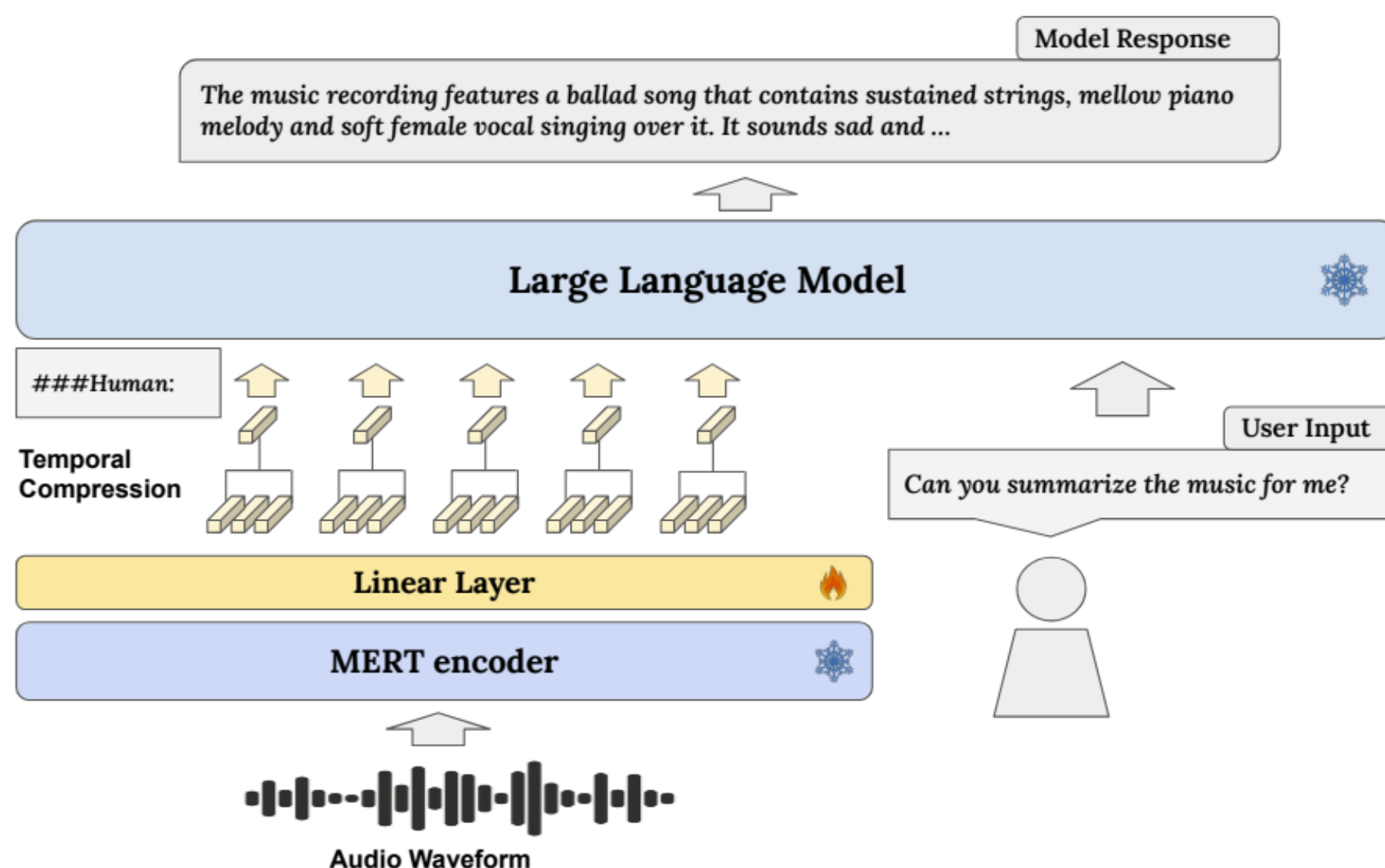


Large Language Model for Music Understanding and Music Generation

Yinghao Ma

2022



Finding

Large Language Models have shown immense potential in multimodal applications. We present MusiLingo, a novel system for music caption generation and music-related query responses. MusiLingo employs a single projection layer to align music representations from the pre-trained music encoder with a frozen LLM, bridging gap between music and textual contexts.

Question

Developing multimodal music-text supervised finetuning (SFT) / instruction finetuning dataset for a better language model finetuning with better generalisation to unseen tasks