ノア：メタバース空間で対話相手に向かって自律移動するAIアバターの開発と評価

小山田 悠人[1,a)]　湯村 翼[1,b)]

概要：メタバース空間で会話可能な AI アバター開発が進む．その多くは対話機能に限定され，アバターの動作を交えたインタラクションができない．そこで本研究では，ユーザーとの会話に加え，呼びかけに応じて自律移動を行う AI アバターを開発した．VRChat を対象プラットフォームとし，ユーザーが発する音声の方向をリアルタイムで算出し，音源へ向かって自律的に接近する機能を Python で実装した．ChatGPT などと連携した対話も行い，AI が能動的に振る舞う，存在感のあるインタラクティブな体験の創出を目指した．評価実験として大学生 9 名を対象に主観評価を実施したところ，会話の円滑さで平均 5.4 点，存在感において平均 5.8 点と肯定的な評価を得た．一方で，問答に答えているだけに感じたといった会話内容に関する課題も見られたが，ユーザーからは人間と話している感覚に近いといった高い双方向性を支持する意見も得られた．本研究は，AI が能動的に振る舞う身体的動作を対話に統合することで，メタバースにおける高い共在感とインタラクティブな体験を創出できることを明らかにした．

キーワード：バーチャルリアリティ，メタバース，AI アバター

1. はじめに

ChatGPT に代表される大規模言語モデル（Large Language Model，LLM）の急速な発展に伴い，AI との対話は身近な技術となった．この潮流はメタバース空間，特にソーシャル VR プラットフォームである VRChat や Cluster においても例外ではなく，キャラクター性を付与された AI と音声で対話するツールやアバターの開発がなされ，報告が行われている．例えば，Cluster 上では不登校生徒の学習支援を目的とした AI カウンセラー Hironya AI[1] が，また VRChat 上では音声認識エンジンを用いて自然な会話を行う AI Avatar AOI[2] が開発されている．ここ数年では，対話型 AI サービス [3][4] を通じて，ユーザーが AI アバターと継続的に会話し，感情的なつながりを築く事例も増えている．これらのサービスは，AI を単なる情報提供者ではなく，個性を持った“会話の相手”として体験させることで，新たなコミュニケーションやエンターテインメントの形を提示している．例えば，ユーザーが AI アバターに話しかけると，そのアバターが固有の声や性格で応答し，まるで仮想空間上で“会話している”ような体験が得られる．一部のシステムでは，AI アバターがユーザーの発言を記憶したり，感情的なリアクションを示したりすることで，まるで友人や仲間と交流しているかのような継続的関係が形成されつつある．このように，AI アバターは単なる情報応答装置ではなく，人格をもった他者として認識される傾向が強まりつつある．

しかし，既存の多くの AI アバターには共通の課題が存在する．その一つは，AI アバターの多くがその場に留まり，ユーザーからの呼びかけに応答するだけの静的な存在である点である．移動能力を持つアバターも存在するが，それらは事前に定められたルートを巡回するワールドギミックとして実装されている場合が多く，ユーザーとの対話に応じて動的に振る舞う機能は持たない．つまり，対話と移動という二つの要素が分断されているのが現状である．この分断により， ユーザーは AI アバターを空間内に存在する相手としてではなく， 画面越しの応答プログラムとして認識する．たとえ音声で会話が成立しても， アバターの接近や視線を向けるといった反応を示さないことで， 相互行為としての自然さが損なわれる．このため， 移動と対話を統合したインタラクション設計が， メタバース空間における人と AI の関係性をより自然にし， 共に同じ場にいるとい

[1] 北海道情報大学
Hokkaido Information University
[a)] s2221031@s.do-johodai.ac.jp
[b)] yumu@yumulab.org

© 2018 Information Processing Society of Japan

う感覚を生む上で重要になる．

そこで本研究では，この課題を解決するため，ユーザーとの音声対話能力と，対話相手に向かって自律的に移動する能力を両立した AI アバターの開発を行う．具体的には，ユーザーが発した声をトリガーとして，AI アバターが能動的にユーザーへ接近し，対話を開始するシステムを構築する．これにより，AI が単なる応答プログラムではなく，空間内にいるという強い存在感をユーザーに与え，より自然でインタラクティブなコミュニケーション体験を創出することを目的とする．最終的には，特定のワールドに依存せず，多くのユーザーが手軽に導入できる汎用的なツールとしての展開を目指す．

2. 関連研究

VRChat 空間におけるアバターの自律移動や AI 対話に関する研究はこれまでにも多く報告されている．クレイマン [5] による VRChat アバターを生物のように操作する人工知能の作成では，原始的な生物の行動原理に基づく移動アルゴリズムが提案されている．また，別のメタバースプラットフォームである Roblox では，DeterminantAI[6] のようなプラグインを用いて，NPC がキャラクター位置など多角的な環境認識に基づいて自律的に行動計画を立てる事例がある．対話システムとの連携に関しては，EmemeAI[7] が AI キャラクターの対話と移動を同時に実現する機能を提供しており，Song らによる LLM-Driven NPCs[8] は，Unity と Discord を横断して一貫した対話体験を可能にするシステムを提案している．

これらの研究は主に AI アバターの自律行動や技術的統合を中心に発展してきたが，近年では AI アバターとの心理的関係性や共在感（social presence）に注目した研究も進む．Zheng ら [9] は，ユーザー自身が音声およびアバター外観などを自由にカスタマイズできる LLM チャットボットシステム ChatLab を提案した．22 名を対象とした調査から，ユーザーは感情的支援や自己探求を目的に多様なペルソナを構築し，AI との継続的な対話を通してオープンで正直な関係を形成することが示された．これは，AI キャラクターのカスタマイズ性が，ユーザーとの信頼と感情的な結びつきを強める要因となることを示唆する．

また，Son ら [10] は，ソーシャル VR 上で 128 名を対象に社会的存在感の要因を分析し，空間的没入感と他者への温かさの印象が共在感を強く予測することを報告した．この結果は，アバターの人間的要素（共感・表情・仕草など）が VR 内での他者との心理的距離を縮める鍵であることを示す．さらに，Zhang ら [11] は，メタバースの人気を具現化された社会的存在感（Embodied Social Presence, ESP）理論の観点からモデル化し，アバターへの自己同一視と他者との相互作用が社会的リアリティ形成に寄与することを明らかにした．

これらの先行研究を踏まえると，AI アバター研究は技術的な自律性だけでなく，ユーザーとの関係性や心理的リアリティを含めた存在としての AI の設計が重要であることが分かる．本研究で開発した AI アバター「ノア」は，こうした知見を背景に，ユーザーの音声をトリガーとして自律的に移動し応答する仕組みを実装することで，メタバース空間における共在感と自然な関係性の構築を目指している．

3. 設計と実装

本システムは，VRChat 上で稼働する AI アバターと，その思考と行動を制御する外部の Python プログラムから構成される．両者は Open Sound Control (OSC) API を介して通信し，連携動作を実現する．

外部プログラムは，音声認識および対話生成を担う会話プログラム（図 1）と，音源方向推定およびアバターの位置制御を行う移動プログラム（図 2）の 2 系統から成る．これらのプログラムは独立して稼働し，アバターがユーザーの音声入力に反応しながら，自律的な接近行動を同時に実行できるよう設計されている．

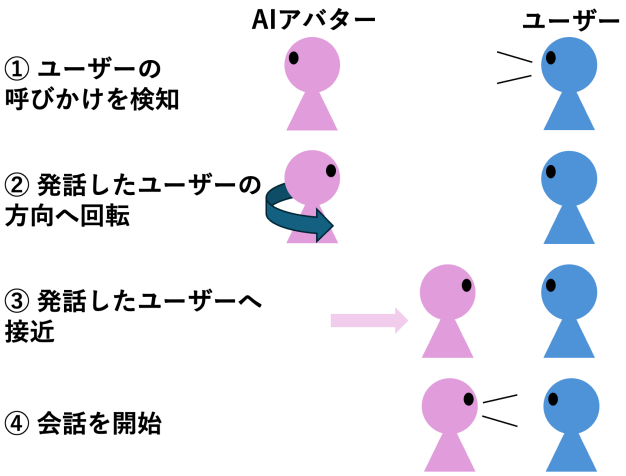


図 1 会話プログラムの概要.

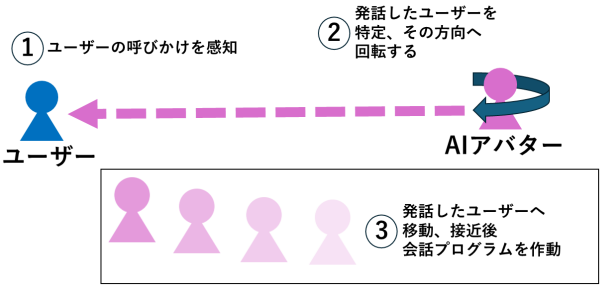


図 2 移動プログラムの概要.

© 2018 Information Processing Society of Japan

3.1 システム構成

本 AI アバター『ノア』は，以下のモジュールを連携させることで機能する．対象プラットフォームは VRChat を選定した．制御プログラムの実装には Python を利用した．

(1) 音声認識・対話生成: リアルタイムで音声を処理し対話文を生成可能な OpenAI の Realtime API を使用, 音声認識及び対話生成の遅延を減少させた．また，待機中に AI アバターが呼ばれたかどうかを判断する為にも利用している．

(2) 音声合成: 対話文を逐一音声合成可能な Aivis Cloud API によって音声合成をし，音声合成の所要時間を減少させた．

(3) アバター制御 (移動・動作): Python から OSC メッセージを VRChat へ送信し，アバターの移動や回転を制御する．

(4) 情報 (座標・移動方向) 取得: VRChat のデバッグログから自己位置や他ユーザーの位置を推定する．

(5) ユーザー記録: 対話終了時に ChatGPT API(gpt-4o-mini) によって内容を処理，一文にまとめ記録する．次回対話時に読み込む事でユーザーの特徴を次回以降も記憶する事を可能にした．

3.2 自律移動の実装

ユーザーへの自律的な接近は，本研究の中核機能であり，以下の手順で実現される．

3.2.1 音源方向の推定

AI アバターが VRChat 内で聞いている音声を，仮想オーディオデバイスを介して Python プログラムに入力する．入力された音声の左右チャンネルの音量差より，発話したユーザーがおおよそどちらの方向にいるかを推定する．ただしこの方法では左右 90 度の範囲しか認識できず，前後の区別が不可能であるため，単純な音量差比較のみによる方向推定には限界がある．そこで，本研究では次の手順を併用して特定精度の向上を図った．まず，待機開始時に音声認識スクリプトを待機モードに設定する．待機状態時に音が聞こえた場合，その方向へアバターを回転させ，その方向に最も近いユーザーを一時的な候補として選択する．次に，そのユーザーが実際に AI アバターを呼んでいるかを判定するために発話を音声認識スクリプトによって Realtime API で処理し，AI アバターが呼ばれていると判断された場合，接近処理を開始する．もし正面のユーザーが発話していない場合や，別方向から新たな音声が検出された場合には，アバターは再び最も近いユーザーの方向へ回転し，音源探索を継続する．このスキャンニングと再判定の反復処理により，発話者の方向をより高精度に特定できるようにし，接近処理の誤認識を防ぎ呼ばれた時のみ接近するように改良する事が出来た．

3.2.2 接近動作の実行

推定された音源方向に対し，アバターをその場に回転させる OSC コマンドを送信する．ターゲットとなるユーザーが正面に来たことを確認後，前進するコマンドを送信し，ユーザーとの距離を縮める．

3.3 対話の開始と終了

アバターがユーザーに接近後，本格的な対話モードに移行する．ユーザーが「さようなら」など特定のキーワードを発話すると，システムは対話終了を認識し，アバターは事前に定められた初期位置へ自動的に帰還する．

3.4 開発プロセスと技術的課題

開発初期段階では，AI アバター自身が視認している VRChat クライアント画面を Optical Character Recognition（OCR）によって解析し，ワールド上に表示されたユーザーの座標を読み取ることで位置情報を取得していた．しかしながら，類似した数字（例：5 と 9 など）の誤認識や，特定のシーンで座標表示が視認困難となる状況が頻発し，情報取得の安定性に課題があった．特に，アバターやオブジェクトが座標表示と重なった場合，文字領域が部分的に隠れてしまい，OCR による正確な読み取りが不可能となるケースが確認された．そこで現在は，AI アバター側のクライアントによって生成されるデバッグファイルをリアルタイムで解析し，ユーザーの座標およびアバター自身の向きを直接取得する方式へと改良した（図 3).



図 3 VRChat クライアント画面．デバッグログからユーザー座標を抽出する方式へと改良した．

この変更により，OCR に起因する認識誤差やフレーム単位での取得遅延を大幅に低減し，情報取得の安定性およびパフォーマンスの向上を確認した．また，MIDI による Python スクリプトから VRChat ワールドへの信号送信により，対話中は他のユーザーの音声が AI アバターへ届かないように改良を行った．

4. 評価実験

本研究で開発した AI アバター『ノア』が，メタバース

© 2018 Information Processing Society of Japan

表 1 アンケート項目

番号	質問	選択肢
Q1	どのような端末でワールドにアクセスしましたか？	PC のみ / PC + HMD / その他
Q2	年齢をお答えください。	～19 歳 / 20～29 歳 / 30～39 歳 / 40～49 歳 / 50 歳～
Q3	性別をお答えください。	男性 / 女性 / その他 / 回答しない
Q4	どのぐらい VRChat にアクセスしていますか？	ほとんどない / たまに / 月 1 回以上 / 週 1 回以上
Q5	ノアとの会話のやり取りはスムーズでしたか？	1: 全くそう思わない ～ 7: 非常にそう思う
Q6	ノアとの会話や動作に煩わしさを感じましたか？	1: 全くそう思わない ～ 7: 非常にそう思う
Q7	ノアの動作や反応が予測しづらいと感じましたか？	1: 全くそう思わない ～ 7: 非常にそう思う
Q8	ノアと会話するための操作は複雑だと感じましたか？	1: 全くそう思わない ～ 7: 非常にそう思う
Q9	ノアはあなたの発話内容を正しく理解していると感じましたか？	1: 全くそう思わない ～ 7: 非常にそう思う
Q10	ノアの行動に不自然さや違和感を感じましたか？	1: 全くそう思わない ～ 7: 非常にそう思う
Q11	ノアに，そこに”いる”という存在感を感じられましたか？	1: 全くそう思わない ～ 7: 非常にそう思う
Q12	ノアを今後も使ってみたいと思いましたか？	1: 全くそう思わない ～ 7: 非常にそう思う
Q13	ノアに独自の個性を感じましたか？	1: 全くそう思わない ～ 7: 非常にそう思う
Q14	ノアを呼びかけてから移動を開始するまでの時間はどう感じましたか？	1: 非常に遅い ～ 7: 非常に速い
Q15	ノアの移動速度はどう感じましたか？	1: 非常に遅い ～ 7: 非常に速い
Q16	会話中のノアの返事が返ってくるまでの時間はどう感じましたか？	1: 非常に遅い ～ 7: 非常に速い
Q17	ノアの特によかった点を教えて下さい（任意回答）	自由記述
Q18	ノアに改善してほしい点があれば教えてください（任意回答）	自由記述
Q19	ノアの動きや反応で印象に残った場面があれば教えてください（任意回答）	自由記述

空間におけるユーザー体験に与える影響を明らかにするため，主観評価実験を実施した．

4.1 実験方法

被験者は 9 名（20 代の大学生）である．実験は VRChat 上の専用ワールドで実施した．被験者は『ノア』に対して音声による呼びかけおよび自由な対話を行い，接近行動と会話のプロセスを体験した．その後，Web フォームを用いたアンケートに回答した．

4.2 評価指標

アンケートは，被験者の属性を調べる為の設問と，7 段階リッカート尺度（1: 全くそう思わない ～ 7: 非常にそう思う）による 12 項目の設問と，自由記述で構成した．主な評価項目は以下の通りである．

(1) 存在感： アバターが空間に「いる」と感じられたか
(2) 自然さ： 動きや反応に違和感がないか
(3) 理解度： 発話内容が正しく伝わっているか
(4) 操作性： システムの利用は容易か
(5) 再利用意図： 今後も継続して利用したいか

アンケートの詳細を表 1 に示す．

4.3 結果

使いやすさ及び操作性を質問した Q5 から Q8 では，会話のやり取りのスムーズさは平均 5.4 点と良好な評価を得た．また，操作の複雑さは平均 1.2 点と極めて低い値を示した．これは，AI アバターと会話する際には専門知識が必要ないという事を示唆している．一方で，動作・反応の予測しづらさは平均 4.3 点とやや高く，エージェントの自律的な挙動に不確実性を感じたユーザーも存在した (図 4)．

システムの理解度と存在感を質問した Q9 から Q11 では，発話内容の理解度は平均 5.4 点，存在感は平均 5.8 点と，いずれも肯定的な評価が得られた．特に「存在感」の中央値は 6.0 点であり，VR 空間内におけるノアのそこにいる感覚（プレゼンス）が高く維持されていたことが伺える．行動の不自然さは平均 2.6 点と低水準であり，大きな違和感を与えることなくシステムが稼働していたと言える (図 5)．

継続利用意向と個性を質問した Q12，Q13 では，今後の利用意向は平均 5.1 点であった．また，独自の個性については平均 6.0 点と非常に高い評価を得た．自由記述の結果とも照らし合わせると，ノアのキャラクター性や独自の返答内容が，ユーザーの継続的な関心を惹く要因となっている可能性が高い (図 6)．

応答時間と移動速度の評価を質問した Q14 から Q16 では，移動開始までの時間 (4.1 点)，移動速度 (平均 4.3 点)，返事の待機時間 (平均 4.0 点) は，いずれも評価尺度の中間値である 4 点付近に集中した．これは，システムの応答速度が致命的な不満には繋がっていないものの，さらなる体験向上のためには改善の余地がある許容範囲内の状態であることを示している (図 7)．

自由記述による定性評価である Q17 から Q19 の自由回

© 2018 Information Processing Society of Japan

答では，以下の 3 つの主要な傾向が見出された．

(1) 対人感の醸成：「人間と話している感覚に近い」「リアルタイムで会話している感覚が良い」といった，エージェントとの高い双方向性を評価する意見が目立った．

(2) 知識と個性の評価：「話題を振ってくれる」「正確な回答と個性を感じる」など，対話内容の質が満足度に寄与していた．

(3) 技術的課題：改善点として「テキストが表示されない場面があった」「エモートの正確性」などが挙げられ，視覚的なフィードバックの確実性が今後の課題として浮き彫りになった．

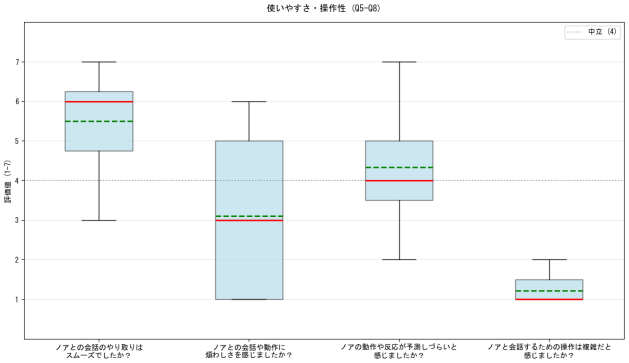


図 4 Q5 から Q8 の箱ひげ図.

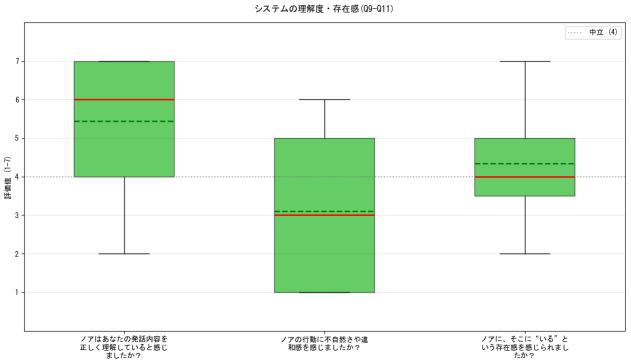


図 5 Q9 から Q11 の箱ひげ図.

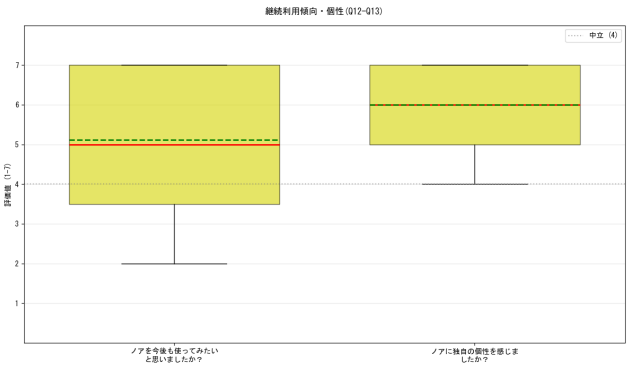


図 6 Q12 及び Q13 の箱ひげ図.

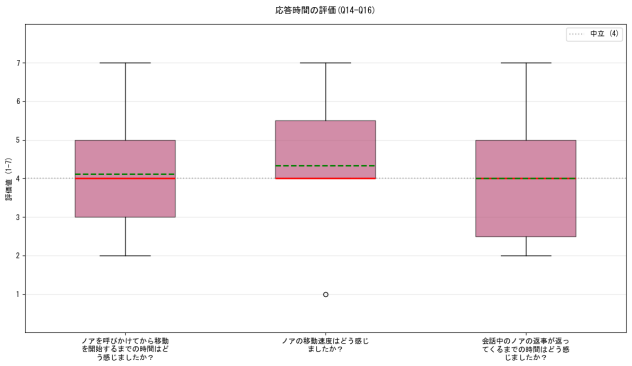


図 7 Q14 から Q16 の箱ひげ図.

5. 考察

プロトタイプの実装と動作検証を通じて，ユーザーの呼びかけに応じて AI アバターが振り向き，接近し，対話を開始するという一連の動作が実現可能であることを確認した．声を発した相手に AI が能動的に関わろうとする挙動は，静的な AI アバターと比較して格段に高い存在感とインタラクティブ性を生み出すことが示唆された．

なお，本システムは EC2025[12] において発表を行った．その後 IVRC2025 メタバース部門にも応募し．SEED Stage では不通過となったものの，審査員から得られたコメントは本研究の今後の方向性に対して極めて有益であった．

特に，アバターを呼び出す際に名前を呼ぶという行為の意味づけについて重要な指摘がなされた．従来の音声アシスタント（例：Siri, Alexa）における呼称は単なる機能起動のトリガーであったが，本研究で扱う AI アバターはメタバース空間における社会的存在として，名前を呼ぶという行為そのものが感情的・非言語的な関係性の表現となりうることが示唆された．すなわち，呼びかけに頼らず，視線の一致や近接動作といった自然なインタラクションを契機として対話を開始する仕組みを設計することで，共に存在する感覚の深化が期待される．

また，技術的な観点からは，AI がいつ話を聞いているのか，音声を正しく取得できているのかをユーザーに視覚的に提示する必要性，さらには過去の会話内容を記憶して継続的な関係を形成する仕組みの重要性が指摘された．これらは単なる応答精度の向上にとどまらず，AI アバターを社会的存在として確立するための本質的要素である．

以上の指摘を踏まえると，今後はアバターとのインタラクション設計を入力手段としてではなく，関係構築の体験として再構成する必要があると考えられた．その後，バーチャル学会 [13] において発表及びデモセッションを行った．その際，多人数が同時に発話した際に AI アバターが処理出来ず，会話しようとしているユーザーへ接近を行わなかった他，全ユーザーの発話を処理した結果，全ての発話に対し返答を実施し会話しようとしていたユーザーが会

© 2018 Information Processing Society of Japan

話を出来なかったという事例が見られた．その為，会話中は対象ユーザーのみの発話を聞くことにより対処した．

6. おわりに

本研究では，VRChat 上で動作する AI アバターを開発し，音声対話およびユーザーの向きや距離に応じた振る舞いによって，ユーザーに対して存在感を与えるシステムの構築を行った．特に，リアルタイム音声対話とアバターの身体動作を組み合わせることで，単なる音声アシスタントとは異なる体験を目指した．

評価実験の結果より，アバターの存在感や今後の利用意図については比較的高い評価が得られ，本システムがユーザーに対して「そこにいる存在」として認識される可能性が示唆された．一方で，会話内容の単調さや応答の安定性については課題が残り，より自然な対話体験を実現するための改善が必要であることが明らかとなった．

今後の課題として，対話生成の高度化や応答遅延の低減を今一度考える必要があると思われる．また，アバターの表情や動作のバリエーションを拡充することで，より豊かなコミュニケーション体験が実現できると考えられる．最終的には booth 上においての配布を実施する．

参考文献

[1] 加納寛子：メタバース空間における AI カウンセラーの開発, 日本教育工学会研究報告集, Vol. 2024, No. 4, pp. 203–210（オンライン）, DOI: 10.15077/jsetstudy.2024.4_203 (2024).
[2] AOI, A. A.: AI Avatar AOI — AmiVoice 製品サービスサイト｜音声認識のアドバンスト・メディア，, 入手先 ⟨https://www.advanced-media.co.jp/products/service/ai-avatar-aoi/⟩（参照 2025-10-24）. (閲覧日: 2025 年 10 月 24 日).
[3] Cotomo: 音声会話型おしゃべり AI アプリ Cotomo(コトモ)，, 入手先 ⟨https://cotomo.ai/⟩（参照 2025-10-24）. (閲覧日: 2025 年 10 月 24 日).
[4] MiraiMind: , available from ⟨https://web-miraimind.hellogroupjapan.com/#/discover⟩ (accessed 2025-10-24). (閲覧日: 2025 年 10 月 24 日).
[5] クレイマン：VRChat アバターを生物のように操作する人工知能の作成, バーチャル学会発表概要集, Vol. 2024, pp. 40–41（オンライン），DOI: 10.57460/vconf.2024.0_40 (2024).
[6] DeterminantAI: Create Intelligent AI Characters For Your 3D Virtual World, , available from ⟨https://www.determinant-ai.com⟩ (accessed 2025-07-16). (閲覧日: 2025 年 7 月 16 日).
[7] EmemeAI: オリジナルの AI キャラクターを外部アプリに接続できる 3D チャット AI 投稿プラットフォーム『EmemeAI』3 月 27 日新機能リリース，, 入手先 ⟨https://www.atpress.ne.jp/news/430788⟩（参照 2025-07-16）. (閲覧日: 2025 年 7 月 16 日).
[8] Song, L.: LLM-Driven NPCs: Cross-Platform Dialogue System for Games and Social Platforms (2025).
[9] Zheng, X., Li, Z., Gui, X. and Luo, Y.: Customizing Emotional Support: How Do Individuals Construct and Interact With LLM-Powered Chatbots, *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, ACM, p. 1–20 (online), DOI: 10.1145/3706598.3713453 (2025).
[10] Son, G., Tiemann, A. and Rubo, M.: I am here with you: an examination of factors relating to social presence in social VR, *Frontiers in Virtual Reality*, Vol. Volume 6 - 2025 (online), DOI: 10.3389/frvir.2025.1558233 (2025).
[11] Zhang, G., Cao, J., Liu, D. and Qi, J.: Popularity of the metaverse: Embodied social presence theory perspective, *Frontiers in Psychology*, Vol. Volume 13 - 2022 (online), DOI: 10.3389/fpsyg.2022.997751 (2022).
[12] 小山田悠人，湯村　翼：メタバース空間で対話相手に向かって自律移動する AI アバター『ノア』の開発，エンタテインメントコンピューティングシンポジウム 2025 論文集，Vol. 2025, 情報処理学会，pp. 378–380 (2025).
[13] 小山田悠人，湯村　翼：メタバース空間で対話相手に向かって自律移動する AI アバター『ノア』の開発，バーチャル学会発表概要集，Vol. 2025, pp. 78–81（オンライン），DOI: 10.57460/vconf.2025.0_78 (2025).

© 2018 Information Processing Society of Japan