Gemini で構築する マルチモーダル "ライブ" アプリケーションの事例と構築の勘所

Google Cloud Next Tokyo

段野 祐一郎

Google Cloud
メディア カスタマー エンジニア

アジェンダ

01 **ライブ AI によるパラダイム シフト**
- インタラクションの再定義
- ライブ AI がもたらす価値
- ユースケース例

02 **Gemini Live API & デモ**

03 **開発者ガイド**
- Gemini API / Vertex AI
- SDK、認証、アーキテクチャ、ツール、会話スタイル

04 **プロトタイプから本番への勘所**
- パフォーマンスとコストの最適化
- 信頼性と堅牢性の確保
- 倫理とプライバシー

05 **まとめ**

01. ライブ AI による パラダイム シフト

ライブ AI の世界観（Project Astra）

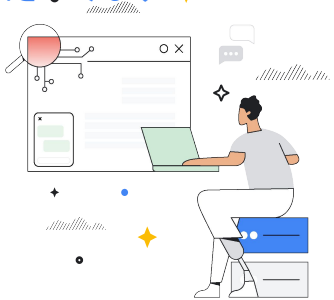
「応答」から「会話」へ。
AI との関係性を再定義するパラダイム　シフト

従来の AI 対話モデル	Gemini Live API 対話モデル
モデル	**モデル**
ターン制の「リクエスト＆レスポンス」 （チェスの対局のような形式）	連続的で流れるような「真のインタラクション」 （人間同士の自然な会話）
課題	**提供価値**
高レイテンシ: 応答が遅く、待たされる 文脈の途絶: 会話の流れが切れやすい 一方通行: ユーザーは AI の応答を待つ	低レイテンシ: 人間のような即時応答 割り込み可能: 相手の話を遮って話せる自然さ ステートフル: 文脈を維持し、深い対話を実現

ライブ AI がもたらすビジネス価値

顧客体験の飛躍的向上

リアルタイムでパーソナライズされたインタラクション（顧客対応）により顧客満足度とロイヤルティを新たな高みへと引き上げます



オペレーション効率の劇的な改善

リアルタイム処理により、より迅速で的確な意思決定を可能に

例えば、マニュアル確認や在庫検索など AI アシスタントが自動的に行うことで、オペレーションのスピードを迅速化できます

全く新しいサービスモデルの創出

リアルタイムの対話型サービスが実現可能になることで、ハンズフリーの現場作業員向け AI アシスタント、没入型の教育プラットフォームなど、新たなビジネス チャンスが生まれます

“目が見える AI との対話”
マルチモーダル ライブ AI の大きな可能性

技術の継承

操作を学ぶ際、マルチモーダル AI は映像から手順の正誤を判断し、リアルタイムでフィードバックを返すことが可能

迅速な問題解決

トラブルに遭遇した際も、画面や映像をマルチモーダル AI に共有することで、迅速かつ的確なサポートが可能になる。

口頭での説明が難しい状況でも、AI が一緒に画面を見てくれる

パーソナライズされた提案

「この服に合うコーディネートを提案して」と AI に尋ねると、AI は色やスタイルを瞬時に分析し、店内の在庫から最適な組み合わせを提案することも可能

ライブ AI のユースケース例

現場作業アシスタント

現場作業員が映す機器や状況を AI が認識し、修理やメンテナンスの手順を音声と映像でガイドする

バーチャル説明員

カメラにかざした商品を AI が認識し、その場で機能や使い方を対話形式で詳しく説明する

ファッション コーディネーター

試着した服を AI が評価し、コーディネートの提案や着こなしのアドバイスをリアルタイムで行う

インタラクティブ料理アシスタント

手元の食材や調理工程を認識し、次の手順を対話形式でリアルタイムにナビゲートする

パーソナル コーチ

ユーザーの動きをカメラで分析し、リアルタイムで的確なアドバイスと音声フィードバックを提供する

会議ファシリテーター

会議中の発言をリアルタイムでテキスト化・要約し、決定事項やタスクを整理して円滑な議事進行を支援

視覚障がい者向けナビゲーション

カメラ越しの風景を AI が音声で描写。標識の読み上げや障害物の警告で、安全な歩行をサポート

AI 語学学習パートナー

実際の会話場面を想定し、ユーザーの発音や表現をリアルタイムで修正しながら自然な会話練習をサポート

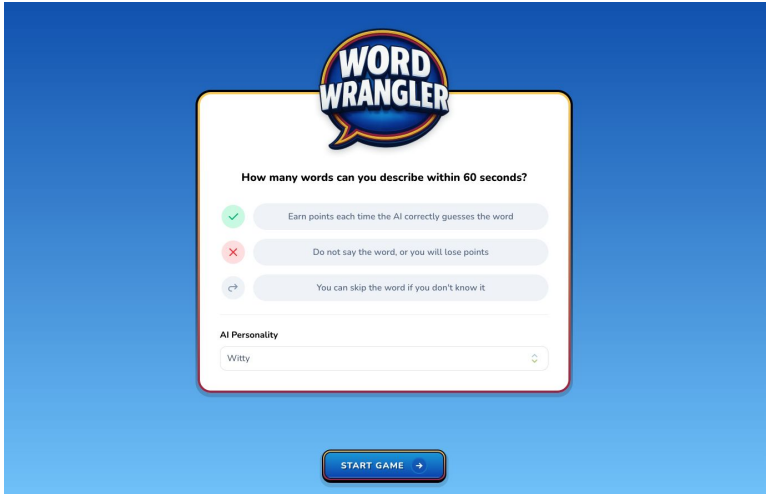
ロールプレイング・トレーニング

営業や面接の練習相手として AI が機能。会話内容だけでなく、表情や声のトーンも分析しフィードバック

ライブ AI の事例

Daily.co

Live API を使って音声ベースのワード推測ゲーム Word Wrangler を作成し、クラシックな単語ゲームに AI でひとひねりを加える



Bubba.ai

トラック ドライバー向けに特別に開発されたエージェント型音声ファースト AI アプリケーション Live API を利用してシームレスな多言語の音声コミュニケーションを実現。ドライバーはハンズフリーで操作可能

・貨物運送案件を検索し、詳しい情報を提供
・ブローカー / 荷主に電話をかける
・市場データに基づいて運賃交渉を行う
・貨物の予約と運賃の確認を行う
・トラックの駐車場を検索して予約する
・ホテルに電話をかけて予約可能状況を確認
・荷主や受取人と受け渡しのスケジュールを調整

引用: Google for Developers - リアルタイム インタラクションの実現 Live API を使って開発する -

02. Gemini Live API & デモ

Gemini

- **マルチモーダル**
  複数のデータ種類で学習を行っており、画像、動画、音声、コード、テキストと様々なフォーマットを入力としてモデルとやりとりできます。
- **長いコンテキスト（文脈）の理解 とキャッシュ**
  100 万トークンの入力に対応。数時間の動画、数十時間の音声。
- **柔軟なモデルサイズとプライシング**
  用途に応じたコスト・品質・応答性のバリエーション
- **安定性**
  可用性に関するSLA を提供しています。
  予めスループットを確保する提供形式も有り。

Gemini 2.5

パフォーマンス

- 1.5 より
  高速かつ高品質

出力フォーマット

- **音声**や**画像**での出力
  （現在 Allowlist で提供中）

ツールへの対応

- Google 検索
- コード実行
- 合成関数の呼び出し

Live API

- **マルチモーダルかつリアルタイム**な入出力に対応する双方向ストリーミング API

Thinking モデル

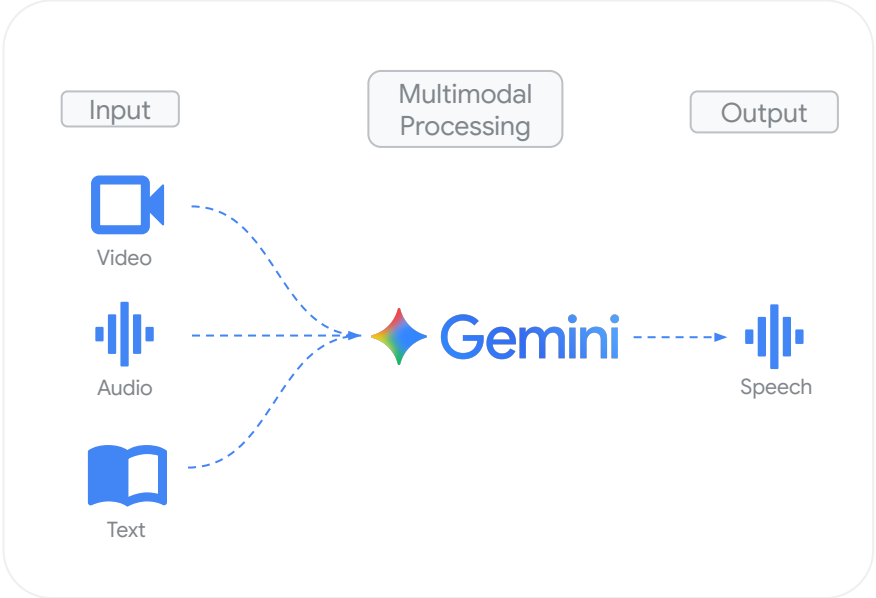
- **「思考プロセス」**を生成し、より強力な推論機能を実現。

現在提供中の Live API モデル (Gemini 2.5 系)

- Gemini 2.5 Flash with Live API (gemini-live-2.5-flash)
- Gemini 2.5 Flash with Live API native audio (gemini-live-2.5-flash-preview-native-audio)

Gemini Live API

- テキストやカメラ、スクリーンから
  音声や動画をリアルタイムに入力 し
  音声を出力 。
- 複数言語、複数音声をサポート
  （**日本語にも対応** ）
- Websocket を使った双方向通信
- セッション時間はデフォルト 10 分
  （**GenAI SDK を利用することで延長可能** ）
- **関数呼び出し、コード実行、ツールとしての検索のサポート：モデルを外部サービスやデータソースと統合可能**



Gemini Live API モデル比較（2025/8 時点）

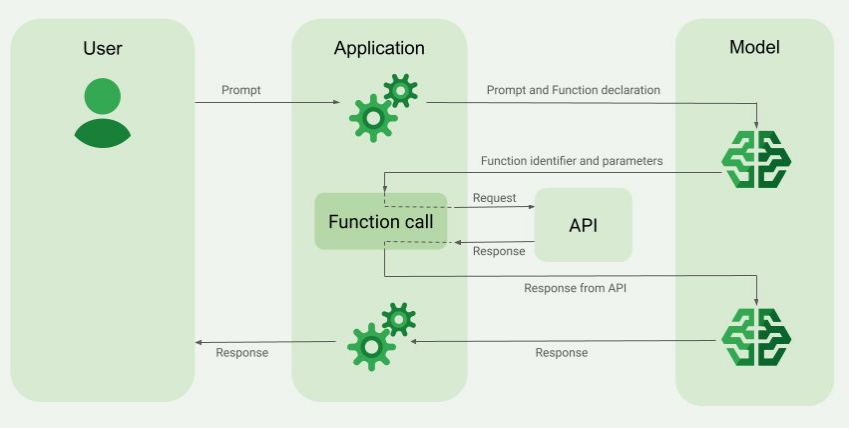
モデル名	**Gemini 2.5 Flash with Live API**	**Gemini 2.5 Flash with Live API native audio**	**Gemini 2.0 Flash with Live API**
モデル ID	gemini-live-2.5-flash	gemini-live-2.5-flash-preview-native-audio	gemini-2.0-flash-live-preview-04-09
公開ステータス	Private GA	Public Preview	Public Preview
主な目的/特徴	汎用性が高く、価格とパフォーマンスに優れる。	Live API での自然な音声対話に特化	Live API 対応の旧世代モデル
入力	音声、動画、テキスト	音声、動画	音声、動画
出力	テキスト、音声	テキスト、音声	テキスト、音声
関数呼び出し/コード実行	⭕	△（関数呼び出しのみ）	⭕
検索 / グラウンディング	⭕（Google検索、RAG Engine）	△（Google 検索のみ）	△（Google 検索のみ）
ネイティブオーディオ機能	❌	⭕ 強化された音声品質、プロアクティブオーディオ、感情認識対話	❌
Provisioned Throughput	⭕	❌	❌

Live API コア機能
組み込みツール（Tools）

問い合わせ内容に応じて、特定のツールを呼び出し、使用可能
→ より正確で関連性の高い情報提供、パーソナライズされたインタラクション、 タスクの自動化
→ **顧客満足度の向上、オペレーション コスト削減、新たなビジネス機会の創出へ**

組み込みツール

- 関数呼び出し (Function Calling)
- Google 検索によるグラウンディング
- Python コード実行 (gemini-live-2.5-flash のみ)
- Vertex AI RAG Engine によるグラウンディング (gemini-live-2.5-flash のみ)



「関数呼び出し」の例

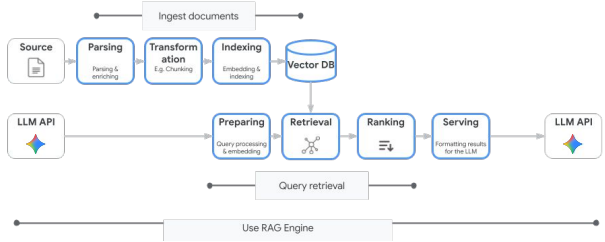
Vertex AI RAG Engine

Vertex AI RAG Engine は独自の RAG * を容易に構築可能なソリューション

** RAG: 関連情報をモデルに提供し、質問に正確に答えるようにする手法*

Vertex AI RAG Engine を Gemini Live API のコンテキスト ストアとして使用して、会話に関連する過去のコンテキストを取得し、コンテキストとして拡充可能。

この機能を利用して、さまざまなGemini Live API セッション間でコンテキストを共有することができるようになる。




```
memory_store=types.VertexRagStore(
    rag_resources=[
        types.VertexRagStoreRagResource(
            rag_corpus=memory_corpus.name
        )
    ],
    store_context=True
)

async with client.aio.live.connect(
    model=MODEL_NAME,

config=LiveConnectConfig(response_modalitie
s=[Modality.TEXT],

tools=[types.Tool(

retrieval=types.Retrieval(

vertex_rag_store=memory_store))]
),
) as session:
```

Gemini 2.5 Flash Live API ネイティブ音声機能
強化された音声品質

24+ の言語、30 種類の HD 音声を備え、より豊かで自然な音声インタラクションを実現

24+ 言語・地域

英語（米国、オーストラリア、英国、インド）
スペイン語（米国、スペイン）
ドイツ語（ドイツ）
フランス語（フランス、カナダ）
ヒンディー語（インド）
ポルトガル語（ブラジル）
アラビア語（汎用）
インドネシア語（インドネシア）
イタリア語（イタリア）
日本語（日本）
トルコ語（トルコ）
ベトナム語（ベトナム）
ベンガル語（インド）
グジャラート語（インド）
カンナダ語（インド）
マラヤーラム語（インド）
マラーティー語（インド）
タミル語（インド）
テルグ語（インド）
オランダ語（地域指定なし）
韓国語（韓国）
中国語（北京語）（中国）
ポーランド語（ポーランド）
ロシア語（ロシア）
スワヒリ語（ケニア）
タイ語（タイ）
ウルドゥー語（インド）
ウクライナ語（ウクライナ）

30 種類の HD 音声

男性
Puck
Charon
Fenrir
Orus
Achird
Algenib
Algieba
Alnilam
Enceladus
Iapetus
Rasalgethi
Sadachbia
Sadaltager
Schedar
Umbriel
Zubenelgenubi

女性
Aoede
Kore
Leda
Zephyr
Autonoe
Callirrhoe
Despina
Erinome
Gacrux
Laomedeia
Pulcherrima
Sulafat
Vindemiatrix
Achernar

Gemini 2.5 Flash Live API ネイティブ音声機能
感情認識対話（Affective Dialog）

ユーザーの感情やニュアンスを理解して応答を自動的に変更

同じ言葉でも話し方が異なれば、Gemini の回答が大きく変わる

（平坦なイントネーション）「この前、試験に合格しました！」

→特に平凡なイントネーションで、祝福の応答が返される

（喜びを込めて興奮気味に）「この前、試験に合格しました！」

→ユーザーの喜びを共有し、祝福するような肯定的なトーンの応答が返される

モデルが単に情報を伝えるだけでなく、感情的なニュアンスを伴って応答し、より自然で人間らしい双方向の音声会話を実現できる

ファッション アドバイザー AI デモ
(Gemini 2.5 Flash with Live API native audio)

GitHub ソース:
GoogleCloudPlatform/generative-ai/gemini/multimodal-live-api/websocket-demo-app

動画あり
アーカイブ動画をご視聴ください

03. 開発者ガイド

Live API アプリ開発用 SDK

特徴	ADK (Agent Development Kit)	GenAI SDK (Generative AI SDK)
主な目的	AI エージェントおよびマルチエージェントシステムの構築	Gemini API の機能を直接利用するための汎用ライブラリ
得意なこと	複数の AI エージェントの連携、 複雑なタスクの自動化、 ワークフローの管理	Gemini の各機能（テキスト、音声、画像など）を柔軟にアプリケーションへ組み込むこと
リアルタイム対話	双方向の音声・映像ストリーミング機能がネイティブで組み込まれている	Live API をサポートしており、ストリーミング通信を実装可能
主なユースケース	- 自律型 AI アシスタント - 複雑な問題解決を行うエージェント	- Web サイトやアプリへのチャットボット搭載 - 既存システムへの AI 機能の統合
開発の自由度	フレームワークの作法に従う必要がある	非常に高い

GenAI SDK

Gemini Developer API と Vertex AI Gemini API を統合した SDK。

6/24 より Vertex AI SDK の生成 AI モジュールは非推奨になりました。

- まずは Python, Go, Java, JavaScript から対応
  - pip install google-generativeai
  - go get google.golang.org/genai
  - import com.google.genai.Client;
  - npm install @google/genai
- **Gemini Live API にすでに対応済**

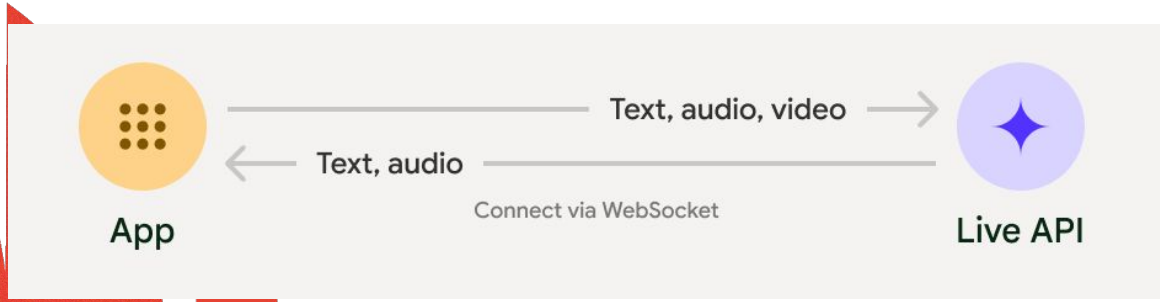
WebSockets 接続

WebSockets は永続的な接続を介して**双方向通信**を行うための通信規格

Live API は、WebSockets を使用するステートフル API

音声、動画、テキストの連続ストリーミングを処理して、人間のような即時音声レスポンスを提供することで、ユーザーに自然な会話エクスペリエンスを提供します

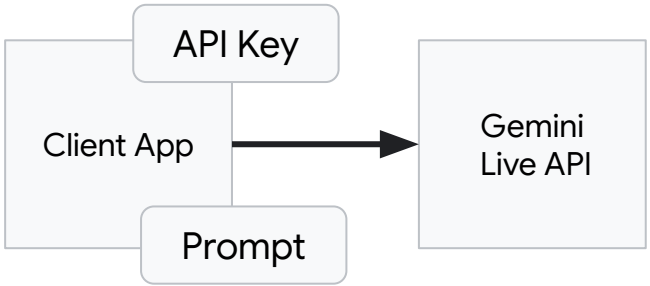
前述の **Gen AI SDK は WebSockets 接続の確立、管理、切断等の複雑な処理をカプセル化し開発者が本質的なロジックの記述に集中できるようにしてくれるため、利用を推奨** します。



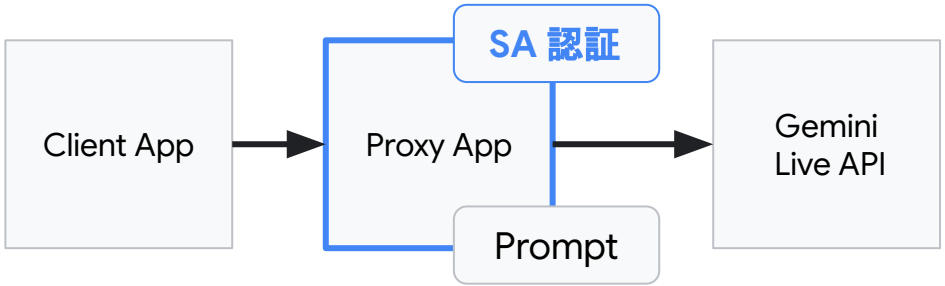
認証

クライアント サイドに直接キーを埋め込むことは、キー漏洩リスクを伴います

エンタープライズ環境では、サービス アカウントとプロキシ サーバーを介した認証を強く推奨。IAM による厳格な権限管理が可能になります。

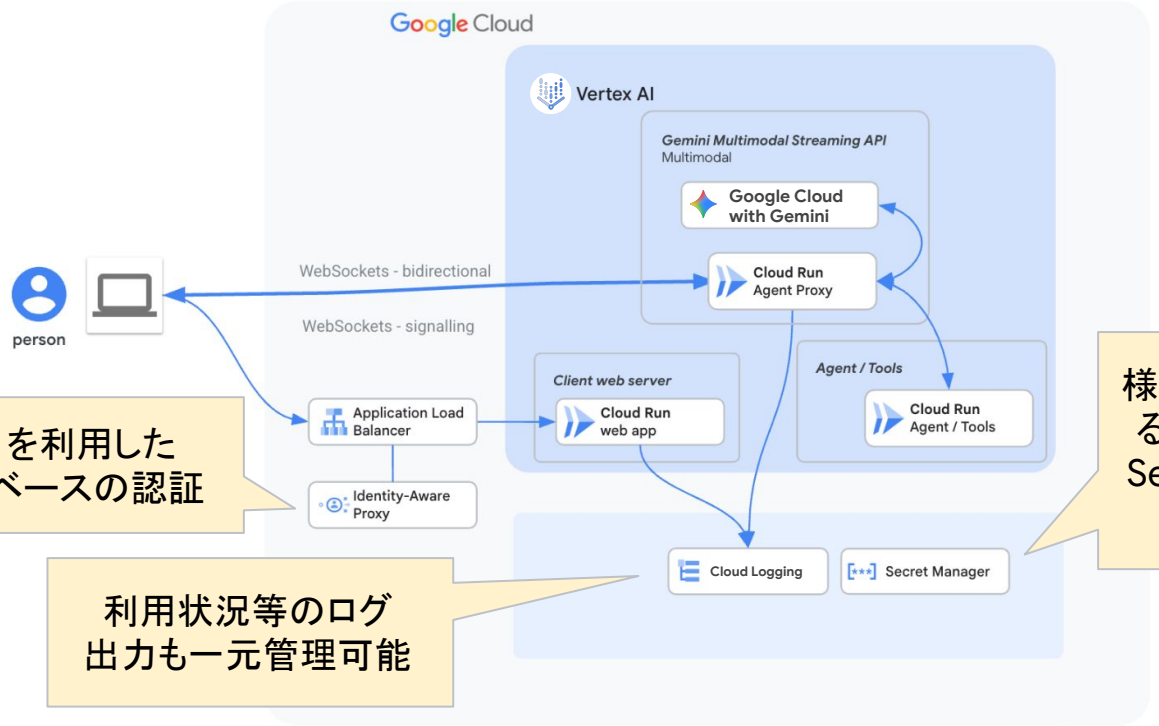


Live API の**直接**利用
（非推奨）



Live API の**間接**利用
（推奨）

機能拡張性も考慮した実装アーキテクチャ



IAP を利用した
IAM ベースの認証

利用状況等のログ
出力も一元管理可能

様々なツールと連携する場合の認証情報を Secret Manager で安全に管理可能

Tool による機能拡張（再掲）

組み込みツール機能を活用することで、企業は単なる自動応答システムを超えた高度な実務能力を持つリアルタイム音声ボットを構築できます。以下例。

関数呼び出し（Function Calling）

ユーザーからの予約問い合わせに対し、予約代理 AI が予約システムを呼び出して最新の空き状況を確認し、予約を即座に完了 させる

Google 検索によるグラウンディング

「来月のハワイの天気と、何か特別なイベントはあるか？」という質問に、旅行代理店 AI が最新の天気予報やイベント情報を探して回答する

Python コード実行

「毎月 3 万円を年利 3% で 20 年間積み立てると、最終的にいくらになりますか？」という質問に資産助言 AI が複利計算の Python スクリプトを実行し、正確なシミュレーション結果を返す

Vertex AI RAG Engine によるグラウンディング

従業員の「新しい経費精算システムの申請方法を教えて」の質問に、社内ヘルプデスク AI が社内マニュアルを検索し、社内ルールに準拠した正確な手順を案内する

モデルの会話スタイルの調整

① **システム指示 (System Instruction)** を指定することで、音声レスポンスのトーンや感情を指定でき、より自然な会話を演出できます。また「できるだけ簡潔に答えてください」「専門用語を使わないでください」等を指示することで、モデルの会話スタイルも調整可能

② **感情認識対話 (Affective Dialog)** を使用し、モデルがユーザーの感情を検出し、モデルに感情に応じた応答を可能にすることで、より共感的で自然な対話を実現

③ **プロアクティブ音声 (Proactive Audio)** を有効にし、ユーザーが意図しない発言に対してあえて応答しないように設定できます。不必要な返答を避け、より洗練された自然な対話を提供可能

④ **音声のカスタマイズ**し、多様な音声と多言語で最適なモデル像を提供

⑤ **temperature, topP, topK** のパラメータを調整し、より創造的・論理的な出力を生成し、自然な会話に貢献できます。**maxOutputTokens** により、生成される応答の最大トークン数を設定し、不必要に長い、または途切れた応答を防ぐことも可能です。

04. プロトタイプから本番への勘所

パフォーマンスとコストの最適化

トークン使用量の監視

トークン使用量を継続的に監視・分析することで、アプリのどの部分が最もコストを消費しているかを正確に把握し、最適化する

出力トークンの制御

最大出力トークン数(maxOutputTokens)を設定することで、モデルの応答長を制限し、出力トークンのコストを直接調整できます

メディア解像度の設定

入力映像ストリームの解像度を調整し、トークン消費量とコスト、そして物体認識精度の最適なバランスを、要件に応じて都度調整する。

料金体系の深い理解

使用するモデル、モダリティ(テキスト, 音声, 映像)、そしてコンテキストの長さによってコストは変動する。事業計画上、コスト構造の理解は極めて重要。

パフォーマンスとコストの最適化

Provisioned Throughput (Gemini 2.5 Flash Live API のみ提供)

固定料金のサブスクリプションを使用して、Live API の利用容量（スループット）を確保することで、安定的なサービス利用の実現と固定費用化が可能になります

保証された スループット

重要なワークロードをカバーするスループットのために PTを注文します。

サービスの可用性

可用性 SLA に基づいて、指定されたスループットまでの予約容量を利用できます。

固定費用

毎月の支払い額を予定しておくことができ、超過料金を完全に管理できます。

信頼性と堅牢性の確保

エラー ハンドリング

API タイムアウト、ネットワーク接続の中断、API からのエラーレスポンスなど、発生しうる様々なエラーシナリオを想定し、それぞれに対する堅牢なハンドリング ロジックを実装する

セッション再開機能

セッション再開機能は、信頼性確保の要

ネットワークが一時的に切断されても、ユーザーが会話を最初からやり直す必要がないように、この機能を積極的に活用した再接続ロジックを組み込む（デフォルト無効）

指数バックオフ

API のレートリミットに達した場合に備えてリクエストの間隔を指数関数的に増加させながらリトライを行う「指数バックオフ」戦略を実装し、API への過剰な負荷を避け、安定したサービス復旧を目指す

AI 倫理とプライバシー

データ プライバシーと同意

顧客の音声データなどの個人情報を扱う場合、関連法規を遵守し、データの収集・利用目的をユーザーに明示した上で、明確な同意を得ることが絶対条件

バイアス軽減

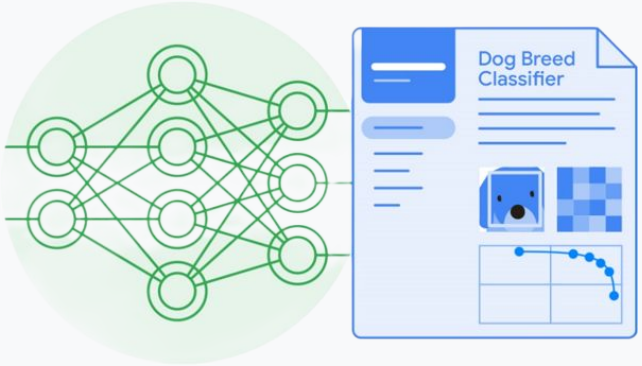
AI の分析対象を、顔認識やプロファイリングといった「個人」の特定ではなく、不正アクセスといった「行動」の検知に利用する

重要な意思決定は、必ず人間のオペレーターによる確認と判断を介在させる

透明性

モデルカード のようなツールを用いて、モデルの能力と限界を文書化し、公開しましょう。

また、ユーザーに対して、対話している相手が AI であることを明確に伝えることも、透明性確保の観点で非常に重要



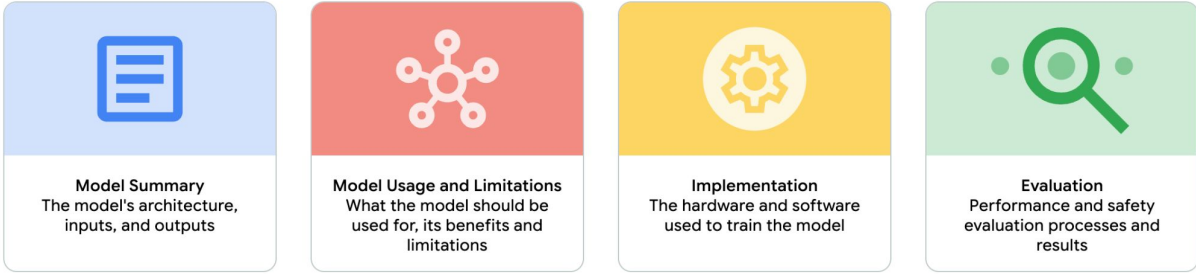
モデルカード (Model Cards)

AI モデルの透明性と説明責任を高めるために利用される、
AI モデルの「栄養成分表示ラベル」のようなもの

サービス提供者は、**自社の AI サービスがどのようなモデルを基にしており、どのような特徴や限界があるのかをユーザーや関係者に示すことで、信頼性を向上させ、責任ある AI の利用を促進することが可能**

AI サービス提供者の記載例

- モデルの概要
- 学習データ
- 性能評価
- 制限事項と潜在的なリスク
- 倫理的考慮事項
- 使用方法と推奨事項



05. まとめ

まとめ

- Gemini Live API は、単なる応答から自然な「会話」へと AI との関係を再定義する技術です
- 顧客体験の向上、業務効率化、そして全く新しいサービスの創出といった価値をもたらします
- Gemini Live API は、リアルタイムのマルチモーダルな入出力とツール連携で開発を支援します
- 本番環境では**パフォーマンスやコストの最適化、信頼性、倫理への配慮が不可欠** です

Gemini Live API を活用して、新しい顧客体験の実現を