



国産AI vs 海外AI 技術比較レポート2026 - 日本企業のための音声AI選定フレームワーク | Leadsia Inc.

2026-02-23

Leadsia Inc.

カテゴリー：導入検討・比較

「AIを導入するなら、国産と海外どちらがいいのか？」

日本企業の生成AI利用経験率は、2023年のわずか3.4%から2025年には38.9%に急上昇しました。総務省の情報通信白書（令和7年）によると、生成AIの活用方針を策定した企業は49.7%に達しています。

しかし、米国や中国では企業のAI活用率が90%を超えている中、日本はまだ試行錯誤の段階です。PwCの2025年調査（5カ国比較）でも、日本の活用率は56%で海外勢に後れを取っています。BCGの調査では、日常的にAIを活用している日本人は51%と、世界平均の72%を大きく下回っています。

この差を埋めようとする中で「国産AIか海外AIか」という議論が浮上します。しかし、この問いの立て方自体に落とし穴があります。

正しい問い：「どの技術要素に、どの選択肢が最適か」

音声AIインテリジェンスは、3つの技術要素で構成されています。

音声認識（STT） - 相手の声をテキストに変換する「聞く力」

大規模言語モデル（LLM） - テキストを理解し応答を生成する「考える力」

音声合成（TTS） - 応答を自然な音声に変換する「話す力」

（詳しくは「AI営業電話の品質を決める3つの要素」をご参照ください）

この3要素で、国産と海外の強みはまったく異なります。「全部国産」でも「全部海外」でもない、要素ごとの最適な組み合わせが答えです。

要素①：音声認識（STT） - 日本語は国産が明確に優位

音声認識は、音声AI全体の精度を左右する土台です。ここでは国産エンジンが海外勢をリードしています。

精度比較：数字で見る差

医療分野での比較検証データが、その差を端的に示しています。

AmiVoice（アドバンスト・メディア社）の医療専用エンジンは、電子カルテ入力の認識精度で97.33%を記録しました。同じテストで、Google STTは87.19%、汎用モデルは84.63%にとどまっています。

約10ポイントの差は、100語の発話で10語以上の聞き間違いが生じることを意味します。Whisperでは「充血」を「無視しや」と誤認識するような事例も報告されており、日本語の同音異義語や専門用語は海外モデルにとって依然として難関です。

会議やビジネス会話でも、AmiVoiceは95%以上の安定した精度を維持しています。Whisperは日常会話レベルでは高い精度を出しますが、業界用語や固有名詞でハルシネーション（もっともらしい誤変換）が発生しやすい傾向があります。

レイテンシー（応答速度）の差

リアルタイム対話で重要なのが、音声認識の処理速度です。

Google STTのレイテンシーは約300ミリ秒。Whisperは1～3秒。営業電話の「会話のキャッチボール」では、この差は体感的に大きい。相手が話し終わってから1～3秒の沈黙が生じると、「聞こえてますか？」と不安にさせてしまいます。

AmiVoiceはストリーミング認識に対応しており、リアルタイム対話に適した応答速度を実現しています。

なぜ国産STTが強いのか

AmiVoiceの強みは、日本語に特化した数十年の蓄積にあります。医療、建設、金融、ITなど業界別の専門用語辞書が充実しており、フィラー（「えーと」「あー」等）の自動削除機能も備えています。コールセンターのような騒がしい環境でのノイズ除去技術も、日本の通話環境に最適化されています。

Whisperの学習データは68万時間を超える多言語データですが、日本語が占める割合は限定的です。英語中心で訓練されたモデルが日本語の敬語構造やビジネス慣用表現を完全に理解するには、構造的な限界があります。

音声認識は国産が合理的。これがデータに基づく結論です。

要素②：大規模言語モデル（LLM） - 海外勢が明確にリード

一方、LLM（会話の「頭脳」）では、海外モデルが圧倒的な優位を保っています。

海外LLMの現在地

AnthropicのClaude、OpenAIのGPTシリーズ、GoogleのGemini - これらの海外LLMは、数兆語規模の学習データと数百億ドルの研究投資を背景に開発されています。日本語を含むすべての言語で、文脈理解、推論能力、長文処理において国産LLMを上回っています。

国産LLMの現状

国産LLMの開発も着実に進んでいます。2026年、デジタル庁がガバメントAI「源内（げんない）」の試用モデルを公募し、NTTデータの「tsuzumi 2」（300Bパラメータ、国産最大級）、Preferred Networksの「PLaMo 2.0 Prime」、KDDI・ELYZAの共同開発モデル、NECの「cotomi v3」、ソフトバンクの「Sarashina2 mini」などが選定されました。

PFNは金融ドメイン向けのツール呼び出し機能に強みを持ち、ABEJAは自律型AIエージェントの研究に注力しています。NTTのtsuzumi 2は300Bパラメータで国産首位、富士通Takane、NEC cotomiも含め、金融・公共・医療への浸透が進んでいます。

全体として、国産LLMは「軽量高性能」と「大規模」の二極化が進行中です。データを海外に出せない用途（政府、防衛、医療、金融の一部）では重要な選択肢ですが、汎用性能で海外モデルとの差は依然として大きいのが現実です。

営業AIで重要なのは「アライメント」

営業電話やカスタマー対応に使うLLMでは、単なる言語能力以上に重要な特性があります。

誠実さ - 相手に嘘をつかないか。実現不可能な約束をしないか。

安全性 - 不適切な発言をしないか。感情を不用意に増幅させないか。

一貫性 - 会話中に矛盾した発言をしないか。

Leadsiaが採用するAnthropicのClaudeは、「Constitutional AI（憲法AI）」というアプローチで訓練されています。明文化された原則に基づいてAI自身が応答を評価・修正する仕組みにより、おべっかを言わず、事実に基づいた誠実な会話を設計レベルで担保しています。

(詳しくは「AIのモデルに性格はあるのか？」をご参照ください)

LLMは海外モデル、特にアライメント品質に優れたモデルを選ぶのが営業AI用途では合理的です。

要素③：音声合成（TTS） - 海外勢が品質トップ

TTS（音声合成）でも、2026年時点では海外勢がリードしています。

業界の音質評価では、ElevenLabsとOpenAI TTSが最高評価（5つ星）、GoogleとAzureが4つ星とされています。最高水準のTTSエンジンは発話品質の誤認率（WER）2.83%を達成し、ストリーミング遅延も約200ミリ秒に短縮されています。吐息、笑い、微細な感情の揺れまで再現可能なレベルです。

ただし、日本語には英語にはない「丁寧さのグラデーション」があります。同じ「はい」でも場面によってトーンが変わる。この微妙な使い分けの自然さは、実環境での検証が不可欠です。

TTSは海外の最高品質エンジンを基本とし、日本語の自然さを実通話で検証すべきです。

ハイブリッド戦略 - 要素ごとに最適を組み合わせる

ここまでの分析を統合すると、音声AIインテリジェンスの技術選定は明快です。

STT（聞く力） → 国産エンジンが日本語精度で優位

LLM（考える力） → 海外モデルが言語能力・アライメントで優位

TTS（話す力） → 海外エンジンが品質トップ

Leadsiaは、このハイブリッド戦略を実践しています。日本語の音声認識には国産技術の精度と安定性を活かし、会話の頭脳にはClaudeのアライメント品質を採用し、音声合成には海外の最先端TTSを活用。各要素の「最強」を組み合わせ、一つのサービスとして統合することで、人間レベルの自然な日本語対話を実現しています。

データ主権とセキュリティ - 冷静に評価する

海外AI採用時に必ず挙がるデータ主権の懸念は正当です。約78%のユーザーがAIへの法規制が必要と感じてい

るという調査結果もあります。

ただし「海外AI＝データ漏洩」は短絡的です。多くの海外AIサービスはAPI経由のデータを学習に使用しないポリシーを採用しています。Anthropicも、APIデータをモデル学習に使用しないことを明言しています。

重要なのは個別のデータポリシーの確認です。データの保存場所、保持期間、第三者共有の有無、暗号化方式 - これらを契約前に確認してください。

金融・医療・法律など規制の厳しい業界では、国産LLMとの組み合わせやオンプレミス型の音声認識エンジン採用を検討する価値があります。

「AIツール乱立」の罅を避ける

2026年の新たな課題として、AIツールの乱立による情報の分断があります。BCGの調査では、AIエージェントを導入している日本企業はわずか7%にとどまっており、多くの企業がツール間の連携に苦慮しています。

この問題を避けるためにも、音声AIインテリジェンスはSTT・LLM・TTSが一つのプラットフォーム上で統合されたサービスを選ぶことが重要です。

Leadsialは、複数の最適な技術要素をバックエンドで統合し、ユーザーには迷わないシンプルなインターフェースを提供しています。導入は最短3分。内部でどの技術を使っているかを意識する必要はありません。

(導入の簡便性については「ゼロタッチ運用とは？」をご参照ください)

まとめ：「国産か海外か」ではなく「要素ごとの最適組み合わせ」

「国産AIか海外AIか」 - この問いに対する答えは、「両方の強みを、要素ごとに活かす」です。

音声認識は国産の精度と安定性を。LLMは海外のアライメント品質を。TTSは海外の最先端技術を。そして、これらを一つのサービスとして統合し、「自然な日本語で会話するAI」として提供する。

日本のAI活用率は世界に後れを取っていますが、それは選択肢がないからではなく、選び方が定まっていないからです。本記事が、その選定基準の一助になれば幸いです。

関連記事

- 音声AIインテリジェンス技術とは？従来の音声システムとの決定的な違い
- AI営業電話の品質を決める3つの要素 | なぜ人間レベルの対話が可能なのか
- AIのモデルに「性格」はあるのか？ - Claudeの魂をつくった哲学者と、AIの人格設計という新領域
- セールステックSaaSの選び方 | 失敗しない5つのポイント

Leadsiaは、AI営業インテリジェンス「ALICE」、AI音声インテリジェンス「SOPHIA」、AI業務インテリジェンス「LYDIA」を通じて、日本のB2B企業の営業DXを支援するセールステックSaaS企業です。各AIエージェントの頭脳にはAnthropicのClaudeを採用し、Constitutional AI（憲法AI）に裏打ちされた安全性と会話品質を両立した営業自動化を実現しています。

詳しくは[[Leadsia公式サイト](#)]をご覧ください。