



Gemini 3.5 Audio (Transcribe, Transcribe Live) Model evaluation

Approach, methodology & results

Approach

Gemini 3.5 Transcribe and Gemini 3.5 Transcribe Live were evaluated using the methodology below:

Capabilities / Benchmarks

- **FLEURS:** This multilingual speech dataset is used as a critical benchmark to evaluate the speech recognition (ASR) capabilities of flagship systems across 26 top locales.
- **Artificial Analysis:** An independent evaluation platform tracking speech-to-text models across transcription accuracy using the AA-WER Word Error Rate index across diverse datasets. It evaluates models across both non-streaming (batch) and streaming modalities.

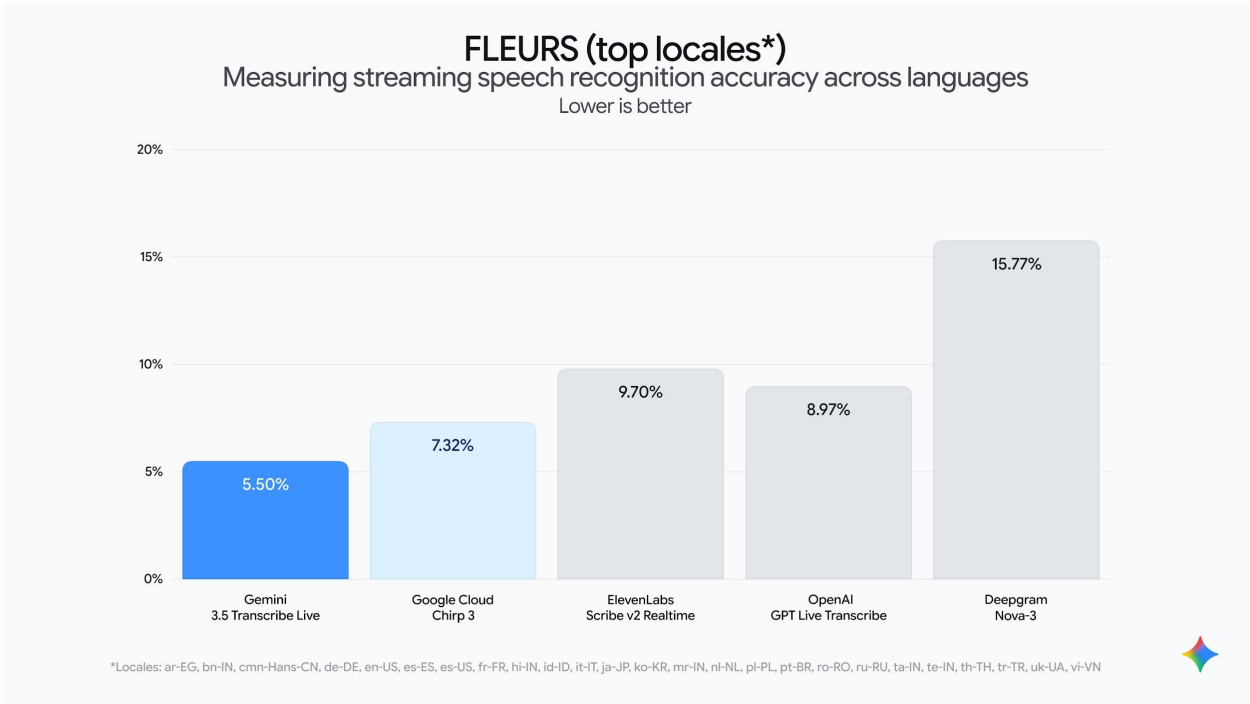
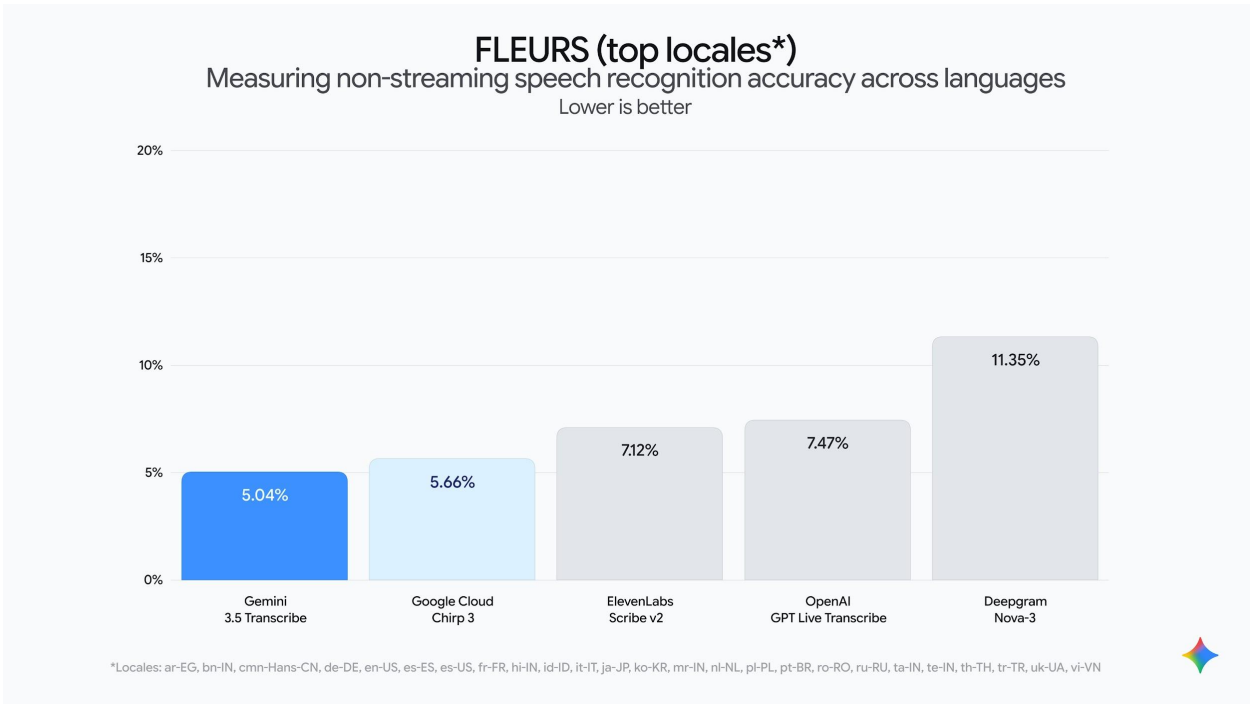
Methodology

We benchmarked our internal systems, including Chirp-3 and the Gemini 3.5 Transcribe and Gemini 3.5 Transcribe Live models against competitors' default APIs. Evaluations were conducted locally across both streaming and non-streaming modes.

Results

Results as of August 2026 are listed below:

FLEURS



Artificial Analysis

As measured by Artificial Analysis, Gemini 3.5 Audio achieves an average Word Error Rate (WER) of 4.0% for streaming (Gemini 3.5 Transcribe Live) and 2.6% (Gemini 3.5 Transcribe) for non-streaming use-cases. It shows strong performance across noisy, real-world environments, accurately capturing alphanumeric entities like postal codes and order IDs.