E-Blob: Explaining Student Engagement in Online Teaching

Kevin Bekker
TU Eindhoven
Eindhoven, NL
k.bekker@student.tue.nl

Huizhong Ye
TU Eindhoven
Eindhoven, NL
h.ye@student.tue.nl

Jing Li
TU Eindhoven
Eindhoven, NL
j.li1@student.tue.nl

ABSTRACT

It is a key topic for educators to obtain students' engagement in offline classrooms. However, in the context of the COVID-19 pandemic, teachers are confronting the challenge of understanding student engagement in an online education environment. Smart agents are able to automatically detect a student's engagement by combining emotion and attention, but the human agency is considered to be an essential role to achieve better interaction. We propose E-blob, a smart system that detects student engagement based on computer vision technology, with a blob on the interface explaining the algorithm in the user's interaction-attention continuum. The system is implemented locally with python and processing. For future work, the accuracy and user experience of the whole system need to be evaluated.

Keywords

XAI(Explainable AI), engagement, online lectures, interation-attention continuum

1. INTRODUCTION

Nowadays, many university courses are transmitted from offline to online, which allows students to attend lectures from all over the world. Especially due to the COVID-19 pandemic, students are suggested or asked to study from home, thus live-stream has become a very popular way of online education. What usually is a 3D has now become a 2D environment, where student engagement can hardly be detected by the teachers. As for teachers, they lose the opportunity to make eye contact and observe the expressions and movements of students which makes it difficult for them to obtain subtle feedback. As a result, they can be confused and anxious [1]. On the other hand, the teachers cannot adjust the pace or content of the course according to the students' engagement, so the students cannot get enough support from the lecturer [2]. Therefore, it is a key topic for teachers to obtain students' engagement in live classrooms.

Fredricks et al. [3] describe the behavioral, emotional, and cognitive engagement and recommend studying engagement as a multifaceted construct. During a live-stream lecturer, behavioral engagement can be detected easily, such as whether the student attends class, or submits required work. However, the emotional and cognitive engagement is usually unpredictable, which is always detected in an offline class by observing facial expression and attention.

To solve this problem, we designed E-Blob, see Figure 1, a smart system that reflects student engagement based on the detection of emotion and attention in an interactive and explainable way. The system captures images of students from the webcam and generates the emotion, attention, and engagement index through a machine learning algorithm. Then a blob on the lecturer's screen provides different levels of information, by making use of the interaction-attention continuum [4] and allows the user to switch within the attention range through cursor interactions.

We implemented the machine learning algorithm with Python and visualized the statistical outcome with Processing. With the working demonstration, evaluation plans were made to test the accuracy of the system, as well as user experience. Our system is supposed to be deployed as a plugin with any live-stream education platforms. The explainable AI will enable teachers to actively assess the lecture, and the quality of the lecture is expected to be improved.

Figure 1. E-Blob working on the teacher's laptop

2. RELATED WORKS

2.1 Automatic recognition of student engagement

Several studies proposed various methods for automatically recognizing student engagement in an e-learning environment. There are three main categories: engagement tracking, sensor data analysis, and computer vision-based detection.

In the engagement tracking, the system gives out short questions to students and assesses the time and accuracy of the replying [5]. Other systems detect engagement based on physiological (EEG, blood pressure, heart rate, etc) and neurological readings (arousal or alertness) [6], which requires students to wear devices or use a special chair when taking online courses. There is no doubt that these two methods are very obtrusive for students and could have a negative influence on student engagement.

In contrast, computer vision-based methods unobtrusively recognize student engagement by analyzing cues from facial expressions [7] [8] [9] [10], hand gestures and body postures [11] [8], and eye movements [2].

However, most of the automatic engagement detection methods are applied in intelligent tutoring systems (ITSs) [11], or massive open online courses (MOOCs) [9] [10]. We found that there are still few studies on applying computer vision technology to real-time feedback on student engagement in remote live classrooms.

2.2 Interactive videoconferencing systems

The trend of shifting to online distance teaching from physical class requires teachers and students to rely on videoconferencing systems for communication [12]. For distance education, it caused an exorbitant amount of stress and anxiety among lecturers who have less experience in online teaching while trying to maintain the same quality as face-to-face learning [13]. The importance of simulating face to face interaction which enables immediate interaction and feedback between lecturers and students becomes the main consideration in this transformation [12]. An ideal interactive online teaching system is supposed to provide real-time interaction, enable instant feedback, and promote engagement [14]. Besides all audio, visual and verbal technologies support in the system, the human agency also plays an essential role to achieve better interaction. Humans act and make decisions based on their experiences, reflections, self-determination, and motivation [12]. In this digital transformation, the design and optimization of online teaching systems should advance toward a more human-centered approach [12]. Further consideration after embedding any AI assistance into online teaching systems, as a tool for enhancing human abilities to conduct education activities, should not take the initiative from humans. For example, the new generation of online teaching systems should give lecturers alike opportunities to make more accurate interaction and feedback and take advantage of the system to improve virtual working ability, instead of limiting it as restricted system functionality [12].

Taking everything into consideration, the goal of our design is designing an AI system to explain engagement situations to humans for supporting them in making decisions, and further improving the quality of online lectures.

3. METHODS AND MATERIALS

The goal of E-Blob is building an Explainable AI system for the end-users to make use of the Interaction-Attention continuum [4] to enhance interaction. Therefore, we not only applied the theory of Explainable AI but also the methods of creating various levels of human attention in interaction design of the system. In order to better interpret computer language to people, the working principles of the AI system are broken down to different levels, users' attention aroused in the means of interacting with different levels of information. In this way, the system can ensure that user attention is kept at the same stage as the progress of the information.

The design process consists of brainstorming, expert consulting, and online user survey. We iterate the designing during the process, see Appendix I.

3.1 Explain your approach/specific methods or theory.

Student engagement plays an essential role in evaluating the quality of online lectures[15]. By examining student engagement, it offers the right feedback to lecturers on organizing the structure and progress of an online lecture, furthermore, building an active interaction between students and lecturers during classes. However, traditional methods to assess students' engagement such as by questionnaire are time and effort consuming[15]. Take the advantage of computer vision that is mentioned in 2.1, intelligent software agents are more applicable to evaluate student engagement. Therefore, the primary step is identifying dimensions of student engagement from a computer perspective.

Social studies listed that Behavioral, Emotional, and Cognitive engagement are three main aspects of student engagement [3]. Other research [15] [16] has pointed out emotions as one of the most significant contextual factors shaping student's engagement. To address these aspects, we propose to use an emotion and attention combination model to evaluate the engagement [17]. Computer vision offers the prospect of unobtrusively estimating a student's engagement by analyzing cues from faces [18]. Thus, our system is designed to evaluate emotional engagement by focusing on facial emotion and cognitive engagement by learning cognitive abilities from gaze tracking.

To achieve the goal of E-blob and gain trust from users from increasing transparency, we applied **the theory of Explainable AI** into our design. The way the machine explains should be related to the human decision-making principle [19]. The decision-making process is supposed to be interpreted in the means of deduction, induction, and abduction [20] to the users. As users, they can know more information beyond a final result to understand certain situations. In our design case, by applying XAI, lecturers can understand an overall engagement conclusion of E-blob from a "bottom-up logic", that means they can see the weights of the emotional engagement and attention engagement and further to the weights different vectors of each aspect step by step. Based on the explainable structure, users' decisions can be influenced to react to various situations even with the same final engagement result.

The theory of Interaction-Attention Continuum was addressed in the design of E-blob for enhancing users' attention while interacting with the different-level of information. The inductive reasoning " bottom-up logic" of XAI requires increasing attention from users when it is processing. Therefore, interactions with computing technology should be available at various levels of attention [16]. E-blob is a blob interface without requiring any focused attention, it shows overall engagement in real-time consistently. In the beginning stage, it asks for implicit interaction with users. Once it gains users' attention, the following peripheral interaction like "mouse hover", and focused interaction like "mouse press" need to be conducted intentionally with increased attention from users. In this way, the series of interactions were designed to give control of XAI to users, making sure they are synchronized.

3.2 Learning algorithm

In order to determine the engagement of the student, a three part system is used. First the face is detected from the webcam feed, which is then cropped and outputted to the convolutional neural network (CNN) to determine the emotion. In parallel, the gaze of the student is detected by calculating the relative position of the pupils in relation to the landmarks of the eyes [21]. In this chapter the individual steps will be elaborated upon.

The system is based on the code found at: https://github.com/sherkhan15/distarctor-Detection_Eyes_Emotions_CI

Face detection

In order to filter the face from the webcam input the widely used Haar-based cascade classifier [22] is applied. Haar cascade is a popular method to detect objects in other images. In order to achieve the face detection, the classifier is trained on positive images (images with the object) and negative images (images without the object). It then compares the pixels of rectangular areas in an image, calculating the average intensity. This can then be matched to Haar-like features, detecting features like edges.

The output of the classifier is the cropped face of the user.

Emotion detection

To estimate the current mood of the student, a 21 layered CNN is applied to analyze the cropped image of the face. CNNs are widely used to identify objects in images and videos. The hidden layers in the CNN are convolutional layers which can be identified as filters. Each layer is set to detect a filter (e.g., texture, edges, shapes, etc.) and as more layers are present, more sophisticated filters can be applied (e.g., detecting specific objects). The output of the CNN are softmax scores, rating the probability of the estimated emotions on a range from 0 to 1.

Gaze detection

As mentioned in [17], another important variable in estimating the engagement of students is the gaze of the eyes. The input is the cropped face from the first step. This image is transferred to grayscale, after which a pre-trained deep neural net (DNN) [23] determines the 68 unique facial landmarks which can be found on the face. This is achieved by mapping the image to a 128 dimension vector space, comparing the face to other faces. The more similar structured faces are placed near the input in the vector space, resulting in a determination of the landmarks [23]. Since the landmarks of the face are constant, this can be used to filter out the eyes from the frame.

Since the image is in grayscale, the pupil is always the darkest part of the eyes [24], this can then be used to calculate the relative position of the pupil in relation to the eye. This is used to determine the gaze of the student.

4. DESIGN AND IMPLEMENTATION

The proposed system E-Blob consists of four layers and connects multiple sides, see Figure1. The system could work as a plugin for video conferencing platforms such as Zoom, Microsoft Teams, BigBlueBotton, etc. On the student side the system collects data through the webcam, and then provides real-time feedback on the teacher's screen.

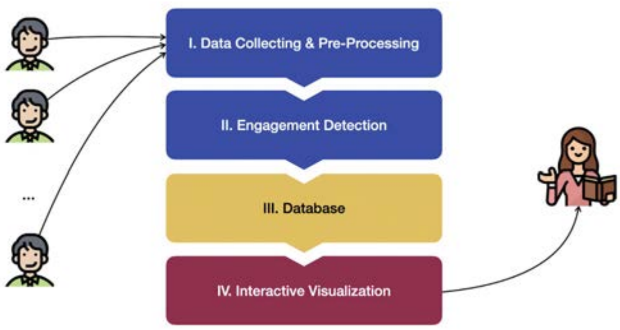


Figure 1. system architecture.

The system architecture is based on computer vision technology, and the design and implementation of different layers are described in subsections below. And the video of our working prototype can be found at:

https://www.youtube.com/watch?v=z22PGFKUOWc

4.1 Layer I&II: Intelligent behavior and embodiment

The first and second layers are the bottom layers of the system and the algorithmic layer of artificial intelligence. In these two layers, three pre-trained machine learning models are implemented, and a formula is set for calculating the concentration index, see Figure 2.

Instead of video clips, we use images captured by webcams with OpenCV as input for the system. There are two reasons, firstly, Whitehill et.at [9] suggested the estimated engagement score of combining each frame had a high accuracy as the engagement score of short video clips (10sec), and the accuracy is much higher than long video clips (60sec). Moreover, we want to provide a real-time feedback of engagement for teachers, so using frames can decrease the delay

To preprocess the input data, we aim to get the range of faces. A machine learning object detection algorithm, Haar cascade, was adopted to detect the facial features in the image. This pre-trained model is implemented with Opencv and Python by loading the pre-trained classifiers.

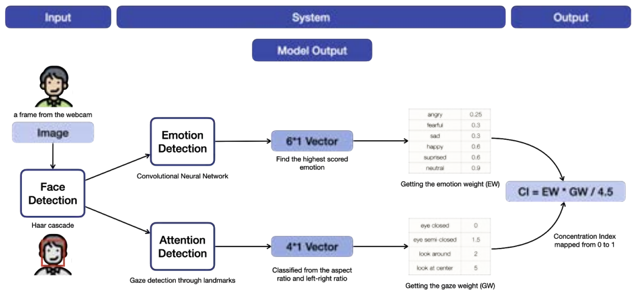


Figure 2. Layer I&II.

After we get the range of the face from the model, we used a 21 layered Convolutional Neural Network to get the softmax scores of the 6 basic emotions. We also implemented this pre-trained model with Python. Based on the study by Sharma et.al. [17], all the emotions get a weight, and the highest scored emotion will be picked for calculating the concentration index.

To further detect the eye movement, we firstly use a model in the Dlib library to generate 68 face landmarks in Python. With the serial numbers representing the left eye, we calculate the length and width of the eye, and obtain the blink ratio by calculating the aspect ratio. We also locate the darkest point in the eye range as pupils. By calculating the ratio of the distance from the pupil to the two ends of the eye socket, we define the direction of the student's sight. The different gaze test results are divided into four categories: closed eyes, semi-closed eyes, looking around, and looking at the center. They also received different weights.

Finally, the concentration index was calculated by multiplying the emotion weights and gaze weights and then is mapped 0 to 1.

4.2 Layer III: database

The database is a transfer station between the algorithm layer and the interface layer. It should be located on a server, storing data and updating every time the system analyses a new frame. Due to time limits, we set up our database in a CSV file and run the system locally. We randomly created 28 data as a start, and then the system algorithm would update the final line of our database every time it analyses a face. This process is conducted in the Python code.

4.3 Layer IV: Interactive visualization

Layer IV is the user interface and relates to our design goal: reflecting student engagement based in an interactive and explainable way. In the final concept, information is divided into three levels which are visualized as floating blobs with different styles and require different interaction-attention phases, see Figure 3.

Overall engagement. The final output, overall engagement is the average score of the concentration index of all the students that is represented by the size of the blob. When students are highly engaged, the blob is small and outside the attentional field. Then if the score decreases, the blob grows bigger to ask for attention.

Explanations. In order to increase transparency and gain user trust on the system, we want to let users know more clearly how E-Blob calculates student engagement. Thus explanations are generated from the middle steps in the system. The explanations are the average of the emotional and gaze weight of all the students, which reveals the engagement score is calculated by combining the emotion and attention detection. When the user places the cursor over the blob area, explanations are visualized in the user's periphery attention. The color of the blob varies from white to red to indicate the degree of negative to positive emotions. And the edges of the blob vary from smooth to jagged, indicating that the student's attention changes from highly concentrated to distracted.

Detailed explanations. To be more detailed, the ratio of every emotion and attention labels among all the students were given out. By displaying the distribution of students with different emotions and attention, the teacher can have a clearer understanding of the class and take different actions based on the explanation. The detailed explanation is visualized as a concentric pie chart. Corresponding to the previous blob, the outer circle explains the concentration distribution of all students, and the inner circle explains their emotional distribution. In order to obtain such complex information, users need to switch from peripheral attention to focused attention by clicking on the blob area.

The concept was developed with processing, see coding in Appendix II.

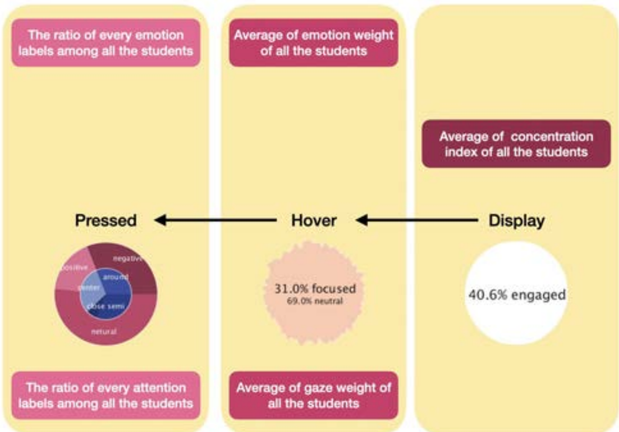


Figure 3.The interactive visualization of information

5. EVALUATION (PLANS)

The model we applied for Face detection has 100% accuracy on Indian face databases with a simple background, 93.24% accuracy with Caltech face database with complex background [25]. And validation for the Emotion detection from haar cascade has 66% accuracy, trained on the FER-2013 emotion dataset [18]. The attention detection via gaze tracking model has a 99.38% accuracy on the standard LFW face recognition benchmark [23].

We plan to set up a system analysis to **evaluate the accuracy of the engagement level of E-blob**. It will be a control test to compare the result of the engagement level assessed by E-blob with lectures. 10 students will be invited to watch an online lecture(5-10mins) in the field at the same time, the process will be recorded by each individual webcam. Then 10 teachers will be requested to observe from recorded video. They need to assess the engagement level of all students every 10 seconds. In the end, we can compare the manual assessing report of the observation with E-blob outcomes.

In this way, we can know if E-blob has close results as a human observation on sensing student engagement and how much the users can trust E-blob.

There is another **user experience testing plan** to evaluate how users understand and experience the interaction / explainable visualizations. A questionnaire to analyze user experience will be sent out to find out users' understanding of different levels of information as well as their trust level towards E-blob. The target participants here are 10 teachers, and they are invited to experience the working prototype during their online classes. Participants fill in the questionnaire after the experience. It aims to evaluate the user's attention, interaction as well as the XAI of E-blob. Based on the results, more further development from users' perspective can be generated.

6. CONCLUSIONS

The challenge of this project comes from the difficulty of detecting emotional engagement of students during online lectures. The move to digital lectures has removed the opportunity for lecturers to easily detect subtle cues to detect engagement, which therefore creates a design opportunity. This information is required for lecturers to support them in making the correct choice regarding the online lecture (e.g., take a break, shift to an interactive session, etc.).

E-Blob, the agent designed for this challenge, supports the lecturer by guiding them through the information required to make this decision. By means of E-Blob the lecturer is able to make an informed decision on what he or she believes is the correct next step, based on the information shown. E-Blob is therefore a support agent, not an agent which will decide what the correct action might be.

Looking at the design, we believe a solid argument can be made that translating the emotional engagement into an understandable and explainable design towards the lecturer was successful. What previously was an unexplained and difficult process, got elaborated and explained by introducing the blob, guiding the lecturer through the available information using the interaction-attention continuum [4].

In relation to the social context, we believe the current state of the world calls for these kinds of solutions in order to make the new way of working benefit the advantages that were found in the pre-corona era. With the already difficult barrier for social connection as a result of corona measures, projects like E-Blob can create these between people. The focus of E-Blob is not merely to socially connect people, but focuses on education. By making sure lecturers have the information regarding engagement of students, they are able to make the correct decisions and improve their lectures where needed. This could lead to more satisfaction for both the students and the lecturer in online education.

All in all, we can conclude that by introducing E-Blob, we have created an agent which through the means of XAI can support lecturers in their daily job of online education.

7. DISCUSSION

In this chapter we will elaborate on the limitations, as well as the feasibility of the system.

This project was performed to increase the awareness for lecturers of emotional engagement of students in online lectures. It is stated that in online lectures the detection of behavioral engagement (e.g., submissions of homework, presence, etc.) is easily detected, and the challenge lies at detecting emotional engagement.

Even with the confidence shown in the conclusion, proper user testing is required to fully validate the user experience, which is explained in the evaluation chapter.

What can be put up for discussion is how the emotional engagement is calculated. Sharma, P. et al. [17] argues that by multiplying the emotion weight by the gaze weight, the concentration index can be calculated. Both these variables can be put up for discussion.

For gaze, the natural behaviour of students during lectures comes into play. In the current system it is expected that the student is focussed on the screen, otherwise the student is not paying attention. Activities such as note taking often require the student to look away from the screen, but are not taken into account. The counter argument can be made that by taking notes the student does not actually pay attention to the lecture anymore but rather focuses on processing the information at his or her own leisure. More research into the design of the algorithm is needed to find the sweet spot for this potential issue.

Using emotion in this calculation has some potential pitfalls as well.People tend to have unique resting face expressions. Ekman, Paul, et al. found that across cultures, people were able to unanimously determine the emotion expressed, however the absolute intensity of the emotion was found to be subject to cultural differences [26]. This could result in a bias when determining the correct weight values for emotion. An elaborate study consisting of a varied background research team, as well as participant group, is required to make this entirely generalizable.

Even when correctly determined by the system (e.g., resting face has sad features;), the student could feel neutral. This mismatch in emotion can therefore lead to a wrong translation of the data towards the lecturer, possibly leading to a wrong action performed.

Sharma, P. et al. [17] have set up a study to provide a scientific approach to overcome the aforementioned pitfall. They created an experiment where 30 students looked at an informative video, after which a short quiz was taken. During the study, the students were recorded and the emotion of the students was analyzed afterwards. The students were then divided into groups according to the majorly present emotion expressed (more than 50% of the duration of the video). This resulted in the emotion weights used in the calculation (e.g., neutral = 0.9, sad = 0.3). Therefore, it can be concluded that less common and less expected emotions when engaged are not contributing as much to the total for the concentration index. In order to fully understand the impact, the system evaluation as proposed in chapter 5 should be performed to determine the future steps for this potential challenge.

Continuing on the topic of emotion, the accuracy of the algorithm is a potential challenge. With 66% accuracy[18], more work is needed to increase this to get closer to 90%. With 90% accuracy the system would be more convincing in determining the emotion, therefore leading to a better concentration index calculation.

Lastly, the privacy of the student is at the core of this system. Since the webcam of the student is used to detect the engagement, it can be compared to proctoring systems used to detect misconduct in exams. Milone, A et al. [27] set up a study to look at the impact of proctoring of online exams on the educational experience. They found that 88.95% of the participants were satisfied with their experience [27]. The complaints of the unsatisfied participants could be categorized in three main themes: long setup, technological issues and personal issues with proctoring.

Since the technological setup of e-blob is at the lecturer's side, these arguments can be dismissed. Unfortunately no more detailed information is given on the personal issues mentioned by Milone, A, et al. NOS op 3 [28] mentions that the Dutch student organisation ISO received at least 300 privacy complaints in April of 2020 regarding the use of proctoring software in exams. This supports the case that E-Blob, if it were widely adapted and used in online lectures, should not be mandatory for the students in order to protect their privacy. When deployed, a proper documentation of how the data is transmitted and what is stored should be accessible to create a safe environment. Besides these future steps, the willingness of students to participate in this system could be evaluated in an online survey. The outcomes of this survey could then influence the future development of e-blob.

REFERENCES

[1] Hodges, C. , Moore, S. , Lockee, B. , Trust, T. & Bond, M. A. (2020, March 27). The difference between emergency remote teaching and online learning. Educause Review .

[2] Musingafi, M. C., Mapuranga, B., Chiwanza, K., & Zebron, S. (2015). Challenges for open and distance learning (ODL) students: Experiences from students of the Zimbabwe Open University. Journal of Education and Practice, 6(18), 59-66.

[3] Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. Review of educational research, 74(1), 59-109.

[4]Bakker, S., & Niemantsverdriet, K. (2016). The interaction-attention continuum: Considering various levels of human attention in interaction design. International Journal of Design, 10(2), 1-14.

[5] Joseph, E. (2005). Engagement tracing: using response times to model student disengagement. Artificial intelligence in education: Supporting learning through intelligent and socially informed technology, 125, 88.

[6] Fairclough, S. H., & Venables, L. (2006). Prediction of subjective states from psychophysiology: A multivariate approach. Biological psychology, 71(1), 100-110.

[7] D'Mello, S. K., Craig, S. D., & Graesser, A. C. (2009). Multimethod assessment of affective experience and expression during deep learning. International Journal of Learning Technology, 4(3-4), 165-187..

[8] Kapoor, Ashish, and Rosalind W. Picard. "Multimodal affect recognition in learning environments." Proceedings of the 13th annual ACM international conference on Multimedia. 2005.

[9] Whitehill, J., Serpell, Z., Lin, Y. C., Foster, A., & Movellan, J. R. (2014). The faces of engagement: Automatic recognition of student engagementfrom facial expressions. IEEE Transactions on Affective Computing, 5(1), 86-98.

[10] Dubbaka, A., & Gopalan, A. (2020, April). Detecting Learner Engagement in MOOCs using Automatic Facial

Expression Recognition. In 2020 IEEE Global Engineering Education Conference (EDUCON) (pp. 447-456). IEEE.

[11] D'Mello, S., Lehman, B., Sullins, J., Daigle, R., Combs, R., Vogt, K., ... & Graesser, A. (2010, June). A time for emoting: When affect-sensitivity is and isn't effective at promoting deep learning. In International conference on intelligent tutoring systems (pp. 245-254). Springer, Berlin, Heidelberg.

[12] Correia, A. P., Liu, C., & Xu, F. (2020). Evaluating videoconferencing systems for the quality of the educational experience. Distance Education, 1-24.

[13] Hodges, C. , Moore, S. , Lockee, B. , Trust, T. & Bond, M. A. (2020, March 27). The difference between emergency remote teaching and online learning. Educause Review .

[14] Kumar, C. B. , Potnis, A. , & Gupta, S. (2015). Video conferencing system for distance education. In Proceedings of the 2015 IEEE UP Section Conference on Electrical Computer and Electronics (UPCON) (pp. 1–6). IEEE.

[15] Subramainan, L., Mahmoud, M. A., Ahmad, M. S., & Yusoff, M. Z. M. (2016). An Emotion-based Model for Improving Students' Engagement using Agent-based Social Simulator. International Journal on Advanced Science, Engineering and Information Technology, 6(6), 952-958.

[16] Skinner, E., Furrer, C., Marchand, G., & Kindermann, T. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic?. Journal of educational psychology, 100(4), 765.

[17] Sharma, P., Joshi, S., Gautam, S., Filipe, V., & Reis, M. J. (2019). Student Engagement Detection Using Emotion Analysis, Eye Tracking and Head Movement with Machine Learning. arXiv preprint arXiv:1909.12913.

[18] Arriaga, O., Valdenegro-Toro, M., & Plöger, P. (2017). Real-time convolutional neural networks for emotion and gender classification. arXiv preprint arXiv:1710.07557.

[19] Wang, D., Yang, Q., Abdul, A., & Lim, B. Y. (2019, May). Designing theory-driven user-centric explainable AI. In Proceedings of the 2019 CHI conference on human factors in computing systems (pp. 1-15).

[20] Peirce, C. S. (1903). Harvard lectures on pragmatism, Collected Papers v. 5.

[21] Canu, S. (2019, July 1). Eye Gaze detection 2 – Gaze controlled keyboard with Python and Opencv p.4. Pysource. https://pysource.com/2019/01/17/eye-gaze-detection-2-gaze-controlled-keyboard-with-python-and-opencv-p-4/

[22] Viola, P., & Jones, M. (2001, December). Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001 (Vol. 1, pp. I-I). IEEE.

[23] http://dlib.net/face_recognition.py.html

[24] Canu, S. (2019a, June 13). Eye motion tracking - Opencv with Python. Pysource. https://pysource.com/2019/01/04/eye-motion-tracking-opencv-with-python/

[25] Singh, P., & Tripathi, M. (2013). HAAR CASCADE CLASSIFIER PROVIDES HIGH ACCURACY EVEN THE IMAGES ARE HIGHLY AFFECTED BY THE ILLUMINATION.)

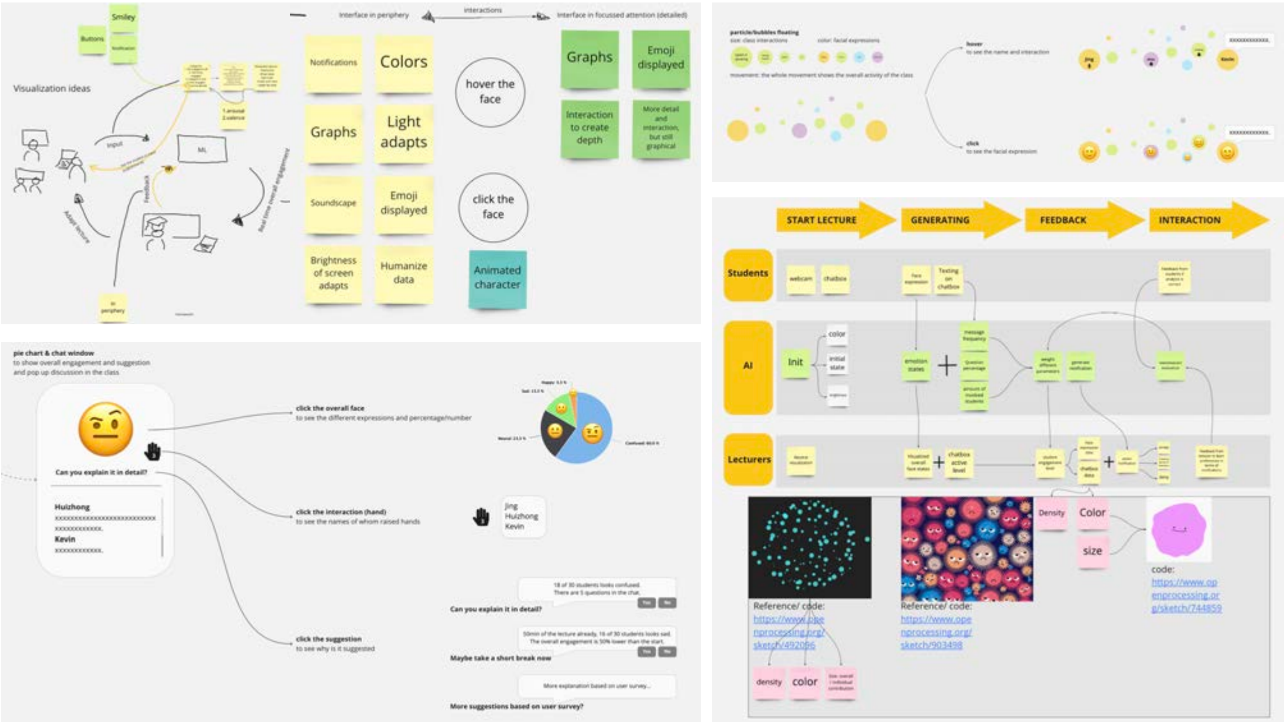
[26] Ekman, P., Friesen, W. V., O'sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., ... & Scherer, K. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. Journal of personality and social psychology, 53(4), 712.

[27] Milone, A. S., Cortese, A. M., Balestrieri, R. L., & Pittenger, A. L. (2017). The impact of proctored online exams on the educational experience. Currents in Pharmacy Teaching and Learning, 9(1), 108-114.

[28] NOS op 3. (2020, April 28). Hoe jouw studie meekijkt via de webcam [Video]. Youtube. https://www.youtube.com/watch?v=N8p8s2u6JKs

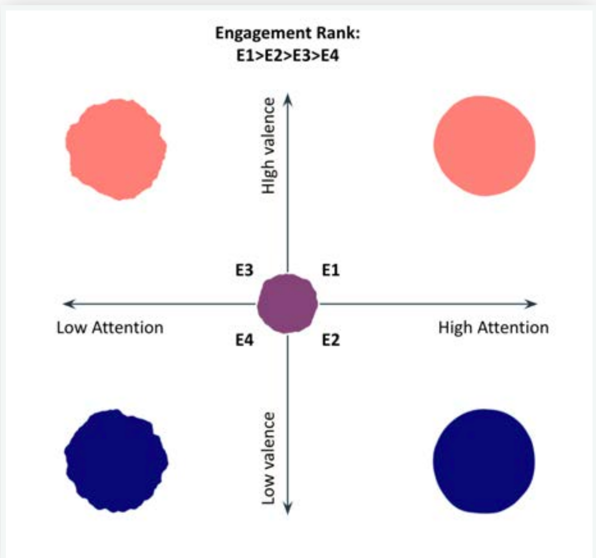
Appendix I: Design Process

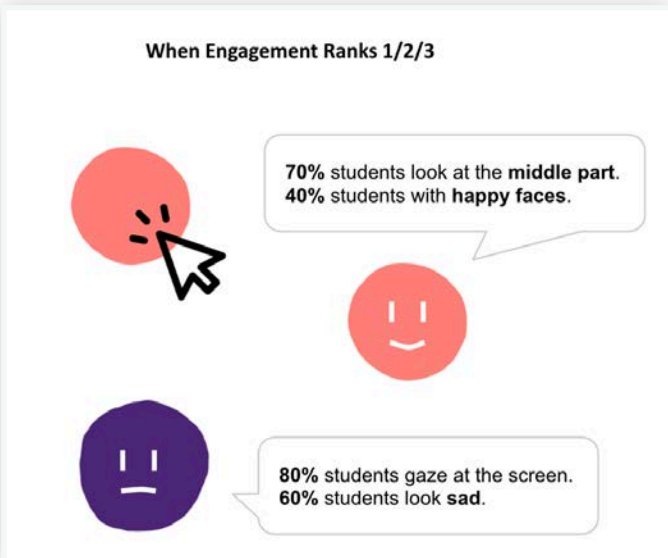
Brainstorm



Iteration 1

- Visualization: mapping attention & valence to color & shape in user's periphery. Demo: https://drive.google.com/file/d/1gnO6X1kjTfYiASiFOarMmefwJ-Za_QfO/view?usp=sharing
- Explanation in focused attention: When user click the blob, it shows the explanations in text.
- Notification calls for attention: The user can choose "ok" or "deny". And in the future more suggestion strategies can be implemented. Demo: https://drive.google.com/file/d/1_h30vvgrnK5CTUwJkdiG1HXE_Kd99W19/view?usp=sharing



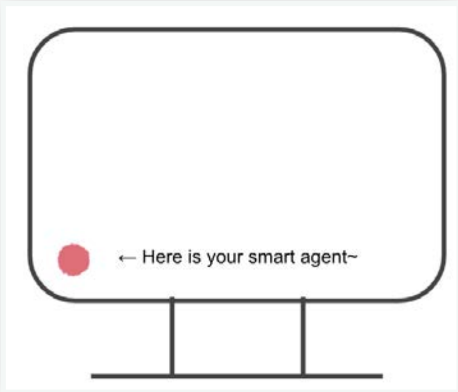


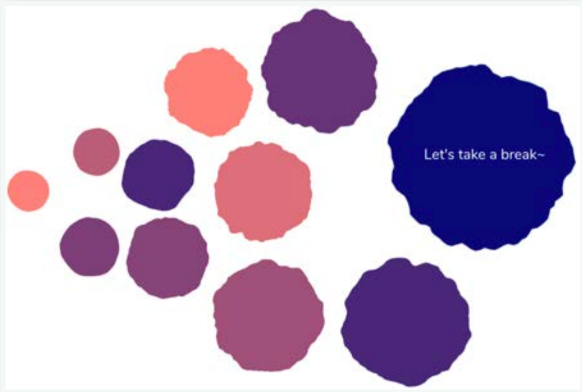
Iteration 2

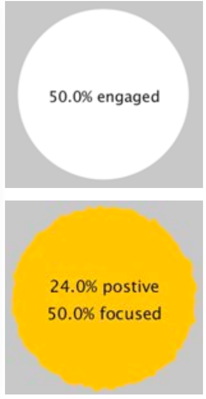
On the basis of the first iteration, use the size to represent the engagement level.

Explain the situation in number.

Two modes of interaction: display & click.







User Survey & Expert Consulting

Link: https://forms.gle/1GRuCuob3ioVeKgM9

Conclusion from the feedbacks gathered:

Improving user experience by giving clear explanations. In the iteration, the visualization of emotion and attention is neither clear nor detailed enough.

Privacy issues can influence the willingness of using. If the personal data is revealed, or individual can be tracked, it could influence the student's willingness to use the system.

Don't interrupt. From the online survey, teachers want to know student engagement but don't like to be notified when they focus on the lecture.

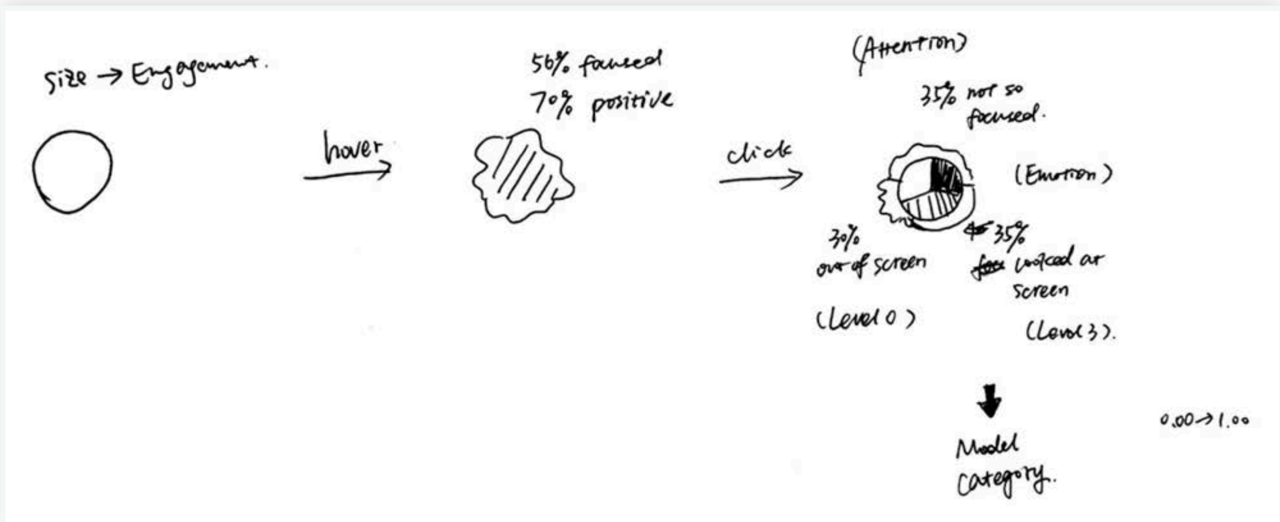
Iteration 3 (final concept)

The information is shown step by step, getting more and more detailed:

Size — engagement: In the beginning, a blob only shows size-changing with a percentage number, it will increase and decrease in size.

Color — emotion; edge — attention: When the mouse stays on it for more than a few seconds, it appears color and edge;

Then Mouse click the blob, it presents a detail explanation



Appendix II: Code

Link: https://github.com/KevinBekker/StudentEngagement_DBM140_G1

README.md

StudentEngagement XAI by Group 1 - DBM140

This project is based on the following repository: https://github.com/sherkhan15/distarctor-Detection_Eyes_Emotions_CI

It was expanded by Kevin Bekker, Huizhong Ye and Jing Li for the DBM140 "Embodying intelligent behavior in social context" course of the Industrial Design department at Eindhoven University of Technology.

This project aims to bring explainability in the process of how the engagement of students is calculated by the agent. It will visualize the overall focus of students in a blob, which will change color and size according to the variables used in the calculation. By interacting with the blob the user can get more detailed information about how the calculation is made, in a visual overview.

Installation and running the program

In order to run the XAI Processing sketch dynamically, both the sketch and the Python script need to be running. The repository contains a *database.csv* file, containing the static data. When running the Processing sketch individually it will use this data to calculate all values and display them in the XAI blob. When running the Python script as well, it will update the final row in the csv in order to facilitate dynamic data to the Processing sketch.

Processing

To run the sketch, please open up and run the file in the blobV5 folder.

Python

In order to run the Python script run the run_local_cv.py file in the root of the folder. Please check which Python version you are using, this might change the first argument in the command (e.g., python3 instead of python).

```
python run_local_cv.py
```

Install all needed libraries in order to run the program, it will prompt to access your webcam. This is needed for the facial analysis.