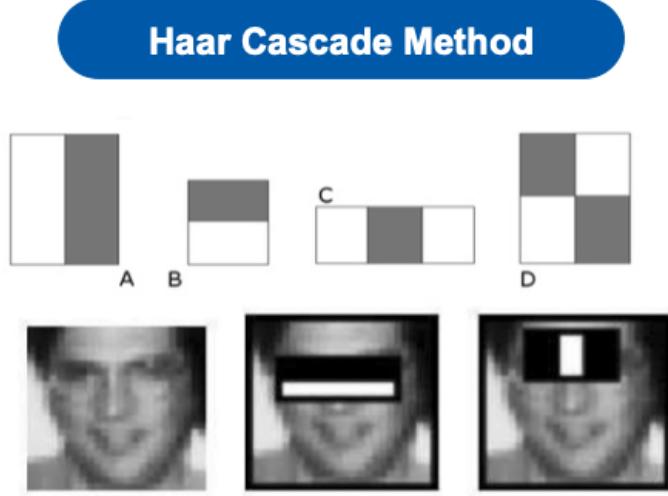


Figure 10: Haar cascade model utilizes edge detection to detect faces.

3.3.1.2 Facial Expression Recognition - Keras CNN Model

As for the facial expression recognition part, a CNN model (with 4 Blocks of 2D Conv and Batch Normalization) is first pre-trained over the facial expression dataset and will be used to predict the facial expression of the candidate at that snapshot. All the results of the prediction of the facial expression recognition will be stored in a data frame for further analysis and score computation.

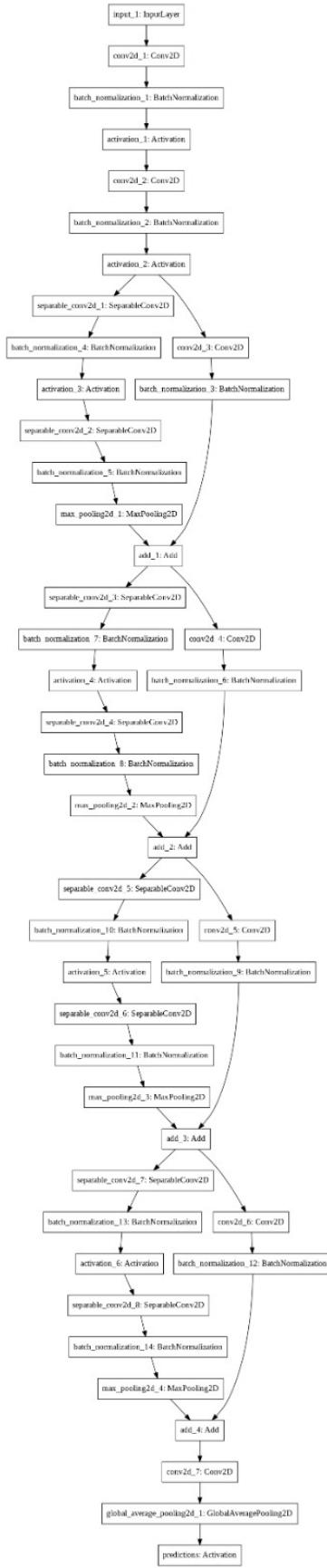


Figure 11: Pre-trained Keras CNN model structure used by AVIA.
10

AVIA Facial Expression Recognition Real-time Prediction

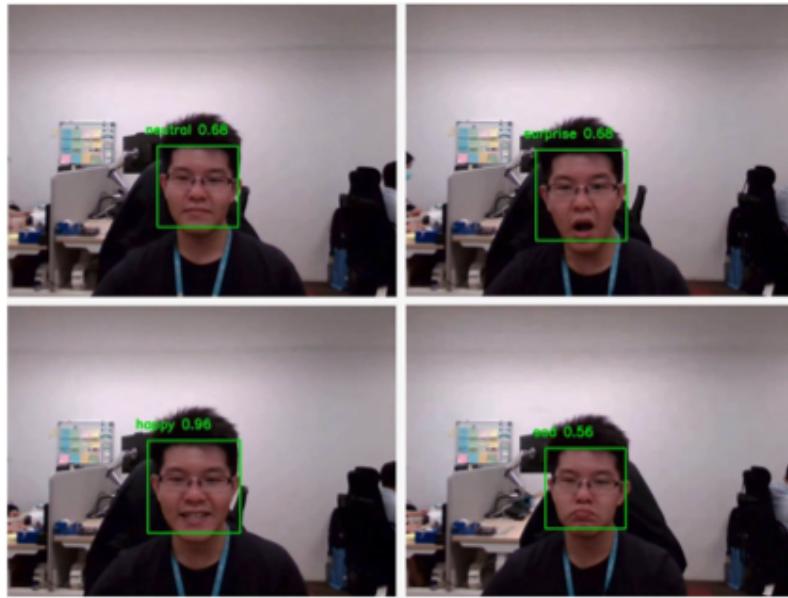


Figure 12: Pre-trained Keras CNN output on predicting facial expresions.

3.3.1.3 Eye Gaze Tracking - GazeTracking

For eye contact detection, we use the Python library **GazeTracking** which is efficient and fast for eye gaze tracking. There are 4 steps, pupil detection, blurring, eroding and binarization, to estimate gaze direction. Our model expects center gaze with direct eye contact with the camera.

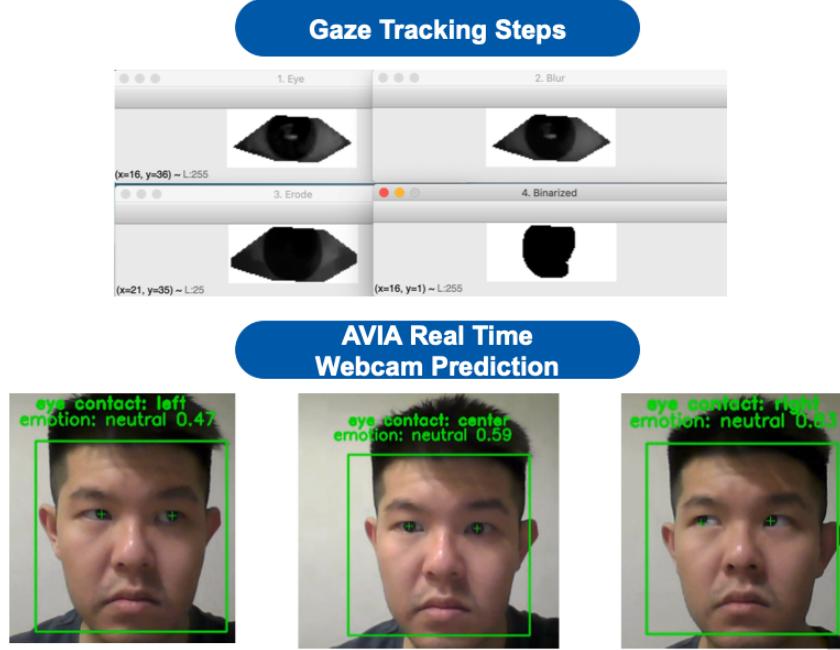


Figure 13: Eye gaze tracking method used by AVIA through GazeTracking Python library.

3.3.2 Text Analytics

3.3.2.1 Sentiment Analysis - TextBlob

Natural language processing (NLP) is a machine learning technique that allows computers to break down and understand text much as a human would. We use this to build our model to understand candidates' type-in responses. Sentiment analysis data using TextBlob models has been trained over more diverse data instead of just twitter data. Polarity is calculated out for the final result.

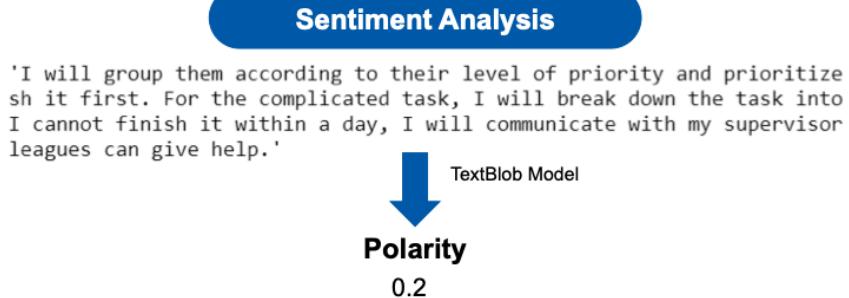


Figure 14: Sentiment analysis on candidate's written answer.

3.4 Data Visualization

For visualizing the web-application, **streamlit** Python library was used and charts were plotted using Python library **altair**.



Figure 15: AVIA Web application built using streamlit and charts plotted using altair.

4 Chapter 5: Discussions

4.1 Video Analytics

From the facial expression prediction, candidates can review their results by each of the questions they answered. Since our goal is to help the candidates to get well-prepared for the AI interview, the summary of facial expression percentage by question can help see their own performance.

Stage 1 - Video Interview

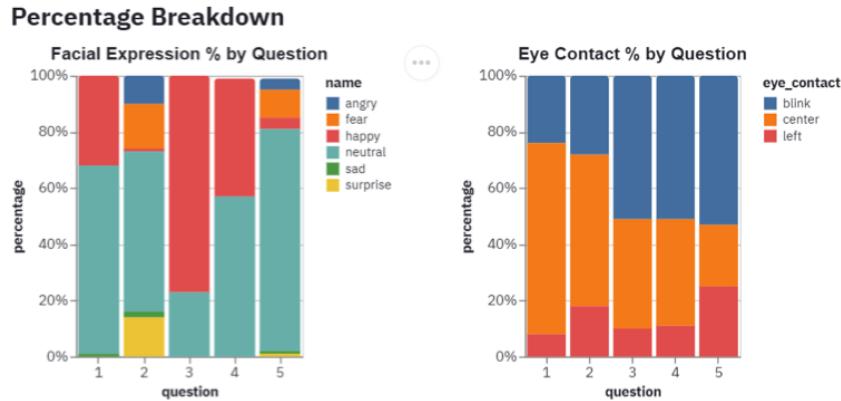


Figure 16: Results from a candidate's video interview.

For example, from the summary above, we can see that the 5 questions are answered with a neutral facial expression. In question 3, it is answered with a happy facial expression. During the interview, the company prefers to hire candidates with more happy facial expressions. Candidates who are confident, passionate, and positive will have a higher proportion of happy facial expressions than the other emotions. For question 2 and 5, only a little proportion predicted happiness. The candidate can adjust their facial expression and

smile more on those questions to give a better impression. Through practicing with AVIA, we believe that our candidates can get used to talking in front of the camera and perform well in the facial expression part.

In the eye contact part, a similar summary chart will be provided. For this new norm, candidates are asked to talk to the AI instead of a real person. When candidates are not comfortable with this new norm, he/she may be nervous and blink a lot or forget to give eye contact to the camera. The summary graph below helps the candidate to see whether they have good eye contact with the camera and carry out proper adjustment. From the summary graph, we can see the eye contact percentage by question. There are three parts of eye contact, “blink”, “center” and “left”. For question 2 and 5, with the least proportion of happiness, we can see that the percentage of “left” on eye contact is higher than those of the other questions. This is exactly reflecting difficulty and weakness when answering question 2 and 5. The higher the percentage on “center”, the more confident the candidate is when answering the question. Therefore, the candidate should practice question 2 and 5 more for better performance.

$$\text{Video Interview Score} = \frac{\sum_1^n(p(emotion) \times score)}{n} + \frac{\sum_1^n(p(eye contact) \times score)}{n}$$

Figure 17: Formula to compute the score for video interview.

4.2 Text Analytics

For the second part of the interview, the text analytics, we aim to match the model answer as close as possible. Therefore, a hit rate will be calculated. In order to increase the hit rate, candidates would like to match as many words as possible in the noun and verb group we extracted out from the model answer. A summary chart of keywords hit rate and sentiment breakdown will be provided. Below is the result for question 2 for the type-in response part.

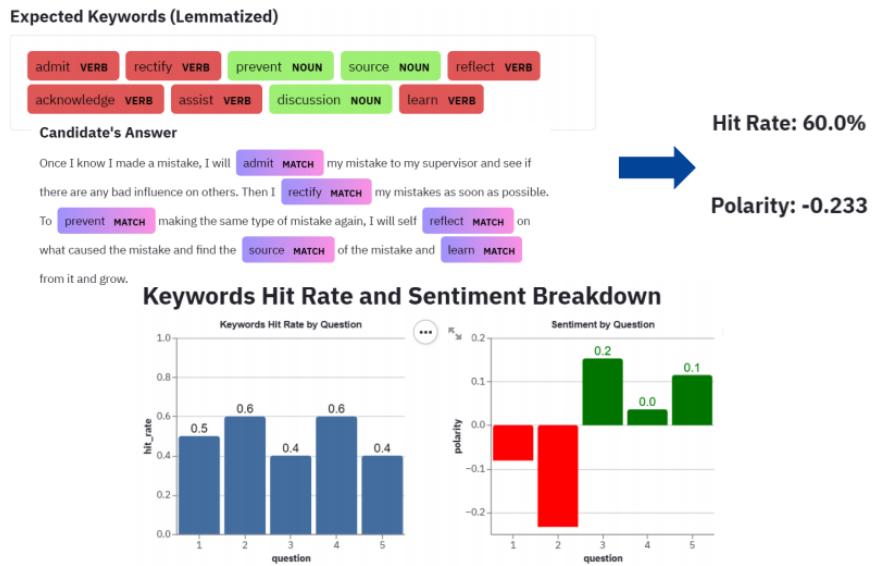


Figure 18: Results of a candidate's response and the ideal model answer keywords (lemmatized nouns and verbs).

We can see that out of the 9 keywords we extracted, the candidates hit 6 of them. The hit rate would be 60% in question 2. Since we also perform a sentiment analysis using the Textblob Model, we can have a polarity for the answers to each question. From the summary chart, we can see that question 1 and 2 are having negative polarity since the questions are about facing difficulties and mistakes. The result for the sentiment

part is not exactly what we expected since the polarity mainly relies on the topic we are discussing instead of the polarity of the type-in responses. Therefore, we decided to make it less weighted when calculating the overall score. Below is a detailed formula for the writing score of the second part of our interview. We used 5 as a base score and the full score is 10, which is the same as the video interview score.

$$\text{Writing Score} = 5 + 4 \frac{\sum_1^n(p(\text{hit_rate}))}{n} + \frac{\sum_1^n(p(\text{polarity}))}{n}$$

5 serves as a base score

Figure 19: Formula used to compute written assessments score.

4.3 Combined Overall

By combining the two parts of the interview, a performance review will be generated as below. Taking the average of the 2 scores discussed, an overall score is out. With the help of AVIA, our candidates can know which questions they are not doing well and keep practicing on them. Our candidates can get a higher score each time after they adjust for their weakness. This is the objective of our model, helping candidates to get prepared for the AI interview.

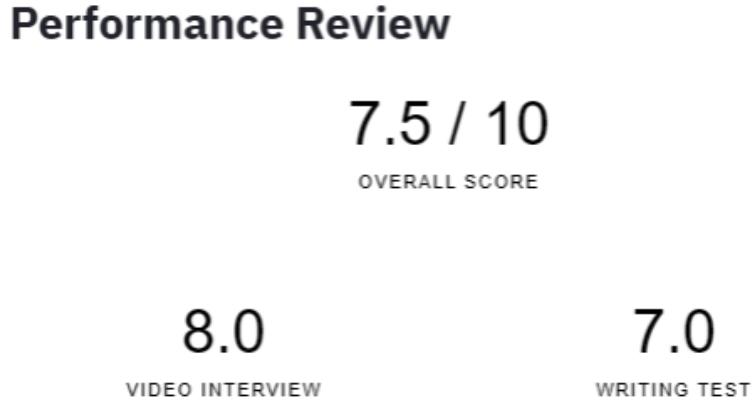


Figure 20: AVIA results of overall score and breakdown of video interview and writing assessment.

5 Chapter 6: Conclusion

Building AVIA, an AI web application that helps prepare candidates for online interview and assessments, is our goal to solve the problem of the new norm. Since the candidates have no resources to get prepared for the online interview, AVIA is definitely a good helper to them. AVIA uses face detection, facial expression recognition, eye gaze tracking for video interview and text mining and sentiment analysis for writing assessment. These are all important grading criteria in AI interview. In order to pass the first round with AI interview, our candidates can practice answering with confidence, meaning with a happy facial expression. Our candidates are also trained up to look at the camera, meaning the eye gaze to be located at the center. With the Demonstrated usage of AVIA web application real time with working functionalities and features, AVIA can help eliminate candidates' nervousness through practicing real time interviews in front of the camera. For the type-in responses, candidates can see what keywords are expected from the interviewer. By practicing the usage of more and more positive keywords, the candidates can answer confidently no matter

what questions they face in the future. We believe that with the help of AVIA, our candidates are able to express themselves confidently when facing the camera and typing in their responses.

We know that there is always room for improvements for AVIA. There are some proposed features we would like to continue to work on in the future, for example, speech analysis. In addition to the video and text analysis, we can also work on the audio part, to analyze the speaking speed or content of the candidates. Other features, such as stutter analysis, filler words analysis, are also helpful in practicing AI interview. Not limited by face detection, we can also add advanced posture detection such as shrug or hand on face/chin. The more analysis added with AVIA, the overall score will be more accurate. We wish to give a thorough training to our candidates in all aspects to get them well-prepared for the AI interview.

6 Chapter 7: References

- Liu, H. (2014). Facial detection and recognition on mobile devices. Morgan Kaufmann.
- Bin Aziz, Fayeem & Mokhtar, Norrima & Arof, Hamzah & Mubin, Marizan & Ibrahim, Zuwairie. (2014). Review of Eye Gaze Tracking Application.
- Nonis, F., Dagnes, N., Marcolin, F., & Vezzetti, E. (2019, September 18). 3D Approaches and Challenges in Facial Expression Recognition Algorithms-A Literature Review. MDPI. <https://www.mdpi.com/2076-3417/9/18/3904>.
- Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E., Scherer, K., Tomita, M., & Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, 53(4), 712–717. <https://doi.org/10.1037/0022-3514.53.4.712>
- Cherry, K. (n.d.). Are Our Emotional Expressions Universal? Verywell Mind. <https://www.verywellmind.com/the-expression-of-emotion-2795180>.
- Pejić Bach, M., Krstić, Ž., Seljan, S., & Turulja, L. (2019, February 28). Text Mining for Big Data Analysis in Financial Sector: A Literature Review. MDPI. <https://www.mdpi.com/2071-1050/11/5/1277>.
- Cavnar, B. (n.d.). N-gram-based text categorization. Ann Arbor MI. https://www.academia.edu/1046696/N_gram_based_text_categorization.
- Talview Hiring Reviews - Pros & Cons, Ratings & more. GetApp. (n.d.). <https://www.getapp.com/hr-employee-management-software/a/talview/reviews/>.
- Sambare, M. (2020, July 19). FER-2013. Kaggle. <https://www.kaggle.com/msambare/fer2013>.
- Microsoft. (n.d.). microsoft/FERPlus. GitHub. <https://github.com/microsoft/FERPlus>.
- MahmoudSabra1. (n.d.). MahmoudSabra1/Facial-emotion-recognition. GitHub. <https://github.com/MahmoudSabra1/Facial-emotion-recognition>.
- Ganguly, S. (2020, August 23). Confident Unconfident Facial Expression. Kaggle. <https://www.kaggle.com/sougatganguly/confident-unconfident>.

7 Chapter 8: Appendix

7.1 Results Summary

Below are the summary results of the prediction outputs:

Performance Review

7.5 / 10

OVERALL SCORE

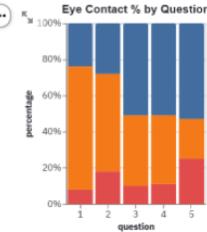
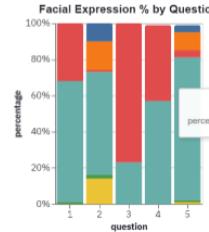
8.0

VIDEO INTERVIEW

7.0

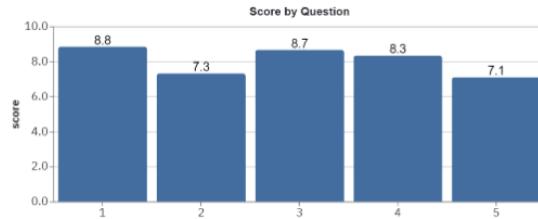
WRITING TEST

Percentage Breakdown



Stage 1 - Video Interview

Score Breakdown



Stage 2 - Writing Test Assessment

Score Breakdown

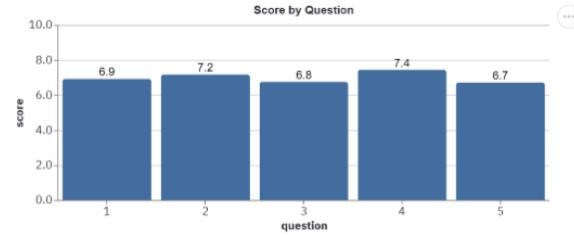
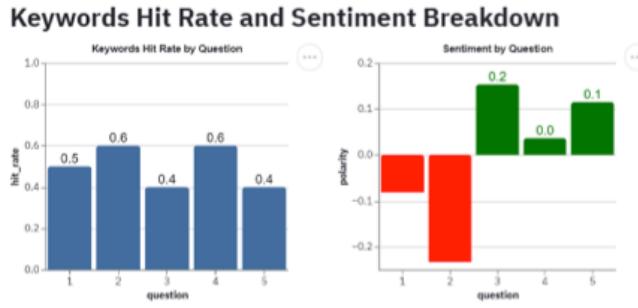


Figure 21: Results of the prediction of the AI techniques applied in AVIA.



Question 1

Expected Keywords (Lemmatized)

priority NOUN	supervisor NOUN	prioritize VERB	help VERB	grouping NOUN
extend VERB	deadline NOUN	report NOUN	liaise NOUN	continue VERB
persist NOUN				

Candidate's Answer

I will group them according to their level of priority MATCH and prioritize MATCH them. If the task is urgent and critical, I will finish it first. For the complicated task, I will break down the task into several pieces and finish it step by step. If I know I cannot finish it within a day, I will communicate with my supervisor MATCH to see if the deadline MATCH can be extended MATCH or if some colleagues can give help MATCH.

Hit Rate: 50.0% **Sentiment Polarity: -0.081**

Figure 22: Results of the prediction of the AI techniques applied in AVIA. (cont.)

Below are the summary of the technology stack the team used to build AVIA:

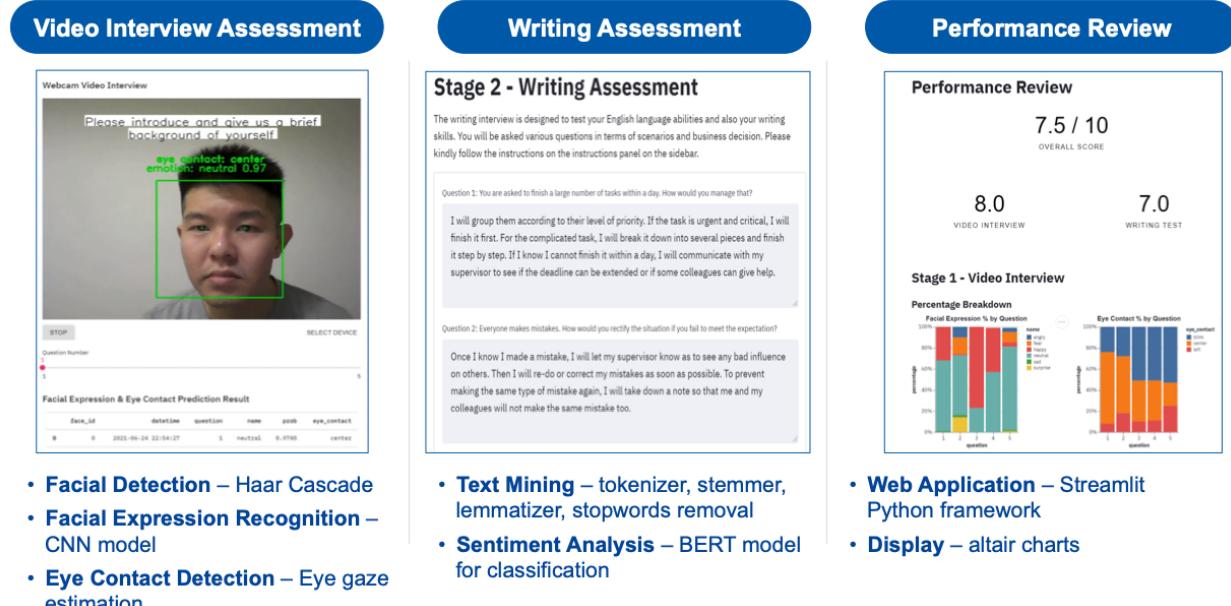


Figure 23: Summary of tech stack used to build AVIA.

7.2 Program Documentation

Below are the screenshots of the AVIA web application and some basic instructions on how to use AVIA.

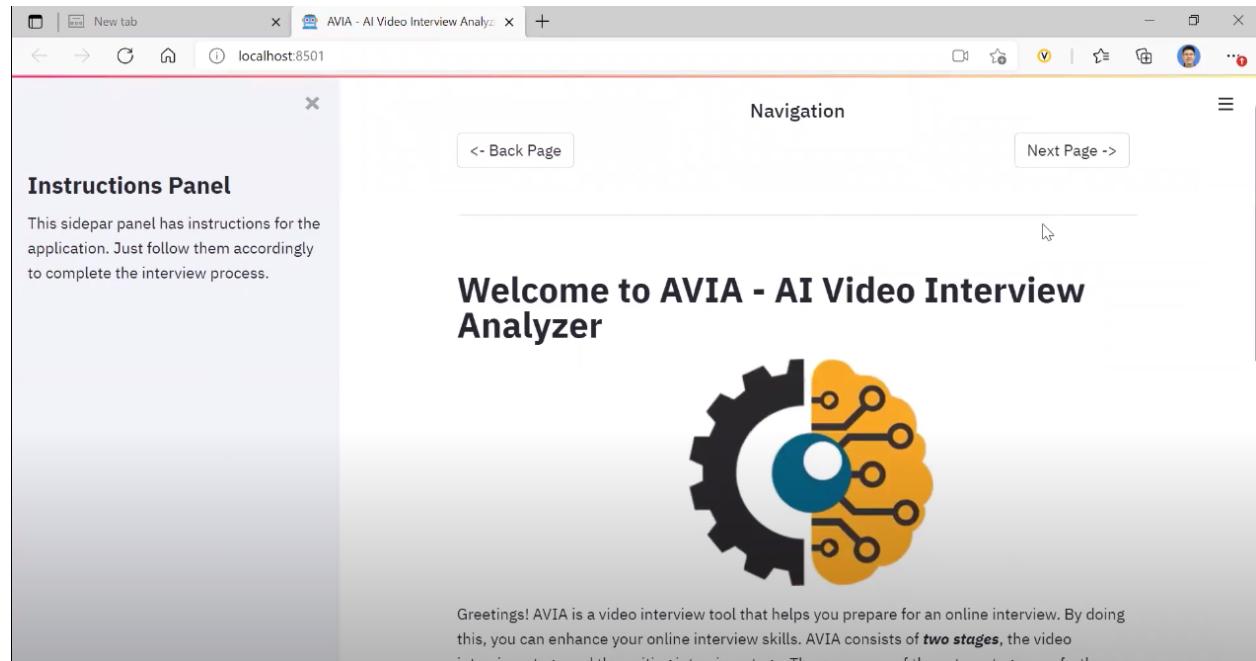


Figure 24: Step 1: Launch AVIA and read greeting message and instruction panel. Click next page after.

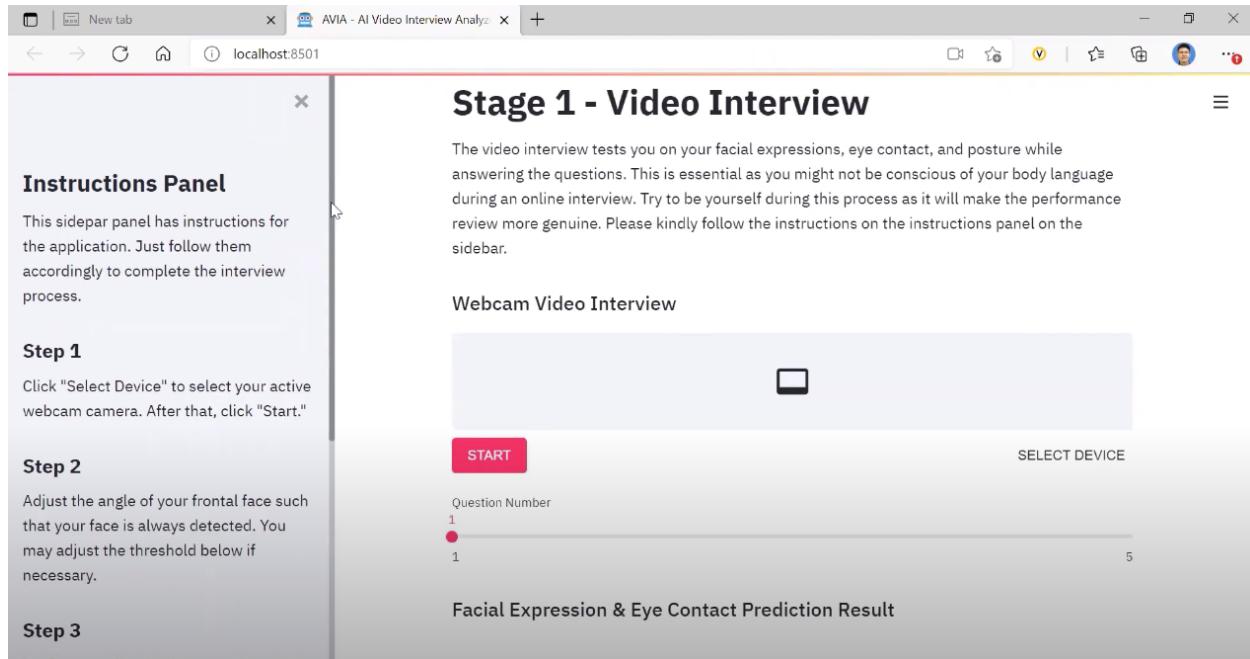


Figure 25: Step 2: Read video interview page instructions panel and select device. After that, click start to begin video interview

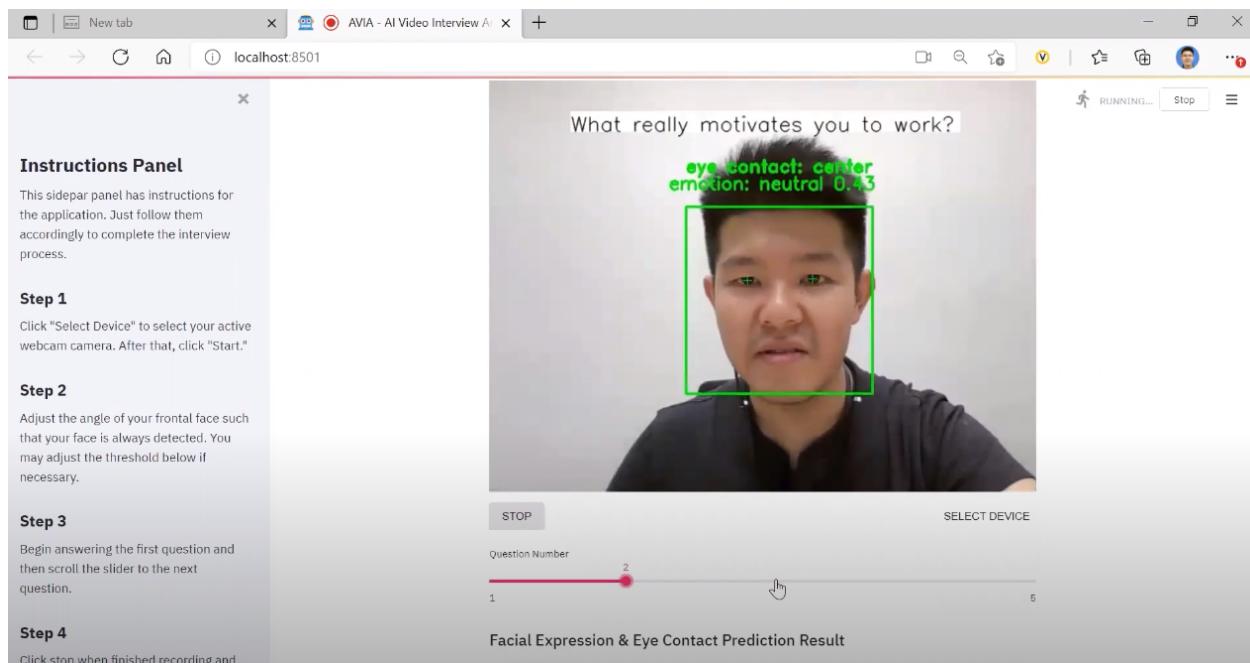


Figure 26: Step 3: Answer the questions and drag the slider to go to next question. Click STOP when finished answering all questions.

Instructions Panel

This sidebar panel has instructions for the application. Just follow them accordingly to complete the interview process.

Step 1

Read the questions carefully and answer them with the best of your abilities. Answer all questions. Please make sure you have at least 3 sentences of answer for each question. Do expand the text box shall you require larger space to write.

Step 2

After you finished writing your answers for all questions, and press submit.

Step 3

Scroll back to the top and click on next page to move on to the performance review section.

Stage 2 - Writing Assessment

The writing interview is designed to test your English language abilities and also your writing skills. You will be asked various questions in terms of scenarios and business decision. Please kindly follow the instructions on the instructions panel on the sidebar.

Question 1: You are asked to finish a large number of tasks within a day. How would you manage that?

I will group them according to their level of priority and prioritize them. If the task is urgent and critical, I will finish it first. For the complicated task, I will break down the task into several pieces and finish it step by step. If I know I cannot finish it within a day, I will communicate with my supervisor to see if the deadline can be extended or if some colleagues can give help.

Question 2: Everyone makes mistakes. How would you rectify the situation if you fail to meet the expectation?

Once I know I made a mistake, I will admit my mistake to my supervisor and see if there are any bad influence on others. Then I rectify my mistakes as soon as possible. To prevent making the same type of mistake again, I will self reflect on what caused the mistake and find the source of the mistake and learn from it and grow.

Question 3: When there is an expansion of your team and many new teammates join in. How would you adjust for this?

I will be friendly to them and give a helpful hand whenever they need it. Then, I will see if my supervisor has any training plan for the new teammates in which I can contribute my experience and knowledge. Clear division of labor and communication is needed since

Figure 27: Step 4: Read the instructions panel of the writing assessment and answer all questions. After that, click submit and then next page.

Instructions Panel

This sidebar panel has instructions for the application. Just follow them accordingly to complete the interview process.

Congratulations on completing all stages of AVIA! You may now review your performance in this section. This section will have a breakdown for Stage 1 - Video Interview and also Stage 2 - Writing Test Interview. Take your time to review these results, gather some insights, and identify areas where you may improve.

Scores Explanation

Below are the explanations of the scores:

- **Overall Score:** Average score of video interview and writing test.
- **Video Interview:** Score of video interview.
- **Writing Test:** Score of writing test.

Performance Review

7.5 / 10
OVERALL SCORE

8.0
VIDEO INTERVIEW

7.0
WRITING TEST

Stage 1 - Video Interview

Score Breakdown

Score by Question

Score	Question
8.8	Q1
7.3	Q2
8.7	Q3
8.3	Q4
7.1	Q5

Figure 28: Step 5: Review candidate's score and breakdown of the scores of each section thoroughly.

7.3 Program Coding

The code for this AVIA project is provided in a file separately, please kindly refer to those files. As AVIA application has several components and scripts, it is easier to refer to separately rather than putting it all in this document.