MAI-Transcribe-1.5 Model Card

Date: June 2, 2026

Model summary

The second iteration of our best-in-class speech-to-text model family. MAI-Transcribe-1.5 is now even more robust for real-world audio. It provides consistently strong transcription across accents, speaking styles, and noisy environments, giving developers a strong foundation for building high-quality voice understanding into their applications. MAI-Transcribe-1.5 now supports entity biasing — domain-aware transcription that better recognizes industry and scientific terms, proper names, and other domain-specific terminology.

Overview

About this model

MAI-Transcribe-1.5 is a speech-to-text model built in-house by the Microsoft AI team, designed to deliver reliable transcription across 43 languages. It powers a wide range of use cases, including video captions, meeting transcription, accessibility tools, call analysis, content creation workflows, and enabling voice agents. The model is optimized to be robust across diverse accents, dialects, and real-world acoustic conditions, giving developers a transcription system they can rely on.

Key capabilities

- Best-in-class transcription accuracy across 43 languages.
- 25 languages already covered by MAI-Transcribe-1: English, French, German, Italian, Spanish, Hindi, Portuguese, Czech, Danish, Finnish, Hungarian, Dutch, Norwegian Bokmål, Polish, Romanian, Swedish, Japanese, Korean, Chinese, Arabic, Indonesian, Russian, Thai, Turkish, and Vietnamese.
- 18 additional languages: Bulgarian, Catalan, Greek, Estonian, Lithuanian, Slovak, Slovenian, Ukrainian, Assamese, Bengali, Gujarati, Kannada, Malayalam, Marathi, Odia, Punjabi (Gurmukhi script), Tamil, and Telugu.
- Robust in noisy, real-world conditions.
- Faster inference: substantially lower latency than MAI-Transcribe-1, especially on long-form audio, with up to 5x faster processing.
- Automatic language identification.
- Keyword/entity biasing (up to 200 keywords) to improve transcription in domain-specific contexts.

Performance and quality

Consistently high quality across 43 languages

MAI-Transcribe-1.5 transcribes with high accuracy across all 43 supported languages, making it particularly interesting for global products. This is reflected by the FLEURS benchmark where MAI-Transcribe-1.5 achieves the lowest average word error rate (WER) across languages when compared to comparable transription systems.

FLEURS Benchmark results: | Model | Avg. WER (across 43 languages) | | --- | --- | | **MAI-Transcribe-1.5** | **4.86%** | | Elevenlabs scribe v2| 5.53% | | OAI-transcribe | 5.73% | | Google Gemini-flash-lite | 5.63% |

Improved noise robustness

Even on common languages like English, there were noteable improvements since the last version of MAI-Transcribe, driven by targeted quality improvements and better noise robustness. For example, this is demonstrated, by the Enlish-only speech-to-text benchmark of Artificial Analysis (AA). Here, MAI-Transcribe-1.5 managed to reduce the error rate further to an overall AA-WER of 2.38%, placing it among the top-3 models on the Artificial Analysis speech-to-text leaderboard.

Boosting transcription quality via keyword biasing

Transcribing and formatting acronyms or names correctly is very challenging. MAI-Transribe-1.5 now features keyword biasing which allows users to improve transcription accuracy by informing the model about domain-specific terminology or names. By adding important terms and names to the model's keyword context, we ensure that the model detects them in the audio and transcribes them correctly. In our evaluations, we were able to achieve up to 30% reduction in WER when using keyword biasing.

Faster long-form transcription

Thanks to improvments on how the model handles long audios, it transcribes long-form much faster, for example, 1 hour of audio in only 15 sec.

Use cases

Key use cases

Use case	Scenario	Solution
Live captions	A virtual event platform provides real-time captions for webinars.	Chunk audio and transcribe spoken content into captions displayed live during the event.
Call center transcription	A call center wants accurate, fast transcriptions of customer calls to empower their customer service agents.	Transcribe calls in real time, enabling agents to better understand and respond to customer queries.
Video subtitling	A video-hosting platform needs to generate subtitles for uploaded videos.	Transcribe the full video audio to produce a complete subtitle track.
Accessibility	An organization needs to make audio content accessible to deaf or hard-of-hearing users.	Transcribe audio from meetings, announcements, or media to provide text alternatives that support compliance and inclusive access.
E-learning	An e-learning platform provides transcriptions for video lectures.	Process prerecorded lecture videos, generating text transcripts for students.
Media archiving	A media company needs subtitles for a large archive of videos.	Transcribe video files in bulk, generating accurate subtitles for each video.
Market research	A research firm analyzes customer feedback from audio recordings.	Convert audio feedback into text, enabling easier analysis and insights extraction.

Out of scope use cases

Diarization is not supported yet; this capability is planned for an upcoming release.

Pricing

$0.36 per hour of audio

Input formats

LLM Speech: WAV, MP3, FLAC

Max file size: 300 MB / 2 hours.

Supported languages

Arabic, Assamese, Bengali, Bulgarian, Catalan, Chinese, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Gujarati, Hindi, Hungarian, Indonesian, Italian, Japanese, Kannada, Korean, Lithuanian, Malayalam, Marathi, Norwegian Bokmål, Odia, Polish, Portuguese, Punjabi (Gurmukhi script), Romanian, Russian, Slovak, Slovenian, Spanish, Swedish, Tamil, Telugu, Thai, Turkish, Ukrainian, and Vietnamese.

Supported Azure regions

MAI-Transcribe-1.5 can be accessed globabally. The model is currently served from three regions

- CUS=Central US
- SEC=Sweden Central
- SEA=Southeast Asia

to which the requests are routed.

Sample JSON response

```
{
    "durationMilliseconds": 4000,
    "combinedPhrases": [
        {
            "text": "Your transcription results will appear here"
        }
    ],
    "phrases": []
}
```

Distribution

You can access MAI-Transcribe-1.5 via Azure Speech SDK. Alternatively, you can also use the REST API directly to access the Speech service. For an example how to use the REST APIs, see LLM Speech.

More information

Learn more in the full Azure Speech Service documentation.