Process Pie Chart

INPUT DATA

MODALITY PROCESSING

CROSS MODEL INTERACTION

SORT TASK SPECIFICALY

OUTPUT PREDICTION Input Data is fed to the Multimodal Transformer

Each Modality is processed by modality specific encoders

The Final Output is Predicted

> Classification, Generation and Translation is done

Involves interaction between different Modalities