

デジタルヒューマン(マルチモーダル技術)が社会にもたらす影響

The impact of digital humans (multimodal technology) on society.

デジタルヒューマン協議会

デジタルヒューマン協議会の参加企業のうち以下のメンバーで執筆した。

デジタルヒューマン株式会社 代表取締役 荒尾和宏

デジタルヒューマン株式会社 制作部 デジタルヒューマン アーティスト 柳島秀行

日本電気株式会社 ビジネスイノベーション統括部 主任 金子瑛士

日本電気株式会社 ビジネスイノベーション統括部 上席プロフェッショナル 兼保圭介

1 はじめに

本稿は、デジタルヒューマン協議会が考えるマルチモーダル技術が可能にするデジタルヒューマンに焦点を当て、その社会実装例と知財・プライバシー・UX（ユーザー体験）上の課題および今後への期待について論じる。

背景として、近年のマルチモーダル技術の発展が著しいことがあげられる。マルチモーダル技術とは、テキスト、音声、画像、動画など複数のモダリティを統合して処理する技術を指す。ChatGPT に代表される生成 AI は本来テキスト生成を得意とするが、画像生成 AI の登場や、テキストと同時に画像・音声をモデルに学習させることによって、テキストから画像・音声を作り出す、あるいは画像を読み取って説明するなど、モダリティ（情報の表現形式や伝達手段）を跨いだ情報処理が可能になりつつある。

マルチモーダル技術と生成 AI の融合によって生まれた象徴的な技術としてデジタルヒューマンがある。デジタルヒューマンとは AI とコンピュータグラフィックスなどを組み合わせたデジタル空間上の人間に似た存在のことであり、高度な対話や人間らしい感情表現を可能とする。

デジタルヒューマン協議会では、音声・画像・センサーデータ・テキストといった複数情報源を駆使し、画像や音声出力により人とやりとりできる点で、デジタルヒューマンをマルチモーダル技術の最先端の活用例と考えている。同時に、デジタルヒューマンの発展は、新たな利便性や価値を生みだし、我々の社会やビジネスに大きな変革をもたらすという視点に立つ。また、今後起こ

る社会的変化として、知的財産権、プライバシー、ユーザー体験上の課題があると想定する。

次章以降では「デジタルヒューマン（マルチモーダル技術）が社会にもたらす影響」というテーマのもと、デジタルヒューマンが各種分野に与えるインパクトにも触れながら、最新動向を概観したい。

2 デジタルヒューマンとは

デジタルヒューマンとは、AI（人工知能）と CG（コンピュータグラフィックス）技術を組み合わせて作られた、人間のように振る舞う 3 次元モデルである。単なる静的な CG キャラクターではなく、AI により人間さながらの自然な対話や感情表現を行える点が特徴だ。近年は画像生成 AI を用いて静止画からフォトリアルなキャラクター動画を生成し、リアルタイムに動かすタイプも登場しており、実在の人物と見分けがつかない水準に達しつつある。言い換えれば、デジタルヒューマンは見た目や動作が人間に近い対話エージェントであり、利用者とのインタラクション（対話・コミュニケーション）をリアルタイムに行うことができる。

デジタルヒューマンを実現するためには様々な技術要素が必要となる。主な要素技術として、以下のようなものが挙げられる。

- 3DCG モデリングか画像生成 AI（写実的な外見を作成）
- レンダリング / 動画生成（表情や口パク、身振りなどの動きをリアルタイム描画）

デジタルヒューマンの技術マップ

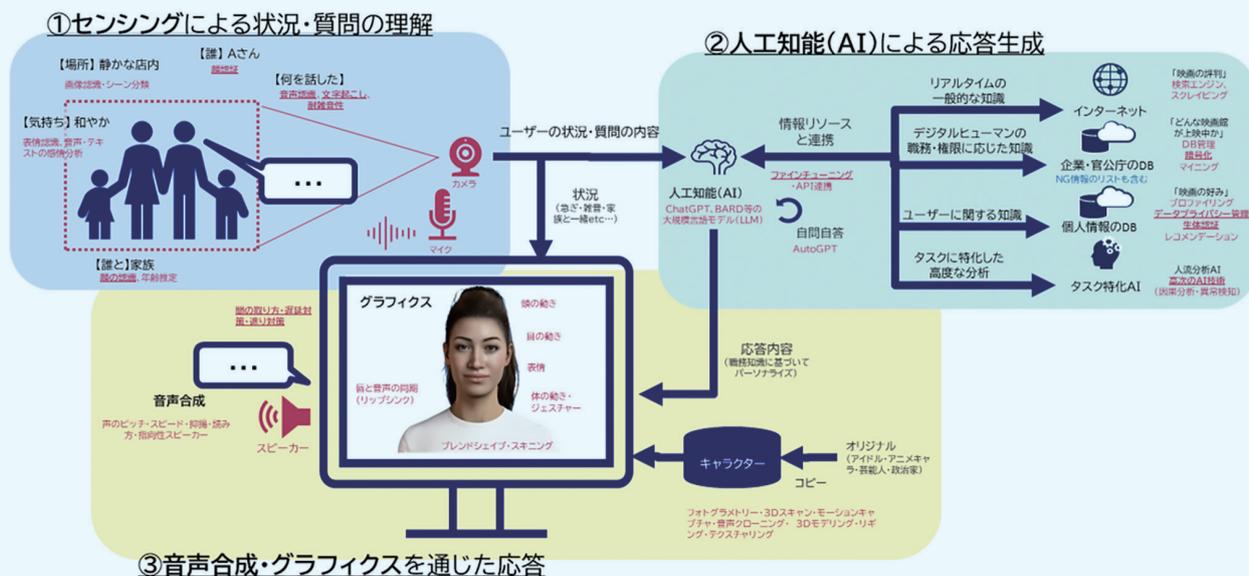


図1 デジタルヒューマンの基本的なシステム構成 出展：日本電気株式会社

- 音声合成（自然で人間らしい発声を生成）
- 音声認識・画像認識（マイクやカメラを通じてユーザの発言内容や表情・ジェスチャーを解析）
- 自然言語処理（ユーザの発言意図を理解し、適切な応答内容をテキスト生成）
- 対話管理（文脈を維持し、会話の流れを制御）
- 外部システム連携 など

複数の先端技術を組み合わせることで、利用者の表情や感情を読み取り、状況に応じた自然な応答や振る舞いを実現する。それゆえデジタルヒューマンの開発には相応のコストと高度な技術力が必要だが、一方で24時間365日の稼働や多言語対応、人手不足の補完といったメリットも期待され、近年各企業で研究開発や導入が進んでいる。

図1は、デジタルヒューマンの基本的なシステム構成を示したものである。

ユーザ側から音声入力やカメラ映像を受け取り、音声認識や表情分析によってユーザの状態・意図や感情を読み取る。その情報をもとにAIが最適な応答内容（テキストや動作指示）を生成し、音声合成やアニメーションによって視覚・聴覚両面へ出力する。このようにマルチモーダル（音声×視覚×言語）インタラクションを可能

にするのがデジタルヒューマンの強みである。従来のテキストチャットボットではどうしても機械的・画一的になりがちだった対話も、デジタルヒューマンであればあなたかも「そこに人がいる」かのような親しみやすいコミュニケーションが実現できる。声色や表情を相手や状況に合わせて変化させる豊かな感情表現が可能であり、ユーザ体験の質を大きく向上させるポイントである。

なお、類似の概念に「バーチャルヒューマン」「AIアバター」などがある。基本的にはデジタルヒューマンと同義で使われることも多いが、SNSなどで一方的な情報発信のみ行うCGや画像生成AIキャラクター（例：バーチャルモデルやバーチャルインフルエンサー）と、AIによる双方向対話が可能でデジタルヒューマンは区別される場合がある。本稿で扱うデジタルヒューマンは後者の、対話型を指すものとする。

3 デジタルヒューマンの社会実装例

デジタルヒューマンは様々な分野で実証実験やサービス実装が始まっている。デジタルヒューマン協議会が公開した『デジタルヒューマン白書 2023』では、社会実装のユースケースとして9つの類型が例示されている。ここではその中から代表的ないくつかを紹介する。

A. 「おもてなしの最適化」- 人間の代わりに顧客対

応を行い、高度な接客サービスを提供するデジタルヒューマンである。コールセンターや銀行ATM、ホテルの受付などで活用が想定されている。相手の表情や声のトーン（感情）に応じて対話スクリプトを柔軟に変更し、きめ細かな対応が可能となる。実際の例として、ドラッグストアの店頭に設置され、顧客からの質問に対して最適な商品の提案を行ったり、ホテルにおいて訪日客向けにデジタルヒューマンの観光コンシェルジュを試験導入する取り組みも行われた。ホテルの案内デスクに設置されたデジタルヒューマンが旅行者と対話しながら観光情報を案内し、24時間365日・多言語対応の接客を実現している。対話の中で顧客の趣味嗜好を聞き取り、最適な観光プランを提案するといった高度なサービスの提供も検討されており、ホスピタリティ向上への効果が期待されている。

B. 「自分の分身」 - ユーザ本人の代理人として活動するデジタルヒューマンである。例えば、忙しいビジネスパーソンに代わって会議に出席し、あらかじめ設定したユーザの意見や意思を代弁して発言する、といった使い方が考えられる。会議後には議論の内容や相手の反応をレポートしてくれるため、物理的にその場になくとも自分の「分身」が場をこなしてくれるイメージである。現時点では実証例が少ないものの、将来的には個人の知識や思考パターンを学習したAIエージェントが自律的に対話できるようになり、オンラインミーティングやイベントで本人の代理を務めるといった応用も期待される。

D. 「反復練習台」 - 対話による練習相手となってくれるデジタルヒューマンである。人間相手だと何度も同じ練習に付き合ってもらうのは難しいが、デジタルヒューマンなら嫌がることなく何度でも付き合ってくれるのが利点だ。例えば、語学学習者が会話相手としてAI講師のデジタルヒューマンを使い、繰り返し英会話の訓練を行うといったケースである。また、就職活動中の学生が模擬面接官のデジタルヒューマンを相手に練習するサービスも登場している。ある企業のAI面接ツールでは、面接官役のデジタルヒューマンが24時間いつでも応募者と会話し、模擬面接の映像や文字起こし、評価レポートまで自動生成してくれる。このように、時間や相手の都合に

縛られず自主トレーニングができる環境は、教育・人材育成の分野で大きな価値を持つだろう。

他にも、「息の合う相方」（人間のパートナーとして即応的に情報提供し場を盛り上げる相棒役）や「秘密の話し相手」（ユーザのプライベートな悩み相談に乗る聞き役）、さらには「デジタルタレント」に至るまで、幅広い可能性が示唆されている。これらはいずれも現実の人間が担ってきた役割をデジタル技術が補完・代替し、新たな価値を生み出すシナリオである。

4 社会実装における課題、効果、期待

デジタルヒューマンの社会実装を進めるにあたり、乗り越えるべき課題がいくつか存在するが、適切に実装できれば社会にもたらす効果は大きく、将来的な期待も非常に高い。本章では課題・効果・期待のそれぞれについて述べる。

まず課題であるが、主に知的財産権、プライバシー、ユーザー体験（UX）の3つがあげられる。

知的財産権の観点では、デジタルヒューマンは生成AIを用いて会話とコンテンツ（音声応答や画像表現等）を生み出すため、著作権や肖像権などの法的問題が顕在化する可能性がある。例えば、生成AIによってキャラクターを生み出した場合、そのキャラクターが特定の俳優や有名人、既存のキャラクターに酷似した容姿・声であれば肖像権やパブリシティ権の侵害が疑われる。

また、応答内容に第三者の著作物に由来するフレーズや知識が含まれていた場合、その二次利用の適法性が問われる。現状では世界的にも法制度の整備が追いついておらず、各企業・団体がガイドライン策定や許諾取得など対応策を模索している段階である。生成AIに「AIで作成」と明示する欧州のような取り組みを代表例として、知財保護と利活用のバランスが検討されている。

一方、生成AIで生み出したキャラクターをオリジナルとして保護する場合、どの程度の人間の創作意図、創作的寄与があれば、著作物として認められるのかについても今もなお議論が続いている。

プライバシーの観点では、デジタルヒューマンの高度な対話には、しばしばユーザに関する個人データの活用が伴う。ユーザの顔映像や音声から感情・属性を推測する機能、あるいはユーザの購買履歴・嗜好をデータベー

スから引き出してパーソナライズ応答する機能は、ユーザに関する深い情報を扱うことになる。このため個人情報保護の観点からは、データの安全管理や目的限定、本人同意の取得など慎重な対応が必要だ。特に、デジタルヒューマンにプライバシー情報を学習させる場合、アクセス権限の厳格管理（認証）やデータの暗号化に加え、機密情報を誤って応答に含めないための制御（いわゆるNGワードリストの実装など）も考慮すべきだと指摘されている。一方で興味深いことに、ユーザ側の意識は二極化しているとの報告もある。すなわち「個人情報はできればAIには預けたくない」という人がいる一方、「人間だと何をするか分からないので、デジタルヒューマンに個人情報を持たせたい」という人もいるという。このように、デジタルヒューマンに対する信頼感やプライバシー許容度はユーザによって異なるため、サービス提供側は透明性を高め選択肢を用意する（利用者がデータ提供範囲を選べる等）といった配慮が求められる。

ユーザー体験（UX）では、デジタルヒューマンとの対話体験をユーザが受け入れ、満足するかどうかも重要な論点である。不気味の谷現象に代表されるように、人間に似せたCGには微妙な違和感がつきまといやすく、特に写実性が中途半端だとかえってユーザに不快感を与える恐れがある。そのため、デザイン面の工夫（あえて非現実的なキャラクター調にする、逆に極限までフォトリアルに近づける等）によって違和感を低減する取り組みがなされている。また対話の質もUXを左右する。現時点のAIは完璧ではなく、聞き間違いや文脈を取り違える場合もあるため、スムーズなエラー修復やオペレータへのエスカレーションなどバックアップ体制を整える必要がある。さらに、デジタルヒューマンが人に代わって対応する際にはそのことを明示する（相手がAIであると利用者に知らせる）ことも信頼性の観点で重要である。適切に設計・運用されたデジタルヒューマンであれば、ユーザは心理的安全性を感じつつ対話に集中でき、人間相手以上に本音を引き出せるケースすらある。「AI面接官の方が公平に判断してくれ、緊張せずに臨める」と感じる応募者が多いとの声もある。このようにUX上の課題をクリアすれば、人との対話では得られない独自の安心感や利便性を提供できるのもデジタルヒューマンの魅力と言える。

次にデジタルヒューマンが社会にもたらす効果を述べる。特に大きな効果として期待されているものをあげると以下3つがあげられる。

- **業務効率化・コスト削減**：デジタルヒューマンは人件費の高い業務（受付業務や問い合わせ対応など）を部分的に代替し、24時間休みなく稼働できるため、企業にとって大きな効率改善が見込める。労働力人口の減少が進む中で、デジタルヒューマンは不足する人手を補う存在として期待されている。
- **サービス向上・顧客満足度の改善**：デジタルヒューマンは表情や声色を用いた親しみやすいコミュニケーションにより、機械的なチャットボットでは得られない温かなユーザ体験を提供する。さらに多言語対応が比較的容易であるため、ホテルや交通機関で外国語対応スタッフの代わりに活躍し、インバウンド対応力を高めることができる。また前述のように、AIならではの公平性・客観性によりユーザの緊張を和らげる効果も報告されている。
- **個別最適化・アクセスの民主化**：一人ひとりのユーザに合わせてパーソナライズされた対応や、専門知識へのアクセス容易化も効果の一つである。例えば、医療や法律といった専門領域の知識を持つデジタルヒューマン相談員がいれば、利用者は時間や場所を問わず専門的な助言を得ることができる。教育分野でも、各学生の理解度に応じて教え方を変えるAI家庭教師のように、個別最適化された学習支援が可能となる。知的財産情報の分野においても、多言語の特許文献や膨大な技術情報をAIが解析し、人に身振り手振り、音声、生成画像を使って分かりやすく説明してくれるデジタルアシスタントが登場すれば、専門家以外でも知財情報にアクセスしやすくなるだろう。

次に、デジタルヒューマンに寄せられる期待について述べる。まず市場の観点からは、デジタルヒューマン市場は2030年に約5,275億ドル（約70兆円）規模に達するとの期待がある。巨大な市場になれば、多くの事業者が切磋琢磨する市場となり、先にあげたリスクは限定され、反対に効果は高まると期待される。

特に接客・教育・医療・介護・行政など、人手不足が深刻な産業でデジタルヒューマンの導入が進めば、先進諸国が抱える労働人口減少の問題が緩和されるばかり

か、高度で均一化されたサービスを誰でも受けられることになり、人々の日常生活はより便利で豊かなものへ変わっていくに違いない。もちろん、その実現には課題克服も必要であるが、技術の進歩に合わせて社会側の受け入れ態勢や制度整備が進めば、人とAIが共生する新たな社会インフラとしてデジタルヒューマンが定着する未来が展望されている。

5 おわりに

マルチモーダル技術と生成AIの発展により、生身の人間に迫る高度な対話エージェントであるデジタルヒューマンの社会実装が進む。本稿ではデジタルヒューマンの定義やユースケース、直面する課題と将来展望について概観した。デジタルヒューマンは情報の流通や利活用の在り方にも少なからず影響を及ぼし、その普及にあたっては知的財産権やプライバシーへの配慮、ユーザ体験の設計が重要となる。一方で、適切なガバナンスのもと社会実装が進めば、人手不足の解消やサービス水準の向上、公平なアクセス機会の提供といった多方面のメリットが期待できる。

デジタルヒューマン協議会としても、産学官の知見を結集しながら望ましいデジタルヒューマン活用社会の実現を目指して活動を続けている。発足以来、本稿で述べた様々なユースケースの早期社会実装に向けた検討を進め、課題解決のための指針策定や実証実験の支援を行っている。幸い国内外で本領域に関心を持つ企業や研究者は増えており、協議会では今後も幅広い仲間との連携を通じてデジタルヒューマンの普及促進とエコシステム構築に寄与していきたいと考えている。

最後に、マルチモーダル技術がもたらす未来は人間とAIの新たな関係性を築く可能性に満ちていることを強調したい。知的財産情報の世界においても、AIが人の良きパートナーとして創造性や知識探求を支援する姿が現れつつある。デジタルヒューマンは有力な実現方法の一つであり、「人類の知的活動を拡張する存在」となり得るだろう。私たちは今、その未来の扉を開く途上にいる。デジタルヒューマンをはじめとするマルチモーダルAI技術を上手に活用し、より豊かで安心な社会を築いていくことを期待する。

参考文献

情報源（一部抜粋）：

- [1] デジタルヒューマン協議会 (2023), “デジタルヒューマン白書 2023”
- [2] デジタルヒューマン協議会 (2024), “デジタルヒューマンの望ましい活用法に関する意見書”
- [3] 文化庁 (2023), “AIと著作権”
- [4] Emergen Research, “Digital Human Avatar Market, By Type (Interactive, Non-Interactive), By Industry Vertical (BFSI, Retail, Healthcare, Automotive, IT & Telecommunication, Gaming & Entertainment, Others), and By Region Forecast to 2030”, (<https://www.emergenresearch.com/industry-report/digital-human-avatar-market>), (2025年6月15日最終閲覧)

マルチモーダル技術が知財情報にもたらす未来