Whisper Projection Multi-Task Multi-Label (MTML) Model

Audio Input
Raw audio of player's utterance

→ Feature Extractor → Whisper-small (244M parameters) → Projection Layer (Last Token's Logits) → Dropout Layer → Multi-Label Classification Head

Previous Utterances
Previous Utterances:
Utterance -> {prev utterance 1}
...
Utterance -> {prev utterance 5}

→ Tokenizer

Last Utterance
Last Utterance:
Utterance -> {utterance}

Multi-Task Multi-Label (MTML) Approach:
One unified model that determines all strategies simultaneously without strategy prompting