

Intelligent transcription with Gemini 3.5 Transcribe

26.8.2026 - Diego Melendo Casado | Google Italia

Our latest speech-to-text model designed for precise and intelligent real-time transcription.

Today, we're introducing Gemini 3.5 Transcribe, our most precise speech-to-text model yet, designed for intelligent voice interactions. Unlike conventional speech recognition models that struggle with background noise, complex jargon, and disfluency cleanup, Gemini 3.5 Transcribe converts raw audio directly into accurate, polished, formatted text.

Across our products like the Gemini app and on Android, we've seen consumers already benefiting from this transcription model with new voice capabilities like Rambler on Android and in the Gemini app on macOS. Now, developers can build similar capabilities with Gemini 3.5 Transcribe in the Gemini API in Google AI Studio and Gemini Enterprise Agent Platform.

We've built 3.5 Transcribe to plug seamlessly into your developer workflows, whether you're building voice agents, real-time captioning tools, or post-call analytics pipelines. The model is available across two separate APIs:

- Real-time streaming: Delivers continuous, bidirectional streaming with sub-second latency for interactive voice apps via the Live API using `gemini-3.5-transcribe-live`.
- Pre-recorded audio processing: Transcribes recorded audio, meetings, call logs, and more with speaker attribution and word-level timestamps via the Interactions API using `gemini-3.5-transcribe`.

Get more precise and intelligent transcription

Gemini 3.5 Transcribe is designed to capture your natural speaking style to better understand your intent and recognize custom vocabulary, so you can execute tasks with your voice.

- Smart transcription: Seamlessly handles self-corrections (like "*let's meet Tuesday—no, Wednesday*"), removes filler words ("ums" and "ahs"), auto-formats your text.
- Function calling: The model can delegate complex tasks (such as image generation and file analysis) to other Gemini models via function calls. Currently available in the Gemini macOS app.
- More precise transcription: As measured by Artificial Analysis, achieves an average Word Error Rate (WER) of 4.0% for streaming and 2.6% for non-streaming use-cases. It shows strong performance across noisy, real-world environments, accurately capturing alphanumeric entities like postal codes and order IDs.
- Custom vocabulary: Recognizes specialized jargon and unique spellings by seamlessly adapting transcriptions to your provided custom vocabulary.
- Global language support: Automatically detects and transcribes over 85 languages, seamlessly handling regional accents and diverse dialects.
- Multi-speaker identification: Accurately attributes speech in pre-recorded audio with timestamps for up to three speakers (support for 3+ speakers is experimental).

Gemini 3.5 Transcribe handles live language switches and seamless streaming transcription

Watch Gemini 3.5 Transcribe clean up speech disfluencies with smart transcription capabilities.

3.5 Transcribe delivers transcription with multi-speaker attribution and word-level timestamps.

Gemini 3.5 Transcribe's performance represents a major advancement from our previous transcription model, Chirp 3, offering new capabilities, improved word error rates, and significantly better latency. As measured by Artificial Analysis, time to final transcription, for example, improves by 70%. On the FLEURS benchmark across a set of top languages and locales, the model delivers precise multilingual performance, improving over Chirp 3, and achieving a 5.50% WER in streaming mode and 5.04% WER in non-streaming use-cases.

Experience smart transcription and advanced dictation

In addition to the Gemini API in the Google AI Studio and Gemini Enterprise Agent Platform, 3.5 Transcribe goes further than standard speech-to-text to make working across Google feel more natural and intuitive. By bringing context-aware understanding directly into everyday surfaces like Gboard, Antigravity, the Gemini app, and Chrome, it captures nuances, intent, and inline edits with ease.

- On Gboard on Android, through the new Rambler feature, 3.5 Transcribe transforms spoken thoughts into well-formatted text, filtering out filler words. You can also use your voice to make edits, correct misspellings, and change the writing style.
- On Google Antigravity, 3.5 Transcribe pairs screen context and chat history, with your permission, to ensure pinpoint transcription accuracy across file names, agent thoughts, and active documents.
- In Google AI Studio, you can access 3.5 Transcribe in Build mode to vibe code apps with your voice on the fly.
- In the Gemini app on macOS, 3.5 Transcribe not only transcribes your free natural speech into clean formatted text, but also enables voice commands that can pair seamlessly with screen context to power complex workflows. By calling on other Gemini models in the background to handle the heavy lifting, the model makes it effortless to summarize local files, repurpose text across apps, or generate images right at your cursor—using just your voice.
- Coming soon to Chrome, you'll be able to talk to type in any web field — making it effortless to dictate replies, draft posts, or prompt Gemini in Chrome more naturally and easily with your voice.

Gemini 3.5 Transcribe lets you analyze files, generate images, and search in the Gemini app on macOS using just your voice.

See how Gemini 3.5 Transcribe uses Rambler on Android to automatically remove filler words and clean up speech.

Gemini 3.5 Transcribe leverages screen context on Google Antigravity to ensure accurate transcription accuracy.

Read the early reviews

By leveraging the Gemini Live API, developer platforms such as Agora, Fishjam, LangChain, LiveKit, Pipecat, Vercel, and Vision Agents enable developers to build and deploy high-performance voice-driven interfaces with ease. These platforms manage complex real-time media streaming

infrastructure behind the scenes, allowing developers to focus entirely on crafting the user experience.

Companies like Vivo, Intellitek Health, and Lingopal have also shared positive feedback on 3.5 Transcribe, highlighting its impressive latency, accuracy, and expansive language support.

Start using 3.5 Transcribe today

- For developers: In public preview in the Gemini API via Google AI Studio and Google Antigravity.
- For enterprises: In public preview via Gemini Enterprise Agent Platform and coming soon to Gemini Enterprise for Customer Experience.
- For everyone: In Gemini app on macOS in English, Rambler on Android in select countries and languages, and coming soon to Chrome.

<https://blog.google/innovation-and-ai/models-and-research/gemini-models/gemini-3-5-transcribe>