

# 対話型人工エージェントは個性を持つか?: Big-5 を付与した大規模言語モデルの応答の観察

## Can Large Language Models Accurately Reflect Assigned Personalities?

田中 葉月<sup>1\*</sup> 飯田 愛結<sup>1</sup> 福田 聡子<sup>1</sup> 中島 亮一<sup>2</sup> 大澤 正彦<sup>1</sup>

Hazuki Tanaka<sup>1</sup>, Ayu Iida<sup>1</sup>, Satoko Fukuda<sup>1</sup>, Ryoichi Nakashima<sup>2</sup>, Masahiko Osawa<sup>1</sup>

<sup>1</sup> 日本大学

<sup>1</sup> Nihon University

<sup>2</sup> 京都大学

<sup>2</sup> Kyoto University

**Abstract:** 本研究では、大規模言語モデル (LLM) にパーソナリティ特性を与えることで、特定のパーソナリティを持つ対話型人工エージェントの創出を試みた。Big-5 尺度の数値を LLM に与えるプロンプトに組み込むことで、そのパーソナリティ特性を反映した応答を生成できるかを検証する。具体的には、特定の数値データを与えられた LLM が Big-5 尺度の項目にどのように回答するかを観察・分析する。結果は、付与する数値を高くするほど TIPI-J の測定スコアも高くなり、LLM が付与された数値に応じて性格特性を反映させる能力を持っていることを示した。ただし、付与した数値の反映しやすさは、特性によって異なることが観察された。

## 1 はじめに

パーソナリティは人間の行動パターンを決める要因の一つとして考えられており、人間同士のコミュニケーションにおいて重要な役割を果たしている。具体的には、パーソナリティは、人間が情報を処理し、感情を表現し、他者との関係を築く方法に影響を及ぼす。人間が様々なパーソナリティを持つことは、個人間の独特な交流や深い感情的結びつきの形成を可能とし、コミュニケーションの質を向上させる。結果的に、コミュニケーションによる満足度が上がる可能性がある。

これは人間と人工エージェントとのコミュニケーションにも当てはまる。現代の技術発展により、私たちは人工エージェントと対話する場面も増えてきており、この対話の質を高めることはますます重要になっている。従来の研究では、コミュニケーションプロセスにおける視覚的特性の利用 [1] や、個人の抱える問題の性質に応じた励ます発話の調整 [2] など、特定の側面に焦点を当てた研究がなされている。

しかしながら、個人固有の属性—性格、価値観、などを包括的に反映することの難しさが課題として指摘されている [2]。本研究では、チャットボットなどの

対話型人工エージェントに、人間のようなパーソナリティを与えることができるかを検討することで、この課題に取り組む。

近年、対話型人工エージェントを実現する技術として、大規模言語モデル (Large Language Model: LLM) が世界的に注目を集めている。LLM においても、人間とのコミュニケーションにおけるパーソナリティの重要性は無視できない。現在の LLM は、多様なテキスト生成タスクにおいて優れた性能を示している [3] が、ユーザーとのインタラクションにおいてパーソナリティを反映させた応答が十分にできるかは定かではない。

LLM にパーソナリティを与える研究として、Serapio-García ら [4] は、Big-5 パーソナリティを活用した検証を行っている。ここでは、自然言語を用いてパーソナリティを形成させることが試みられている。一方、もし LLM にパーソナリティを数値として付与することができるならば、パーソナリティに関するより定量的な議論を行うことができるだけでなく、大規模言語モデルをモジュールとして様々なアーキテクチャと組み合わせる方法論 [5, 6] の応用も広がる。しかし、数値化されたパーソナリティ特性を割り当てることが可能かの検証を試みた研究は著者らの知る限り存在しない。

本研究では、LLM へ組み込むパーソナリティ特性を数値として付与する。その際に、Big-5 パーソナリティ特性の得点を付与する数値として採用する。Big-5 とは人

\*連絡先: 日本大学文理学部  
〒156-8550 東京都世田谷区桜上水 3-25-40  
E-mail: chha21048@g.nihon-u.ac.jp

間のパーソナリティを5つの因子に分けて考える方法である。日本語版 Ten Item Personality Inventory (TIPI-J) は Big-5 モデルに基づいたパーソナリティ特性を簡易的に測定する尺度であり、本研究では、TIPI-J で得られる数値から事前に定めたパーソナリティ特性の数値データを LLM に入力し、その LLM に TIPI-J 尺度に回答させ、その結果を出力させる。これらの値を比較することで、数値データに基づいたパーソナリティ付与が実現可能かを検証する。このプロセスを通じて、LLM が人間らしいコミュニケーションを実現するために必要とするパーソナリティ特性を、いかに正確に反映できるかを確かめる。

## 2 背景

### 2.1 大規模言語モデルのコミュニケーション能力の現状と課題

近年の LLM の開発と応用は、人工知能 (AI) 分野における注目すべきトピックの 1 つである。これらのモデルは、膨大なテキストデータの学習を通じて人間の言語を理解し自然な応答テキストを生成する能力を持つように設計されている。これにより、翻訳、要約、質問応答、チャットボットといった多様なアプリケーションが可能になり、教育、エンターテインメント、プログラミングの補助など、幅広い領域での応用が期待されている。さらに、与えられたコンテキストに基づく応答生成能力により、LLM は人間とのインタラクションにおいても重要な役割を果たすようになってきている。

しかしながら、LLM には未解決の課題も多く存在する。現状の LLM は、人間関係の把握や複雑な状況把握についてはあまり高い精度を発揮できない [7]。また、LLM が提供する情報がしばしば誤っていたり、発話内容が一貫しないことがあるという問題もある [3]。それにより、ユーザーの混乱を招くだけでなく、LLM の信頼性に対する疑念も引き起こしている。

#### 2.1.1 心理学における Big-5 パーソナリティおよびその測定方法としての TIPI-J の活用

Big-5 とは、パーソナリティ特性を 5 つの主要な因子で表すものであり、個人の行動、嗜好、感情反応の違いを理解するための心理学における基礎的な枠組みとして広く受け入れられている。5 つの因子を表 1 に示す。

TIPI-J [8] は、短縮型パーソナリティの質問紙である。10 個の質問項目で、外向性、協調性、勤勉性、神経症傾向、開放性の 5 つの主要な性格因子を簡易的に測定する。各パーソナリティ特性について肯定的および否

表 1: Big-5 モデルの 5 つの因子

性格特性	
外向性	社交性、活動的であること、積極性などを示す。外向的な人は、人との交流を好み、エネルギッシュで楽観的な傾向がある。
協調性	他人に対する共感や友好性を示す。協調性が高い人は、他人との調和を重んじ、協力的である。
誠実性	計画性、責任感、自己規律などを示す。誠実性が高い人は、組織的で信頼性があり、仕事や目標達成に真剣に取り組む。
神経症傾向	感情の安定性とストレス耐性の度合いを示す。神経症傾向が高い人は、ストレスや不安を感じやすく、感情的に不安定になりやすい。
開放性	新しい経験への開放性、創造性、好奇心などを示す。開放性が高い人は、新しいアイデアや活動に対して好奇心が強く、創造的である傾向がある。

定的な表現の両方を含む 2 つの質問で構成されている。自身が各項目にどの程度当てはまるかを 7 点リッカート尺度で評価する。各パーソナリティ特性について、肯定的な質問の得点と否定的な質問（逆転項目）の得点を 8 から引いたものを平均することで、その特性を数値で表すことができる。つまり、TIPI-J において、各パーソナリティ特性は 1 から 7 の範囲の値をとる。この数値は、特定のパーソナリティ特性をどの程度強く持ち合わせているかを示している。

### 2.2 LLM にパーソナリティを与える既存研究

LLM にパーソナリティを与える研究として、Serapio-García ら [4] は、プロンプティングに基づくシミュレーションを通じて、LLM のパーソナリティを独自の手法で評価している。

彼らの研究では、Big-5 特性を分類するために IPIP-NEO と BFI の 2 つの尺度を使用してパーソナリティを測定した。それによって、LLM のシミュレートされたパーソナリティの信頼性を定量的に評価することが可能となった。結果として、LLM は設定に基づくパーソナリティを再現できたことを報告している。Serapio-García らの研究 [4] の特徴として、LLM に対して性格

特性を自然言語で記述したプロンプトで与えていたことが挙げられる。

### 3 数値データによる TIPI-J のパーソナリティ特性を付与

#### 3.1 実験方法

本実験では、パーソナリティ特性を数値データとして LLM に与えることで、その LLM がパーソナリティを反映した応答ができるかどうかを調べる。

##### 3.1.1 実験における大規模言語モデルの選定

本研究では、LLM として OpenAI 社の提供する API である GPT-4 Turbo を採用する。この選定は、GPT-4 Turbo がその前身であるモデルに比べて、コミュニケーションの自然さと応答の適切性を大幅に向上させているためである。GPT-4 Turbo は、先進的な自然言語処理技術を駆使し、大量のテキストデータから学習を行った結果、人間らしいテキストを生成する能力が特に強化されたモデルである。また、GPT-4 Turbo は、以前のモデルよりも幅広い知識と、文脈に敏感な応答能力を備えていることが特徴である。

##### 3.1.2 数値データの用意

パーソナリティ特性の項目 (外向性・協調性・勤勉性・神経症傾向・開放性) について、LLM に与える数値を用意する。具体的には数値 2(低), 4(中), 6(高) の全組み合わせを用いる。すなわち全体で  $3^5 = 243$  通りの異なるパーソナリティプロファイルを生成する。

##### 3.1.3 LLM への数値付与と TIPI-J 診断テストの実施

本実験では、Python で実装した実験実施プログラムを利用する。プログラムから読み込んだ 243 件のパーソナリティプロファイルの数値データを 1 件ずつ取り出し、OpenAI 社が提供する API を利用して、以下に示すプロンプトとともに LLM に入力した。

これは BIG5 の 5 項目のパーソナリティの数値です。  
これらの数値は 1 から 7 のスケールで、低いほどその特性が低いこと、高いほどその特性が高いことを意味します。  
あなたは付与された数値を基にパーソナリティを持ち合わせる設定です。

次に、日本語版 TIPI-J の質問項目 [8] と併せて以下プロンプトを与える。

今与えられた質問を、あなたの設定に基づいて [1: 全く違うと思う 2: おおよそ違うと思う 3: 少し違うと思う 4: どちらでもない 5: 少しそう思う 6: まあまあそう思う 7: 強くそう思う] から、最も近いと思う番号を選んでください。それぞれの質問に対して、「質問番号、あなたの回答番号」の形式で教えてください。  
例えば、「1,3」と答えることになります。私が指示した内容以外は出力してはいけません。それでは、質問に対するあなたの回答をお願いします。

なおこれらのプロンプトは、LLM に与えるプロンプトとして一般に適切と思われるように作成した。

#### 3.2 結果

各パーソナリティ特性の測定結果を図 1 に示す。

全てのパーソナリティ特性において、数値 6 を付与した場合の中央値と平均値が最も高く、数値 2 を付与した場合は最も低いことが読み取れる。特に数値 2、数値 4 を付与した場合、中央値は各特性で付与値よりもスコアの方が高くなっている。数値 6 を付与した場合の中央値は協調性以外全て 6 である。

四分位範囲 (第 1 四分位数から第 3 四分位数までの範囲) は特性によって異なり、特に数値 2 を付与した外向性、勤勉性、神経症傾向のスコアが広範囲に分布していることが観察される。協調性については、付与値ごとの中央値の差が小さい。全付与値の四分位範囲で重なる部分が存在し、数値 2 の中央値は 4、数値 6 の中央値は 5.5 である。一方で、開放性と勤勉性の四分位範囲は全て重なりを持たず、特に開放性は範囲が小さくスコアの振れ幅が小さいことが分かる。数値 6 を付与した場合のばらつきは協調性以外では他の数値を

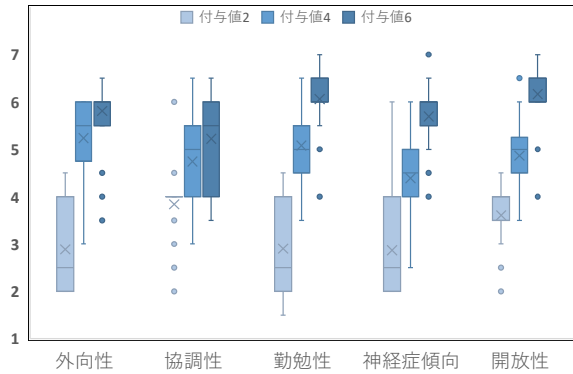


図 1: TIPI-J 測定結果の付与値ごとのばらつき

ひげの上端から下端までを4つの区間に区切られていて、それぞれの区間が全体の1/4のデータを収容している。箱中の線は中央値、箱中の×印は平均値を示している。ひげの外側にプロットされた点は外れ値を示しており、「第3四分位数+四分位範囲×1.5」以上のデータ、「第1四分位数-四分位範囲×1.5」以下のデータと定義する。

付与した場合よりも小さい。前述の通りこれらは中央値が設定した値と同じ6であることも踏まえると、正確に値が反映されているといえる。

また、いくつかのパーソナリティ特性において外れ値が観察されており、これはLLMが時々付与値とは離れたスコアの反応を示す可能性があることを示している。

## 4 考察

本研究により得られた結果は、LLMにパーソナリティ特性の数値を付与することが、TIPI-J尺度の質問に対するLLMの回答に影響を及ぼすことを示唆している。各パーソナリティ特性に付与した数値が高いほど、LLMのTIPI-Jの質問項目への回答から算出される各特性のスコアが高くなる傾向が観察された。これは、LLMが付与された数値を基にして一貫したパーソナリティ特性を表出できる能力を有していることを示しており、パーソナリティの数値付与が有効であると考えられる。

しかしながら、特に数値2、数値4を付与した場合、ほとんどのパーソナリティ特性で付与値よりもスコアの方が高くなっている、つまり回答が上振れする傾向があり、低い数値の特性を強調するのは困難かもしれない。一方で、数値6を付与した場合、協調性以外の全パーソナリティ特性の中央値が6であったことから特性の高い傾向は反映しやすいと考えられる。また、協調性の全付与値の四分位範囲で重なる部分が存在し、付与値ごとのスコアの差が比較的小さいことから、本研究で行った数値の付与によってLLMに協調性を反映させるのは難しい可能性がある。

本研究では、LLMに数値を用いたパーソナリティ特性を付与し、そのLLMのTIPI-Jへの回答を測定をした。これはパーソナリティ特性数値の付与が、質問の回答に反映されるかを検証するものであった。本研究では、その有効性はある程度確認された。LLMがユーザーとのインタラクションをする際に、LLMが数値によるパーソナリティ特性を保持することで、それを反映したテキストの生成ができるかどうかについては、さらに検証する必要がある。また、LLMと人間とのより自然なインタラクションを実現するためには、パーソナリティ特性のより洗練された数値付与や、パーソナリティ特性をより正確に反映させるための追加的なパラメータやアルゴリズムの開発が求められる。これらの取り組みは、LLMが人間のパーソナリティの多様性をより再現できる存在になるための基盤を築くことに貢献すると期待できる。

## 5 おわりに

本研究では、LLMに対する数値によるパーソナリティ特性の付与が、そのLLMのTIPI-Jへの回答結果にどのように影響を与えるかを検証した。実験結果は、LLMが与えられた数値に基づいてパーソナリティ特性をある程度反映した出力ができることを示したが、完全ではないことも明らかになった。この研究は、人間のパーソナリティの複雑さを模倣するLLMの可能性を示す出発点となる。LLMの応用範囲を拡大し、より人間らしいインタラクションを実現するための研究は、今後も継続される必要がある。今後の研究では、文脈への敏感さ、パーソナリティ特性数値付与によるテキスト生成への影響、LLMをモジュールとして様々なアーキテクチャとの組み合わせなどを検証する必要があるだろう。本研究の知見は、人間とLLMのコミュニケーションの質を向上させるシステム開発に貢献すると考えている。

## 参考文献

- [1] 藤堂 健世, 高橋 聡, 吉川 厚, 山村 雅幸, “対話エージェントの特性の相互作用による自己開示への影響の実験”, HAIシンポジウム, 2023.
- [2] 向 凌萱, 菊池 浩史, 楊 潔, 菊池 英明, “悩みの種類に応じた対話ロボットによる励まし発話の励まし効果の調査”, HAIシンポジウム, 2023.
- [3] 杉崎 光一, 阿部 雅人, 全 邦釘, “大規模言語モデルの専門領域への適用に関する検討”, AI・データサイエンス論文集 4 巻 3 号, 2023.

- [4] Greg Serapio-Garcia, Mustafa Safdari, Clement Crepy, Luning Sun, Stephen Fitz, Peter Romero, Marwa Abdulhai, Aleksandra Faust, Maja Mataric, “Personality Traits in Large Language Models”, arXiv:2307.00184v3 [cs.CL], 2023.
- [5] 飯田 愛結, 奥岡 耕平, 福田 聡子, 大森 隆司, 大澤 正彦, “ChatGPT を用いた認知アーキテクチャの構想—ユーザーの発話と発話意図に乖離があるケースを例に”, 情報処理学会研究報告, 2023.
- [6] 飯田 愛結, 阿部 将樹, 奥岡 耕平, 福田 聡子, 大森 隆司, 中島 亮一, 大澤 正彦, “意図を読む AI の実現に向けて: 対話型生成 AI と他者モデルの統合を例に”, HAI シンポジウム, 2024.
- [7] 狩野 芳伸, “生成系 AI と自然言語処理研究”, デジタル関係制度改革検討会 デジタル法制ワーキンググループ 第一回会合, 2023.
- [8] 小塩真司, 阿部晋吾, カトローニ ピノ, “日本語版 Ten Item Personality Inventory (TIPI-J) 作成の試み”, パーソナリティ研究, 2012.