Google

Gemini 3.5 Audio (Live Translate) Model evaluation

Approach

Gemini 3.5 Live Translate was evaluated across three main quality dimensions: translation quality, latency, and speech naturalness.

Capabilities / Benchmarks

- **Translation Quality**: We evaluate translation quality using AutoMQM - an error-based automatic metric which identifies and categorizes translation errors to produce a fine-grained quality score. These metrics are applied across a variety of language pairs and content types to assess the accuracy and adequacy of the translated output.
- **Latency**: For a real-time translation system, latency (how far behind the translation lags during a live session) is a critical quality dimension. We measure latency at multiple granularities:
    - *Initial latency*: the time between the start of speech in the input stream and the start of speech in the translated output stream. This is measured using standard voice activity detection.
    - *Word-level latency*: to capture the most complete picture of translation delay, we align words in the input to their corresponding words in the output and measure the average time between the end of a source word and the start of its corresponding translated word.
- **Speech Naturalness**: We evaluate the naturalness and quality of the synthesized output audio using established speech synthesis quality metrics. These capture issues such as choppy or discontinuous audio, voice drift (where the output voice characteristics shift over the course of a session), and unintended artifacts. Maintaining natural-sounding, consistent output audio is essential for a usable real-time translation experience.

Methodology

Gemini 3.5 Live Translate was evaluated using internal implementations of these benchmarks on outputs generated using the Gemini Live API.