

# 視聴覚 AI がもたらすビジネス変革

～人間と AI が自然に会話できる時代の到来～

ライフデザイン研究部 主席研究員/テクノロジーリサーチャー 柏村 祐

## 1.目と耳をもつAIとは

視聴覚 AI とは、人間のように視覚と聴覚を備えた人工知能である。これまでの AI は主にテキストデータを処理することに特化していたが、視聴覚 AI は画像や音声データも理解し、より人間に近い感覚で情報を処理することができる。たとえば、OpenAI が新たに発表した言語モデルは、テキストだけでなく画像や音声も受け入れることができ、より自然な対話を可能にしている。

この新しい言語モデル（通称 GPT-4o）は、テキスト、音声、画像、ビデオのあらゆる組み合わせを入力として受け入れ、それらを組み合わせて生成し出力することができる。音声入力に対しては、平均応答時間がわずか 320 ミリ秒と人間の会話スピードとほぼ同等の速さで応答でき、人間との自然なコミュニケーションを実現している。

こうした視聴覚 AI の登場により、これまで人間にしかできなかった様々なタスクを自動化できるようになり、ビジネスに大きな変革をもたらすことが期待されている。

本レポートでは、視聴覚 AI の最新技術動向を確認するとともに、それがビジネスにもたらす変革の可能性について展望する。

## 2.視聴覚AIの実態を示す 16 動画の概要

本節では、視聴覚AIが実際にどのような能力をもっているのかを示すため、OpenAI が作成し、その公式ホームページ上に提示されている 16 本の動画を事例として紹介する（注 1）。これらの動画は、最新の視聴覚AI技術がどのように機能するのかを視覚的かつ聴覚的に体験できるものであり、2024 年現在の技術の進展とその実用例を確認することができる。動画には、音声認識から画像解析、さらには複合的な情報処理を行う事例まで幅広く網羅されており、視聴覚AIの多様な応用可能性を感じ取れる。特に、動画内で示される技術の具体例は、ビジネスにおける活用方法を模索している企業や研究者にとって非常に参考になるだろう（図表 1）。

図表 1 視聴覚 AI の活用を紹介する動画と内容

動画タイトル	動画内容
2つのGPT-4oの対話と歌唱	2つのGPT-4oが対話しながら歌う様子。AIが自然に会話を交わし、同時に歌唱パフォーマンスを行うことができることを示しています。
インタビュー準備	インタビュー準備のための対話。AIがユーザーの受け答えをサポートし、インタビューの準備を円滑に進める様子を紹介。
皮肉の理解	皮肉を理解し返答する様子。AIが複雑な言語表現、特に皮肉を理解し適切に応答する能力を示しています。
サルとイムラン・カーンによる数学の議論	サル・カーンとイムラン・カーンによる数学の議論。AIが高度な数学的議論をサポートし、学習支援を行う例。
リアルタイム翻訳	リアルタイムの音声翻訳。AIが複数の言語をリアルタイムで翻訳し、異なる言語を話す人々のコミュニケーションを助けます。
子守唄	AIが音声合成を用いて優しい子守唄を歌い、リラックスした環境を提供する様子。
速い話し方への対応	AIが速い話し方にも対応し、正確に理解し応答する能力を紹介。
指さしで学ぶスペイン語	画像を指し示してスペイン語を学ぶ方法。ユーザーが画像を指し示すだけで、その画像に関連するスペイン語を学ぶことができる。
ジャンケン	AIがジャンケンの手の形を認識し、対戦相手としてゲームを進行。
誕生日祝い	AIが画像を用いて誕生日祝いのシーンを生成し、パーソナライズされたお祝いメッセージを提供。
会議 AI	会議の音声と映像を統合して処理するシステム。会議の内容を自動的に記録し、重要なポイントを抽出する。
犬	AIが犬の鳴き声や動きを認識し、適切な応答を行う。
ダジャレ	AIがジョークを理解し、それに対する適切なリアクションを返す。
BeMyEyesとの連携	視覚障害者支援アプリ「BeMyEyes」との連携。視覚障害者が必要とする視覚情報をAIが提供する。
顧客サービスの概念実証	AIが複数のチャネルを通じて顧客サポートを行う。
2つのGPT-4oのハーモニー	複数のAIが協力して音楽を作り出す能力を示す。

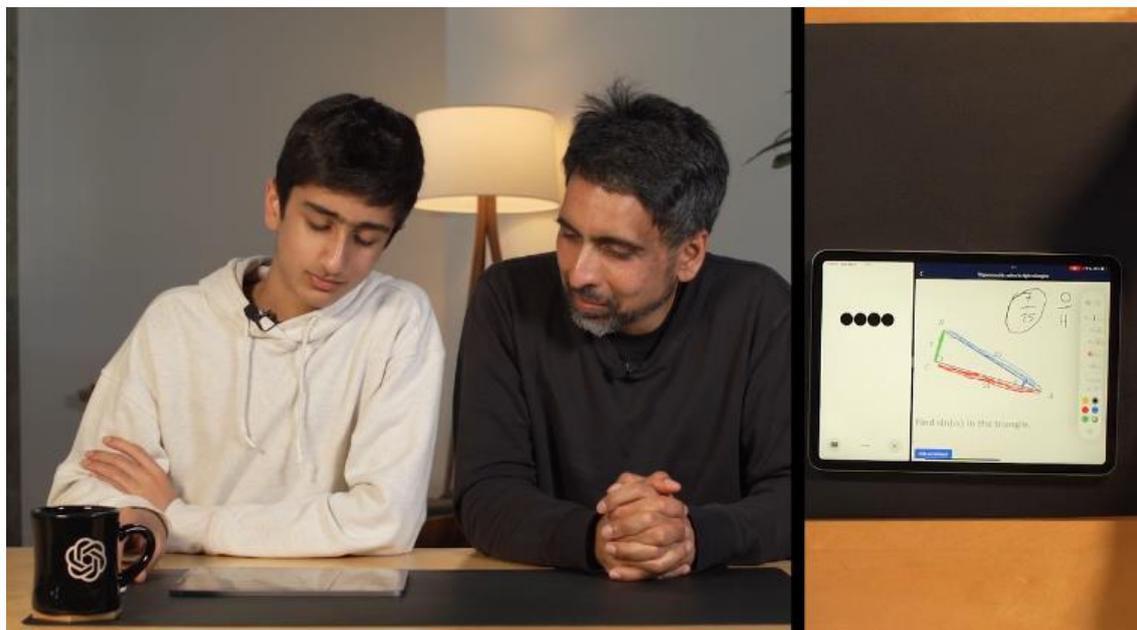
資料: OpenAI「<https://openai.com/index/hello-gpt-4o/>」の動画をもとに筆者作成

### 3.動画の詳細をしてみる

視聴覚AIの能力を具体的に理解するために、本節では3つの注目すべき動画を取り上げる。これらの動画は、AIが日常生活やビジネスシーンでどのように活用されているかを具体的に示している。

まず、「サルとイムラン・カーンによる数学の議論」では、父親であるサルが息子イムランに数学の問題を解かせるため、AI家庭教師を利用する様子が描かれている。AI家庭教師はイムランに適切な質問を投げかけ、ヒントを与えながら彼自身で答えを導き出すよう促す。イムランはAIの助けを借りて、三角形の各辺の名称（対辺、隣辺、斜辺）を理解し、サインの公式を使って問題を解決することに成功する（図表2）。

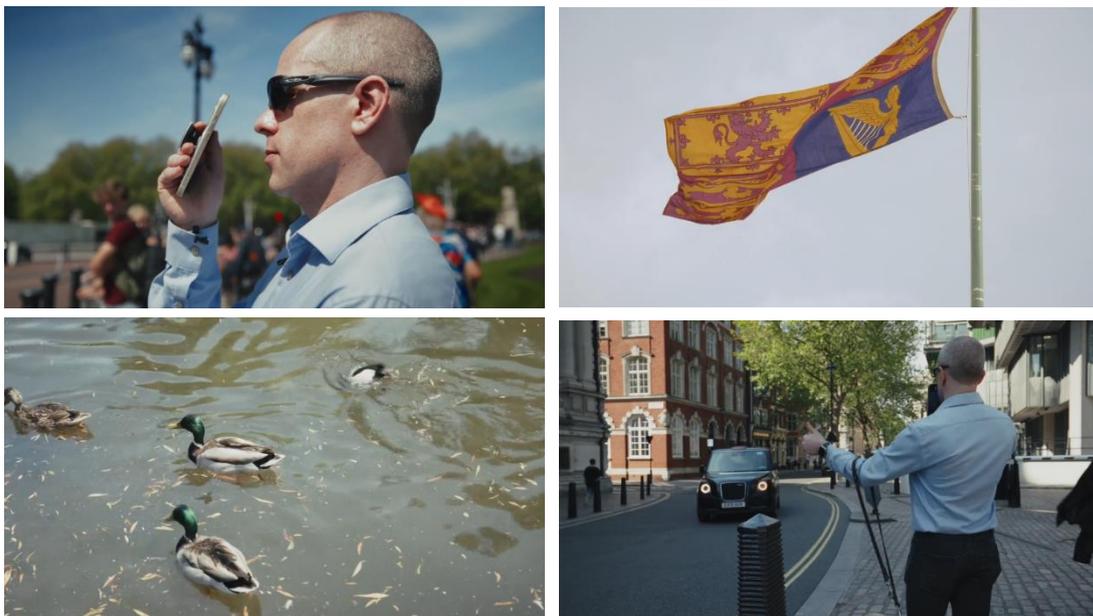
図表 2 AI 家庭教師と会話しながら問題を解いていく様子



資料: OpenAI HP より「<https://openai.com/index/hello-gpt-4o/>」

次に、「BeMyEyesとの連携」では、視覚障害者が視覚障害者支援アプリ「BeMyEyes」を通じて、AIガイドから視覚情報を提供してもらう様子が紹介される。バッキンガム宮殿の上空に掲げられた王室旗から国王の在宮を説明し、宮殿前の池でアヒルが泳ぐ様子をリアルタイムで音声によって伝える。また、男性がタクシーを探す場面では、AIガイドがオレンジ色のライトを点灯したタクシーを発見し、男性にタクシーを止めるよう指示して無事に乗車できるよう支援する（図表3）。

図表 3 視覚障害者が AI ガイドと会話しながら行動する様子



資料: OpenAI HP より「<https://openai.com/index/hello-gpt-4o/>」

最後に、「顧客サービスの概念実証」では、男性がAIアシスタントに不具合のある新しいiPhoneの交換手続きを依頼するシーンが描かれる。AIアシスタントは男性の代わりに事業会社に電話をかけ、オペレーターに状況を説明する。オペレーターは注文番号を確認し、返品手順と配送ラベルを男性にメールで送ることを約束する。AIアシスタントはオペレーターに男性のメールアドレスを伝え、男性にメールが届いたことを確認する。この一連のプロセスを通じて、AIが顧客サービスの効率化に大いに役立つことが示されている（図表4）。

図表 4 AI アシスタントがオペレーターと話す様子



資料: OpenAI HP より「<https://openai.com/index/hello-gpt-4o/>」

#### 4.視聴覚AIがもたらすビジネスへの変革

以上の事例から明らかなように、視聴覚 AI は様々な場面で活用され、人々の生活やビジネスに大きな変革をもたらす可能性を秘めている。特に、AI 家庭教師の事例からは、AI が個々の学習者に適した指導を行うことで、教育の質を向上させられることがわかる。また、視覚障害者支援の事例からは、AI が障害者の日常生活をサポートし、社会参加を促進する役割を果たせることが示唆される。さらに、顧客サービスの事例では、AI が人間に代わって顧客サポートを行うことで、業務の効率化と顧客満足度の向上を同時に実現する可能性が明らかになった。

視聴覚 AI の活用が想定される分野は多岐にわたる。図表 5 は、主要な分野における視聴覚 AI の活用方法と将来展望をまとめたものである。

図表 5 視聴覚 AI の主要な活用分野と将来展望

活用分野	活用方法	将来展望
教育	- 個別指導AI家庭教師 - 学習進捗のリアルタイム分析	- 学習効率の向上 - 教育格差の縮小
医療	- リアルタイム診断サポート - 音声および画像データの解析	- 診断精度の向上 - 早期発見と治療
顧客サービス	- AIチャットボットによる 24 時間 365 日対応 - 音声認識による問い合わせ処理	- カスタマーサポートの効率化 - 顧客満足度の向上
障害者支援	- 視覚障害者向けのAIガイド - 聴覚障害者向けのリアルタイム字幕生成	- 障害者の社会参加促進 - 日常生活の質の向上
マーケティング	- 顧客の声や行動の分析 - 個別化されたマーケティング戦略の構築	- マーケティングの精度向上 - 顧客体験の向上
セキュリティ	- 画像解析による不審者の検知 - 音声認識による脅威の識別	- 防犯対策の強化 - リアルタイムのリスク管理
エンターテインメント	- インタラクティブなコンテンツの提供 - 音声合成によるキャラクターの創出	- ユーザー体験の向上 - 新しいコンテンツの創造

資料:筆者作成

これらの動画から明らかなのは、視聴覚 AI が単なる技術的なブレイクスルーではなく、社会のあり方そのものを変える力をもっているということである。教育、医療、顧客サービスなど、あらゆる分野で AI が人間の能力を拡張し、新しい価値を生み出す可能性がある。特に、AI が人間の感覚や感性を理解し、きめ細かなサポートを提供できるようになったことは、画期的な進歩だといえる。

ただし、AI の活用にはリスクも伴う。AI が人間の仕事を奪うのではないかという懸念や、AI に依存しすぎることによって人間の能力が低下するのではないかという危惧もある。また、AI による差別や偏見の助長、プライバシーの侵害など、倫理的な課題にも注意が必要である。

したがって、視聴覚 AI を活用するには、技術的な側面だけでなく、社会的・倫理的な側面にも十分に配慮しなければならない。AI と人間がうまく協働し、お互いの長所を生かしながら、持続可能な形でビジネスを変革していくことが求められる。企業は、AI を単なるコスト削減の手段ではなく、イノベーションを創出するための戦略的なツールとして位置づけ、積極的に活用していくべきである。

同時に、AI と向き合うための環境整備として、以下のような取り組みが必要である。まず、教育現場において AI リテラシーを向上させるためのカリキュラムを導入し、学生が AI の基礎知識や倫理的な考え方を身につけられるようにすることである。次に、企業内での AI 活用に関するガイドラインを設定し、社員が AI を適切かつ効果的に利用できる環境を整備することである。さらに、政府には AI に関する法整備を進め、プライバシーや安全性に配慮しながら、AI の社会実装を後押しする施策を打ち出すこと

が求められる。加えて、AIの研究開発に従事する機関は、倫理的な課題について継続的に議論し、責任ある形でAI技術を発展させていく必要がある。

社会全体で、ビジネスにおけるAIのあるべき姿を模索し、新しい時代に適応していくことが肝要である。視聴覚AIは、人間がAIと真剣に向き合うための重要なきっかけとなるだろう。AIの登場は、人間の能力を拡張し、新たな可能性を切り開く一方で、私たち自身のあり方を根本から問い直す契機にもなる。AIとの共生を通じて、人間は自らの役割や価値観を見直し、より豊かな社会を築いていくことができるだろう。

**【注釈】**

1) OpenAI HP より

<https://openai.com/index/hello-gpt-4o/>