A person and a baby

Description automatically generated

Figure 1: Each frame of video is labelled with body position keypoints, parental speech, participant identity and facial expression.

A screenshot of a computer screen

Description automatically generated

Figure 2: We use the FiftyOne platform to allow visualisation filtering and analysis of whole dataset and model predictions.