**Whisper モデルとは?**

[アーティクル] • 2024/10/16

Whisper モデルは、⾳声ファイルの⽂字起こしに使⽤できる OpenAI の⾳声テキスト変換モデルです。 このモデルは、英語の⾳声とテキストの⼤規模なデータセットでトレーニングされます。 このモデルは、英語の⾳声を含む⾳声ファイルの⽂字起こしに最適化されています。 このモデルを使⽤して、他の⾔語の⾳声を含む⾳声ファイルを⽂字起こしにも使⽤できます。 このモデルの出⼒は英語のテキストです。

Whisper モデルは、Azure OpenAI Service または Azure AI Speech 経由で使⽤できま

す。 これらのオファリングの機能は異なります。 Azure AI ⾳声 (バッチ⽂字起こし) において、Whisper は⾳声テキスト変換に使⽤できる複数のモデルの 1 つにすぎません。

次のような質問があるかと思います。

Whisper モデルは⾃分のシナリオに適していますか、それとも Azure AI Speech モデルの⽅が良いでしょうか? 2 種類のモデル間の API の⽐較は何ですか?

Whisper モデルを使⽤する場合は、Azure OpenAI Service または Azure AI Speech 経由で使⽤する必要がありますか? どちらか⼀⽅を使⽤するようにガイドされるシナリオは何ですか?

# Whisper モデルまたは Azure AI Speech モデル

シナリオに応じて、Whisper モデルまたは Azure AI Speech モデルが適切です。 Azure AI Speech を使⽤する場合は、Whisper モデルを含む複数のモデルから選択できます。次の表は、使⽤するシナリオに関する推奨事項とオプションを⽐較しています。

| **シナリオ** | **Whisper モデル** | **Azure AI Speech モデル** |
| --- | --- | --- |
| 音声とビデオのリアルタイムの文字起こし、キャプション、字幕。 | 使用不可 | 推奨 |
| 事前に録音された音声とビデオの文字起こし、キャプション、字幕。 | 個々の音声ファイルの高速処理には、[Azure OpenAI](https://learn.microsoft.com/ja-jp/azure/ai-services/openai/whisper-quickstart) を介した Whisper モデルを推奨します。 [Azure AI 音声 (バッチ文字起こし)](https://learn.microsoft.com/ja-jp/azure/ai-services/speech-service/batch-transcription-create#use-a-whisper-model) を介した Whisper モデルは、大きなファイルのバッチ処理で推奨されます。 詳細については、「[Azure AI 音声バッチ文字起こしを介した Whisper モデルか Azure OpenAI Service を介した Whisper モデルか?](https://learn.microsoft.com/ja-jp/azure/ai-services/speech-service/whisper-overview#whisper-model-via-azure-ai-speech-or-via-azure-openai-service)」を参照してください | 大きなファイル、ダイアライゼーション、およびワード レベルのタイムスタンプのバッチ処理に推奨されます。 |
| 通話記録のトランスクリプトと通話の概要、センチメント、主要なトピック、カスタム分析情報などの分析。 | 利用可能 | 推奨 |
| リアルタイムの文字起こしと分析により、お客様の質問に対応するコール センター エージェントをサポート。 | 使用不可 | 推奨 |
| 会議記録のトランスクリプトと、会議の概要、会議のチャプター、アクション アイテムの抽出などの分析。 | 利用可能 | 推奨 |
| 音声ディクテーションによるリアルタイムのテキスト入力とドキュメント生成。 | 使用不可 | 推奨 |
| コンタクト センターの音声エージェント: コール センター向けの通話ルーティングと対話型音声応答。 | 利用可能 | 推奨 |
| 音声アシスタント: セットトップ ボックス、モバイル アプリ、車内、その他のシナリオ向けのアプリケーション固有の音声アシスタント。 | 利用可能 | 推奨 |
| 発音評価: 話者の声の発音を評価する。 | 使用不可 | 推奨 |
| ライブ音声をある言語から別の言語に翻訳する。 | 使用不可 | [音声翻訳 API](https://learn.microsoft.com/ja-jp/azure/ai-services/speech-service/speech-translation) を介して推奨される |
| 他の言語から事前に録音された音声を英語に翻訳する。 | 推奨 | [音声翻訳 API](https://learn.microsoft.com/ja-jp/azure/ai-services/speech-service/speech-translation) を介して利用可能 |
| 事前に録音された音声を英語以外の言語に翻訳する。 | 使用不可 | [音声翻訳 API](https://learn.microsoft.com/ja-jp/azure/ai-services/speech-service/speech-translation) を介して推奨される |

# Whisper モデルは、Azure AI Speech または

**Azure OpenAI Service を介していますか?**

Whisper モデルを使⽤する場合は、2 つのオプションがあります。 Azure OpenAI と

Azure AI ⾳声 (バッチ⽂字起こし) のどちらを介した Whisper モデルを使⽤するかを選択することができます。 どちらの場合も、⽂字起こしされたテキストの読みやすさは同じです。 混合⾔語の⾳声を⼊⼒することができ、出⼒は英語です。

Azure OpenAI Service を介した Whisper モデルは、次の場合に最適な場合があります。

⾳声ファイルを⼀度に⼀つずつすばやく⽂字起こしする他の⾔語の⾳声を英語に翻訳する

出⼒をガイドするためのプロンプトをモデルに提供する

サポートされているファイル形式: mp3、mp4、mpweg、mpga、m4a、wav、 webm

Azure AI ⾳声バッチ⽂字起こしを介した Whisper モデルが適しているのは以下の場合です。

25 MB (最⼤ 1 GB) を超えるファイルの⽂字起こし。 Azure OpenAI の Whisper モデルのファイル サイズの制限は 25 MB です。

⾳声ファイルの⼤規模なバッチの⽂字起こし。 会話に参加している異なる話者を区別するためのダイアライゼーション。 ⾳声サービスは、⽂字起こしされた⾳声の特定の部分を話していた話者に関する情報を提供します。 Azure OpenAI を介した Whisper モデルでは、ダイアライゼーションはサポートされていません。ワードレベルのタイムスタンプ

サポートされているファイル形式: mp3、wav、ogg。

地域的なサポートも考慮する必要があります。

Azure OpenAI Service を介した Whisper モデルは、⽶国東部 2、インド南部、中北部、ノルウェー東部、スウェーデン中部、スイス北部、⻄ヨーロッパのリージョンで利⽤できます。

Azure AI Speech 経由の Whisper モデルは、オーストラリア東部、⽶国東部、⽶国中北部、⽶国中南部、東南アジア、英国南部、⻄ヨーロッパのリージョンで利

⽤できます。

**関連するコンテンツ**

Azure AI Speech のバッチ⽂字起こし API を介して Whisper モデルを使⽤する

Azure OpenAI を介した Whisper の⾳声テキスト変換のクイックスタートを試す

Azure AI Speech 介したリアルタイム⾳声テキスト変換のクイックスタートを試す

**フィードバック**

**このページはお役に⽴ちましたか?**  **Yes** **いいえ**

製品フィードバックの提供 | Microsoft Q&A でヘルプを表⽰する