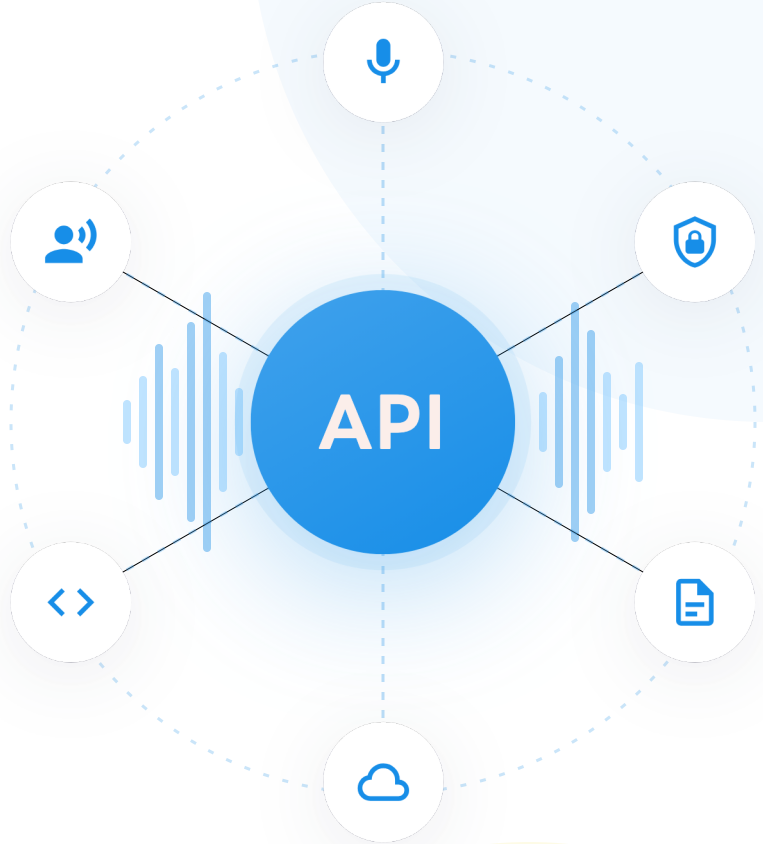
Citoras

Cloud-based Inference TTS
On Realtime Accelerator Server

自社モデルを、セキュアに、本番運用へ。

多キャラクター音声に対応する、本番向け音声合成 API サーバー。

Citoras ／ Aivis Project



Citoras とは？

自社管理 GPU サーバーで動く、本番向け音声合成 API サーバー

セキュア

入力テキスト・生成音声が社外に出ることはありません。全ての処理が、お客様の自社環境内で完結します。

高品質

Aivis Cloud API と同じ音声合成エンジンを、自社環境でそのままご利用いただけます。

スケーラブル

400 以上の音声合成モデルを、
1台のGPU サーバーで運用可能。
将来の多キャラクター展開も低コストで。

Aivis Cloud API の内部基盤として実運用中 — 11か月の実績

ご提案の背景

キャラクターが増えるほど、運用基盤の選び方が将来を左右する

モデル資産は、作るだけでは終わらない

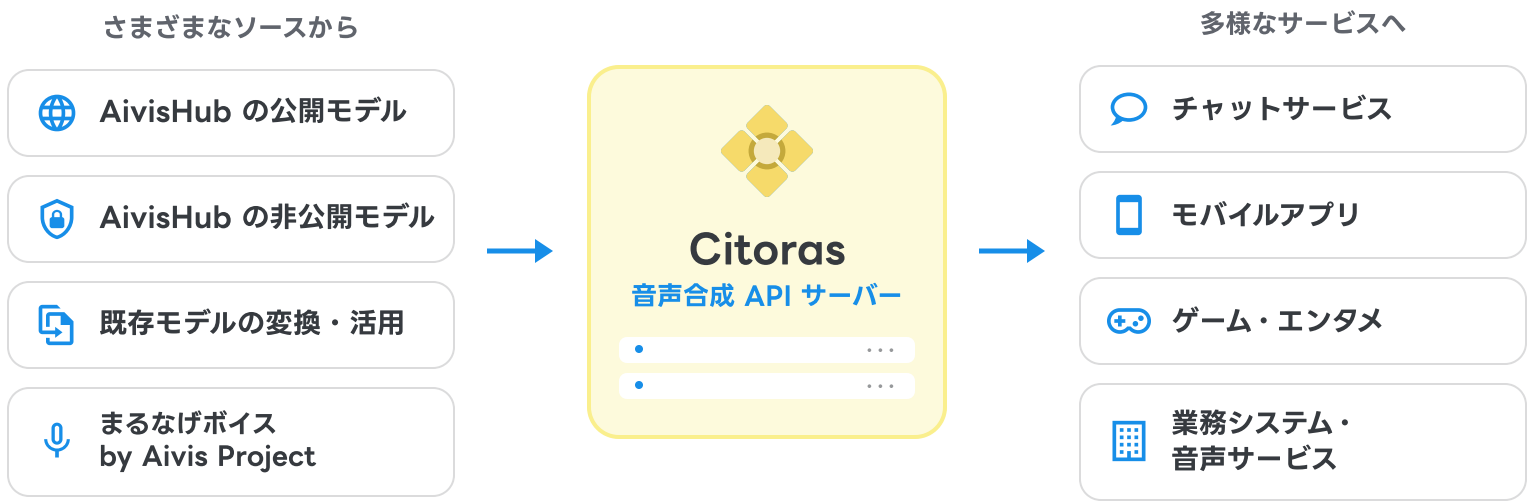
キャラクターや音声合成モデルは増え続け、更新・管理には継続的な運用が必要です。

複数モデルを継続的に運用できる基盤が必要

スケール・バージョン管理・リソース最適化まで、まとめて支える基盤が欠かせません。

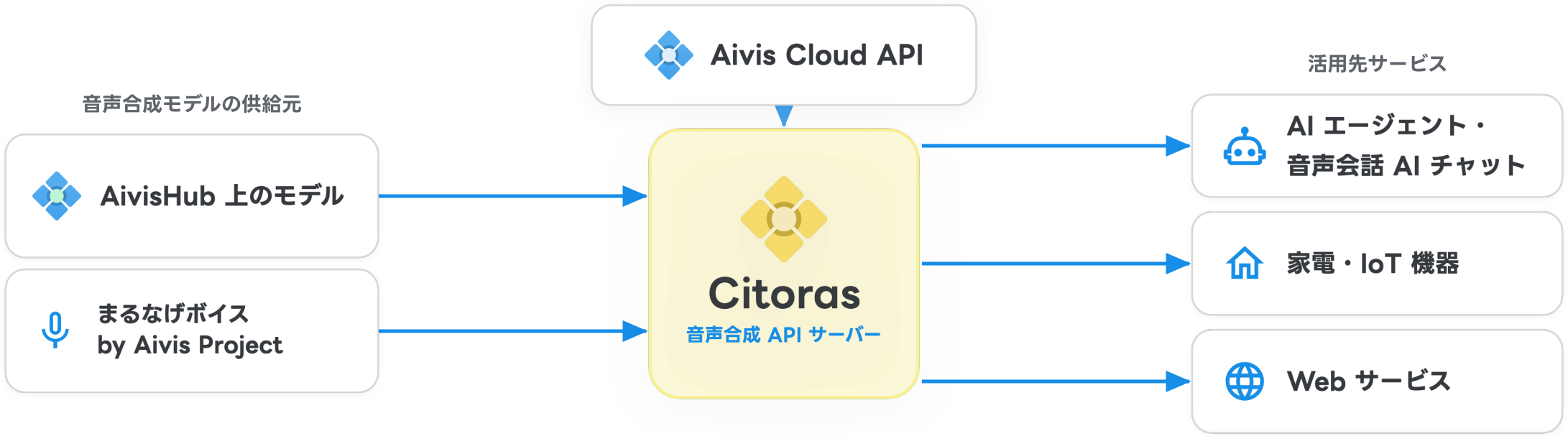
低遅延と安定性が、音声体験を左右する

リアルタイム性が体験を大きく左右します。安定品質で届ける仕組みが求められます。



Citoras の位置づけ

Aivis Cloud API と同じ音声合成基盤を、お客様の環境で運用



Citoras は、**Aivis Cloud API の内部で実運用されている音声合成 API 基盤**。
まずは Cloud API で性能・品質をお確かめ頂いてから、**品質をそのままに自社環境で運用**できます。

役割分担

同じ音声技術を、異なる運用形態へ

	AivisSpeech Engine	Citoras
主な用途	ローカル利用・検証・アプリ同梱	**GPU サーバーでの本番 API 運用**
実行環境	PC・Docker（CPU / GPU 選択可）	**NVIDIA GPU サーバー**
想定する利用	単発・対話的な利用	**本番環境での多数リクエストの同時処理**
モデル形式	AIVMX	**AIVM**
ストリーミング	主用途としては想定しない	**低遅延ストリーミング対応**
長文の処理	短文・手動分割を想定	**長文を自動で適切に分割**
モデル運用	ローカル配置	**S3 互換ストレージから自動検出・再起動不要**
監視・運用向け機能	最小限	**ヘルスチェックやメモリやモデル状態を取得できるモニタリング API を完備**

いずれも Aivis Project の持つ音声合成技術を、**異なる運用形態に向けて実装した製品**です。

2026年6月時点の情報です。

セキュリティ

入力テキストと生成音声を、お客様環境内で処理

生成処理はすべてお客様環境内で完結。性能評価は Cloud API で、本番は自社環境で。

処理はお客様環境内で完結

入力テキストの解析から音声生成まで、
すべてお客様環境内で実行します。
入力テキストや生成音声は弊社サーバーに一
切送信されません。

モデル保管先を選択可能

S3 互換ストレージに対応。Cloudflare R2
など外部クラウドから、MinIO などの
完全オンプレミスまで、ポリシーに合わせて
保管先を選べます。

多層の認証・アクセス制御

複数 API キー・CORS 制御・IP ベースの
レート制限機能を標準で備えます。

段階導入によるリスク低減

まず Aivis Cloud API で性能をお確かめ
頂いた上で、段階的に導入いただけます。

応答品質

リアルタイム AI 対話に耐える低遅延

高速なレスポンス速度

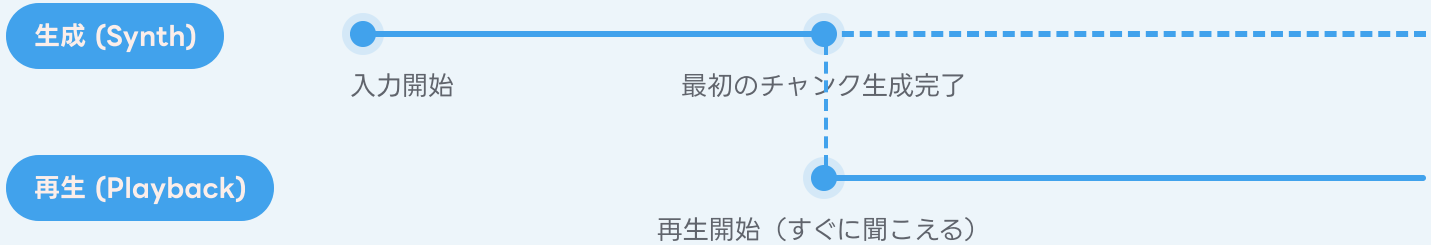
ストリーミング生成で、最初に生成された音がすぐ耳に届きます。
最速0.3秒で再生を開始できます。

音声生成と再生を並行

音声生成と再生を同時に進めることで、待ち時間を最小限にとどめ、
ユーザーの会話体験を向上できます。

長文も適切に分割生成

文末や句点で自然に区切って生成し、
途切れのない再生を実現します。

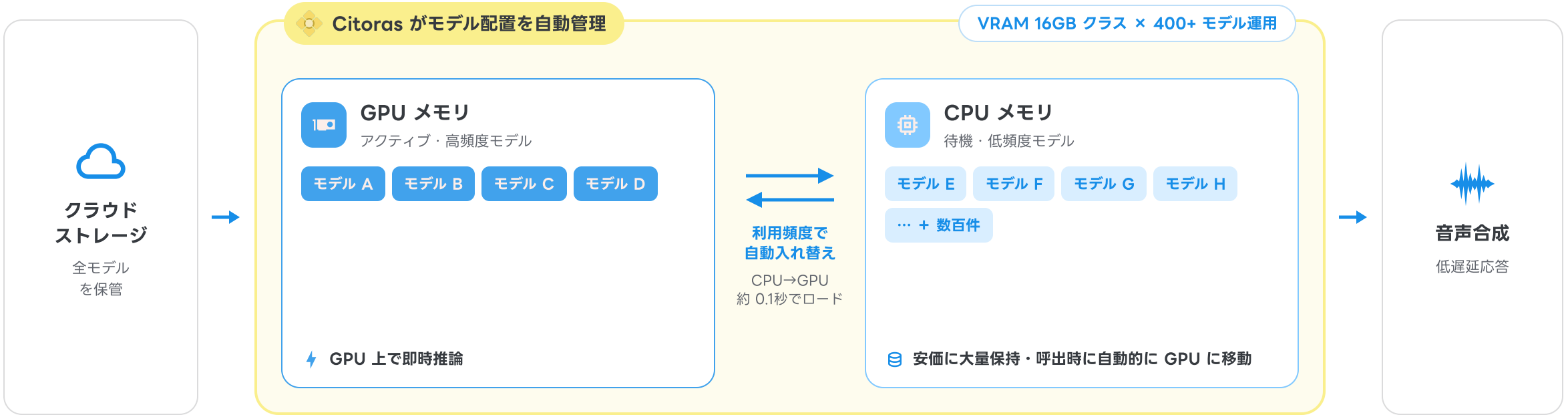


実測条件：NVIDIA RTX A4000 ／ 30秒 (230文字) の音声 ／ 対応出力形式 WAV・FLAC・MP3・AAC・Opus

拡張の経済性

400件以上のモデルを、1台の GPU サーバーで運用

VRAM 16GB クラスの GPU でも、将来の運用モデル数の拡大に応えます。



AivisHub 公開モデル
61 件

実運用モデル
400 件 +

- **全モデルを GPU に載せる必要はありません**
  社内では VRAM 16GB の GPU サーバー2台で、400件以上のモデルを1年近く運用している実績があります。
- **よく使うモデルは速く、待機モデルも約 0.1秒で復帰**
  利用頻度に応じて GPU／CPU の配置を自動調整。CPU 側に保持されているモデルも、約 0.1 秒で GPU にロードして以降は高速に生成できます。
- **将来の多キャラクター展開を低コストに支える**

2026年6月時点の情報です。

日本語の読み品質

英単語・数値・固有名詞も、自然な読みに

テキストをそのまま渡すだけで、自然な読み方に整えて読み上げます。

英単語混じりでも自然に読み上げ

ChatGPT などの製品名をはじめ、サービス名・ブランド名といった英単語の固有名詞を、文脈に沿うカタカナ英語に自動で変換して読み上げます。

数値・記号・単位を文脈で判断

電話番号は桁読み、住所は番地読み、日付・時刻・単位付きの文章も適切に読み分けます。

固有名詞・人名・地名の読み方は、次ページの「ユーザー辞書」で解説します。

入力テキスト		実測の読み
iPhone 15 Pro Max (256GB)	→	アイフォンジューゴプロマックス、ニヒャクゴジューロクギガバイト
Wi-Fi (5GHz/2.4GHz)	→	ワイファイ、ゴギガヘルツ／ニーテンヨンギガヘルツ
03-1234-5678	→	ゼロサン、イチニーサンヨン、ゴーロクナナハチ
〒100-0001	→	郵便番号イチゼロゼロの、ゼロゼロゼロイチ
東京都港区六本木1-2-3	→	東京都港区六本木イチのニのサン
¥1,234,567	→	ヒャクニジュウサンマン ヨンセンゴヒャクロクジュウナナ円
2025/7/5(土) 7:05	→	2025年7月5日 土曜日 シチ時ゴ分

ユーザー辞書

お客様ごとに、独立したユーザー辞書を運用

サービスの先にいる、一人ひとりのお客様に正しい読み方を届けます。

ユーザー辞書 ID ごとに完全に分離

エンドユーザーや取引先ごとに専用の辞書を割り当て、読みの横断混入を防ぎます。

読み・アクセント・品詞まで細かく指定

表層形・読み方・アクセント型・品詞・優先度を単語ごとに設定できます。

AivisSpeech / Aivis Cloud API と互換

各アプリのユーザー辞書管理画面からエクスポートした辞書データを、そのままインポートできます。

データはお客様システム側で一元管理

辞書はステートレスに保持。お客様のデータベースを正本とし、変更の都度反映する運用を推奨します。

お客様ごとに独立したユーザー辞書 ID を発行

横断読みの混入なし

単語ごとに指定できる項目

表層形	読み	アクセント型	品詞	優先度
担々麺	タンタンメン	3	固有名詞	8
新田 ／ 真剣佑	アラタ ／ マッケンユウ	1／3	人名	10

開発予定：ユーザー辞書の管理や、モニタリング状況をリアルタイムに確認できる、管理者向けの GUI ダッシュボードも後日開発予定です。

Coming Soon

想定ユースケース

Citoras が活きるユースケース

AI エージェント・音声会話 AI チャット

人が話すように自然な返答音声を、低遅延で再生。
キャラクターボイスの追加・変更も、サーバー再起動なしで反映できます。

家電・IoT 機器の音声インターフェイス

入力を社外に出すことなく、高速に音声化。
テレビやスマートスピーカーなど IoT 機器のバックエンドサーバーに組み込めます。

コールセンター・IVR の自動応答

キャンペーン音声を即日更新、録音スタジオ不要。
商品名・地名は辞書登録で正確に読み上げできます。

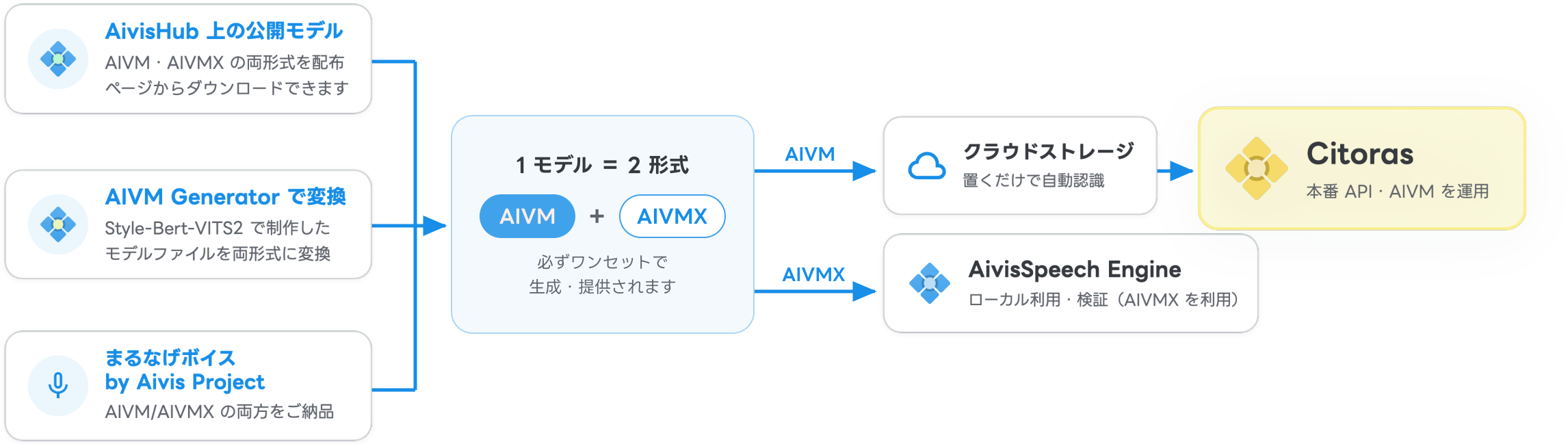
コンテンツ制作・ナレーション

大量の原稿を、同時並行で高速に生成。
多数のキャラクター音声を使い分け、制作期間を短縮できます。

モデル管理

モデルを置くだけで運用

S3 互換ストレージに置くだけで自動認識。サーバー再起動なしでモデルの追加・変更・削除が可能。



モデルの追加・削除は再起動なし
Citoras はストレージを常時監視し、モデルの追加・変更削除を自動で認識します。

AIVM と AIVMX は必ずペア
どの経路でも両形式が揃うため、Citoras 用の AIVM が欠けることはありません。

きめ細やかな音声制御

LLM が生成した発話を、表現豊かな音声へ

SSML で音声の話速・音量・無音区間などを細かく制御可能。

間の挿入

全体の無音長さとは別に任意の長さの無音区間を挿入でき、会話のリズムを作れます。

話速・ピッチ・音量

全体の話速や音量などとは別に、一文の中で部分的に変えることができます。

読み方の指定

特定の単語の読みを、その場で指定できます。
ユーザー辞書機能でも対応可能ですが、特定単語の読みをその場で指定したい場合などにも便利です。

感情表現の制御

音声合成モデルに含まれるスタイル（喜び・落ち着き等）と、その強さを指定できます。

SSML での入力例

```
<aivis:emotion style="Happy" intensity="1.5">
  <prosody rate="110%">こんにちは。<break time="300ms"/>今日の天気は<sub alias="ハレ">晴れ</sub>です。</prosody>
</aivis:emotion>
```

業界標準である SSML のサブセットに対応。
他の音声合成サービスからの移行コストを最小化できます。

本番運用機能

監視・認証・アクセス制限を備えた運用基盤

既存のサービス基盤と、API でつながります。

監視 API

ヘルスチェック、メモリ使用状況、モデル状態、利用統計を Citoras のモニタリング API から取得できます。
Prometheus / Datadog などの既存の監視基盤に、これらの API から取得したメトリクスを取り込む運用も可能です。

認証・アクセス制御

複数 API キーの発行、CORS 設定、IP ベースのレート制限機能を標準利用できます。
もちろん、複数台構成のためのロードバランサや、認証サーバー・リバースプロキシを挟む構成も想定。

Tailscale の標準サポート

アプリケーションサーバー と GPU サーバー間の通信用内部ネットワークとして Tailscale を標準サポート。
GPU サーバーを直接インターネットに公開することなく、アプリケーションサーバーと GPU サーバーを異なるクラウドに配置し、GPU サーバーのランニングコストを抑えることが可能です。

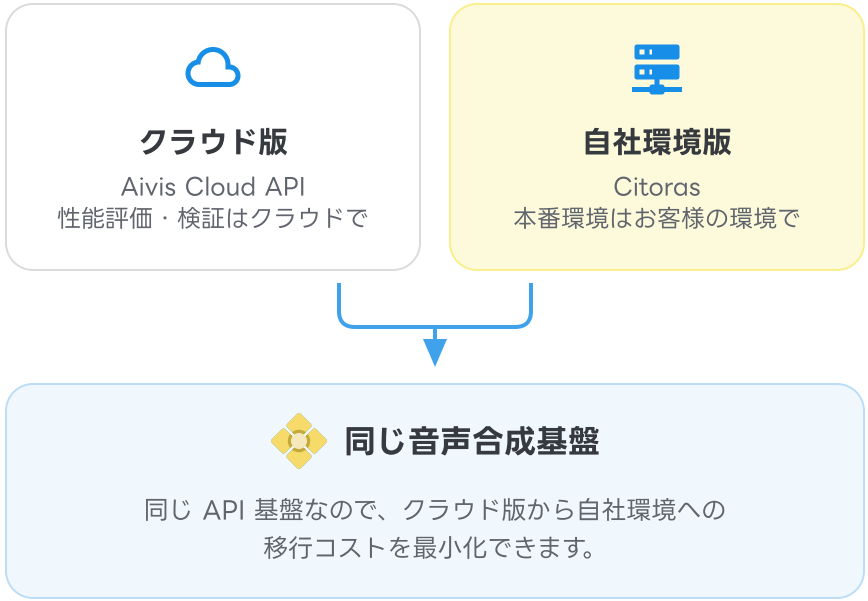
ログとエラー追跡

構造化された直感的に読みやすいログを出力します。
ログは自動で1日ごとにジャーナリングされます。
任意で Sentry によるエラー収集を許可いただきますと、予期せぬエラーが発生した際に弊社側で迅速に修正と更新を行えます。

実運用実績

Aivis Cloud API のバックエンドとして実運用中

同じ基盤を、約11か月にわたり実運用中。



- **Citoras は、Aivis Cloud API の音声合成基盤**として実際に運用されています。
- **2025年7月23日から無料ベータ運用、2026年2月1日から正式課金**を開始しました。
- 実ユーザーからのフィードバックを受け、**読み品質・安定性・運用性を継続的に改善**しています。
- まずはクラウド版で性能を評価し、要件が合い次第**自社環境への展開へ移れます**。

2025年7月23日
無料ベータ版として運用開始

2026年2月1日
正式リリース・課金開始

2026年6月時点
約11か月の実運用

2026年6月時点の情報です。

料金

月額 10万円 の、わかりやすい定額制

構成やモデル数によらず、まずは定額でご導入いただけます。

月額ライセンス

¥ 100,000

円 ／ 月（税別）

サーバー台数・モデル数に関わらず
一律料金でご利用いただけます。

モデル数・API リクエスト数を問わず定額

運用するモデル数やリクエスト数が増えても追加費用は発生しません。定額でお使いいただけます。

監視・認証・ログ機能を標準搭載

追加料金なしで、本番運用に必要な機能一式をご利用いただけます。

機能追加・内蔵辞書の読み精度改善アップデートを継続的にご提供

ご契約期間中、機能追加・修正を行った新バージョンや内蔵辞書の更新を継続的にご提供します。

AI導入補助金など公的支援制度の活用についても、あわせてご相談いただけます。
対象制度の選定から申請の進め方まで、私たちがアドバイスいたします。

導入ステップ

試用から運用まで、段階的に

5 ステップで、低リスクに自社環境運用へ。

1 モデル資産の準備

AivisHub の公開モデルから利用するモデルを選ぶか、「まるなげボイス」またはお客様ご自身で自社独自のモデルを準備します。

- 公開モデル・自社モデル
- 変換・新規制作の手配

2 クラウド版で試用

Aivis Cloud API の音声合成デモで試用いただき、生成される音声の品質と挙動を確認します。

- 用途別の音声評価
- パラメータの調整

3 サーバー準備

GPU サーバー・ネットワーク環境・S3 互換ストレージを整えます。

- GPU・ネットワークの手配
- モデル保管先の決定

4 コンテナで構築

プライベート GHCR から配信されるコンテナイメージを pull し、Docker コンテナを構築します。環境に合わせ、認証やロードバランサを設定し、本番運用へ進みます。

- API キー・CORS の設定
- 監視・ログ基盤の接続

5 継続アップデート

プライベート GHCR に公開される新しいイメージタグへに適宜更新し、機能改善を継続的に適用します。

- 機能・読み品質の改善
- セキュリティ更新の適用

推奨構成

GPU	GPU メモリ	システムメモリ	ストレージ
NVIDIA RTX A4000 クラス以上	16GB 以上	16GB 以上	50GB 以上

2026年6月時点の情報です。

次の進め方

まずは、お手元の用途で試していただけます

1 Aivis Cloud API で オンライン試用

音声品質・速度・API の使い勝手を、ブラウザ上のデモページで低リスクに確認できます。

2 モデル資産の 調達計画をご相談

AivisHub 上の公開/非公開モデル、既存の Style-Bert-VITS2 モデルからの変換、
新規制作（**まるなげボイス by Aivis Project**）のいずれにも対応します。

3 Citoras 採用時の 構成検討

GPU サーバーの構成・ネットワーク・ストレージなどをお客様環境に合わせてご提案します。

お問い合わせ

https://forms.gle/sTsZGfX7aR8ox8Rs7

Citoras は月額 10万円（税別）でのご提供。導入補助金のご相談も承ります。

Appendix A

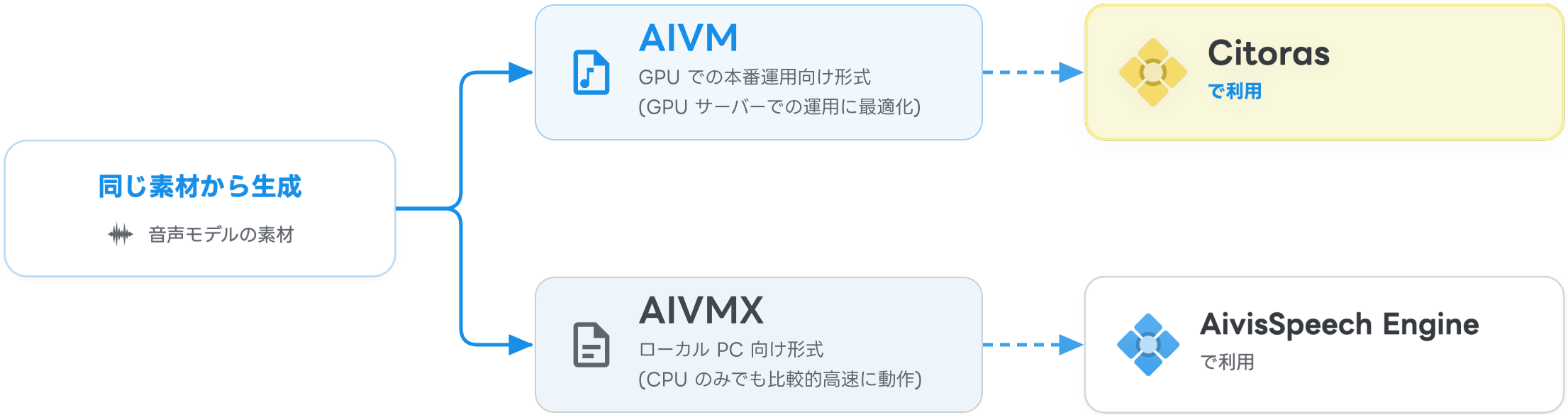
付録 A：音声制御（SSML）の対応範囲

制御	仕様
無音区間	0〜60秒（ミリ秒指定、もしくは medium 等文字値で指定できます）
単語の読み指定	30文字以内のカタカナで読みを指定できます
話速	0.5〜2.0 倍
ピッチ	−1.0〜+1.0
音量	0.0〜2.0 倍
テンポの緩急	0.0〜2.0
スタイル・感情表現の強さ	0.0〜2.0
段落・文の区切り	デフォルト: 0.4秒

ⓘ 未対応の SSML タグは**無視され、エラーにはなりません**。仕様詳細は Aivis Cloud API の API ドキュメント をご確認ください。

Appendix B

付録 B：AIVM / AIVMX 形式の関係



AivisHub からは、AIVM / AIVMX の両形式の
モデルファイルをダウンロードできます。

まるなげボイス by Aivis Project は、
AIVM / AIVMX の両形式で納品いたします。

Style-Bert-VITS2 で制作したお手元のモデルを、
AIVM Generator で両形式に変換できます。

Appendix C

付録 C：読み方サンプル集（一部）

英単語

Bluetooth	→	ブルートゥース
Wi-Fi	→	ワイファイ
AI	→	エーアイ
API	→	エーピーアイ
HyperAI	→	ハイパーエーアイ
Project	→	プロジェクト

123 数字・単位

03-1234-5678	→	ゼロサン、イチニーサンヨン、ゴーロクナナハチ
256GB	→	256ギガバイト
5GHz	→	5ギガヘルツ
2.4GHz	→	ニーテンヨンギガヘルツ
100%	→	100パーセント
100km/h	→	100キロメートル毎時

日付・時刻

2026年6月	→	ニセンニジューロクネンロクガツ
10時30分	→	ジュージサンジュップン
14:05	→	ジュウヨジゴフン
7月5日(土)	→	シチガツイツカドヨウビ
AM 9:00	→	エーエム クジ
B2F	→	チカ ニカイ

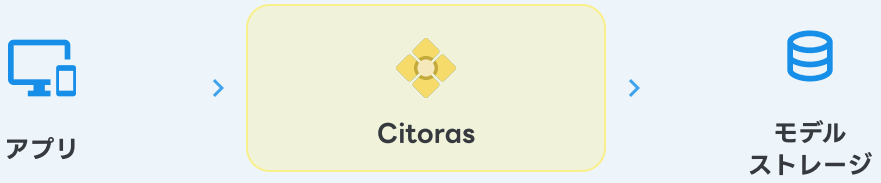
そのほかの固有名詞の読み方は**辞書や設定でカスタマイズ可能です**。

Appendix D

付録 D：接続構成パターン

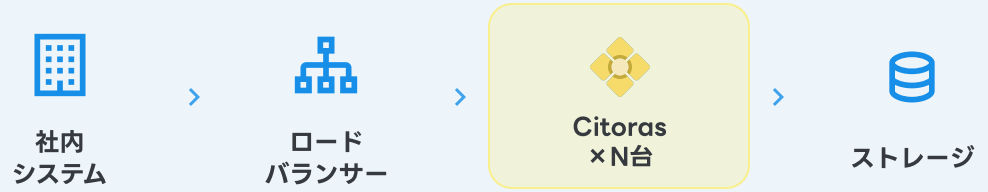
既存システムに合わせて配置できます。

パターン 1：シンプル接続（ローカル・小規模利用向け）



アプリ/静的サイトから直接つなぐ最小構成。検証や小規模運用に適しています。

パターン 2：高可用・スケール接続（本番向け）



複数台構成による**負荷分散・冗長化**が可能。大規模・高可用に適しています。

Citoras は特定のクラウドやインフラ構成を強制せず、お客様の既存構成に合わせて配置できます。

Appendix E

付録 E：API リクエストのパラメーター一覧

POST /v1/tts/synthesize

```
{
  "model_uuid": "xxxxxxxx-xxxx-…",
  "speaker_uuid": "xxxxxxxx-xxxx-…",
  "style_name": "Happy",
  "user_dictionary_uuid": "xxxxxxxx-xxxx-…",
  "text": "こんにちは！今日の天気は晴れです。",
  "use_ssml": true,
  "speaking_rate": 1.15,
  "emotional_intensity": 1.5,
  "tempo_dynamics": 1.1,
  "volume": 1.0,
  "leading_silence_seconds": 0,
  "line_break_silence_seconds": 0.3,
  "output_format": "mp3",
}
```

カテゴリ	指定項目	役割
音声合成モデル	モデル UUID・話者 UUID・スタイル指定	声を選択
辞書	ユーザー辞書 UUID	ユーザー辞書を適用
テキスト	読み上げテキスト（SSML 対応）	最大 3000 文字
表現制御	話速・ピッチ・音量	音声を調整
感情表現	感情表現の強さ・テンポの緩急	表現を調整
出力形式	WAV・FLAC・MP3・AAC・Opus	ビットレート等

リクエストパラメータの詳細は API ドキュメント をご確認ください。

オプションで API キー認証を追加できます。
より高度な認証が必要な場合は、前段に認証サーバーを挟む構成を推奨します。