TIPI-J に基づく大規模言語モデルを用いたゲーム NPC 人格対話生成に関する研究

高 信哲　小林 篤史

近年，大規模言語モデル（LLM）の発展に伴い，ゲーム NPC に自然な対話能力を付与する研究が活発に行われている．しかし，NPC の人格を一貫して表現することは依然として課題であり，同一の人格設定であっても，使用する LLM によって対話内容や人格表現に差が生じる可能性がある．本研究では，TIPI-J に基づいて NPC の人格特性を定量的にモデル化し，その人格情報をプロンプトに反映した対話生成手法を提案する．さらに，DeepSeek, Gemini，Claude を対象として，同一条件下で対話生成実験を実施し，各モデルについて 100 回の対話実験を行い，NPC の人格一致率を評価する．これらの実験結果を基に，各 LLM における人格表現の一貫性および特徴を比較・分析することで，TIPI-J を用いた NPC 人格設計の有効性を検証するとともに，ゲーム NPC 対話生成に適した LLM の特性を明らかにし，NPC 人格設計に関する基礎的知見を提供することを目的とする．

With the recent development of large language models (LLMs), research on providing game NPCs with natural conversational abilities has become increasingly active. However, maintaining consistent personality expression in NPCs remains a challenging issue, as differences in dialogue content and personality representation may occur depending on the LLM used, even when the same personality settings are applied. In this study, we propose a dialogue generation method that quantitatively models NPC personality traits based on the Japanese version of the Ten Item Personality Inventory (TIPI-J) and incorporates the obtained personality information into prompts. Furthermore, we conduct dialogue generation experiments under identical conditions using three LLMs: DeepSeek, Gemini, and Claude. For each model, 100 dialogue experiments are performed to evaluate the personality consistency of NPCs. Based on the experimental results, we compare and analyze the consistency and characteristics of personality expression among different LLMs. Through this analysis, we aim to verify the effectiveness of TIPI-J-based NPC personality design and clarify the characteristics of LLMs suitable for game NPC dialogue generation, providing fundamental insights into NPC personality design.

1　はじめに

近年，大規模言語モデル（LLM）の発展により，自然言語処理技術は大きく進歩している [1] [2]．また，LLM を対話システムや知的エージェントへ応用する研究も進められている [3] [4]．特にゲーム分野では，LLM を利用してキャラクターらしい行動や対話を生成する試みが行われている [4]．

一方で，本研究では，LLM を用いた NPC 対話生成において，キャラクター固有の人格を一貫して表現することを課題として扱う．ゲーム NPC には自然な文章生成だけではなく，固有の性格や行動傾向を維持する能力が求められる．同一の人格設定を与えた場合でも，利用する LLM によって人格表現に差が生じる可能性を考慮し，心理学的な人格モデルに基づいて人格情報を定量化した上で，異なる LLM の人格表現能力を比較する．

人格を定量的に表現する手法として，Big Five 人格モデルが広く利用されている．Big Five は，外向性（Extraversion），協調性（Agreeableness），勤勉性（Conscientiousness），神経症傾向（Neuroticism），開放性（Openness）の 5 因子によって人格特性を表現するモデルである [5]．また，Big Five を簡略化し

A Comparative Analysis of Personality Expression in Game NPC Dialogue Generation Using Large Language Models Based on TIPI-J.

Xinzhe Gao, Atsushi Kobayashi, Daiichi Institute of Technology, Department of Information and Data Science, Daiichi Institute of Technology.

た人格尺度として TIPI が提案されており [6]，日本語版である TIPI-J についても信頼性および妥当性が検証されている [7].

対話生成における話者の一貫性を扱う研究や，心理尺度を用いて LLM ベースのロールプレイエージェントの人格忠実度を評価する研究も行われている [8] [9]. しかし，同一人格設定を異なる LLM へ入力した場合に，各モデルがどの程度人格特性を維持できるかについては十分に明らかになっていない.

著者らはこれまで，TIPI-J を用いて NPC 人格を定量化し，LLM による対話生成へ反映する手法について基礎的な検討を行った [10]. しかし，異なる LLM 間で人格表現の一貫性を比較した検証は十分ではない.

そこで本研究では，TIPI-J に基づいて NPC 人格をモデル化し，その人格情報を LLM へのプロンプトへ反映することで，DeepSeek，Gemini，Claude における NPC 人格対話生成能力を比較分析する．各 LLM について 100 回の対話生成実験を行い，人格一致率に基づいて人格表現の一貫性を評価する.

2 関連研究

2.1 LLM を用いたゲーム NPC 対話生成

ゲーム NPC の対話には，ゲームデータに基づく分岐型の対話木が用いられる例があり [11]，近年では LLM を利用してゲーム世界の設定やキャラクター情報を反映した対話を生成する研究も進められている [11] [4].

また，Generative Agents や Character-LLM など，LLM を利用して人間らしい行動やキャラクター性を生成する研究も行われている [3] [4]．一方で，ロールプレイエージェントが対象キャラクターの人格をどの程度忠実に再現できるかは重要な評価課題である [9].

2.2 人格モデルに基づく対話エージェント

対話エージェントや仮想キャラクターに一貫した人格を付与するため，心理学に基づいた人格モデルを利用する研究が行われている.

代表的な人格モデルである Big Five は，外向性（Extraversion），協調性（Agreeableness），勤勉性（Conscientiousness），神経症傾向（Neuroticism），開放性（Openness）の 5 因子によって人格特性を表現する [5]．また，TIPI は少ない質問項目から Big Five を推定できる簡易人格尺度であり [6]，日本語版 TIPI-J についても信頼性および妥当性が検証されている [7].

対話生成における話者の一貫性を扱う研究 [8] や，心理尺度を用いて LLM ベースのロールプレイエージェントの人格忠実度を評価する研究 [9] が行われている.

2.3 LLM における人格表現比較

LLM ベースのロールプレイエージェントを複数のモデル条件で評価した研究では，人格忠実度に差がみられることが報告されている [9]．そのため，同一の人格情報を入力した場合に，各 LLM がどの程度一貫した人格表現を生成できるかを評価することは重要である.

しかし，心理学的な人格モデルに基づいて NPC 人格を定量化し，同一条件下で複数 LLM 間の人格表現を比較する研究は十分ではない．特に，TIPI-J によって定義された人格特性を用いて，異なる LLM 間の人格一致率を定量的に比較する研究は限られている.

そこで本研究では，TIPI-J に基づく人格情報を DeepSeek，Gemini，Claude へ入力し，各 LLM による NPC 対話生成結果を比較することで，LLM ごとの人格表現の一貫性および特徴を分析する.

3 提案手法

3.1 システム概要

本研究では，TIPI-J に基づいて NPC 人格を定量化し，その人格情報を LLM へのプロンプトへ反映することで，異なる LLM における NPC 人格対話生成の比較分析を行う．提案手法の構成を図 1 に示す.

まず，TIPI-J の評価結果から NPC 人格特性を算出し，自然言語形式の人格情報へ変換する．その後，同一の人格設定および対話条件を DeepSeek，Gemini，Claude へ入力し，生成された対話結果から人格一致率を算出することで，各 LLM の人格表現能力を比較する.

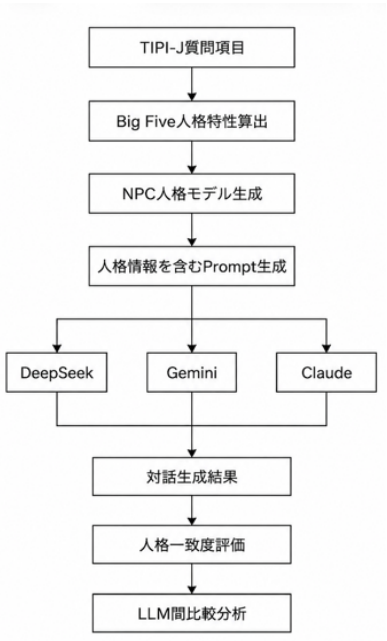

図 1　提案手法の全体構成

3.2　TIPI-J に基づく NPC 人格モデル

本研究では，NPC 人格を定量的に表現するため，TIPI-J を利用する．TIPI-J は，Big Five 人格モデルに基づき，10 項目の質問から 5 つの人格因子を推定する簡易人格尺度である [7]．TIPI-J では，各質問項目について 7 段階評価を行い，外向性（E），協調性（A），勤勉性（C），神経症傾向（N），開放性（O）の 5 因子を算出する [7]．本研究では，算出された 5 因子を NPC 固有の人格パラメータとして利用する．これにより，主観的な文章設定ではなく，心理学的尺度に基づいた定量的な人格モデルを構築することが可能となる．

$$E = \frac{Q_1 + (8 - Q_6)}{2}, \quad A = \frac{(8 - Q_2) + Q_7}{2}, \tag{1}$$

$$C = \frac{Q_3 + (8 - Q_8)}{2}, \quad N = \frac{(8 - Q_4) + Q_9}{2}, \tag{2}$$

$$O = \frac{Q_5 + (8 - Q_{10})}{2}. \tag{3}$$

3.3　人格情報を用いたプロンプト生成

LLM へ NPC 人格を反映させるため，TIPI-J から得られた人格情報を自然言語形式へ変換し，プロンプトとして入力する．例えば，外向性が高い NPC の場合は「積極的に発言し，他者との交流を好む性格」，神経症傾向が高い NPC の場合は「慎重で不安を感じやすい性格」として表現する．生成プロンプトには，NPC 名，Big Five 人格特性，人格特徴，対話条件，役割設定を含め，各 LLM へ同一形式の人格情報を入力する．

3.4　LLM を用いた対話生成方法

本研究では，異なる LLM における人格表現能力を比較するため，DeepSeek，Gemini，Claude の 3 種類を対象とする．DeepSeek-V3 は Mixture-of-Experts 構成を採用し，効率的な推論と学習を重視して設計された LLM である [12]．Gemini は Google が開発したマルチモーダル LLM であり，テキストに加えて画像，音声，動画を扱い，複雑な推論課題を含む評価が報告されている [13]．Claude は Anthropic が開発した LLM ファミリーであり，言語理解や推論を含む複数の能力について評価結果が報告されている [14]．特に複数ターンの対話において，初期設定や会話履歴を維持する能力が期待されるため，本研究の人格維持評価に適した比較対象とした．各 LLM へ同一の NPC 人格設定および対話シナリオを入力し，生成された対話ログを人格一致率の算出に利用する．

4　実験

4.1　実験目的

本実験では，TIPI-J に基づいて設定した NPC 人格情報を各 LLM へ入力し，DeepSeek，Gemini，Claude における人格表現の一貫性を比較することを目的とする．

4.2　実験条件

本実験では，使用する LLM 以外の条件による影響を避けるため，NPC 人格設定，対話シナリオ，プロンプト形式，評価方法を統一した．実験条件を表 1 に示す．

また，実験手順を図 2 に示す．各 LLM に対して同一条件で 100 回の対話生成を実施し，生成された対話ログを人格一致率の算出に利用した．

表 1　実験条件

項目	内容
LLM	DeepSeek, Gemini, Claude
人格モデル	TIPI-J
人格因子	Big Five 5 因子
NPC 数	5 名
対話シナリオ	地震発生後の避難状況
対話形式	複数 NPC による逐次対話
実験回数	各 LLM 100 回
評価指標	NPC 人格一致率

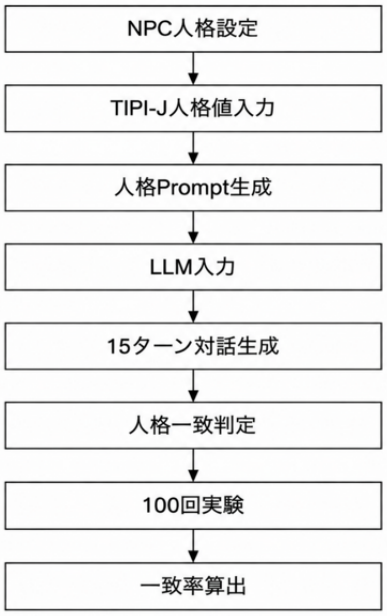

図 2　実験手順

表 2　人格設定例

NPC	E	A	C	N	O
Aki	0.83	0.75	0.50	0.17	0.83
Mika	0.33	0.75	0.83	0.50	0.17
Yui	0.83	0.50	0.50	0.83	0.83
Rina	0.17	0.17	0.83	0.17	0.50
Kana	0.50	0.75	0.17	0.83	0.83

4.3　NPC 人格設定

本研究では，5 種類の NPC を設定し，各 NPC に異なる人格特性を付与した．表 2 に NPC ごとの人格設定を示す．

各 NPC の人格パラメータは，研究者が TIPI-J の 10 項目に対する回答値を 1〜7 段階で設定し，TIPI-J の計算方法 [7] に基づいて Big Five の 5 因子を算出することで生成した．算出された外向性（E），協調性（A），勤勉性（C），神経症傾向（N），開放性（O）の値を NPC 固有の人格パラメータとして利用した．これにより，主観的な文章による人格設定ではなく，心理学的尺度に基づいた定量的な人格モデルを構築した．

4.4　対話生成条件

本実験では，地震発生後の避難状況を想定した対話シナリオを設定した．具体的には，地震直後に複数の NPC が建物内に残された状況を想定し，周囲の安全確認，避難経路の判断，負傷者への対応，今後の行動方針について議論する．各 NPC は設定された人格特性に基づいて発言を行い，例えば外向性が高い NPC は積極的な発言を行い，協調性が高い NPC は他者への配慮を示す発言を行う．各 LLM に同一のシステムプロンプト，NPC 人格情報，シナリオ条件を入力し，15 ターンの対話を生成した．

4.5　人格一致率の評価方法

生成された対話について，各 NPC の人格特性が維持されているかを評価した．各 NPC について 15 回の発話を対象とし，設定された人格特性と一致する発話数を確認した．15 回の発話のうち 8 回以上で人格特性との一致が確認された場合，当該 NPC の対話を成功と判定した．また，初回発話において設定人格と明確に矛盾する場合は，不一致として判定した．各 LLM について 100 回の実験を実施し，NPC ごとの成功回数を集計した．NPC i の人格一致率 R_i は，以下の式により算出した．

$$R_i = \frac{S_i}{N} \times 100 \tag{4}$$

ここで，R_i は NPC i の人格一致率（%），S_i は NPC i が人格一致と判定された実験回数，N は総実験回数を表す．本研究では，各 LLM について $N = 100$ とした．さらに，各 LLM の全体的な人格一致率 R_{LLM} は，5 名の NPC の人格一致率の平均値として，以下の式により算出した．

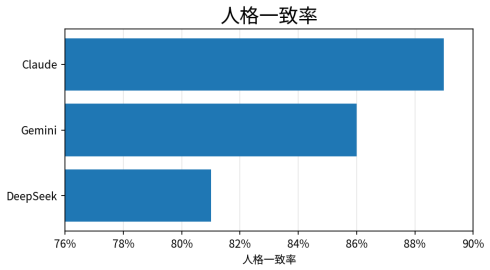

図 3　各 LLM における人格一致率

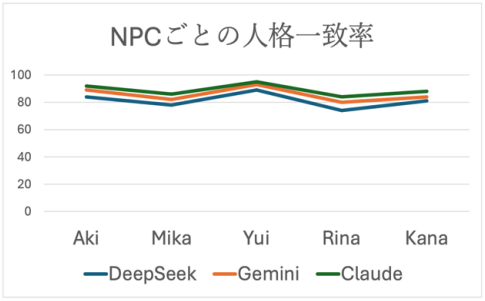

図 4　NPC ごとの人格一致率

$$R_{\mathrm{LLM}} = \frac{1}{M}\sum_{i=1}^{M} R_i \tag{5}$$

ここで，R_{LLM} は各 LLM の全体的な人格一致率（%），R_i は NPC i の人格一致率（%），M は NPC 数を表す．本研究では $M = 5$ とした．算出結果は小数第 1 位を四捨五入して整数値で示した．

5　実験結果

5.1　人格一致率の比較

本実験では，DeepSeek，Gemini，Claude の 3 種類の LLM を対象として，各モデルについて 100 回の対話生成実験を実施した．生成された対話結果に対して，第 4 章で述べた評価基準に基づき人格一致率を算出した．図 3 に各 LLM における人格一致率を示す．

図 3 より，すべての LLM において 80%以上の人格一致率が確認された．また，LLM 間で比較すると，Claude が最も高い人格一致率を示し，次いで Gemini，DeepSeek の順となった．

5.2　NPC ごとの人格一致率分析

次に，NPC ごとの人格一致率について分析した．図 4 に NPC ごとの人格一致率を示す．

図 4 より，NPC によって人格一致率に差異が確認された．また，すべての NPC において Claude が最も高い人格一致率を示した．一方で，Rina および Mika では，他の NPC と比較して人格一致率が低下する傾向が確認された．

5.3　LLM ごとの人格表現特徴

各 LLM における人格表現の特徴を分析した．DeepSeek では 81%，Gemini では 86%，Claude では 89%の人格一致率を示した．DeepSeek では一部の対話において人格特徴が弱まる傾向が確認された一方，Claude では複数ターンにわたって NPC 人格を維持した発話が確認された．Gemini では両者の中間的な結果を示し，設定された人格情報を反映した対話生成が確認された．

6　考察と結論

6.1　LLM による人格表現能力の差異

本研究では，同一の TIPI-J 人格情報および対話条件を用いて，DeepSeek，Gemini，Claude の人格表現能力を比較した．その結果，すべての LLM において 80%以上の人格一致率が確認され，TIPI-J に基づく人格情報をプロンプトとして利用することで，一定程度の人格制御が可能であることが示された．

特に，Claude は 89%の人格一致率を示し，本実験条件では最も高い結果となった．この要因として，15 ターンの対話における文脈理解能力や初期設定の維持能力が影響した可能性が考えられる．

6.2　NPC 人格特性による影響

NPC ごとの結果から，設定される人格特性によって人格一致率に差異が生じることが確認された．外向性や開放性など，発話として表現しやすい人格特徴は LLM によって反映されやすい一方，慎重な判断や思考傾向などの特徴は表現が難しい傾向が確認された．

このことから，LLM による人格表現能力は，モデル性能だけではなく，設定する人格特性の種類にも影響を受けることが示された．

6.3 TIPI-J を用いた NPC 人格設計の有効性

本研究では，TIPI-J を利用して NPC 人格を定量化し，LLM への入力情報として利用した．実験結果から，心理学的尺度に基づく人格モデルを用いることで，異なる LLM に対して比較可能な NPC 人格設定を実現できることが示された．

また，TIPI-J による定量的な人格表現は，LLM を利用したゲーム NPC 設計において，キャラクター性を管理するための有効な手法であると考えられる．

6.4 今後の課題

本研究では，地震発生後の避難状況という単一シナリオを対象として評価を行った．今後は，日常会話やイベント進行など複数の状況を対象とした評価が必要である．

また，本研究では静的な人格設定を対象としたが，今後は記憶，感情変化，プレイヤーとの関係性変化などを導入し，時間経過によって変化する動的な NPC 人格モデルへの発展を目指す．

参 考 文 献

[1] Brown, T. B., Mann, B., Ryder, N., et al.: Language Models are Few-Shot Learners, *Advances in Neural Information Processing Systems*, Advances in Neural Information Processing Systems, Vol. 33, 2020, pp. 1877–1901.

[2] Ouyang, L., Wu, J., Jiang, X., et al.: Training language models to follow instructions with human feedback, *Advances in Neural Information Processing Systems*, Advances in Neural Information Processing Systems, Vol. 35, 2022, pp. 27730–27744.

[3] Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., and Bernstein, M. S.: Generative Agents: Interactive Simulacra of Human Behavior, *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 2023, pp. 1–22.

[4] Shao, Y., Li, L., Dai, J., and Qiu, X.: Character-LLM: A Trainable Agent for Role-Playing, *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 2023, pp. 13153–13187.

[5] Goldberg, L. R.: An Alternative "Description of Personality": The Big-Five Factor Structure, *Journal of Personality and Social Psychology*, Vol. 59, No. 6(1990), pp. 1216–1229.

[6] Gosling, S. D., Rentfrow, P. J., and Swann, W. B. J.: A Very Brief Measure of the Big-Five Personality Domains, *Journal of Research in Personality*, Vol. 37, No. 6(2003), pp. 504–528.

[7] Oshio, A., Abe, S., and Cutrone, P.: Development, Reliability, and Validity of the Japanese Version of Ten Item Personality Inventory (TIPI-J), *The Japanese Journal of Personality*, Vol. 21, No. 1(2012), pp. 40–52.

[8] Li, J., Galley, M., Brockett, C., Spithourakis, G., Gao, J., and Dolan, B.: A Persona-Based Neural Conversation Model, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2016, pp. 994–1003.

[9] Wang, X., Xiao, Y., Huang, J.-t., Yuan, S., Xu, R., Guo, H., Tu, Q., Fei, Y., Leng, Z., Wang, W., Chen, J., Li, C., and Xiao, Y.: InCharacter: Evaluating Personality Fidelity in Role-Playing Agents through Psychological Interviews, *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2024, pp. 1840–1873.

[10] 高信哲, 小林篤史: ゲーム NPC 対話生成のための TIPI を用いたパーソナリティモデルの基礎検討, 情報処理学会第 *88* 回全国大会講演論文集, 2026. 講演番号 5V-03.

[11] Weir, N., Thomas, R., d'Amore, R., Hill, K., Van Durme, B., and Jhamtani, H.: Ontologically Faithful Generation of Non-Player Character Dialogues, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 2024, pp. 9212–9242.

[12] DeepSeek-AI: DeepSeek-V3 Technical Report, Technical report, DeepSeek-AI, 2024. arXiv:2412.19437.

[13] Gemini Team: Gemini: A Family of Highly Capable Multimodal Models, Technical report, Google, 2023. arXiv:2312.11805.

[14] Anthropic: Claude 3 Model Card, Technical report, Anthropic, 2024.