

生成 AI を用いた状況認識による社会的行動促進手法

岩崎 玲^{1,a)} 寺田 努^{2,b)} 塚本 昌彦^{2,c)}

概要：普段、人は歩行時に「ゴミを拾う」、「交通ルールを守る」といった社会の秩序や協調を保つために必要な社会的行動を行える状況であってもそれを行わない場合がある。社会的行動をしない理由としては気づかない、あるいは面倒であるといった理由が考えられる。そこで本研究では、生成 AI モデルにリアルタイムで周囲の状況を解析させ、その場で可能な社会的行動や、行動を行ったことによる点数をフィードバックするシステムを提案する。生成 AI モデルの使用により、その場に応じた柔軟な回答を得ることができ、プロンプトに応じて出力内容を容易に変更できる。

1. 研究の背景と目的

普段、人は歩行中に社会的行動ができる場面があったとしても、気づかない、あるいは面倒であるといった理由から社会的行動をしないことが多い。ここで社会的行動とは、交通ルールの遵守やゴミの収集といった、社会の秩序や協調を保つために必要な行動を指す。

社会的行動を促す行動変容の手法の 1 つとしてゲーミフィケーションが挙げられる。ゲーミフィケーションとはゲーム要素を非ゲーム分野に応用することで、目標達成に向けたユーザの意欲向上を図る手法である。実際に、Xu らの研究では、普段の活動にポイントや目標を与えることで活動意欲が促進されることが明らかになっている [1]。

こういった背景から、本研究ではリアルタイムでその場に応じた社会的な行動を提示しつつユーザの行動を点数化し、行動変容を促すシステムを提案する。

2. 提案システム

2.1 システム要件

システム要件として、周囲の状況を認識できること、状況に応じた社会的行動を出力できること、ユーザの行動を認識して点数化できること、社会的行動と点数を可視化できることの 4 点が挙げられる。要件を満たすツールとして、生成 AI モデルと頭部装着型ディスプレイ (HMD: Head Mounted Display) の利用を考える。

生成 AI モデルは従来の AI と比較して用途が限定されて

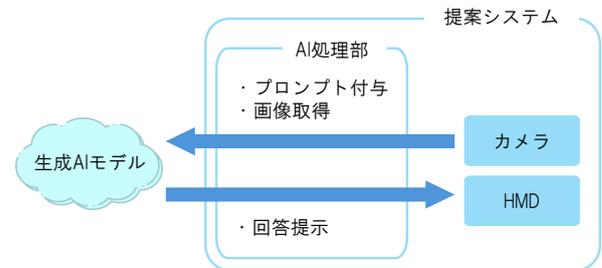


図 1 提案システムの概要



図 2 デバイスの外観

おらず、社会的・倫理的行動という汎用的で抽象的な内容の判断ができる。そのため、状況認識や適切な社会的行動及び点数の出力を、一つのモデルで行うことができる。加えて、一般に公開されている生成 AI モデルを使用する場合は、使用時に新たな学習を必要とせず、モデルに特定のタスクを指示するためのテキスト入力であるプロンプトを変更することで出力内容を容易に変更可能である。

Kadiyala らは Gemini や Chat GPT などの生成 AI モデルに「洪水により倒壊した建物」や「汚染された水域に浮かぶ魚」といったシチュエーションの画像を送信し、それぞれ「建物の数や状態」、「水質の状態」という内容について質問した結果、モデルによって精度に差があるものの、画像に写っている物や数、状態についての回答を得ること

¹ 神戸大学工学部

² 神戸大学大学院工学研究科

a) rei-iwasaki@stu.kobe-u.ac.jp

b) tsutomu@eedept.kobe-u.ac.jp

c) tuka@kobe-u.ac.jp

ができた [2]. このように、生成 AI モデルの使用による周囲の状況認識が可能であることが示された。

また、社会的行動と点数の可視化を可能にするために、デバイスには歩行時にユーザが周囲の状況を見れる程度の解像度であるカラーパススルー HMD を使用する。

2.2 システム構成

前節のシステム要件を踏まえた、システム構成を図 1 に示す。システムの動作は大きく 4 段階に分けられる。

まず、API を通して生成 AI に返答内容を指示するためのプロンプトを送信する。プロンプトを要約すると、「送信した画像を参考にしてその場でできる社会的行動を検出してください。社会的行動をしたと判断した場合は加点してください。」という内容である。その場で行うべき行動の検出だけでなく、ユーザの行動に対する評価も生成 AI が行う。

次に HMD のカメラでユーザの視点を定期的に撮影する。そしてその画像を生成 AI に送信して解析する。最後に生成 AI から得られた解析結果を、HMD を用いてユーザに提示する。

例えば人とゴミが映った画像が撮影された場合、「人に挨拶をする」、「ゴミを拾う」といった回答が得られる。加えて、撮影した画像からユーザが社会的行動をしたと判断されると点数が加点され、それぞれユーザに提示する。プロンプトの送信はシステム起動後の 1 回のみで、以降は視点の撮影からユーザへの提示を定期的に繰り返す。

2.3 プロトタイプ

使用する HMD としては、空間コンピューティングによる生活利用を想定しているため、装着性の良さ、視野角の大きさ、カメラの解像度高さ、パススルー性能の良さ、バッテリーの持続時間を考慮して Apple Vision Pro を選んだ。Apple Vision Pro にはユーザ視点の写真を撮影できるカメラが内蔵されているが、今回は Web カメラである logicool C505e を Apple Vision Pro にとりつけた。Web カメラは別のノートパソコンに接続されており、処理を円滑にするため高性能なノートパソコンを使用する。

Meta Quest シリーズや Pico, Apple Vision Pro や HTC など現在のカラーパススルーゴーグルでは、悪意のあるプログラムがカメラ映像を勝手に取得するというようなことを防ぐプライバシー配慮のために、基本的にカメラ映像が取得できない。使用できるのは法人契約された AVP で Enterprise API を用いる必要がある。そのため、システムの汎用性を高める目的で外部カメラを用いた。

また、使用した生成 AI モデルは Gemini-1.5-Pro である。

プロトタイプにおけるシステムの動作については、図 3 に示す。ノートパソコン内のプログラムでは、前節で述べ

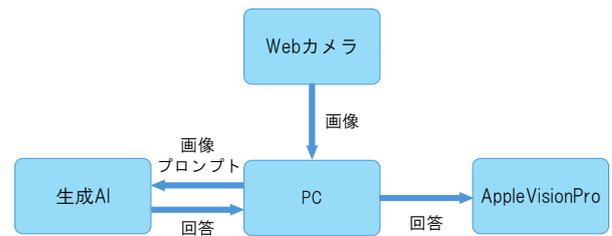


図 3 プロトタイプにおけるシステムの動作

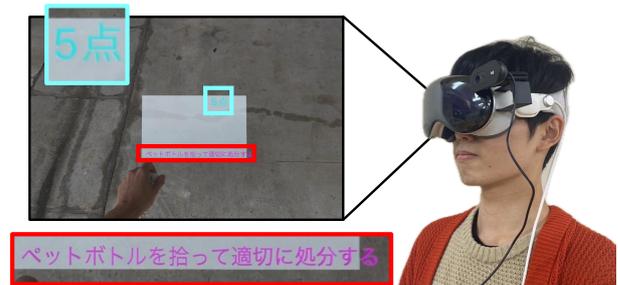


図 4 システムの UI

た生成 AI とのやりとりを行い、回答を取得し、その回答を Apple Vision Pro に UDP 通信で送信する。回答は「『その場ですべき社会的行動』***『累計点数』」となるようにプロンプトで指示してあるため、Apple Vision Pro では受信したテキストをアスタリスクの前後で 2 つのテキストに分割し、それぞれのテキストをユーザの目の前にある UI にテキストベースで表示する。UI は図 4 に示す。

プロトタイプ実装では、ノートパソコンでは Python, Apple Vision Pro では Swift を使用している。

実際に使用すると、空のペットボトルが写っているシチュエーションでは、「ペットボトルを拾って適切に処分する」という内容を含む回答が表示され、ペットボトルを拾うことで表示されている累計点数が追加された。一方で、ペットボトルが写った段階で点数が入るといった不安定な挙動も見られた。

3. 実験計画

様々なシチュエーションにおいて、複数の生成 AI モデルが送信したプロンプトに対して、どのような回答をするのか比較するための実験を計画する。その結果を基に、提案システムで使用する生成 AI モデルを選定する。ここでは今後の計画について述べる。

実験の手順を図 5 に示す。まず、各シチュエーションにおける歩行者視点の映像を用意する。今回は、ごみが道端に落ちているシチュエーションや、歩行者用信号機がある横断歩道を渡るシチュエーションなどを想定する。次に、その映像から 5 秒ごとのフレーム画像を取得する。それらを生成 AI モデルに送信し、その場ですべき社会的行動と、歩行者の行動に対する点数が含まれた回答を得る。そ

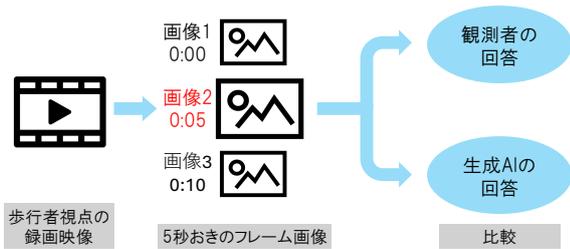


図 5 実験手順

	Gemini	CHAT GPT	Claude
コスト	\$1.25/100万トークン (Gemini 1.5Pro)	\$2.5/100万トークン (GPT-4o)	\$3/100万トークン (Claude 3.5 Sonnet)
画像入力	○	○	○
動画入力	○	○	○
音声入力	○	○	○
音声出力	X	○	X

図 6 使用する生成 AI モデルの特徴

それぞれの画像に対して、生成 AI モデルの回答と筆者の回答を比較する。また、ユーザの視点を撮影してからどの程度の遅れが発生するかを把握するため、画像を送信してから回答が返ってくるまでの応答時間も測定する。

この結果をもとに、シチュエーションごとの生成 AI の回答を比較し、システムの使用に適した生成 AI を検討する。比較する生成 AI モデルは Google 社の Gemini, OpenAI 社の ChatGPT, Anthropic 社の Claude の三つである。モデルごとの特徴の比較内容を図 6 に示す。全てのモデルが画像入力に対応しているため、本実験で使用可能である。動画の送信も可能であるが、リアルタイム性を重視して今回は画像を扱うこととする。

4. 考察

プロトタイプ実装では外部の Web カメラとノートパソコンを使用し、本来はプライバシーの観点から産業用でしか使用できない機能であるユーザ視点のカメラの画像取得を個人的に行った。Apple Vision Pro においても Enterprise API を使用することで、Web カメラとノートパソコンを使用せずに、Apple Vision Pro のみで提案システムを実装することができる。しかし、これではやはりプライバシーに関する問題があり、一般的に普及させることは難しい。この問題を解決したり、ユーザ視点のカメラの画像取得の制限を緩和したりするためには、顔や個人情報の検出による自動モザイク処理といったプライバシー保護技術を搭載する必要がある。

また社会的行動の誘発に対する有効性についてはゲーミフィケーションは有効であるが、単に点数を与えるだけではなく、点数に対する報酬を用意することができれば強く

促進でき、これを実現できるビジネスモデルの構築が必要と考えられる。報酬の用意だけでなく、他のシステム利用者の点数も可視化して点数で競うことも有効であり、地域のコミュニティでシステムを導入したり、教育現場で利用したりすることについて検討する。

そして、プロトタイプの使用で発生した不安定な挙動に関しては、使用する生成 AI モデルの変更をしたり、プロンプトの内容を修正したりすることで改善を図り、また、生成 AI の最近のアップデートで高度なマルチモーダル機能が追加されているため、今後はそれを使うことを検討する。

5. まとめ

本稿では、生成 AI を用いてその場に応じた社会的行動の提示と点数化により、社会的行動を促すシステムを提案した。今後は実験計画に基づいて実験を行い、システムを実装したのち、システムを搭載したデバイスの使用により社会的行動が促されるか検証する。

謝辞 本研究の一部は、中山隼雄科学技術文化財団研究助成「カラーパススルー HMD を用いた人間性・社会性育成ゲーム」の支援によるものである。ここに記して謝意を表す。

参考文献

- [1] L. Xu et al.: The Effects of mHealth-Based Gamification Interventions on Participation in Physical Activity: Systematic Review, *Journal of JMIR MHealth and UHealth*, Vol. 10, No. 2: e27794, pp. 1–18 (Feb. 2022).
- [2] L. A. Kadiyala et al.: The Implementation of Multimodal Large Language Models for Hydrological Applications: A Comparative Study of GPT-4 Vision, Gemini, LLaVa, and Multimodal-GPT, *Journal of Hydrology*, Vol. 11, No. 9: 148, pp. 1–20 (Sep. 2024).