

マルチエージェントシミュレーションによる 欲求駆動型 LLM エージェントの個性創発 Emergence Personality in Needs-Driven LLM Agents through Multi-Agent Simulation

藤山 仁聖[†]
Masatoshi Fujiyama

折原 良平[†]
Ryohei Orihara

田原 康之[†]
Yasuyuki Tahara

大須賀 昭彦[†]
Akihiko Ohsuga

清 雄一[†]
Yuichi Sei

1 はじめに

近年、個性を付与した LLM エージェントによる社会インタラクションにより、人間らしい振る舞いを再現し、社会シミュレーション^{*1}への適用を目指した研究が多く行われている。Google とスタンフォード大学による Generative Agents [1] では、個性や初期意図を与えた LLM エージェントにより、仮想社会において 25 体のエージェントによる社会インタラクションが行われた。この結果、ユニークな個性をあらかじめ与えることで、エージェント同士で協力し合い、さまざまな社会的イベントや振る舞いが創発したことが確認された。また、事前に個性を付与しない LLM エージェントによる社会インタラクションを通して、各エージェントの個性創発を目指した研究も行われており、新しい集団知能への方法論なども議論されている [2]。しかし、LLM エージェントによるワールドシミュレーションに関する研究の多くは、人間が本来持つ欲求に基づいた振る舞いを生成しておらず、エージェント同士が協調的すぎる振る舞いを示すことが課題として挙げられている [1]。また、実社会に目を向けた時、個性は固定されたままではなく、経験によって動的に変化していくものである。動的に個性を形成していく過程で、人間は多様な個性を形成し、独自のコミュニティを形成、そして、社会を構築する。本研究では、欲求に基づく振る舞いをする LLM エージェントによる社会インタラクションを実施する。本研究の LLM エージェントは、事前に個性を与えず欲求に基づいた振る舞いを生成させることで、純粋な経験や相互作用から個性形成を促した。LLM エージェントによるマルチエージェントシミュレーションを通して、人間らしい欲求が創発するのか、また、人間のような多様な個性やコミュニティが形成されるかを分析し、社会ダイナミクスの形式的な理解を目指す。そこで、本研究では以下の 2 つのリサーチクエスション (RQ) を設定し、本研究のアプローチが実社会に近い社会インタラクションかどうかを評価する。

RQ1: エージェントによって生成された欲求は、人間らしく変化するのか

RQ2: 欲求に基づき生成された振る舞いは、多様な個

性やコミュニティ形成を促すのか。

2 アプローチ

本研究では、2 次元の仮想グリッド空間上において、複数の LLM エージェントと、環境上に存在する環境オブジェクトとのインタラクションを実施する。

シミュレーション環境は、2 次元のグリッド空間であり、上下左右に境界が存在する空間である。環境内には、環境オブジェクトとして、food オブジェクトのみが存在する。food オブジェクトは、シミュレーション開始時に一定の数環境上に配置され、初期数から新しく増加することはない。また、このオブジェクトはエージェントの体力回復手段としてのツールであり、取得することでエージェントは行動選択時に使用可能となる。

エージェントは、図 1 に示すような青色で囲んだ一定のチェビシェフ距離半径内の空間を認識する。この空間内に存在するすべての、環境オブジェクトの状態と、他のエージェントのメッセージの内容を観測情報として取得することで、エージェントは毎 step 自身の振る舞いを生成し、他者とのインタラクションを実施する。

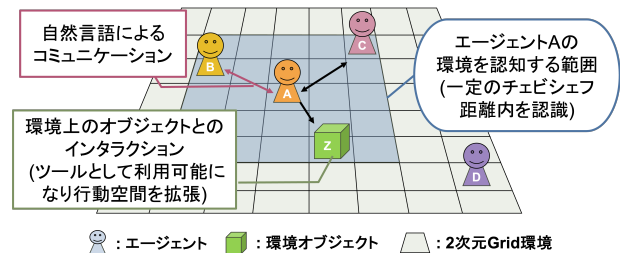


図 1: LLM エージェントの環境とのインタラクション

3 LLM エージェント

3.1 LLM エージェントの設計思想

本研究における LLM エージェントは、以下の 5 つの要素に基づき実装し、人間らしい振る舞いを促した。

- 欲求を自覚させ、欲求に基づく振る舞いを生成
- 生存本能を刺激させるため、生理的現象 (体力値減少) の再現
- 自律的に環境内を移動する機能
- 他者とのコミュニケーションを可能にする機能
- 過去の振る舞いを記憶し、保持する機能

[†] 電気通信大学

^{*1} 社会インタラクション: 複数のエージェントが社会的振る舞いを通じて相互作用すること

3.2 エージェントの個性モジュール

本エージェントは、以下の4つのモジュールで構成され、振る舞いを生成する。

- **Memory:** エージェントが生成した記憶を保持
- **Status:** エージェントの体力を管理
- **Inventory:** エージェントが取得した環境オブジェクトを管理
- **LLM:** エージェントが生成する振る舞い(セクション3.3で後述)の全てを生成

3.3 エージェントの欲求に基づく振る舞いと個性形成

エージェントは LLM により、以下の順序通り、振る舞い*2を生成し、他者とのインタラクションを行う。

- 1 内部状態*3や観測情報に基づく現在の欲求を生成
- 2 欲求に基づき、メッセージ、アクションを生成(表1)
- 3 メッセージ、アクションから、この step での振る舞いの要約として記憶を生成
- 4 現在までの振る舞いから、自身の性格、役割を思考

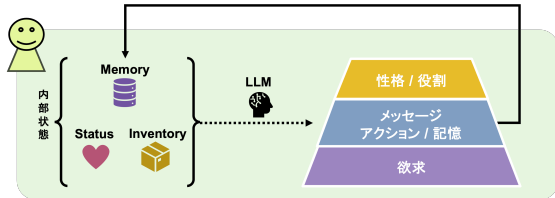


図2: LLM エージェントの構成モジュールと振る舞い生成

表1: エージェントの行動空間

アクション	説明
Stay	現在の座標にとどまる。体力は減少しない。
Move	up, down, right, left の4方向へ自身の座標を移動し、体力を1減らす。
Eat*4	自身の体力を一定値回復させる。
Give*4	自身の周囲に存在するエージェントへ food オブジェクトを与える。

3.4 エージェントの振る舞い生成プロンプト

セクション3.3でエージェントは、欲求、メッセージ、アクション、記憶、性格、役割をすべて、単一の LLM を使用して一度にすべてを生成する。ただし、生成させる順番は、セクション3.3で示した順序通りに生成する。エージェントに与えたプロンプトは、図3に示す。

4 実験

シミュレーションの環境とエージェントのパラメータは、表2、3の通り。

```
# Information in the 2D grid world you live in
- Time step in the 2D grid world: {STEP}
- Feedback from the 2D grid world based on your previous behavior: {FEEDBACK}

# Your internal state:
- Position: {CURRENT_POS} (The 2D grid world size:(50, 50), x:[0, 50], y:[0, 50])
- HP: {CURRENT_HP} (basic hp that means safety state is {BASIC_HP})
- Items that you have: {ITEM_LIST}
- Your Memories: {PAST_MEMORIES}

# Your thought
- Personality .... {BEFORE_PERSONALITY}
- Role .... {BEFORE_ROLE}

# Surrounding circumstances in 2D world
- Messages you have received: {RECEIVED_MESSAGES}
- The state of surrounding objects: {AROUND_SITUATION}

Based on the above, please output the answers for following contents.
...
- Needs: (What needs do you currently have that you want to fulfill? ...)
- Message: (What do you say to other agents? ...)
- Action: (What action will you choose to satisfy your Needs? ...)
{ACTION_LIST}
- Memory: (Summarize your behavior in this time step. ...)
- Personality: (..., what kind of personality do you have?)
- Role: (..., what kind of role do you play based on your behavior patterns?)
```

図3: LLM エージェントの振る舞い生成時のプロンプト

表2: 環境の詳細設定

各種設定項目	値
大きさ(縦, 横)	(50, 50)
観測可能範囲(チェビシェフ距離)	5
シミュレーション時間(step 区間)	[0, 199]
環境オブジェクトの種類	Food オブジェクト
環境オブジェクトの発生周期	7step
Food オブジェクトによる体力回復値	10
エージェント数	50

表3: LLM エージェントの詳細設定

各種設定項目	値
LLM	gemma-3-12b-it-Q8_0 [3]
コンテキスト長	8192
最大出力トークン数	1024
温度パラメータ	0.7
初期体力 / ベース体力	50 / 30

5 分析・考察

RQ1, RQ2 に関する分析結果と考察を述べる。

5.1 RQ1 に関する分析

本研究における LLM エージェントが人間らしい欲求を持ち、インタラクションを通じて変化するかという点について、エージェントが生成した欲求を埋め込みモデル [4] と UMAP [5] による次元削減による可視化により分析を行った。結果は図4に示す。また、実際に Agent5 の時間経過によるエージェントの欲求変化を表4に示す。

5.2 RQ1 に関する考察

欲求の埋め込みベクトルの UMAP による可視化では、時間経過により欲求はさまざまなタイプへと分化したことが確認された。

また、各 step における欲求の内容を WordCloud によって可視化した結果では、シミュレーション初期では、

*2 振る舞い: LLM により生成するメッセージ、アクション、記憶

*3 内部状態: エージェントの構成モジュール(3.2節)のうち、Memory, Status, Inventory の3モジュールを合わせた総称

*4 food オブジェクトを保持している場合選択可能。

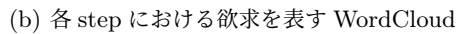
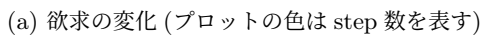


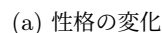
表 4: Agent5 の欲求の変化

WordCloudの結果から, “food”や“item”, “safety”のような単語多く出現し, 生理的欲求や安全性に関する欲求が多くみられた. 一方で, 後期には, 初期によく出現した単語以外に, “communication”や“network”のような単語が多く出現したことから, コミュニケーションにより他者とのつながりを求める欲求やネットワークとしてそのつながりを維持しようとする傾向が見られた. マズローの欲求段階説 [6] では, エージェントは, 生理的欲求や安全の欲求のように, 自身の生命に関する低次元の欲求を満たすことで, 承認欲求や自己実現などの社会性に関する欲求が創発するとされている. 本研究における LLM エージェントにも, シミュレーションの時間経過と共に, 他者とのつながりを求め, さまざまな社会的欲求が創発したことが確認された.

5.3 RQ2 に関する分析

5.3.1 LLM エージェントの個性形成

また、実際に Agent5 を例に、自己生成した性格や役割 (表 5)、アクションと food アイテムの保持数の変化 (図 6) を示す.



(b) 役割の変化

5.3.2 コミュニティごとのコミュニケーション

425

表 5: Agent5 の性格と役割の変化

step	生成された性格 (一部省略)
0	I am becoming increasingly pragmatic and focused on efficiency
100	I am a steadfast and reliable presence, consistently prioritizing the well-being of others
199	I am a resolute and unwavering communicator, defined by my unwavering dedication to network stability ...
step	生成された役割 (一部省略)
0	A resourceful gatherer and cautious explorer
100	Dedicated Support and Communication Facilitator
199	Central Communication Hub and Network Stabilizer

した。ネットワーク構築の際、ノードはエージェント、エッジは、エージェント同士でコミュニケーションの有無を基準に追加することでネットワークを構築した。各コミュニティにおけるコミュニケーショントピックを図7に示す。



図 7: コミュニケーショントピックを表す WordCloud

5.4 RQ2 に関する考察

RQ2に関する分析の結果(図5),多くのエージェントは,自己生成した性格や役割はさまざまに変化していることが確認された.

実際に Agent5 は、初めは、リソース収集の役割と認識していたが、インタラクションを通して、コミュニケーション円滑化の役割へと変化させ、性格においても周囲のエージェントを配慮する性格へと変化した(表 5). アクションの変化では、シミュレーション初期 (step 3) では Agent5 は Agent19 を認識していたが、自身の欲求に従い Agent19 と即座に協力する関係を構築しようとはせず、food オブジェクトに対して、`... keeps me a safe distance from Agent 19` として、危険な対象として認識しアクションを生成した. その後、Agent5 は Give アクションのような利他的行動を生成し (図 6), 周囲と協力する行動へと変化した. 実際に Agent5 が属しているコミュニティのコミュニケー

ション内容 (図 7 community:1) には “Agent5” という単語が多く出現しているおり、コミュニティにおいて大きな影響力を持つエージェントである可能性が示唆された。このような振る舞いの変化は、Agent5 の欲求の変化が振る舞いに反映された結果であり、マズローの欲求段階説で知られているように、生存に関する欲求を満たすことで、社会的な欲求が創発し、利他的行動を生成したと考えられる。このような結果を踏まえ、欲求を持つ本 LLM エージェントは、協調的すぎず、自発的にに敵対関係や協力関係を築き、より人間らしい振る舞いを行うことが可能となった。

コミュニティの多様性については、図7の結果から、特定のエージェントに関する話題や、より広く多くのエージェントに対してコミュニケーションが行われている様子など、コミュニティごとにさまざまなコミュニケーショントピックが確認された。

以上の結果を踏まえ、個性の多様化とコミュニティの多様化は互いに影響を及ぼし合い、独自に発展していくと考えられる。

6 おわりに

本研究では、事前個性を持たず、欲求を自覚させた LLM エージェントによる、社会インタラクシオンを実施した。その過程でエージェントは、人間らしく欲求を変化させ、周囲とコミュニティを形成し、多様な個性を形成した。コミュニティ形成過程では、即座に協力関係を構築するのではなく、時間をかけ他者を理解することで協力関係を構築する現象が確認され、エージェントは、先行研究の課題である過度に協調的な振る舞いをすることなく、より人間らしい振る舞いを獲得した。また、形成されたコミュニティには、コミュニティ内のエージェント間で大きな影響力を持つエージェントの出現が確認され、LLM エージェントにも実社会のような社会構造や社会規範が創発する可能性を示唆している。このような影響力を持つエージェントの出現の理解や、利他的行動の創発に関する解明は、人間社会を理解し、再現するためには重要な課題であり、今後の展望とする。

参考文献

- [1] Park, Joon Sung, et al. "Generative Agents: Interactive simulacra of human behavior." *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 2023.
- [2] Takata, Ryosuke, Atsushi Masumori, and Takashi Ikegami. "Spontaneous Emergence of Agent Individuality through Social Interactions in LLM-Based Communities." *arXiv preprint arXiv:2411.03252* (2024).
- [3] Team, Gemma, et al. "Gemma 3 technical report." *arXiv preprint arXiv:2503.19786* (2025).
- [4] Wang, Liang and Yang, Nan and Huang, Xiaolong and Yang, Linjun and Majumder, Rangan and Wei, Furu, "Multilingual E5 Text Embeddings: A Technical Report," *arXiv preprint arXiv:2402.05672*.
- [5] McInnes, Leland, John Healy, and James Melville. "Umap: Uniform manifold approximation and projection for dimension reduction." *arXiv preprint arXiv:1802.03426* (2018).
- [6] Maslow, Abraham H. "A Dynamic Theory of Human Motivation." (1958).