**Introduction**

Socially Assistive Robots (SARs) are robots designed to support people with disabilities such as Attention Deficit Hyperactivity Disorder (ADHD) and Autism Spectrum Disorder. At the USC Interaction Lab, Blossom – an expressive and accessible SAR (Fig. 2c) – is being developed as study companion for ADHD students, aiding them in accomplishing work by acting as a robotic body double. Often used by the neurodivergent community, body doubling is a method for those with executive dysfunction that involves having a second presence that helps people stay on task. By developing Blossom into an interactive study companion, it can greatly improve the performance of those with executive function regulation by providing them with a body double whenever needed.

**Research Objectives**

I developed an automated vision-based engagement detector to enable Blossom to classify whether the user is on or off-task based on their position on/offscreen. I modified a robotic study companion that prompts the user to continue working based on their facial position on/offscreen.

**System Design**

1. **Facial Tracking**

To determine whether a person is looking on and off screen, I used a combination of Dense Head Pose Estimation (DHPE) and MediaPipe to track the position of the user’s face. DHPE is used to detect the user’s face from the camera input (visualized as the white square in Fig. 2b) and determines three different measures of a person’s head rotation: pitch (up/down), yaw (left/right), and roll (head tilt). In my program, I use the pitch and yaw data of the user’s head orientation (Fig. 7) – measured in degrees from a forward facing position – to determine whether a user is off-task. If the user’s pitch or yaw exceeds a specific threshold, the program classifies the user as off-task. Although this method works well when DHPE can recognize the user’s face, when the user’s face is not fully visible (for example, when the user turns away from the webcam), the program is no longer able to detect the user to determine if they are on-task. To solve this problem, I used MediaPipe’s pose detection to determine if the user’s body is visible in the camera frame. If the user is in front of the camera but their face is obscured, it is classified as off-task. If the user is not in view of the camera at all, the program stops all engagement classification until the user returns to the frame.

2. **Blossom Re-Engagement Animation**

When designing Blossom’s work re- engagement behavior, I wanted to make sure that Blossom’s movements were expressive and clear without having to use verbal communication. With these movements, I hoped to evoke a lifelike presence that enabled Blossom to be an effective body double. Aside from that, I designed Blossom’s movements to be friendly and gentle, as to remind the user to continue working without bothering them.

3. **Timer system**

I used a sliding window algorithm to determine when Blossom should issue a re-engagement behavior to regain the user’s attention. The system uses three main components: a sliding time frame, a stored list of off-task periods, and an adjustable sensitivity variable. First, the sliding time frame is the window in which user off-task/on-task data is collected. The time frame represents the previous $n$ seconds in the session, where $n$ is a duration determined by the user. During this time, off-task data (detected by the prior mentioned face tracking) is collected and stored as an object consisting of a start and end time. I decided to only count a duration period as being longer than one second, to avoid causing Blossom to prompt the user too often because of brief face detection failure. The time window continually updates throughout the session, making sure that the data being gathered is relevant to the user’s recent activity status, and deleting any off-task data that is no longer in the given time frame. This system is fully visualized in Fig. 3. Throughout the study session, the ratio of off-task time in the window is checked against the sensitivity decided by the user (Fig. 4). If the ratio exceeds the sensitivity, Blossom initiates a re-engagement behavior to prompt the user to get back on task, and the sliding window – including the off-task data – is reset.

**Future Work**

For future steps in this project I think it would be beneficial to include a more precise measure for when a user is off-task. Incorporating eye tracking to see when the user is looking off screen or staring at one place for too long could lead to more accurate results when judging whether a person is on-task. Body language can also be a large indicator of a user’s attention. With enough data, I think that an AI model could be trained to more accurately recognize whether a user is engaged, and thus improve the reliability of my program.

Aside from improved engagement detection, adding small noises to Blossom to get the user’s attention could make the meaning behind Blossom’s re-engagement behaviors more clear. When conducting the aforementioned in-dorm study, it was found most students got used to Blossom’s motor noises over time. If Blossom’s prompting animation were to include small noises, it would be better able to get the user’s attention and would give Blossom a more lively feeling.

**Acknowledgements**

I would like to thank Prof. Majka Mataric, my mentor Amy O’Connell, other PhD students who helped me, and my labmates that were with me this summer. USC SHINE has been a wonderful experience that stayed very fun whilst providing me a great opportunity to develop new skills in research.

**Citations**