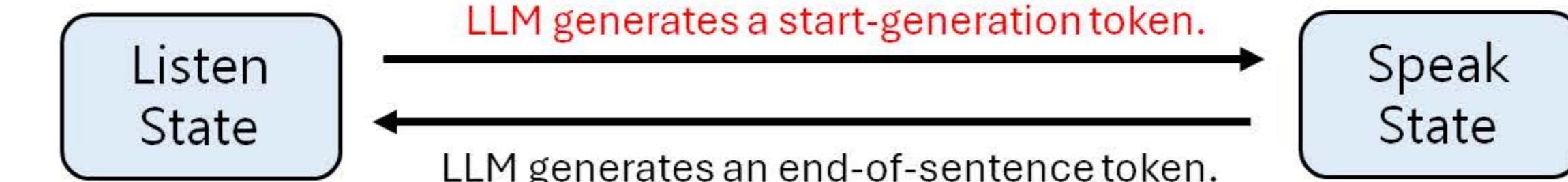


- Accept token inputs typed by the user.
 - Do not produce any tokens.
- The LLM generates tokens.
 - User input is not allowed.



- Accept audio token as inputs.
 - Do not produce any tokens.
- The LLM generates tokens.
 - **The input is still coming.**

Turn-based Text-based Conversation

Full-Duplex Speech Conversation