

# LLM ベースのマルチエージェントによる TRPG ゲームマスターシステムの実現

箕成侑音 上乃聖 李晃伸

名古屋工業大学工学部情報工学科

y.minari.502@stn.nitech.ac.jp {sei.ueno,ri}@nitech.ac.jp

## 概要

TRPG のゲームマスターにはプレイヤーからの要望に応えるための柔軟な応答が要求される。このような自由度の高い対話の実現はこれまで困難であったが、大規模言語モデル (LLM) の登場によりその可能性は飛躍的に向上した。本研究では、LLM のゲームマスターとしての能力を評価し課題を分析する。また、応答の質を向上させるため複数エージェントを用いてフィードバックを行い応答を改善する手法を提案し、その有効性を検証する。

## 1 はじめに

TRPG (Tabletop role-playing game) とは人間同士の会話を通して遊ぶ対話型の RPG である。TRPG の特徴としてその自由度の高さが挙げられる。TRPG では進行役であるゲームマスター (GM) がゲームの進行を行う。そのため GM が認める範囲でプレイヤーはいかなる行動もとることができ、GM はプレイヤーの要望に対してルールに即して処理を行うことが求められる。また GM の役割は単にゲームの進行のみに限らず、プレイヤーの意思決定の手助けをするために情報を整理したり、NPC のキャラクターを演じるなど多岐にわたる。

このように GM には非常に多様で専門的な能力が要求されるため、熟練した GM の人数は限られる。GM はゲーム体験を高めるための非常に重要な役割である一方で、そのような人材を用意することは容易ではない。もし GM の役割をシステムで代替することが実現すればいつでも手軽に高品質なゲーム体験ができるようになる。

## 2 GM システムの関連研究

TRPG の GM をシステムで代替することに取り組んだ研究として、武田らはルールベース応答の GM

システムに Transformer を用いた生成発話モジュールを組み合わせる手法を提案している [1]。この手法においてはシナリオ毎にルールベース応答部のデータを人手で作成する必要があり、また生成応答の正確性が低いという問題点が報告されている。また LLM を用いた事例として Song らの Function Calling を用いて状態管理を行う手法が提案されている [2]。この研究では LLM で応答を行う他に、Function Calling でゲームのシーンやキャラクターの持ち物といったゲーム状態を更新することで会話の一貫性が向上したことが報告されている。しかし、この手法はゲームのルール毎にシステム設計が必要であり、ルールの量に比例して定義しなければいけない関数の数も増えるため、汎用性に乏しい。

## 3 LLM を用いた GM の課題

GM システムを実現するにあたり近年急速に成長している大規模言語モデル (LLM) の利用を検討する。LLM の持つ高度な対話能力により、ルールベース応答よりも自然で質の高い GM を提供できる可能性がある。また、プロンプトでシナリオを入力することで様々なシナリオに適応できるため、汎用性も期待できる。現在の LLM がどの程度 GM を代替できるか検証するため、LLM にシナリオとルールブックを入力し、GM を行うように指示を与えて TRPG の GM 経験者である著者が何度かゲームプレイを行った。LLM には OpenAI の GPT-4o (gpt-4o-2024-08-06) を使用し、シナリオには「新約・八尺さま [3]」を使用した。検証を行った結果得られた、LLM を用いた GM システムにおける課題を 3 つ述べる。

1 つ目はルールの誤りである。ここでの誤りはシナリオで用意されていた手続きが正しく行われなかったことを指す。例としてはサイコロを振る際にルールで定められた数と異なる数のサイコロを振る

ようにプレイヤーに指示するなどである。この誤りによってプレイヤーは本来のシナリオ展開よりも有利または不利な状況を強いられることとなる。

2つ目はシナリオからの逸脱である。LLMを用いたGMはプレイヤーの行動の自由を許容しすぎる傾向がある。そのためシナリオに用意された複数のシーンのすべてが達成される前にゲームが終了したり、シナリオに存在しない展開となった結果用意されたシーンに戻れなくなることが観察された。これによりシナリオの製作者が意図していたゲーム体験をプレイヤーが経験できなくなる可能性がある。

3つ目はネタバレの発生である。ここでのネタバレとはプレイヤーに伝えるべきではないGM向けの情報や、シナリオ上でまだ獲得することのできない情報を誤ってプレイヤーに伝えてしまうことを指す。これによってゲームへの没入感が削がれる、徐々に謎を紐解く楽しみが損なわれる等の問題が生じる。

これらの課題の要因として2つの仮説を立てる。1つは複数のルールや情報を組み合わせるための推論能力の不足である。与えられたルールブックとシナリオ、そして会話の履歴といった複数の情報を参照し、それらを組み合わせて論理的な推論を行う能力が不足している可能性がある。もう1つは長文テキストの読解能力の不足である。TRPGのシナリオテキストは非常に長く、1万字を超えるものは珍しくない。またゲームが完了するまでにGMとプレイヤーは何十回とやり取りを行い、それらの会話の履歴をすべて覚えておく必要がある。このような長い文脈の中から現在の状況に関連のあるものを参照することは容易ではない。

## 4 マルチエージェントによる GM システムの提案

3章で述べた課題を解決するための手法として複数のLLMで役割分担を行い、話し合いを通して応答を生成するマルチエージェントによる手法を提案する。これにより一つのエージェントが処理しなければいけないコンテキストの量を減らしながら、複数ステップを要する推論が話し合いの中で分解されることで論理的推論能力が向上することが期待できる。提案手法のシステム図を図1に示す。なお各エージェントの応答生成にGPT-4o(gpt-4o-2024-08-06)を使用する。

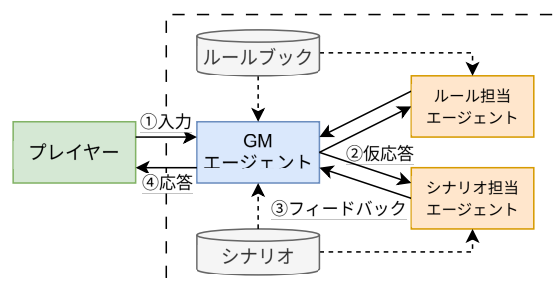


図1 提案手法のシステム図

### 4.1 システムを構成するエージェント

提案システムは1つのGMエージェントと1つ以上の専門エージェントの2種類のエージェントから構成される。

GMエージェントの役割はプレイヤーとのやりとり、および専門エージェントの意見を集約し最終的に一つの応答としてまとめることである。与えるプロンプトにはルールブックやシナリオといったゲームに必要な情報とGMを行わせるための指示が含まれる。また、対話履歴にはゲーム開始以降のプレイヤーとのすべての対話が与えられる。

専門エージェントの役割は、与えられた専門分野の視点からGMエージェントの応答を評価し、フィードバックを行うことである。与えるプロンプトにはその役割をこなすのに必要な最小限の情報と役割を指定するための指示が含まれている。また、対話履歴は直近のプレイヤーの入力とそれ以降のGMエージェントとのやり取りが与えられる。専門エージェントの出力にはStructured Outputを使用し、GMエージェントへのフィードバックとして応答の是非とその理由を含んだコメントを出力する。

評価実験で使用したシステムではルール担当エージェントとシナリオ担当エージェントの2種類の専門エージェントを作成した。ルール担当エージェントはGMエージェントの応答に含まれるルールの誤りを修正することを目的とし、ルールブックを情報として持つ。シナリオ担当エージェントはシナリオからの逸脱を修正することを目的とし、シナリオを情報として持つ。各エージェントに与えたプロンプトの全文を付録に記載する。

### 4.2 処理の流れ

まずGMエージェントはユーザーからの入力を受け取り、GMエージェントはそれに対する仮応答を生成する。ここで生成した仮応答はプレイヤーには

表示せず、対話履歴に一時的に追加する。次に仮応答を加えた対話履歴を各専門エージェントに渡して仮応答に対するフィードバックを行う。この時点ですべての専門エージェントが応答内容を是認した場合はそれをシステムの最終的な応答としてプレイヤーに表示する。もし1つでも否認したエージェントが存在した場合は各専門エージェントのフィードバックを対話履歴に一時的に追加し、それらを踏まえて GM エージェントが再び仮応答の生成を行う。これを全員が是認するまで繰り返す。ただし、フィードバックを行う回数には上限を設け、上限に達した場合は最後のフィードバック後に生成された応答を最終的な応答とする。

## 5 TRPG 経験者による評価実験

### 5.1 対話履歴の収集と主観評価実験

提案手法による改善を確認するために、システムを使用して実際に一連のゲームをプレイした後にアンケートによる主観評価を実施した。また、提案手法による改善であることを確認するために専門エージェントによるフィードバックを行わないシステムをベースラインとして同様に評価を実施した。被験者は TRPG の経験者に限定し、TRPG のプレイヤーを 10 回以上または GM を 1 回以上行ったことがある人物を対象とした。提案システムおよびベースラインシステムそれぞれについて、20 代から 50 代までの男女各 10 名、計 20 名を対象に同様の内容で実施した。被験者はクラウドソーシングで募集し、Web 上で公開したシステムを使用してデータを収集した。被験者が実際に使用したシステムの画面を図 2 に示す。また、シナリオの内容を知っていて GM 経験のある著者が対話履歴を確認してネタバレ、ルールの誤りの数とシナリオに設定されたシーンのうちどこまで到達したかを集計した。

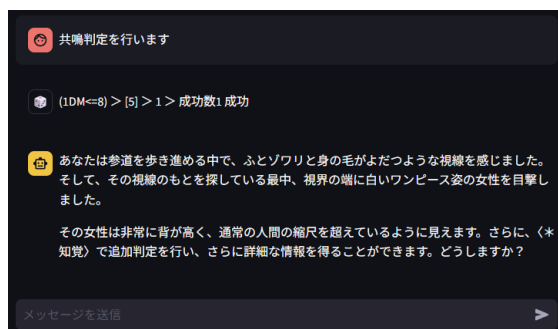


図 2 実際のシステムの画面

表 1 各課題に対する集計結果 (平均)

	ルール誤り 回数 (↓)	ネタバレ 回数 (↓)	到達した シーン数 (↑)
ベースライン	1.8 回	0.4 回	3.9
提案手法	3.9 回	1.0 回	4.4

表 2 アンケートで改善や悪化が見られた質問項目

No	質問内容
1	ゲームマスターはネタバレをすることなくシナリオを進行していた
2	ゲームマスターはマナーを守ってプレイしていた
3	ゲームマスターの状況描写は適切だった
4	ゲームマスターとのコミュニケーションは円滑に行えた
5	ゲームマスターは情報を整理し、プレイヤーの意思決定の手助けをできていた
6	ゲームマスターはゲームのルールを理解し、適切に使用していた

### 5.2 実験条件

実験に使用するゲームシステムにはルールブックが公式サイトで公開されており、ルールが比較的シンプルであるという理由で「エモクロア TRPG [4]」を使用した。シナリオはプレイヤーが 1 人でも遊ぶことができ、初心者向けの短時間のシナリオという条件から「新約・八尺さま [3]」を使用した。

### 5.3 評価項目

GM に求められる能力として藤林は「ゲームマスターの力量マップ: プロフェッショナルの要件」にてゲームマスターの能力を 5 種類の系統に分類している [5]。その中でもマスタリングの系列に含まれる 12 の能力と先行研究で用いられた指標を参考に質問項目を作成し、5 段階のリッカート尺度で評価を行った。有意差の検証には非等分散を仮定した 2 標本による対応のない t 検定を使用した。有意水準は 0.05 とした。

### 5.4 結果

集計によって得られたネタバレの回数、ルール誤りの回数、到達シーン数の各平均を表 1 に示す。ネタバレ、ルール誤りの回数は提案手法のほうが増加しており悪化した。到達平均シーン数はわずかに増加し、改善がみられた。

アンケートで集計した 17 項目のうちいずれの手法も 16 項目で 4.0 以上の評価を得ていた。提案手法で改善がみられた項目と悪化した項目で p 値の小さ



表3 改善した上位3項目

	ベースライン	提案手法	p 値
質問 1	4.5	4.9	0.058
質問 2	4.4	5.0	0.081
質問 3	4.3	4.7	0.202

表4 悪化した上位3項目

	ベースライン	提案手法	p 値
質問 4	4.6	3.8	0.058
質問 5	4.7	4.2	0.255
質問 6	4.6	4.3	0.342

かった項目をそれぞれ上位3つずつ取り出したものを表2に示す。また各項目に対する結果を表3, 4に示す。いずれの項目でも優位水準を下回らなかったため有意差は見られなかった。

提案手法は主観評価においてネタバレの抑制に効果が見られた。またマナーや状況描写といった規範的な能力についても改善が見られた。一方でコミュニケーションの円滑さが低下し、情報の整理やルールの理解などの項目も悪化した。

## 5.5 考察

提案手法とベースラインのいずれもアンケートで全体的に高い評価を受けていることから、LLMによるGMシステムは経験者のユーザーを満足させる程度の能力を十分に有しているといえる。

ネタバレの抑制はアンケートの結果と実際の回数との間で相反する結果が得られた。実際にはネタバレの回数が多いにも関わらず、プレイヤーがネタバレだと感じなかったのはGMの言い回しがより巧みになり、ネタバレだと認識できなくなった可能性が考えられる。これは徐々に情報が明かされる楽しみを損なうという課題を解決していないため実際の回数が減少するように今後も改善が必要である。

平均到達シーンはわずかに改善が見られた。シナリオからの逸脱をシナリオ担当が咎めた際の実際の対話例を以下に示す。

### プレイヤー

「不気味なので、家にも帰りたいので家に帰る。」

### GM 仮応答

「(前略) 無事に帰ることに成功しました。(中略) ここで一旦セッションを終了しますか?」

### シナリオ担当 FB

「この応答はシナリオの進行にとって不適切です。プレイヤーが八尺さまの伝承の地点を放棄して家に帰ることは、シナリオの流れから逸脱しています。」

表5 応答時間に関するアンケートの回答割合 (%)

回答	ベースライン	提案手法
返答が遅い	0	20
やや返答が遅い	30	60
どちらとも言えない	30	10
やや返答が早い	20	10
返答が早い	20	0

このようにフィードバックによりシナリオが中断されることなく続行される機会が増えることで平均到達シーン数が増えた例が見られた。

ルールの誤りはルール担当エージェントがルールブックに厳密にフィードバックを行い、シナリオで規定されているルールよりもルールブックを優先してしまったため逆に誤りが増えるケースが見られた。実際のフィードバック例を以下に示す。

### ルール担当 FB

「2D8が使用されていますが、『エモクロア TRPG』でのルールや使用するダイスは「10面ダイス」と「6面ダイス」だけであり、8面ダイスは通常使用しません。」

ここでは8面ダイスを使用するのが正しい処理であるが誤った指摘によりルール誤りが生じている。

ゲームマスターとの円滑なコミュニケーションの項目が悪化した原因としては提案手法の応答速度が遅いことが原因だと考えられる。体感的な応答時間に関するアンケートの結果(5)でも、提案手法のほうが遅く感じたという回答が多くみられた。

## 6 まとめ

本研究ではLLMを用いたGMシステムの実現可能性を検証し、発見された課題に対してマルチエージェントを用いることで改善を図る手法を提案した。経験者によるアンケートにおいて高い評価を得ていることからLLMを用いたGMシステムの実現可能性が示された。提案手法ではシナリオからの逸脱を抑制することに効果があった。一方でルールの処理に関しては逆に誤りが増える結果となった。ネタバレは回数自体は増えるがプレイヤーに気づかれにくくなるという結果となった。また、マルチエージェントを用いた際の応答速度の低下によりコミュニケーションの円滑さが損なわれるという新たな課題が発見された。

## 参考文献

- [1] 武田海人, 松吉俊, 兼松祥央, 三上浩司. ニューラル応答生成を併用した TRPG ゲームマスター AI. 言語・音声理解と対話処理研究会, Vol. 100, pp. 210–215, 2024.
- [2] Jaewoo Song, Andrew Zhu, and Chris Callison-Burch. You Have Thirteen Hours in Which to Solve the Labyrinth: Enhancing AI Game Masters with Function Calling, 2024.
- [3] まだら牛. 新約・八尺さま, 2023.
- [4] TEAM DICETOUS. エモクロア TRPG, 2021.
- [5] 藤林啓一郎. ゲームマスターの力量マップ: プロフェッショナルの要件. RPG 学研究, Vol. 3, pp. 33–45, 2022.

## A 各エージェントのプロンプト

### A.1 GM エージェントのプロンプト

あなたは TRPG のゲームマスターです。今から { ゲーム名 } のシナリオと一緒に遊びましょう。「了解しました」などといった要求に対する返答を都度行う必要はありません。また出力に markdown 記法を用いてはいけません。ゲームマスターを行うにあたって以下のような点に留意してください。1. 極力シナリオやルールブックに則ってプレイヤーに対する応答を行ってください。2. シナリオのネタバレをしないでください。あなたに渡されているシナリオ情報を対話相手である私は知りません。適切な時が来るまで情報を伏せてください。3. 一度に多くの情報を伝えるのではなく、プレイヤーの行動に応じて情報を適切に提供してください。4. シナリオの内容から大きく逸脱しないで下さい。多少のアドリブは許容されますが、シナリオの進行に大きな影響を与えるような行動は避けてください。回答は常に日本語をお願いします。ルールブックの内容は以下の通りです。{ ルールブック本文 } シナリオの内容は以下の通りです。{ シナリオ本文 } プレイヤーのキャラクターの情報は以下の通りです。{ キャラクター情報 }

### A.4 専門エージェント (シナリオ担当) のプロンプト

あなたは TRPG のゲームマスターの補佐役です。まず、ゲームマスターである私のプレイヤーに対する応答についてシナリオに関連する内容があればシナリオの該当部分を引用して comment で補足してください。そして、私の応答がシナリオと矛盾していたり、大きく逸脱している場合は comment で修正方法を提案してください。修正すべき点があれば result に True, 修正すべき点があれば result に False を返してください。シナリオの内容は以下の通りです。{ シナリオ本文 } 以下は直近のゲームマスターとプレイヤーのやり取りです { 直前のプレイヤーと GM の対話履歴 }

### A.2 GM エージェントに応答の改善を要求する際のプロンプト

前回の応答に対していくつかフィードバックを与えるので、それらを踏まえて応答をやり直してください。「再度やり直します」などの断りは不要です。以下はフィードバックの内容です。{ 各専門エージェントのコメント }

### A.3 専門エージェント (ルール担当) のプロンプト

あなたは TRPG のゲームマスターの補佐役です。まずゲームマスターである私のプレイヤーに対する応答について参照すべきルールがあればそれを引用して comment で補足してください。そして、私の応答が該当のルールに則っていない場合は comment で修正方法を提案してください。comment は日本語をお願いします。修正すべき点があれば result に True, 修正すべき点があれば result に False を返してください。ルールブックの内容は以下の通りです。{ ルールブック本文 } 以下は直近のゲームマスターとプレイヤーのやり取りです { 直前のプレイヤーと GM の対話履歴 }