

経済実験における実験参加者を大規模言語モデルで代替できるのか  
 —ペルソナ情報によるシミュレーション精度向上の検証—  
 Can large language models replace participants in economic experiments?

-Validation of simulation accuracy improvement by persona information-

日室 聡仁<sup>†</sup> 後藤 晶<sup>‡</sup>  
 Akihito Himuro Akira Goto

## 1. はじめに

行動経済学はさまざまな経済実験によって発展してきた。経済実験は行動経済学にとって重要な研究手法であるが、実施には多くの時間や費用がかかるため、実験回数を増やすことが難しいという課題がある。この課題の解決策として大規模言語モデル (LLM) の活用があげられる。具体的には、経済実験の実験参加者を LLM で代替することで、実験参加者を集めることなく、短時間かつ低費用で経済実験を実施できる可能性がある。

少数の LLM が活用された経済実験は、LLM が人間の経済行動をある程度忠実に再現できる可能性を示している [1]。しかしながら、特定の経済実験に対する報告について、LLM を活用した経済実験全般に一般化することは不可能であり、妥当性の検証は依然として不十分である。

上記の状況を踏まえて、経済実験の方法論的革新に寄与するため、繰り返し公共財ゲームを題材に、LLM 同士でゲームを実施させることで、人間がゲームを行った際の行動と LLM の行動を比較し、実験参加者を LLM で代替することが可能かどうかを検討した [2]。検証の結果、一部再現できていない点はあるが、LLM が人間の意思決定を再現できる可能性があり、実験参加者を代替できる可能性がある。

本研究では、LLM による実験参加者の代替精度をさらに高めるために、LLM にペルソナ情報として実験参加者の属性情報を付与することによって、シミュレーション精度が向上するかどうかについて繰り返し公共財ゲームを用いて検証した。

## 2. 繰り返し公共財ゲームについて

繰り返し公共財ゲーム (Repeated Public Goods Game) は、公共財の供給における協力行動を理解するための重要なモデルである。このゲームは、社会科学、経済学、進化生物学など多くの分野で研究されており、個人の意思決定や集団行動のメカニズムを探るための基盤となっている。

公共財ゲームは、複数人で行うゲームであり、各個人が自身の資金を共同の資金に投資するかどうかを決定するものである。公共財への投資額は、それぞれの実験参加者が自由に決めることができ、集まった資金は一定の倍率で増加し、すべての実験参加者に均等に配分される。このゲームは、集まった公共財が多いほど、集団全体の利益が増加するという構造を持つ。しかし、公共財は均等に分配されるため、各自が投資を控えても、ほかの実験参加者の投資

によって自身もその恩恵を受けることができる。そのため、公共財への投資額が多い実験参加者は協力的、少ない実験参加者は非協力的と評価できる。この公共財ゲームを同一グループで一定回数繰り返し実施するものが、繰り返し公共財ゲームである。

## 3. 実験設計

本研究では、繰り返し公共財ゲームを題材とし、3つの実験条件を比較した。実験条件は、人間を実験参加者とした Human 条件、LLM を実験参加者とした LLM ペルソナなし条件、LLM を実験参加者とし、ペルソナ情報を付加したペルソナあり条件を比較した。

### 3.1 Human 条件

既発表の 2022/7/11 に Yahoo クラウドソーシングを活用して実験参加者を募集して集めたデータを利用する。有効実験参加者数は 87 名 (男性 58 名, 女性 29 名, 年齢  $M=50.33$ ,  $SD=18.36$ , 年齢未回答者を除く) である。ゲームの繰り返し回数は 10 回とし、各ラウンド開始時 20 ポイントが付与され、公共財ゲームへの支出は 0~20 を選択する方式とした。また、配分時の乗数は 2 とし、実験参加者 3 名で 1 グループを構成してゲームを実施した。

### 3.2 LLM ペルソナなし条件

OpenAI 社の ChatGPT-4o を用いて、繰り返し公共財ゲームの実験参加者をシミュレーションした。ゲームのルールは 3.1 節と同様とし、実験参加者 3 名で 1 グループを構成した。実験参加者数は人間対照実験である 3.1 節に合わせて 87 名とした。ChatGPT-4o のオプション設定については、回答の多様性を調整する temperature のみ最大値である 2 に変更した。プロンプトは図 1 に示すものを使用したが、ペルソナに関する記載は削除して動作させた。

### 3.3 LLM ペルソナあり条件

3.2 節と同じ条件でデータを収集した。ペルソナ情報として性別、年齢、所在地、学歴、個人所得、世帯所得、結婚の有無、子供の有無をプロンプトに設定した。ペルソナ情報は 3.1 節の実験参加者から収集した情報をランダムに付与したプロンプトは図 1 を使用した。

## 4. 実験結果

### 4.1 支出推移

図 2 に、公共財への支出の推移を示す。LLM ペルソナなし条件と LLM ペルソナあり条件は、Human 条件より支出が低い傾向が見られた。また、すべての条件で支出が右肩上がりに増加する傾向も確認された。さらに、LLM ペル

<sup>†</sup> NEC ソリューションイノベータ株式会社

NEC Solution Innovators, Ltd.

<sup>‡</sup> 明治大学 Meiji University

ソナなし条件は LLM ペルソナあり条件に比べて支出が低い傾向が見られた。

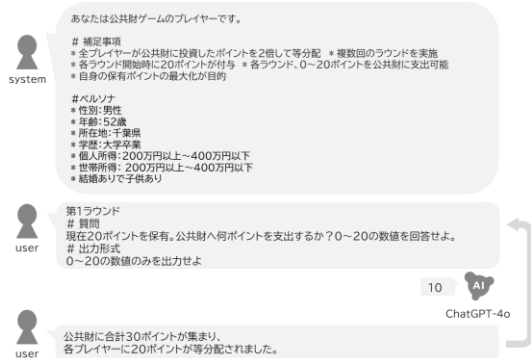


図 1: プロンプト

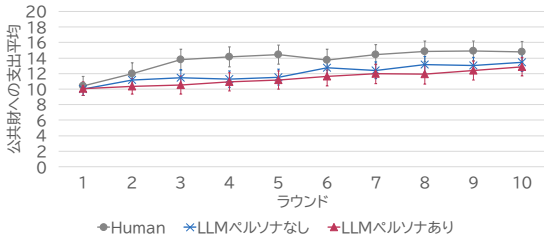


図 2: 公共財への支出の平均推移

| Predictors                 | 公共財への支出                  |        |  | 公共財への支出                  |        |  |
|----------------------------|--------------------------|--------|--|--------------------------|--------|--|
|                            | estimates                | p      |  | estimates                | p      |  |
| (Intercept)                | 13.76<br>(12.78 - 14.75) | <0.001 |  | 11.59<br>(10.51 - 12.67) | <0.001 |  |
| 条件<br>[LLMペルソナなし]          | -1.73<br>(-3.12 - -0.34) | 0.015  |  | -1.48<br>(-3.01 - -0.05) | 0.057  |  |
| 条件<br>[LLMペルソナあり]          | -2.37<br>(-3.76 - -0.98) | 0.001  |  | -1.87<br>(-3.40 - -0.34) | 0.016  |  |
| ラウンド                       |                          |        |  | 0.40<br>(0.31 - 0.48)    | <0.001 |  |
| 条件<br>[LLMペルソナなし]<br>×ラウンド |                          |        |  | -0.05<br>(-0.16 - 0.07)  | 0.439  |  |
| 条件<br>[LLMペルソナあり]<br>×ラウンド |                          |        |  | -0.09<br>(-0.21 - 0.02)  | 0.121  |  |
| Conditional R2             | 0.599                    |        |  | 0.632                    |        |  |

表 1: 群間比較

| Predictors           | Human条件<br>公共財への支出    |        | LLMペルソナなし条件<br>公共財への支出 |        | LLMペルソナあり条件<br>公共財への支出 |        |
|----------------------|-----------------------|--------|------------------------|--------|------------------------|--------|
|                      | estimates             | p      | estimates              | p      | estimates              | p      |
| (Intercept)          | 6.73<br>(5.38 - 8.18) | <0.001 | 5.49<br>(4.22 - 6.77)  | <0.001 | 4.95<br>(3.61 - 6.30)  | <0.001 |
| 前期自身の<br>公共財への支出     | 0.45<br>(0.39 - 0.51) | <0.001 | 0.44<br>(0.37 - 0.50)  | <0.001 | 0.44<br>(0.38 - 0.51)  | <0.001 |
| 前期メンバーの<br>公共財への支出合計 | 0.05<br>(0.01 - 0.09) | 0.022  | 0.07<br>(0.02 - 0.11)  | 0.002  | 0.07<br>(0.03 - 0.12)  | 0.002  |
| Conditional R2       | 0.570                 |        | 0.453                  |        | 0.562                  |        |

表 2: 支出額決定ロジック比較

## 4.2 群間比較

繰り返し公共財ゲームにおける条件比較手法として一般的な一般線形混合モデルを活用して、公共財への支出を目的変数として群間比較した結果を表 1 に示す。

表 1 左のモデルは、純粋な群間比較の結果を示しており、Human 条件に比べて LLM ペルソナなし条件では支出が 1.73 ポイント低いこと、LLM ペルソナあり条件では支出が 2.37 ポイント低いことが確認された。

表 1 右のモデルはラウンドと群間の交互作用項も説明変数に含めたモデルである。Human 条件に比べて LLM ペルソナなし条件では支出が 1.48 ポイント低い、LLM ペルソナあり条件では支出が 1.87 ポイント有意に低いことが確認された。また、ラウンドが経過するに従い支出は有意に増

加すること確認された。最後にラウンドとの交互作用項については有意な差が見られなかった。

## 4.3 支出額決定ロジック比較

公共財への支出は「前ラウンドの自身の支出」と「前ラウンドのメンバーの支出合計」によって決まると仮定し、一般線形混合モデルで分析した結果を表 2 に示す。すべての条件で「前ラウンドの自身の支出」と「前ラウンドのメンバーの支出合計」の係数が概ね同じ値であることが確認された。この結果は、LLM が人間と同様に自身と他者の行動を踏まえて支出額を決定していることを示唆している。

## 5. 考察

### 5.1 シミュレーション精度は向上するのか?

支出の推移において LLM ペルソナあり条件は LLM ペルソナなし条件よりも Human 条件の推移から大きく乖離している点や、群間比較において LLM ペルソナあり条件が LLM ペルソナなし条件よりさらに支出が低い点から、ペルソナ情報を付与することによってシミュレーション精度は向上せず、むしろ悪化することが確認された。

### 5.2 なぜシミュレーション精度が悪化したのか?

支出額決定ロジック比較においては、各条件間に大きな差はみられなかったことから、群間比較での差異がシミュレーション精度悪化の要因であると考えられる。群間比較における主な違いは、公共財への支出のベースラインが有意に低い点である。性別や年齢などのデモグラフィックなペルソナ情報を付与した結果、支出のベースラインが低下し、シミュレーション精度が悪化したと考えられる。デモグラフィック情報だけでなく、互惠性などのサイコグラフィック情報も付加した先行研究は<sup>[3]</sup>、シミュレーション精度の向上が認められている。本研究では、デモグラフィック情報のみを使用したことがシミュレーション精度悪化の一要因と考えられる。

## 6. おわりに

本研究では、LLM による実験参加者の代替精度をさらに高めるために、LLM にペルソナ情報として実験参加者の情報を付与することで、シミュレーション精度が向上するかどうかを検討した。その結果、ペルソナ情報を与えることで精度は向上せずに悪化することが確認された。この結果は、ペルソナ情報として性別や年齢などのデモグラフィックな情報のみを付与したために、精度が悪化した可能性が考えられる。今後は、デモグラフィックな情報に加えてサイコグラフィックな情報を LLM に付与することによって精度が向上するかどうかを検討したいと考えている。

### 参考文献

- [1] Qiaozhu Mei, Yutong Xie, Walter Yuan, Matthew O. Jackson, "A Turing test of whether AI chatbots are behaviorally similar to humans", In Proceedings of the National Academy of Sciences, 121(9), e2313925121 (2024)
- [2] 日室聡仁, 後藤晶, "実験参加者を大規模言語モデルで代替できるのかー繰り返し公共財ゲームを用いた人間と大規模言語モデルの比較ー", 行動経済学会第 18 回大会
- [3] 北代絢大, 深澤祐, 西野成昭, "最後通牒ゲームの大規模言語モデルを用いたシミュレーション: バイアス補正による実験結果の再現性検証", 行動経済学会第 18 回大会