2021 ASEE ANNUAL CONFERENCE
Virtual Meeting | July 26–29, 2021 | Pacific Daylight Time
Paper ID #34758

Engage AI: Leveraging Video Analytics for Instructor-class Awareness in Virtual Classroom Settings

Mr. Jeremy Stairs, University of Toronto

Jeremy Stairs is an undergraduate student in computer engineering and artificial intelligence at the University of Toronto (graduating May 2021).

Mr. Raman Mangla, University of Toronto

Raman Mangla is an undergraduate student studying computer engineering with focus on software and artificial intelligence at the University of Toronto (graduating May 2021).

Mr. Manik Chaudhery , University of Toronto

Manik Chaudhery is an undergraduate student studying Computer Engineering at the University Of Toronto with a focus on Artificial Intelligence and Business. Manik will be graduating in May 2021.

Mr. Janpreet Singh Chandhok, University of Toronto

Janpreet Singh Chandhok is an undergraduate student in computer engineering and artificial intelligence at the University of Toronto (graduating May 2021)

Dr. Hamid S. Timorabadi, University of Toronto

Hamid Timorabadi received his B.Sc, M.A.Sc, and Ph.D. degrees in Electrical Engineering from the University of Toronto. He has worked as a project, design, and test engineer as well as a consultant to industry. His research interests include the application of digital signal processing in power systems.

©American Society for Engineering Education, 2021

Engage AI: Leveraging video analytics for instructor-class awareness in virtual classroom settings

Abstract

A difficulty for teachers in COVID-era online teaching settings is assessing engagement and student attention. This has made adapting teaching to the responses of the class a challenge. We developed a system called Engage AI for assessing engagement during live lectures. Engage AI uses video-based machine learning models to detect drowsiness and emotions like happiness and neutrality, and aggregates them in a dashboard that instructors can view as they speak. This provides real-time feedback to instructors, allowing them to adjust their teaching to keep students engaged. There is no video data transmitted outside of students' web browsers, and individual students are anonymous to the instructor. Testing in undergraduate engineering lectures resulted in 78.2% reporting feeling at least potentially more engaged during the lecture and at least 34.4% of students reporting feeling more engaged during the lecture. These approaches could be applicable to many forms of remote and in-person education.

Background

Billions of people are studying, working and socializing remotely in the current COVID-19 pandemic. The videoconferencing market has grown 20 fold some over the past few years, and is expected to grow from less than $14 billion in 2019 to over $50 billion in 2026 [1][2]. In the CMC (computer mediated communication) literature, the richness of a medium is defined as its capacity to change understanding, and it is commonly asserted that media with more information are richer [3]. For instance, Zoom and Google Meet are richer than text messages [3]. CMC researchers have shown richer media to facilitate more fluent conversation, interpersonal awareness, interpersonal bonding, oxytocin release, and perception of understanding [4][5][6][7]. Video is used to quickly communicate nonverbal cues for turn-taking, understanding, and attention [5][8][9].

State of the art videoconferencing reflects the findings of the literature on richness; Zoom and Bb Collaborate feature simultaneous videoconferencing, emoji reactions, text chat, screen sharing, and breakout rooms. However, they are limited in parallel communication; for sound over CMC media such as Zoom and Bb collaborate, only one person can be speaking at a time. For video over these media, the number of people one can simultaneously observe is limited by the size of their screen. Neither of these platforms provide a way to communicate body language and verbalizations of many people simultaneously. Some attempts have been made to do this; a 2006 US patent describes automated multisensory emoticons to increase richness of communication in the context of video games [10].

The inability of current CMC media to convey group verbal and body language responses can have a negative impact on activities that require rich group communication, such as dynamic teaching and learning in classroom settings. As an anecdotal example, a professor at University of Notre Dame describes teaching virtually during COVID-19 lockdown as follows:

> "I am continually repressing my lifelong, trained habit of uttering simultaneous encouragement through 'continuers,' those back-channel cues that encourage the speaker to go on." [11]

Group feedback including laughter, fatigue, visual engagement, head tilting, and auditory cues are no longer accessible to teachers during COVID-19. This may have a serious negative impact on the ability of teachers to adapt their teaching to the responses of the class. For instance, in physical classrooms, when a teacher notices students becoming drowsy, they may opt to ask questions of the class or add some humour to engage their students. The loss of information channels exists beyond education; businesses, conferences, and families have all had to imperfectly adapt to less-rich communication.

Problem Statement

A gap in live virtual communication in large group settings is a feedback mechanism for speakers to gauge the engagement and reactions of other people. This gap widens when there are 100 videos for the speaker to analyze at the same time. Oftentimes people do not want to share live video with the speaker, further minimizing feedback.

This paper describes Engage AI, a machine learning software system that can help speakers adapt to their audiences by providing real-time audience engagement feedback. The project integrates with existing video conferencing applications to provide automated live feedback for virtual classroom settings.

Technical Design

As shown in Figure 1, Engage AI performs 3 tasks: collect participation data, process data using machine learning algorithms and showcase engagement metrics. Data collection is performed using a student participation frontend that runs in the browser for ease of use and client side processing. The student video is analyzed on the client side using an open source ML framework called Google MediaPipe. Then information of facial landmarks, such as lip shape and eye shape are shared with the application backend. This allows the lecturer to receive engagement feedback without students sharing their video streams with the speaker. This also ensures that no student video data is transferred or stored by the application.

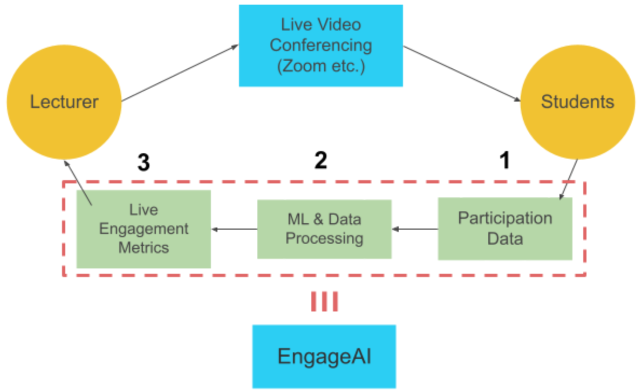

Figure 1: System Context Diagram

The data is processed using custom ML algorithms and open source frameworks, such as PyTorch and OpenCV. The ML models were trained on 1000 facial images from the Flickr FFHQ dataset [12]. The lecture metrics are calculated and anonymously aggregated in real-time. The anonymity of facial landmark data helps alleviate privacy concerns. The metrics are displayed on the lecturer dashboard using intuitive charts and graphics. A representation of the dashboard is shown in Figure 2. Furthermore, Figure 3 illustrates a simplified system diagram showing various technical components.

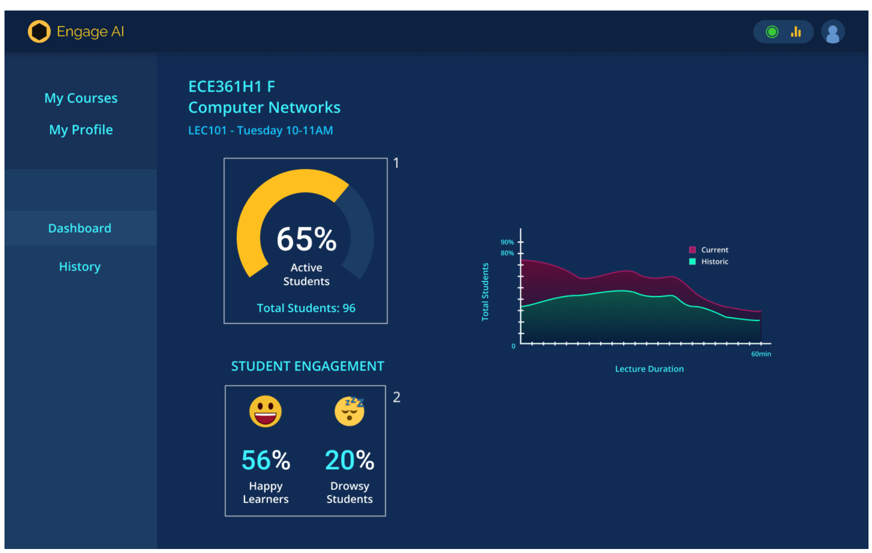

Figure 2: Lecturer Dashboard Mockup

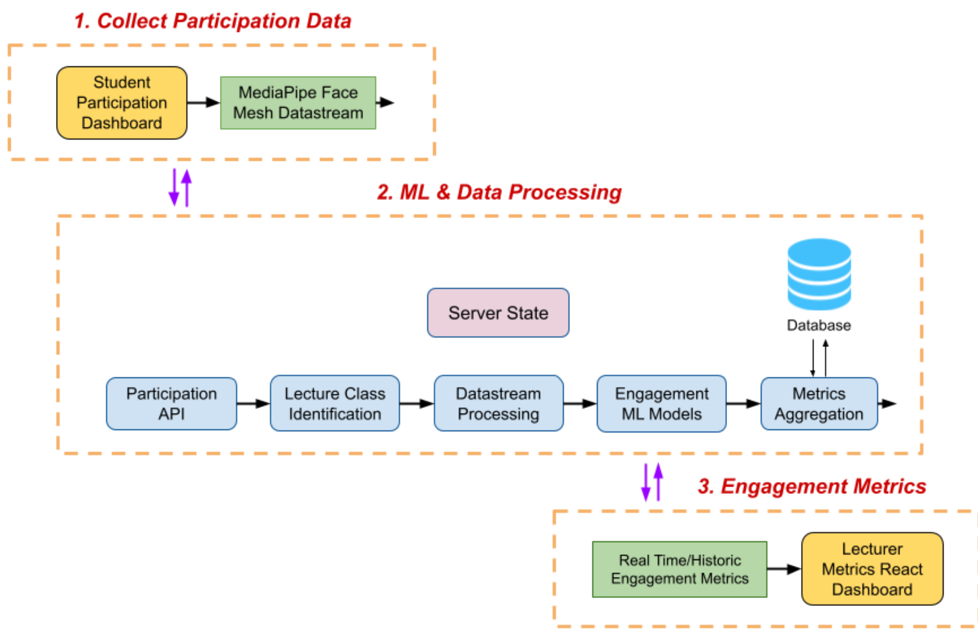

Figure 3: Engage AI System Diagram

Testing and Verification

Prior to developing the product, we surveyed 14 people working/studying from home, the results of which can be found in Figure 4 and 5. Over 71% of the participants surveyed approved of the problem we are solving. The verification table (Table 1) below describes the system requirements with testing criteria.

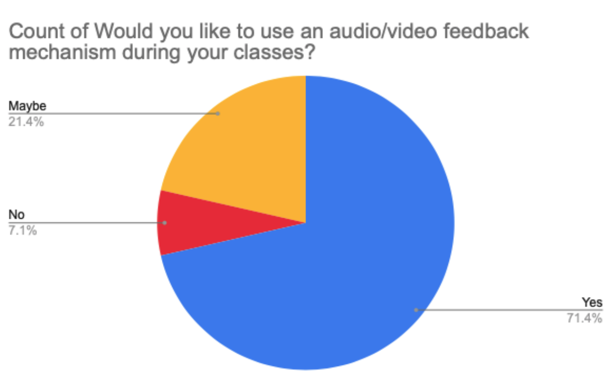

Figure 4: Survey Response 1

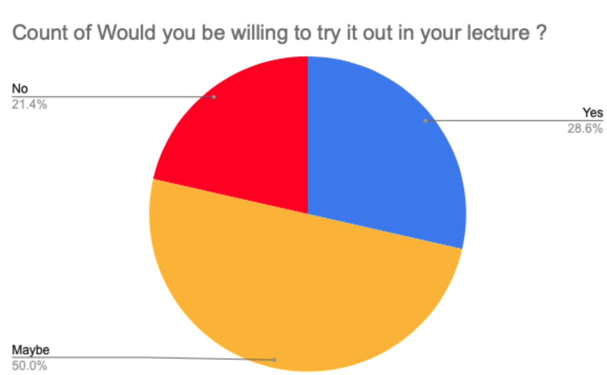

Figure 5: Survey Response 2

Table 1: Verification Table

ID	Requirement	Description
1.0	Function: Measure participants' engagement in virtual communication.	**TEST:** User experience score test. This will be evaluated using a user experience test Pass: If user experience score > 50%
2.0	Function: Provide quantitative and qualitative feedback to the participants in < 1s.	**TEST:** The feedback graph and emotions on the dashboard should be updated every second. Pass: If real time metrics update every second.
3.0	Function: Assist the speaker in understanding the audience's reactions. User satisfaction rate should be > 80%	**TEST:** A survey of participants will be used to obtain feedback on their attention levels. The survey results will be compared with the engagement metrics from our machine learning models. Pass: ML model results are over 80% accurate as compared to the manual survey results.
5.0	Constraint: Must integrate with existing video conferencing systems. Work with speeds >= 384Kbps	**TEST:** Pass if it works on networks with speed >=384Kbps
6.0	Constraint: Must require informed consent from users	**TEST:** Direct measurement
7.0	Cost effective, cost less than $457/month per 1000GB of data processed	**TEST**: Direct Measurement.

Testing Procedure

Testing was carried out in collaboration with undergraduate engineering students and professors. The classes are hosted on Blackboard Collaborate and the students are requested to open the student webpage. Once the webpage is opened by the students, they click participate in order to acknowledge their voluntary participation in our system. No personally identifiable information is recorded by the webpage.

Once students have loaded the webpage and began participation, instructors monitor their students' engagement by visiting the instructor dashboard webpage as shown in Figure 2.

For the testing procedure the following steps were taken:

- Instructor directs students to the webpage
- Instructor directs students to click the "participate" button.
- Instructor teaches while using Engage AI to adapt to class fatigue, happiness etc.
- At random points during the lecture, the instructor asks students to act tired, asks students to smile, or makes jokes to see if genuine laughter is captured.
- After the lecture, students are requested to fill out a survey, and the instructor is interviewed.

Testing Results

Tests were carried out in two classes in February 2021. *Actual emotions* are the emotions that the instructor asked students to show, or those that were inferred by students reporting their emotions. *Happy* and *Drowsiness* are the percentages of students expressing emotions recognized by the Engage AI dashboard. *Total Active Students* is the number of students recognized by Engage AI at the time of the measurement.

Lecture #1

Date: 2021-02-24

During this lecture a small number of students were invited as this allowed a more controlled and manageable environment to conduct the test. The instructor, while teaching, instructed students to emulate different emotions at different timings of the lecture. For instance, in test number 2 and 3 all students and then ⅔ of students were encouraged to smile respectively. Test number 5 shows the result when half of the students were instructed to pose as sleepy. The remaining test results are pulled from the natural course of the lecture where students react. Engage AI can also determine the number of students who are actively maintaining a good connection with the app. This is also shown in Table 2 as the Total Active Students. We found out that few students were experiencing slow Internet connections.

Table 2: Lecture #1 observation and data collected

Test Number	Actual Emotion	Happy	Drowsiness	Total Active Students (Out of 15)
1	Happiness	87%	-	11
2	Happiness	93%	-	12
3	Happiness	75%	-	11
4	Happiness	63%	-	13
5	Drowsiness	-	50%	15
6	Drowsiness	-	79%	14
7	Drowsiness	-	58%	11
8	Drowsiness	-	86%	13

Lecture #2

Date: 2021-02-24

Extra observations were made in lecture 2, and are shown in Table 3. The emotions at a random time interval is shown in row 1. It can be noted that as the professor was asking questions, class engagement increased (row 2), with drowsiness decreasing from 35% to 8%. The last row shows that when students were asked to smile if they knew the answer to a question, 3 knew the answer. *Actual # students expressing happiness* gives the actual number of students that were expressing happiness when measured. Note that the happiness percentages do not perfectly match the actual # students expressing happiness / total active students.

Table 3: Lecture #2 observations

Test Scenario	Happy	Drowsiness	Total Active Students	Actual # students expressing happiness
Professor lecturing	23%	35%	14	4
Professor asks question	34%	8%	14	6
Students who knew answer instructed to smile	21%	25%	14	3

Feedback

We also conducted multiple surveys among students and instructors. The student surveys ask students how comfortable they were using the app, whether the app caused any performance issues on their computer, whether they felt the app made them more engaged in the lecture, and whether they had any suggestions for improvement to the app. We also conducted surveys and interviews with each professor who used the app. Results from the student survey can be found in Figures 6, 7, and 8. The most common request for improvements to the app was to show, in real-time, what data is being sent to the instructors within the student webpage. Students reported that this would make them feel more comfortable using the system. The most common requests by instructors were to simplify the dashboard so that only happiness and drowsiness are shown, and to make it small so that the instructor can use most of their screen for teaching. Instructors reported the system being easy to administer during the lecture.

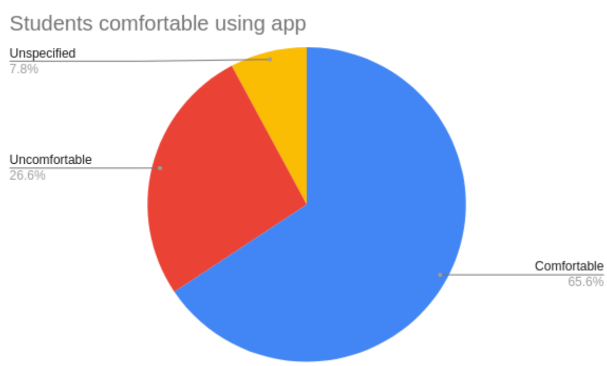

Figure 6: Student Comfort

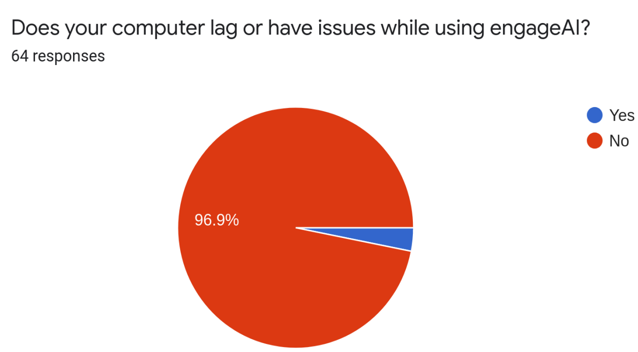

Figure 7: Performance

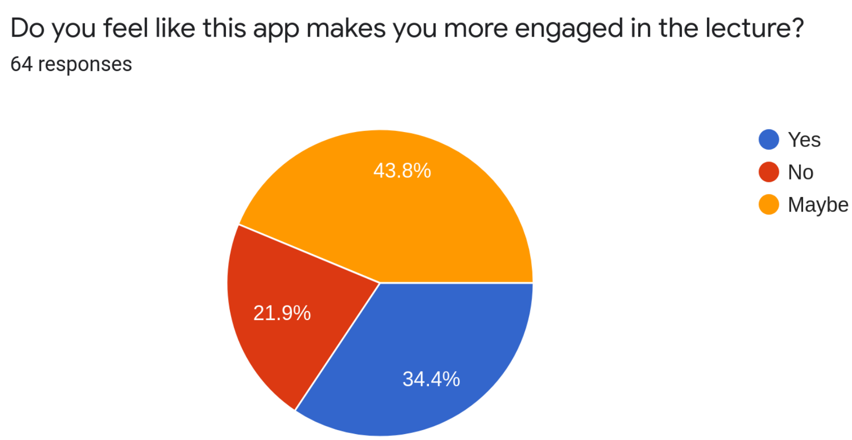

Figure 8: Engagement

As a result, we implemented a simplified dashboard as shown in Figure 9. Instructors can choose either to employ a complete dashboard, Figure 2, or this simplified dashboard that provides only the essential information during the lectures. The simplified dashboard leaves off the analytical tools and hence allows the instructor to focus on the lecture delivery. Then following the lecture the instructor can take advantage of the tools to evaluate the lecture delivery and performance. The simplified dashboard could also be integrated as part of an online multimedia platform such as Bb Collaborate or Zoom. This would make the use of Engage AI seamless to both students and instructors.

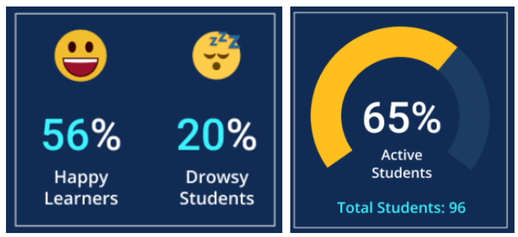

Figure 9: Simplified dashboard

Our project has received positive feedback from the instructors. While teaching a class of over a hundred students, it is virtually impossible to realize students' reactions to lecture delivery especially if they have their cameras turned off. Upon surveying the instructors, they conveyed that the feedback received from Engage AI is helpful. It allows them to determine students' engagement during the lectures. For instance, the drowsiness metric allowed one of the instructors to tell a joke when Engage AI reported an 80% level of drowsiness. This made a change in students' moods and hence allowed them to refocus for the rest of the lecture. Figure 10 shows that 100% of the instructors surveyed revealed that Engage AI helped them understand engagement of their students.

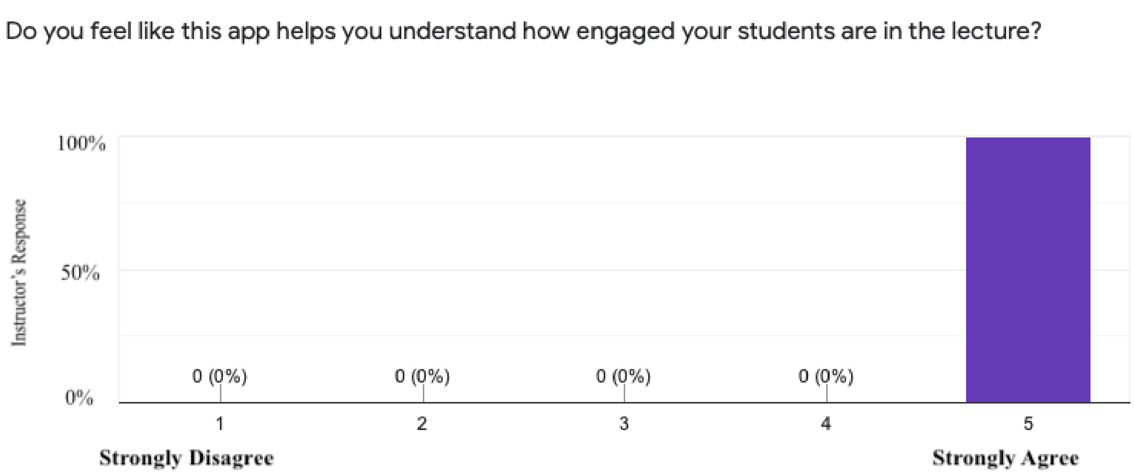

Figure 10: Instructor Survey

It is at the discretion of instructors to take action with respect to the instantaneous feedback provided by Engage AI. This is similar to an in-person classroom where each instructor reacts to dynamics of the lecture differently. One instructor may choose to tell a joke if students are becoming sleepy/drowsy while another instructor may not react. For instance, one of the participating instructors has been using Engage AI since the initial test was conducted. The attending students range between 76 to 91. The topic taught employs PowerPoint slides that can become unexciting at times. The instructor, with the help of Engage AI, has managed to introduce appropriately-timed jokes and 1-minute breaks to increase students' engagement.

Conclusion

A system for providing real-time student engagement feedback to instructors in remote lecture settings was developed and tested on undergraduate engineering classes. The system could be used to help instructors make their computer-mediated teaching more adaptive, e.g. to find a good time to give students a break or add humour. It did not require students to share their video in the lecture, and provided anonymity to students by transmitting only facial expression features and aggregating the features across all students. The system was able to accurately detect drowsiness, smiling and laughter, and presence. Feedback was promising: 34.4% of students reported feeling more engaged during the lecture, and a total of 78.2% reported either feeling or potentially feeling more engaged during the lecture. Most students reported feeling comfortable using the system during lectures, and instructors found the system easy to use while teaching. These results may be justification for longer-term and more rigorous studies of automated engagement feedback in classrooms. It is conceivable that in the future an automated feedback mechanism like Engage AI could be integrated into teaching software like Bb Collaborate. Because systems like Engage AI can protect anonymity, the choice of making the use of such a system required or recommended could be up to the instructor.

The primary goal of Engage AI was to provide real-time feedback to instructors or presenters. This is very similar to an in-person classroom setting where one may choose to react to dynamics of the classroom or not. As Engage AI progresses, we will be not only adding further emotions to assist instructors to understand their audience better but we will provide suggestions to the instructors as well. For instance, we could suggest by popup window a joke or 1-minute break. The advances in machine learning can improve teachers' abilities to understand their students. AI-enhanced real-time teacher feedback is not only applicable to COVID-era learning. It is a promising direction for both virtual and in-person classroom settings.

References

[1] *Video Conferencing Market Size, Share, Trends and Growth, 2026*, Dec-2019. [Online]. Available: https://www.fortunebusinessinsights.com/industry-reports/video-conferencing-market-100293. [Accessed: 22-Sep-2020].

[2] P. Wadhwani and S. Gankar, “Video Conferencing Market Trends: Statistics Report 2026,” *Video Conferencing Market Trends*, May-2020. [Online]. Available: https://www.gminsights.com/industry-analysis/video-conferencing-market. [Accessed: 22-Sep-2020].

[3] R. L. Daft and R. H. Lengel, “Organizational Information Requirements, Media Richness and Structural Design.” *Management Science*, vol. 32, no. 5, pp. 554-571, 1986, doi: 10.1287/mnsc.32.5.554.

[4] L. E. Sherman, M. Michikyan, and P. M. Greenfield, “The effects of text, audio, video, and in-person communication on bonding between friends.” *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, vol. 7, no. 2, 2013, doi: 10.5817/cp2013-2-3.

[5] O. DALY-JONES, A. MONK, and L. WATTS, “Some advantages of video conferencing over high-quality audio conferencing: fluency and awareness of attentional focus.” *International Journal of Human-Computer Studies*, vol. 49, no. 1, pp. 21-58, 1998, doi: 10.1006/ijhc.1998.0195.

[6] L. J. Seltzer, A. R. Prososki, T. E. Ziegler, and S. D. Pollak, “Instant messages vs. speech: hormones and why we still need to hear each other.” *Evolution and Human Behavior*, vol. 33, no. 1, pp. 42-45, 2012, doi: 10.1016/j.evolhumbehav.2011.05.004.

[7] “Challenges in Virtual Collaboration: Videoconferencing, Audioconferencing, and Computer-Mediated Communications” https://www.rand.org/content/dam/rand/pubs/monographs/2004/RAND_MG273.pdf (accessed: Sep. 19, 2020).

[8] Slovák, P, “Effect of Videoconferencing Environments on Perception of Communication.” *Cyberpsychology: Journal of Psychosocial Research on Cyberspace,* vol. 1, no. 1, Article 8, 2007

[9] J. K. Burgoon, J. A. Bonito, A. Ramirez, N. E. Dunbar, K. Kam, and J. Fischer, “Testing the Interactivity Principle: Effects of Mediation, Propinquity, and Verbal and Nonverbal Modalities in

Interpersonal Interaction." *Journal of Communication*, vol. 52, no. 3, pp. 657-677, 2002, doi: 10.1111/j.1460-2466.2002.tb02567.x.

[10] J. D. MacAuley and E. J. Arthur, "MULTI-SENSORY EMOTICONS IN A COMMUNICATION SYSTEM ," 16-Jan-2006.

[11] "A professor explores why Zoom classes deplete her energy (opinion)." https://www.insidehighered.com/advice/2020/04/22/professor-explores-why-zoom-classes-deplete-her-energy-opinion (accessed: Sep. 19, 2020).

[12] GitHub. 2021. *NVlabs/ffhq-dataset*. [online] Available at: <https://github.com/NVlabs/ffhq-dataset> [Accessed 7 March 2021].