**Scientific background**

Multimodal models integrate data from multiple sources, such as audio, video, and text, offering a richer and more comprehensive dataset for analysis than single-modality systems. This integration enables the leveraging of diverse data characteristics, enhancing the model’s ability to capture complex patterns and nuances that might be missed when using only one type of data. Fusion techniques, whether early (combining data at the feature level) or late (integrating decisions from separate models), enhance predictive performance by providing a holistic view. Such approaches capitalize on the unique strengths of each data type, facilitating more accurate and robust predictions, particularly in tasks requiring rich insights like lie detection.

**Preprocessing**

Every modality is preprocessed using its tailored pipelines, each modality outputs its prediction as a probability or a percentage to be fused using late fusion approaches.

**Model Training**

There are no specific models in this stage, specifically in late fusion, This approach basically performs an output level fusion on the resulting predictions of the other modalities, and these probabilities are combined into one probability using weighted averaging or simply a majority vote is done to predict the final label based on the most recurring prediction label among the single modalities.

**Important remarks and discoveries**

Only late fusion is done so far, but later we will experiment with early fusion or even hybrid fusion.