

# Agent Agora: エージェントの選定と放置による アイデア生成システムの提案と評価

三瓶智輝<sup>†1</sup> 宮下芳明<sup>†1</sup>

アイデアの生成には多大な労力を要する。近年では LLM を用いてアイデア生成に係る負荷を軽減するアプローチが広く取られているが、こちらもプロンプトの工夫などで労力を要するものである。本研究では、質の高いアイデアを楽に得ることを目的として、アイデア生成システム『Agent Agora』を提案する。固有のロールを持つ LLM 搭載エージェントを集め、自律的にブレインストーミングを行わせることでアイデアを生成する。ユーザは頃合いを見て状況を確認し、より議論に貢献しそうなエージェントを選定する。選定と放置を繰り返すことで、ユーザは自ら発想することなく多様なアイデアを得ることができる。さらに、実験にて得られた利用例をもとに、システムおよび生成されたアイデアの有用性を評価・考察する。

## 1. はじめに

アイデアの生成には多大な労力を要する。独自性と実現性を兼ね備えた質の高いアイデアを生み出すには、多様な視点のもと発想することが望ましい。ただ、そうした環境を用意するには、多様な視点を確保しアイデアを捻出するだけの時間やコストが必要となる。

近年では、ChatGPT[1]のような LLM (Large Language Models) を用いることにより、アイデア生成に係る労力を軽減するアプローチが広く用いられる。一方で、LLM を用いたアイデア生成では多様性の欠乏が課題となる。単一の LLM による生成では独自性の低いアイデアが生まれやすく、改善にはプロンプトに工夫を施す必要がある。そのため、技能を習得したり、逐一プロンプトに入力を行ったりすることもまた労力を要するものである。

本稿では、質の高いアイデアを楽に得ることを目的として、『Agent Agora』というアイデア生成システムを提案する。本システムは、固有のロールを持つ LLM 搭載エージェントを集めてブレインストーミング (以下「プレスト」) を行い、アイデアを生成するものである。システム上の仮想空間 (以下「広場」) にエージェントを集め、テーマを書き込むことにより、エージェントたちによるプレストが自律的に進行する。生成されたアイデアは要約・評価し、それも含めて表示に反映することでユーザのアイデア選定の目安とする。ユーザに求められる操作は「エージェントの引き入れ・追い出し」と「テーマの書き込み」のみであり、あとは放置していても勝手にアイデアが生成・集積される。

さらに、実験での利用例をもとに、システムおよびアイデアの有用性を評価・考察した。生成されたアイデアの評価を参加者にさせたところ、全員が一定数のアイデアを採用可能と評価した。一方で、事後インタビューから、テーマの微調整やアイデアの追及といった機能の追加や、アイデアの生成傾向の改善といった展望が示された。



図 1 メイン画面

## 2. 関連研究

### 2.1 水平思考とシックス・ハット法

アイデア生成のような答えが複数存在し得る課題に対しては、Bono が考案した思考法である水平思考[2]が広く用いられる。水平思考は多角的な発想を重視しており、論理的整合性に拘泥しない。そのため、従来の垂直思考と比べ、答えが一意に定まらない様な課題の解決に適している。

水平思考の一種として、Bono はシックス・ハット法を考案している[3]。視点や発想のパターンを 6 種類に分類し、それぞれを固有の色の帽子に喩えたものである。6 種類の思考を意識的に使い分けることで、多様な視点からの議論・意思決定を可能としている。Bono は各色の視点や発言傾向について以下のように示している。

- 白: 客観的・中立的。事実に基づき情報提供する。
- 赤: 主観的・直感的。感情や直感のもと意見する。
- 黒: 客観的・否定的。問題点やリスクを指摘する。
- 黄: 客観的・肯定的。提案の良い面を引き出す。
- 緑: 創造的・革新的。新しい解決策を提案する。
- 青: 統括的・俯瞰的。広い視点で全体をまとめる。

<sup>†1</sup> 明治大学  
Meiji University

## 2.2 アイデア生成に他者視点を取り入れる試み

アイデア生成のような発散的思考を用いる場面では、多様な視点からの発言を取り入れることが望ましいとされる。そこで、他者の視点・発想を取り入れることで、より独自性の高いアイデアを生成する試みが多数行われている。

趙らは、子供のアイデアスケッチを成人の発想に取り入れる技法を提案している[4]。独自性および実現可能性の2軸でアイデアを評価したところ、子供の発想の独自性を維持したまま実現可能性も維持させられることが示唆された。

下村らは、飲酒者のアイデアをヒントに非飲酒者にアイデアを生成させる技法を提案している[5]。独自性や実現可能性などの軸で評価を行ったところ、独自性は高く、実現可能性は低いアイデアが生まれやすいことが示された。

## 2.3 LLM 搭載エージェント同士の自律的な相互作用

複数の LLM 搭載エージェントを自律的に動かすことで相互作用を促し、タスクの精度向上や集団行動のシミュレーションを行う試みが複数存在する。

Chan らは、エージェント同士が互いの発言を参考にしてつづ文章の品質評価を行うシステム『ChatEval』を提案している[6]。正確性および人間の評価との一致度の2軸で評価を行い、いずれも従来手法と比較して高い結果が得られた。

Park らは、LLM 搭載エージェントを用いて人間の社会行動をシミュレーションしている[7]。25体の LLM 搭載エージェントで仮想の村社会を形成させることにより、エージェント間での情報拡散など社会的行動が確認された。

さらに、エージェント毎に異なるロールを設定することで効果を検証する研究も複数存在する。

Wang らは、固有のロールを持つエージェント同士を干渉させることで、LLM の専門知識を引き出す手法を提案している[8]。タスクに有効なロールを動的に生成し、相互にフィードバックさせることで、正確性の向上が示された。

Chen らは、マルチエージェントでの協働作業を促進するアーキテクチャを提案している[9]。会話能力や論理的思考力など4種類のベンチマークタスクで評価を行い、シングルエージェントでの実行結果よりも高いスコアを記録した。

# 3. 提案システム

## 3.1 システム概要

提案システム『Agent Agora』は、固有のロールを持つ複数の LLM 搭載エージェントにブレストを行わせ、アイデアを生成・評価するシステムである。本システムのメイン画面を図1に、システムの処理概要を図2に示す。

本システムの機能は、「広場のマネジメント」と「エージェントたちによるブレスト」に二分される。前者は、ユーザがシステム内の仮想空間にエージェントを引き入れ、テ

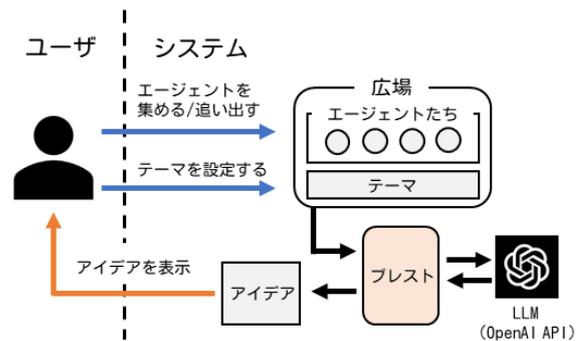


図2 処理概略図



図3 エージェント選択画面

ーマを書き込むことを指す。マネジメントが完了すると、広場のエージェントによって自律的にブレストが進行する。

なお、インターフェースは Unity を、LLM には GPT-3.5-turbo-0613[10]を用いて実装した。なお、リクエストやレスポンスを処理するにあたって、必要に応じ文章を自動で和英訳している。翻訳には DeepL API[11]を用いた。

## 3.2 インタフェース

### 3.2.1 エージェントおよびテーマの設定

メイン画面左下の「エージェントを集める」ボタンを押下するとエージェント選択画面(図3)が表示される。本画面では任意のエージェントを追加・除外することが可能である。エージェントの見た目やロール名、発言傾向、評価軸を確認でき、広場にエージェントを追加した状態でテーマを入力することにより、ブレストを開始する。

### 3.2.2 アイデアの確認

アイデアの生成状況はシステムのメイン画面に動的に反映される(図1)。アイデアの生成が開始されると、対応するエージェントの頭上にフキダシが表示される。フキダシの色はエージェントによって異なる。生成完了までのフキダシ内のメッセージは「(考え中...)」としている。生成が完了すると、対応するフキダシにアイデアのキャッチコピーが表示される。



図 4 アイデア一覧画面

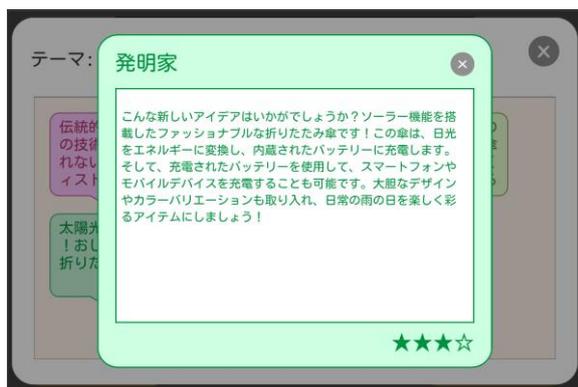


図 5 アイデア詳細画面

さらに、メイン画面右下の「アイデアを見る」ボタンを押下すると、アイデア一覧画面が表示される。本画面にて、過去に生成されたアイデアを確認可能である。各フキダシの大きさはアイデアの評価スコアに比例する。グリッド表示は、アイデアを生成順で左上から並べて配置する形式である。いずれの形式でも、未読のアイデアはフキダシの右上にピンク色の円を表示することで強調している。

なお、いずれの画面においても、フキダシを押下することで対応するアイデアの詳細を確認できる(図5)。アイデア詳細画面では、アイデアの発言者(ロール)や、アイデアを提案する詳細な文章、および評価が確認可能である。

### 3.3 広場のマネジメント

#### 3.3.1 エージェントの種類

エージェントはそれぞれ見た目および「ロールデータ」が異なる。ロールデータはロール名、発言傾向、評価軸などの要素から構成される。発言傾向はアイデアの生成時、評価軸はアイデアの評価時に用いる。

システムの試作にあたり、14種類のロールデータを作成した。表1にその一覧を示す。12種類はシックス・ハット法[3]をもとに、各色の帽子に近い性質を持つと感じる職業を筆者の主観で2種類ずつ選定した。残りの2種類は先行研究の子ども[4]および飲酒者[5]を参考に作成した。

表 1 ロールデータの一覧

ロール名	発言傾向	評価軸	参考
科学者	科学的な整合性を重視。実用性や可能性を論じる	- 論理性 - 複雑性	[3]:白
ジャーナリスト	報道の探究心を反映。感情や主観的意見は排除	- 信頼性 - 影響力	[3]:白
芸術家	芸術的なひらめき、オリジナリティを反映	- エモさ - 先鋭性	[3]:赤
詩人	感性と直観を重要視。独創的な発想	- 共感性 - 美しさ	[3]:赤
評論家	主観を排し分析的に話す。穿った視点を持つ	- 論拠 - 多角性	[3]:黒
弁護士	法的論拠や客観性を反映させる	- 正しさ - 公益性	[3]:黒
コーチ	動機付けの観点から、他者の成長を促す提案をする	- 適応性 - 動機付け	[3]:黄
看護師	ケアや健康増進を中心に考えを述べる	- 対話 - 協調性	[3]:黄
コック	食材や調理法、味覚の知識を駆使する	- 調和性 - 豊かさ	[3]:緑
発明家	既存の枠組みにとらわれない大胆な発想	- 革新性 - 実用性	[3]:緑
船長	柔軟性と広い視野を持ち、リーダーとして発言	- 持続性 - 社会価値	[3]:青
戦略家	徹底的な洞察のもと、戦略策定の観点から提案	- バランス - 競争力	[3]:青
子ども	好奇心や冒険心を刺激するアイデアを提案	- 簡単さ - 楽しさ	[4]
酔っ払い	酔狂で支離滅裂、突飛なアイデアを提案	- 妄想 - 奇抜さ	[5]

#### 3.3.2 マネジメントの目安

エージェントの選択およびテーマの書き込みが完了すると、エージェントたちは各々のロールに沿って自律的にブレストを行う。ブレストの最中、ユーザはインタフェースを通じていつでもその様子を確認できる。さらに、それに応じてエージェントを入れ替えることにより、ユーザ好みのエージェントによるブレストの進行が可能となる。

#### 3.4 エージェントたちによるブレスト

提案システムにおけるブレストは、アイデアの生成・要約・評価の3要素からなる。それぞれ対応するプロンプトでLLMにリクエストを送信することにより実現している。

ブレストの過程で生成されたアイデアは「アイデアモデル」として保存し、システム内での処理に用いる。アイデアモデルはロール名、アイデア本文、キャッチコピー、評価データの4つの要素から構成される。

##### 3.4.1 アイデアの生成

一定時間おきにアイデア生成を行う。生成時には、広場のエージェントや入力されているテーマを基に入力文を作成し、以下のプロンプトと共に送信・生成リクエストを行っている。

与えられたルールに沿って、テーマに対するアイデアを1つ考えなさい。

- ・入力文について
  - #role, #tone, #theme の3要素からなる。
  - #role を基にアイデアを考えなさい。
  - #tone に沿ってアイデアを説明しなさい。
  - #theme はユーザが入力したテーマ。このテーマに対して、あなたのルールならではの回答をしてください。

なお、選択したエージェントが既に同じテーマに対するアイデアを1つ以上生成している場合は、それらのアイデアと内容が被らないようプロンプトにて制限をかける。さらに、選択したエージェント以外が同じテーマに対するアイデアを生成している場合は、一定確率でそれらのアイデアから派生した内容のアイデアを生成するよう促す。

### 3.4.2 アイデアの要約

生成されたアイデアは、アイデアモデルとしてロール名およびアイデア本文をシステムに格納した後、LLM で要約を行う。以下にアイデア要約のプロンプトの要旨を示す。

与えられた入力文をもとに、以下のルールに従いキャッチコピーを出力しなさい。

- ・入力文について
  - #idea, #theme の2要素からなる。
  - #idea は要約対象のアイデア。
  - #theme はアイデアの元となったテーマ。
- ・生成のルール
  - #idea を約10文字の日本語に要約しなさい。
  - #theme の文面と重複する内容は避けること。

キャッチコピーが生成されたら、先述のアイデアモデルに格納する。本文およびキャッチコピーは格納直後に英訳を行い、それらも同様にアイデアモデルに格納している。キャッチコピーはアイデアの一覧表示時や、既存のアイデアに対する重複防止・派生の際に用いる。

### 3.4.3 アイデアの評価

新しいアイデアが生成され次第、LLM によって評価を行う。評価は、対象のアイデアを生成したエージェントを除き、広場にいる全てのエージェントによって行われる。リクエストを受けたエージェントたちは、各々の評価軸をもとに対象のアイデアが魅力的か LLM を用いて判断する。以下にアイデア評価のプロンプトの要旨を示す。

与えられた入力文をもとに、以下のルールに従いアイデアを評価しなさい。

- ・入力文について
  - #idea, #theme, #juries の3要素からなる。
  - #idea は評価対象のアイデア。
  - #theme はアイデアの元となったテーマ。
  - #juries はアイデアを審査する審査員。以下の形式で与えられる。
    - ロール A: A の評価軸1, A の評価軸2。
    - ロール B: B の評価軸1, ...
  - 各評価軸の観点から、#idea を5点満点で評価しなさい。

各エージェントは2つの評価軸を持ち、評価軸ごとに5点満点、計10点満点で採点を行う。採点結果が7点を超えた場合、星を1つつける。この評価は、アイデア一覧画面(図4)にて、各アイデアの表示サイズに反映させることで、ユーザがアイデアを選定する際の目安としている。

## 4. 実験

### 4.1 実験概要

提案システムによって生成されるアイデアの有用性や、ユーザが行う操作の目的について調べる為、実験を行った。参加者は情報系学部に通う20~22歳の大学生4名である。

実験は参加者にシステムを利用させた後、操作に関する事後インタビューやアイデアの評価を行うものである。システム利用中やアイデアの評価中など、実験の最中に感じたことはその場で声に出して呟くように指示した。

### 4.2 実験手順

目的や概要について参加者に説明した後、以下の手順のもと実験を行った。なお、参加者には事前にアンケートを行い、全参加者が週に1回以上 LLM を用いてアイデア出しをしていることを確認した。

1. 提案システムに入力するテーマを考えさせた。
2. 提案システムを用いてアイデアを生成させた。
3. 手順2での戦略や操作に関する事後インタビューを実施した。質問事項は以下の通りである。
  - エージェントを追加・継続・除外した理由
  - 欲しいアイデアを得るために意識していたこと
4. 手順2にて生成されたアイデアについて評価させた。

#### 4.2.1 提案システムの利用

手順 2 にて、参加者に提案システムでのアイデア生成を行わせた。エージェントおよびテーマの設定はユーザの自由とした。生成されるアイデアは随時確認させ、マネジメントの参考にするよう指示した。詳細な手順を以下に示す。なお、手順 c を行うにあたり、参加者自身がエージェントの変更を望まない場合は、エージェントを入れ替えず生成を再開してもよいものとした。

- エージェントを最大 5 人まで広場に追加させた。
- テーマを入力させた。
- 広場のエージェントが 1 つずつアイデアを生成したら、エージェントを入れ替えるよう参加者に促した。
- 手順 b・c を、アイデア総数が 16 件以上になるまで繰り返した。

#### 4.2.2 参加者によるアイデアの評価

手順 4 では、手順 2 でのアイデアを参加者に評価させた。評価軸は実現性と独自性の 2 軸である。評価には、x 軸を実現性で-3.00~3.00, y 軸を独自性で-3.00~3.00 とした座標平面を用い、参加者の主観でアイデアを配置させた。

参加者には、ランダムに並び替えナンバリングしたアイデアと、座標平面が描かれた紙を提示した状態で、紙上の対応する位置にアイデアと同じ番号が書かれた付箋を貼り付けさせた。評価に際し、参加者に以下の 3 点を説明した。

- 参加者の主観のもと、実現性が高いアイデアほど右に、独自性が高いアイデアほど上に配置すること。
- 第 1 象限に配置したアイデアのうち、採用したくないものがある場合、その付箋に×印をつけること。
- 第 3 象限に配置したアイデアは、各軸の評価値を細かく設定する必要はないこと。

参加者による評価の終了後、各評価軸で最高値・最低値を取ったアイデアや、×印のついたアイデアなどに対し、その理由や所感について簡易的なインタビューを行った。

### 5. 利用例

各参加者の選定したエージェントや、生成されたアイデアの一部を示す。さらに、事後インタビューをもとに、各参加者のシステム利用に伴う評価や感想を示す。

#### 5.1 参加者 A の利用例

テーマは「理系論文のタイトルを考えるコツ」で、システムの利用時間は約 25 分だった。表 3 は参加者 A が選定したエージェントの一覧である。

生成された 20 件のアイデアに対し、参加者 A は図 6 のように評価した。各軸で評価が特徴的なアイデアについては、参加者 A にその評価理由を訊ねた (表 4)。

エージェントの選定について、参加者 A は「自分に刺さる提案をする者を残しつつ、なるべく多くのエージェントを選んだ」と述べており、実際に全てのロールを 1 回以上選定していた。また、「他のロールより一般的でない提案だった」という理由から、4 回とも評論家を選んでいた。

アイデアの評価について、独自性では A-6 や A-7 を特に高く評価していた。ロールの影響から詩や料理といった要素を取り入れたため独特なアイデアとなったが、理系論文に適さないため実現性は低くなった。一方、実現性では A-9 を最も高く評価していた。「具体的かつ取り入れやすいアイデアを複数挙げていた」ことが高評価の理由となった。

A-18 については、実現性・独自性ともに高評価だったが、「医学寄りであり取り入れづらい」ため不採用としていた。また、A-10 は「アクロニム」という語を初めて知ったことや、他に無い提案だったことから独自性を高く評価していた。

表 3 参加者 A のエージェントの選定

回数	エージェント
初回	評論家, 発明家, 科学者, 芸術家, 戦略家
入替①	評論家(2), 発明家(2), 科学者(2), 詩人, コック
入替②	評論家(3), 発明家(3), コーチ, 船長, 酔っ払い
入替③	評論家(4), ジャーナリスト, 弁護士, 看護師, 子ども

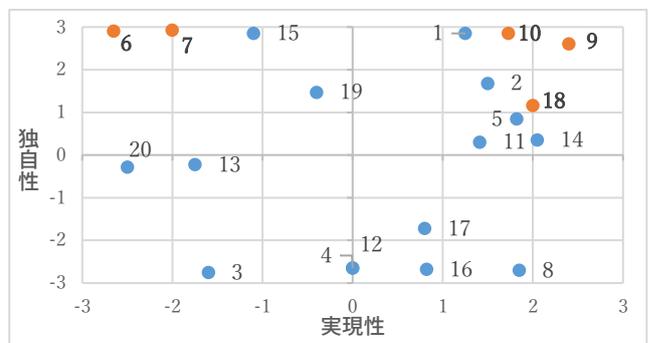


図 6 参加者 A によるアイデアの評価  
(データラベルは ID, 文中では A-n と表現)

表 4 参加者 A の評価が特徴的だったアイデア

ID (A-n)	ロール	アイデア要旨	実現性	独自性	備考
6	詩人	小説や詩のタイトルを参考に	-2.65	<b>2.90</b>	
7	コック	料理や食材に関する単語を使う	-2.00	<b>2.92</b>	
9	評論家	客観的・論理的にアプローチ	<b>2.40</b>	2.60	採用可
10	発明家	アクロニムやメタファを活用	1.73	2.85	採用可
18	看護師	医学論文での例	2.00	1.16	不採用

## 5.2 参加者 B の利用例

テーマは「風邪をひかずに楽しめる冬のお出かけスポット」で、システムの利用時間は約 15 分だった。表 5 は参加者 B が選定したエージェントの一覧である。なお、参加者 B はルール「科学者」を一度も選定しなかった。

生成された 18 件のアイデアに対し、参加者 B は図 7 のように評価した。各軸で評価が特徴的なアイデアについては、参加者 B にその評価理由を訊ねた (表 6)。

エージェント選定の戦略について、「前半は斬新さ、後半は実現性を重視した」と述べた。実際に、入替②や③では客観的なアイデアを生成するロールが多く選定されていた。

アイデアの評価について、独自性では B-5, B-9 を特に高く評価していた。B-5 および B-8 は実現性が低く、参加者 B はこの理由を「スポットを『作る』提案であり、コストがかかるため」とした。一方、実現性では B-4 を最も高く評価している。独自性の評価は低い、「個人的に行きたいと感じた」と述べた。また、B-1 については、「自分でやろうとは思わない」という理由から不採用としていた。

表 5 参加者 B のエージェントの選定

回数	エージェント
初回	子ども, コック, 芸術家, 詩人, 発明家
入替①	子ども(2), ジャーナリスト, 看護師, 船長, 酔っ払い
入替②	コック(2), 弁護士, 評論家, 評論家, コーチ
入替③	子ども(3), 弁護士(2), 戦略家(2)

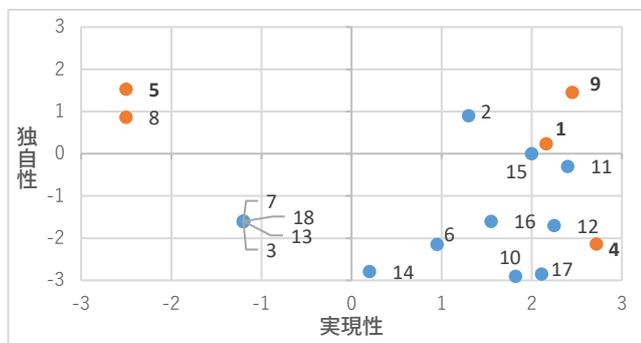


図 7 参加者 B によるアイデアの評価  
(データラベルは ID, 文中では B-n と表現)

表 6 参加者 B の評価が特徴的だったアイデア

ID (B-n)	ロール	アイデア要旨	実現性	独自性	備考
1	詩人	屋内ワークショップを開催	2.16	0.23	不採用
4	子ども	ビニールハウスでイチゴ狩り	2.72	-2.14	
5	発明家	氷上アトラクションを作る	-2.50	1.52	
8	ジャーナリスト	室内スキーリゾートを開発	-2.50	0.86	
9	子ども	おうちキャンプ	2.45	1.45	採用可

また、エージェントのアイデア評価について、参加者 B は「自身とずれているように感じた」と述べた。実際に、参加者 B が特に高く評価した B-9 はシステム内評価では星が 1 つも付いていなかった。他にも、B-4, B-8 などエージェントの評価が参加者 B の評価に反して低くなっていた。

## 5.3 参加者 C の利用例

テーマは「人の感情を生起させるタンジブルなインタフェース」で、システムの利用時間は約 12 分だった。表 7 は参加者 C が選定したエージェントの一覧である。

生成された 20 件のアイデアに対し、参加者 C は図 8 のように評価した。各軸で評価が特徴的なアイデアについては、参加者 C にその評価理由を訊ねた (表 8)。

エージェント選定の戦略として、「選択画面の説明を参考に、『感情』関連の提案をしそうなロール (詩人, 芸術家) を多く選んだ」と述べた。また、参加者 C は研究活動の一助とする為に本テーマを設定しており、ロール「科学者」や「発明家」はその観点から選定したものと説明づけた。

表 7 参加者 C のエージェントの選定

回数	エージェント
初回	発明家, 芸術家, 科学者, コック, 子ども
入替①	発明家(2), 芸術家(2), 詩人, 評論家, 看護師
入替②	発明家(3), 詩人(2), コーチ, 戦略家, 酔っ払い
入替③	芸術家(3), 詩人(3), ジャーナリスト, 弁護士, 船長

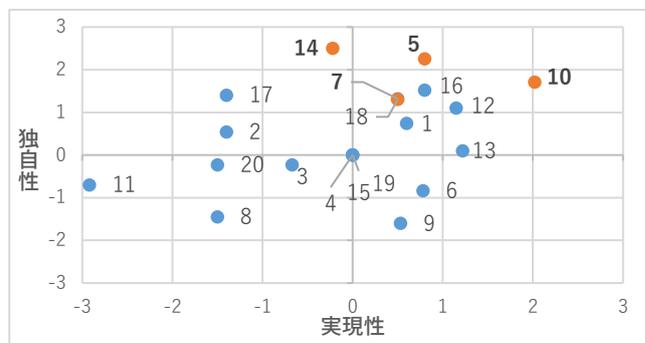


図 8 参加者 C によるアイデアの評価  
(データラベルは ID, 文中では C-n と表現)

表 8 参加者 C の評価が特徴的だったアイデア

ID (C-n)	ロール	アイデア要旨	実現性	独自性	備考
5	子ども	感情に反応するプレゼント箱	0.80	2.25	不採用
7	看護師	患者の感情を絵文字で表現	0.50	1.32	不採用
10	評論家	触れると当時の感情が伝わる写真立て	2.02	1.70	採用可
14	酔っ払い	お酒が入った感情ボトル	-0.22	2.50	

アイデアの評価について、独自性で C-14 を最も高く、実現性では C-10 を最も高く評価した。特に C-10 については独自性のスコアも高く、インタビューでも「自身の意向に近いアイデアである」と述べていた。C-5 および C-7 については、実現性・独自性とも高評価だが不採用としていた。これらの理由について、参加者 C は「面白いアイデアだが研究には向かないと感じた」と述べた。

## 5.4 参加者 D の利用例

テーマは「新しい SNS プラットフォーム」で、システムの利用時間は約 25 分だった。表 9 は参加者 D が選定したエージェントの一覧である。なお、参加者 D はロール「評論家」および「弁護士」を一度も選定しなかった。

生成された 20 件のアイデアに対し、参加者 D は図 9 のように評価した。各軸で評価が特徴的なアイデアについては、参加者 D にその評価理由を訊ねた（表 10）。

表 9 参加者 D のエージェントの選定

回数	エージェント
初回	発明家, 芸術家, 詩人, 子ども, 戦略家
入替①	発明家(2), 芸術家(2), 詩人(2), コーチ, コック
入替②	発明家(3), 芸術家(3), ジャーナリスト, 看護師, 船長
入替③	発明家(4), ジャーナリスト(2), 子ども(2), 科学者, 酔っ払い

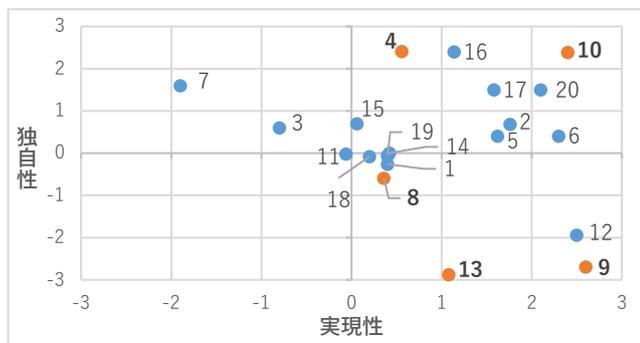


図 9 参加者 D によるアイデアの評価  
(データラベルは ID, 文中では D-n と表現)

表 10 参加者 D の評価が特徴的だったアイデア

ID (D-n)	ロール	アイデア要旨	実現性	独自性	備考
4	発明家	仮想の宇宙空間でコミュニケーション	0.56	2.40	採用可
9	コック	料理の写真やレシピを共有	2.60	-2.70	
10	発明家	個々の感情を色で表現	2.40	2.38	採用可
13	船長	新しいコミュニケーションの形を目指す	1.08	-2.88	

エージェント選定の戦略としては、「具体的で新しいアイデアを期待して選んだ」とし、「ありきたりなアイデアを出すロールは入替時になるべく除外した」と述べていた。4 回連続でロール「発明家」を選んだ理由としては、「自分に刺さるアイデアが多く、引き続き意見を聞きたいと思った」と述べた。後のアイデア評価においても、ロール「発明家」で生成されたアイデア 4 件のうち 3 件を実現性・独自性ともにスコア 1.50 以上と高く評価している。

アイデアの評価について、独自性ではアイデア 4 を最も高く評価していた。この理由として、参加者 D は「『宇宙空間』をベースにした発想が SNS というテーマに対し斬新だと感じた」と述べた。実現性では D-9 を最も高く評価しているが、一方で実現性はスコア -2.70 と非常に低い。理由として、参加者 D は「『料理を共有する』というコンセプトが、既存のサービスに酷似していたため」と述べた。

実現性の評価が最も低い D-13 について理由を訊ねたところ、「アイデアが抽象的な目的の提示に留まっており、具体的な特徴が掴みづらかった」と述べた。参加者 D は本評価における実現性を「実装方法を想定しやすいか」という軸で考えており、要件や目的を抽象的に述べるアイデアは実現性・独自性ともに低くなる傾向にあった。

## 6. 考察

### 6.1 インタフェース

実験でのシステム利用や事後インタビューを通じて、UI デザインに関するポジティブな反応が多数確認できた。参加者 B はシステムの利用中にエージェントやカメラの移動を多く行っており、事後インタビューでも「アイデアの生成中にシステムを触って暇つぶし出来た」と述べていた。また、参加者 C はデザインのポップさに着目しており、「アイデア出しそのものに対するストレスが和らぐ」と述べた。これらの反応からも、本システムのデザインはアイデア出しに係るストレスの緩和に役立つ可能性も考えられる。

一方で、アイデア一覧画面（図 4）については、「アイデアの発言者が一目で分かりにくい」という改善点が挙げられる。現状はフキダシの色が各ロールに対応しているが、エージェントのアイコンなどを追加することでより分かりやすくなると思われる。

### 6.2 アイデアの評価

#### 6.2.1 ロールデータとの関連

実験で生成されたアイデアには、ロールの影響から実現性の低いアイデアが生成されることが複数あった。特にその傾向が顕著だったのがロール「コック」であり、料理の要素が過度に影響した結果、テーマから逸れたアイデア（A-7, C-14 等）が生成されることが多々あった。

また、ロール「船長」「戦略家」が生成したアイデアは、全体的に独自性の評価が低い結果となった。この結果は、プロンプトに「統括的」「俯瞰的」といった語を取り入れたことが原因として考えられる。これらの語はシックス・ハット法[3]の青色の帽子を参考に設定したもののだが、結果としてアイデアが抽象的なコンセプトに留まりやすくなったものと考えられる。

### 6.2.2 テーマの向き・不向き

利用例から、提案システムは「現実に存在しない、創造的な答えが求められるテーマ」での利用に適すと考える。

まず、「現実に存在しない」とした理由は、参加者 B の例において意向と異なる趣旨のアイデアが多く見られたためである。参加者 B は実在する場所に出かける様な提案を求めているが、生成されたアイデアには「冬のイベントを企画・主催する」旨のものが多く見られた (B-1・5・8・18 等)。原因としては、アイデア生成時の共通プロンプトにある「新しいアイデアを提案してください」という記載が優先されたためであると考えられる。

次に、「創造的な答えが求められる」とした理由は、参加者 A の例において一部アイデア (A-4・8・12 等) に内容の重複が見られたためである。アイデアを大量に生成できる本システムは、答えが全く定まっていなような、創造性が求められるテーマでの利用が特に適していると考えられる。なお、これらの傾向はあくまで現状の提案システムにおけるものであり、生成傾向を調整することでより多様なテーマに沿ったアイデアを生成可能になると考えられる。

### 6.3 機能面の改善

提案システムではテーマを途中で再設定でき、変更前のテーマとは別のテーマとして処理される。そのため、古いテーマで生成されたアイデアから派生したアイデアが生成されることもない。しかし、これは一文字でもテーマを変えると派生処理が機能しなくなるということでもある。参加者 B や C の例では、参加者の意向と異なる趣旨のアイデアが複数生成されており、これはテーマを微修正することで改善が可能だった。そこで、システム内では同一のテーマとして扱いつつ、文面を微修正する機能を追加することが望ましいと考える。

また、参加者 C・D はインタビューにて「好みのアイデアについて追加で説明を聞いたかった」と述べていた。提案システムではエージェントが一度生成したアイデアは閲覧以外の干渉が出来ない。アイデアの説明を追求する機能があれば、説明不足だが興味深いコンセプトのアイデアなどを取りこぼす可能性が低減できると考える。

## 7. 結論

本論文では、様々なロールの LLM 搭載エージェントを用いたアイデア生成システム『Agent Agora』の提案を行った。提案システムにおいて、ユーザに求められる操作は「エージェントの引き入れ/追い出し」と「テーマの書き込み」のみであり、放置していてもアイデアが生成される楽なシステムであるといえる。提案システムを用いて利用例を集めたところ、参加者全員が一定数のアイデアに対し採用可能と評価した。一方で、インタビューやアイデアの評価を通じて、システム利用時におけるテーマの微調整やアイデアの追求機能の実装、またアイデアの生成傾向の改善といった展望が示された。これらの課題を改善することで、アイデアの質や選定のしやすさを向上させ、より楽に良いアイデアを得られるシステムにできると考える。

## 参考文献

- [1] Introducing ChatGPT. <https://openai.com/blog/chatgpt>. (2024/02/19 確認)
- [2] Edward de Bono (2014). *Lateral Thinking: An Introduction*. Vermilion. (エドワード・デボノ 藤島みさ子 (訳) (2015). 水平思考の世界: 固定観念がはずれる創造的思考法 きこ書房)
- [3] Edward de Bono (1999). *Six Thinking Hats*. Back Bay Books. (エドワード・デ・ボノ 川本英明 (訳) (2015). 6つの帽子思考法——視点を変えると会議も変わる パンローリング株式会社)
- [4] 趙曉婷, 高島健太郎, 西本一志. 「子供の発想」を利用するアイデア生成技法の提案とその有効性の検証, 情報処理学会研究報告, 2018-GN-104, 1, pp.1-6. (2018)
- [5] 下村賢人, 高島健太郎, 西本一志. 飲酒による認知機能への影響を活用する発散的思考技法の検討, 情報処理学会研究報告, 2020-GN-110, 9, pp.1-7. (2020)
- [6] Chi-Min Chan, Weize Chen, Yusheng Su, Jianxuan Yu, Wei Xue, Shanghang Zhang, Jie Fu, Zhiyuan Liu. ChatEval: Towards Better LLM-based Evaluators through Multi-Agent Debate. In arXiv preprint arXiv:2308.07201
- [7] Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang and Michael S. Bernstein. Generative Agents: Interactive Simulacra of Human Behavior. In arXiv preprint arXiv:2304.03442
- [8] Zhenhailong Wang, Shaoguang Mao, Wenshan Wu, Tao Ge, Furu Wei and Heng Ji. Unleashing Cognitive Synergy in Large Language Models: A Task-Solving Agent through Multi-Persona Self-Collaboration. In arXiv preprint arXiv:2307.05300
- [9] Weize Chen, Yusheng Su, Jingwei Zuo, Cheng Yang, et al. AgentVerse: Facilitating Multi-Agent Collaboration and Exploring Emergent Behaviors in Agents. In arXiv preprint arXiv:2308.10848
- [10] Models - OpenAI API. <https://platform.openai.com/docs/models/continuous-model-upgrades>. (2024/02/19 確認)
- [11] DeepL. DeepL API. <https://www.deepl.com/ja/docs-api>. (2024/02/19 確認).