



HYOUI: マルチモーダル大規模言語モデルを活用した 他者の思考体験 AR システム

河野真有香¹⁾, 平尾悠太郎¹⁾, ペルスキアエルナンデスモニカ¹⁾, 内山英昭¹⁾, 上垣外英剛¹⁾, 清川清¹⁾

1) 奈良先端科学技術大学院大学 (〒 630-0192 奈良県生駒市高山町 8916-5, kono.mayuka.ki4@is.naist.jp)

概要: 大規模言語モデル (LLM) は工学的観点から人の模倣可能性が期待されている。LLM にペルソナを与え、ペルソナに応じた振る舞いをさせる研究が盛んである。本展示では、マルチモーダル大規模言語モデル (MLLM) にペルソナと実世界の視覚情報を入力したときのそのペルソナの視覚や言語処理を再現し、テキスト形式で提示する AR システム「HYOUI」を紹介する。本システムを発展させることで自閉スペクトラム症 (ASD) の個人を支援する人に彼らの視覚や言語処理特性に応じた効果的な関わり方を考える上で役立つと期待する。

キーワード: MLLM, AR, 生成 AI, ペルソナ

1. はじめに

他者理解の促進目的で XR を活用する取り組みがある [1]。これらでは、ユーザに他者の特性の理解を促せると示されているが、個人差を考慮できていないという課題がある。急速な発展を続ける大規模言語モデル (LLM) は工学的観点から人を模倣できる可能性が注目されている [2]。LLM にペルソナを付与して、人間の行動や振る舞いを模倣する取り組み [3] が進められている。さらに、LLM を活用してユーザに対して情報を提示する XR システムも開発されている [4]。以上より、他者理解のための XR システムの開発に LLM を活用することで従来研究で課題となった個人差の考慮が可能になると考えた。

本稿では、開発中のマルチモーダル大規模言語モデル (MLLM) を活用した他者の思考体験が可能なモバイル AR システム「HYOUI」を紹介する。本システムは、ユーザが関わる他者が同じ視覚情報を受容したときの思考を理解することを目的とする。ユーザの基本情報に基づき、既存のペルソナデータのクラスターのどの位置に属するかを推定し、その点から最遠のデータを MLLM のペルソナとして設定する。MLLM にユーザが見ている実世界の画像を入力することで、思考をテキストで出力させ、モバイル画面上の実世界に重畳して提示する。これにより、ユーザは自らの属性から最遠の他者の思考を体験できるようにする。

2. 関連研究

2.1 他者理解のための XR シミュレータ

長井らは、自閉スペクトラム症 (ASD) の知覚鈍麻・過敏の発生過程をモデル化・実装し、ASD 知覚体験シミュレータを開発した [1]。ヘッドマウントディスプレイ (HMD) 上のカメラからユーザの視野に相当する画像と音声を取得し、ASD 当事者の視覚世界に変換し、リアルタイムで提示できる。Tsujita らは ASD に対するスティグマに着目し、知覚体験シミュレータによりスティグマを軽減できると示した [5]。

実験により、他者理解に繋がるだけでなく ASD 当事者の自己理解にも繋がるのが明らかになった。

2.2 LLM にペルソナを持たせる手法

Jiang らは GPT-3.5 と GPT-4 に Big Five に基づくパーソナリティを付与する PersonaLLM を提案した [3]。パーソナリティに合った言語学的特徴を持つ文章を生成できると示した。さらに、人間とペルソナを持つ LLM の間には文章作成時の単語使用に共通点があるとも明らかにした。Tu らは、ChatGPT を用いて MBTI (Myers-Briggs Type Indicator) に基づくペルソナ分類を活用し、個人とペルソナが適合するバーチャル対話エージェントをマッチングする CharacterChat を提案した [6]。彼らは、この研究を通じ、様々なプロフィールを持つバーチャルキャラクタで構成された MBTI-1024 Bank を構築した。

2.3 LLM を活用した XR システム

Lee らは、音声アシスタントでの代名詞の曖昧性問題の解消を目指したウェアラブル AR である GazePointAR を開発した [7]。リアルタイムなコンピュータビジョンと LLM を活用し、視線追跡、指差しジェスチャ認識、代名詞の具体物への置換により、曖昧性解消を実現した。Dogan らは、MLLM を活用してリアルタイムに物体の情報を取得し、AR 上に提示する XR-Objects を開発した [4]。リアルタイムな物体セグメンテーションと分類を MLLM と組み合わせ、物体の事前登録なしに意味情報を自動認識し、詳細な情報の提示を実現した。

従来の他者理解のための XR シミュレータでは、個人差が考慮されていない、見え方に限局しているという課題がある。これに対し、LLM にペルソナを持たせる手法と LLM を活用した AR システムを踏まえ、MLLM を活用することで個人差を考慮した AR システムを実現することを着想した。加えて、MLLM を活用すれば見え方に限局することなく、視覚情報を入力した際の思考を体験できるシステムを

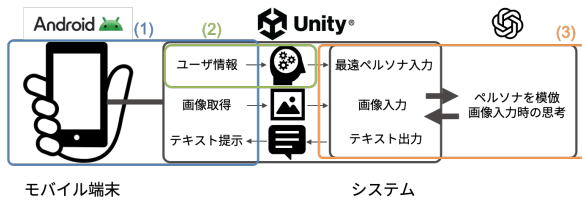


図 1: システム構成。

実現できると考えた。本研究では、MLLM を活用した他者の思考体験 AR システムを開発する。

3. システムの概要

システム構成を図 1 に示す。システムは Unity 環境に構築する。Unity のスクリプトで OpenAI API (GPT-4.1) を利用して、スマートフォン向け AR システムとして実装する。システムは (1) ユーザとのインタラクションモジュール、(2) 最遠ペルソナ推定モジュール、(3) 言語モデル処理モジュールに大別される。(1) はユーザの入力情報や実世界の画像取得、ユーザへの思考体験の提示をする。(2) はユーザの入力情報からユーザがペルソナデータのどの位置に属するかを推定し、その点から最遠のペルソナを 1 つ選択する。(3) は OpenAI API の GPT-4.1 を MLLM として使用する。特定のペルソナとして画像のシーンを入力したときの思考を出力する。

モジュール (1) では、システム実行開始時にユーザから MBTI、性別、年齢の情報を取得する。実世界の任意の場所で画面のボタンを押下してその時点での画像を取得する。モジュール (2) へ取得したユーザ情報を送信する。モジュール (2) は、オープンなペルソナデータである MBTI-1024 Bank¹[6] とユーザ情報でクラスタリングを行い、ユーザのペルソナがどの位置に属するかを推定する。ユーザを示す点から最遠に位置するペルソナをモジュール (3) に入力するペルソナとして 1 つ選択する。モジュール (3) へ取得した画像を送信する。モジュール (3) では、画像を OpenAI API で利用可能な Base64 形式に変換する。GPT-4.1 にプロンプトエンジニアリングにより、モジュール (2) で選択したペルソナとして振る舞うような指示と画像、「この画像を入力したときの思考を述べてください」という指示を与える。モジュール (3) からモジュール (1) に思考内容のテキストを出力する。最終的にモジュール (1) がテキストをスマートフォンの画面に映る実世界上に提示する。

4. 今後の展望

本プロジェクトで使用したペルソナデータセットには MBTI を中心とした小児以外のデータが含まれている。今後はペルソナを拡張する必要がある。例えば、小児、MBTI 以外の個人を特徴づける情報、定型発達・非定型発達の情報などの追加である。これにより、従来の他者理解のための XR システムの個人差を考慮していないという課題を解消する

ことに繋がると期待する。さらに、現時点では指定したペルソナの視覚と言語処理のみに焦点を当てている。今後は、指定した視覚・聴覚・言語処理の再現やある状況下での行動の個人差の再現などにも取り組みたいと考えている。なぜなら、個人差は視覚だけでなく聴覚にも認められるからである。対象モダリティを拡張することで、よりあるペルソナの個人の特性を模倣することに繋がると考える。将来的には、定型発達者の相互理解に留まらず、ASD のような非定型発達者の支援者が当事者個人の特性に合わせた効果的な支援方法を考える支援ができるシステムに発展することを目指す。

5. むすび

本稿では、MLLM を活用した他者の思考体験システムを紹介した。Unity 環境で MLLM 利用したスマートフォン向け AR システムとして開発中である。MLLM はオープンなペルソナデータを用いてペルソナを付与され、ユーザ情報に応じてペルソナを自動切り替え可能なものとする。本プロジェクトの成果は、将来的に他者の相互理解だけでなく、ASD などの非定型発達者の支援者が支援の適切性を判断する際にも役立つと期待する。

参考文献

- [1] 長井ら. 自閉スペクトラム症の特異な視覚とその発生過程の計算論的解明: 知覚体験シミュレータへの応用. 日本認知科学会第 32 回大会発表論文集, pp. 32–40, 2015.
- [2] Chen et al. From persona to personalization: A survey on role-playing language agents. *arXiv preprint arXiv:2404.18231*, 2024.
- [3] Jiang et al. Personallm: Investigating the ability of large language models to express personality traits. *arXiv preprint arXiv:2305.02547*, 2023.
- [4] Dogan et al. Augmented object intelligence with xr-objects. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–15, 2024.
- [5] Tsujita et al. Comprehensive intervention for reducing stigma of autism spectrum disorders: Incorporating the experience of simulated autistic perception and social contact. *Plos one*, Vol. 18, No. 8, p. e0288586, 2023.
- [6] Tu et al. Characterchat: Learning towards conversational ai with personalized social support. *arXiv preprint arXiv:2308.10278*, 2023.
- [7] Lee et al. Gazepointar: A context-aware multimodal voice assistant for pronoun disambiguation in wearable augmented reality. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pp. 1–20, 2024.

¹<https://github.com/morecry/CharacterChat>