

B-04

対話生成 AI エージェントによるグループコミュニケーションの自動評価に基づく振り返り支援システムの検討

Reflection Support System Based on Automatic Evaluation of Group Communication by a Dialogue-Generating AI Agent

中川 翔瑛†
Shoei Nakagawa

佐野 睦夫†
Mutsuo Sano

1. はじめに

現代社会では、多くの企業において「コミュニケーション能力」が重要視されている。一般社団法人日本経済団体連合会による 2018 年度の新卒採用に関するアンケート調査 [1]によると、新卒学生の選考において最も重視される能力は、16 年連続で「コミュニケーション能力」が第 1 位となっている。この傾向は今後も続く予想され、特にグループでの仕事一般的なとなった現代においては、チーム内の円滑なコミュニケーションによるグループ活動が、成果を左右する重要な要素となっている。

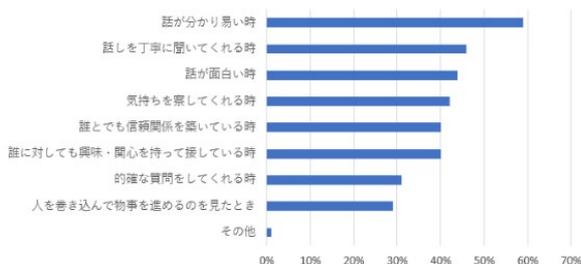


図 1. コミュニケーション能力に関するアンケート (2019 4350 名)[3]

しかし、コミュニケーション能力やグループコミュニケーション能力を手軽に鍛えることができる、自動化されたシミュレーションシステムは、現時点ではまだ少ない。

そこで山部ら[2]は、「相手に対してグループコミュニケーション能力が高いと感じるのはどのようなときですか?」といったアンケート[3]を用いて、グループコミュニケーション能力を「意思伝達能力」「ファシリテーション力」「アサーション」の 3 つに分類した。アンケートの結果は、以下の通りである。

「意思伝達能力」には、話の内容を分かりやすく、かつ的確に伝える力が求められる。また、「ファシリテーション力」には、周囲の気持ちを察し、関係者を巻き込みながら物事を推進する能力が含まれる。さらに、「アサーション」には、相手の話を丁寧に聞き、誰とでも信頼関係を築く能力が求められる。山部ら[2]は、これら 3 つの能力に焦点を当てた 2 人の NPC との対話を通じて、自動評価を行う VR トレーニングシステムを開発した。しかし、同実験では特定のシナリオに基づいたトレーニングのみが実施されており、現実のコミュニケーション場面が多様であることを踏まえると、より実践的なシナリオに柔軟に対応可能なシステムが求められる。

また、NPC の発話は事前に生成された音声を用いており、対話の即時性や柔軟性に課題があるため、NPC 自体の対話能力の向上も必要である。

そこで本研究では、重要なシナリオの一例として、就職活動における面接試験を対象とし、複数の LLM (大規模言語モデル) を搭載した NPC と対話することにより、グループコミュニケーション能力を鍛えるとともに、自動評価に基づく振り返りが可能なシステム手法を提案する。また、本稿では、その実現に向けた予備実験として、システムに必要な要素を明らかにすることを目的に、1 対 1 の面接試験を体験可能なシステムを開発し、その有効性について考察を行う。本手法により、重要な場面の予行演習やグループでの会話のシミュレーションが可能となり、グループコミュニケーション能力の向上が期待される。

2. 関連研究

近年の LLM の技術的発展により、面接試験の練習は容易になりつつある。しかし、Daryanto ら[4]は、面接試験の練習に LLM を用いることの利点や課題を検討した研究はほとんど存在しないとし、その必要性を指摘している。

また、従来の研究では、実験参加者が自身の能力について一方的なフィードバックを受けるものが大半であり、対話型のフィードバックを取り入れた研究はほとんど見られない。一方で、実験参加者が対話を通じてフィードバックに関与することで、面接の現実感が高まり、より深い思考が促されるとともに、認知的負荷や過度な自己批判が軽減されることが示されている。

そこで本研究においても、対話型フィードバックを取り入れる。特に、グループコミュニケーションにおいて対話型フィードバックを導入することで、NPC ごとの評価を得ることが可能となり、グループコミュニケーション能力をより的確に測定・向上させることが期待される。本稿では、まず 1 対 1 の面接試験における実験を行うため、面接官に相当する NPC からの評価を対話形式で提供できるように設計している。

さらに、近年では LLM の信頼性が問題視されている。Nofal ら[5]は、LLM を用いた面接試験において、モデルごとの評価傾向を分析し、LLM の種類によって参加者の背景情報に基づくバイアスが生じる可能性を指摘している。したがって、面接試験に LLM を用いる場合には、LLM によるバイアスの存在を考慮する必要がある。あわせて、LLM が生成する内容の信頼性に対する評価も重要である。

本稿で扱う 1 対 1 の面接試験では、LLM によるバイアスの影響を排除するため、実験参加者の背景情報を非開示の状態で行う。一方で、今後のグループコミュニケーション

ョンに関する実験では、背景情報を開示したうえで公正な評価が行えるよう、適切な評価指標の構築が求められる。

3. 提案方式

3.1 研究概要

本研究では、VR を用いて重要な場面の予行演習を可能にし、手軽にグループでのコミュニケーションを行うことができるシステムを提案する。また、本システムを用いることで、グループコミュニケーション能力の自動評価が可能となる。評価方法には、山部ら[2]のシステムを活用し、LLM を用いて項目ごとの採点を行う。さらに、振り返り支援として、対話終了後に NPC からのフィードバックを提示する。その後、アンケート調査を実施し、本システムの有効性を検証する。

本稿における予備実験では、1 対 1 の面接試験を対象とし、面接終了後に面接官役の NPC からフィードバックを受け取る。その後、システムに関するアンケート調査を実施し、その結果に基づいて本システムの有効性を検討する。

3.2 提案システム

図 2 に本研究で提案するシステムの構成図を示す。

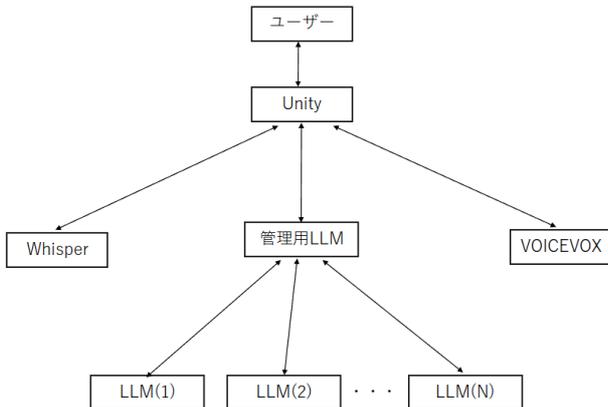


図 2. 提案システム図

図 2 に示すように、本システムは Unity 上で稼働する。実験参加者の発話音声は、OpenAI が開発した Whisper を用いてテキスト形式に変換される。NPC が生成した応答テキストは、Hiroshiba によって開発された VOICEVOX を用いて音声に変換され、実験参加者に提示される。変換されたテキストは、管理用 LLM に送信され、その内容に基づいて、該当する NPC を担当する LLM へ転送される。近年では、Google の Agent2Agent プロトコルのように、管理用 LLM を中心として各 LLM に役割を割り振ることで、複数の LLM による協働を実現する手法が提案されている。本研究においても、このような枠組みを参考に、管理用 AI をシステムに組み込むことで、実験参加者と複数の NPC との対話を可能とする。

Unity 上では、図 3 に示すように、実験参加者に対話画面が提示される。図 3 では、2 人の NPC が表示されているが、シナリオに応じて NPC の人数を変更可能なシステムとして開発する。また、実験参加者についても、1 人だけでなく最大 2 人まで同時に参加可能となるよう、拡張を行う。

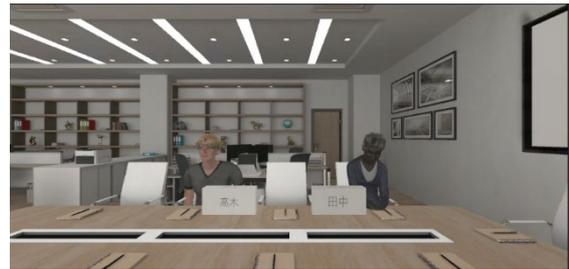


図 3. 提案システムの画面

なお、本稿で実施する予備実験は 1 対 1 の面接試験を対象としており、図 2 で示した構成とは異なる。そこで、予備実験におけるシステムの構成を図 4 に示す。

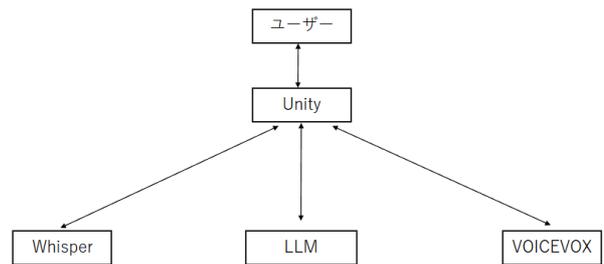


図 4. 予備実験の構成図

予備実験では、LLM として、東京科学大学情報理工学院および国立研究開発法人産業技術総合研究所によって開発された「Llama 3.1 Swallow 8B Instruct v0.3」を使用する。本モデルは、日本語生成に対応したモデルの中でも、特にロールプレイング性能において高い評価を得ている。

また、実験参加者には、図 5 に示すような画面が表示される。



図 5. 予備実験の画面

また、予備実験では本実験とは異なり、VR ゴーグルは使用せず、ディスプレイ上に表示される面接官と面接試験を行う。

3.3 評価項目と定義

本研究では、グループコミュニケーション能力を測定するにあたり、山部ら[2]の評価方法を採用する。同研究では、グループコミュニケーション能力を 3 つの観点に分類しており、それを表 1 に示す。

表 1.評価項目と定義[2]

項目	定義
意思伝達能力	相手に伝わりやすいよう、話が端的にまとめられ、自身の意見を明確に伝えられている。 話しの道筋の立った、論理的な意見が言えている
アサーション	相手の意見を尊重した上で、率直な自己主張ができている。 自分の意見だけでなく、相手の意見に対しても肯定的に聞くことができる。 相手の目を見て話している
ファシリテーション力	人を巻き込んで物事を進める 議論の進め方などの提案をする。 相手の意見よく聞き、それに対して的確な質問を行えている

グループコミュニケーション能力は、「意思伝達能力」、「アサーション」、「ファシリテーション力」の3つに分類されるが、予備実験では複数人でのコミュニケーションを扱わないため、本稿ではこの評価方法を採用しない。本評価方法は、今後グループを対象とした本実験を行う際に取り入れる予定である。

3.4 予備実験のシステム内容

本稿で実施する予備実験では、面接試験を行う。その流れを図6に示す。

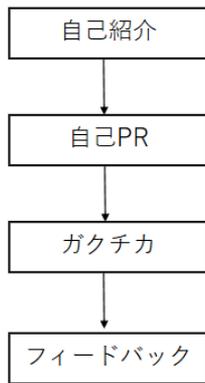


図 6.面接試験の流れ

図6に示すような流れで、本システムを体験することが可能である。面接中には、面接官(NPC)から「自己紹介をお願いします」などの案内が提示されるため、実際の面接試験と同様の流れで進行する。また、自己PRや「学生時代に力を入れたこと(いわゆるガクチカ)」に対しては、面接官から内容の深掘りが行われる。例えば、参加者が「プロジェクト活動に力を入れました」と回答した場合、面接官から「どのようなプロジェクトで、どのように力を入れたのですか」といった追加の質問が行われる。さらに、面接の終了時には、面接官から実験参加者の強みや、会話内容における良かった点についてのフィードバックが提示される。本システムでは、実験参加者は自由に面接官に質問などを行うことが可能である。

本システムでは、実験参加者が面接官に自由に質問を行うことも可能である。通常は、図6のような遷移を経て、実験全体は5~10分程度で完了する。

4. 実験

4.1 実験方法

本実験の実施にあたっては、大阪工業大学における人を対象とする研究倫理審査委員会の承認(審査番号:2023-10-1)に基づき実施した。実験方法としては、参加者に本システムを5~10分程度体験してもらった後、アンケート調査を実施した。

4.2 参加者

実験は、就職活動を終えた大学生3名と、就職活動を控えた大学院生3名を対象に実施した。

4.3 結果

実験後に、12項目からなるアンケート調査を実施した。本システムの使用感に関するアンケート結果を図7に示す。図7の結果から、実験参加者全員が「少し使いやすい」または「使いやすい」と評価しており、本システムは概ね高い評価を得たことがわかる。

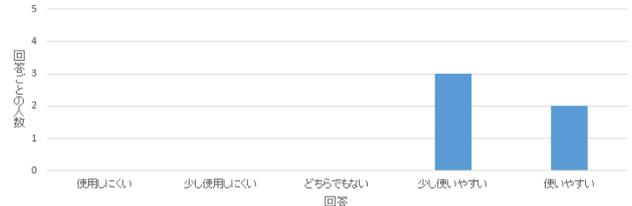


図 7.システムの使用感について

本システムを使用した際に、ストレスを感じたか、あるいはリラックスできたかに関するアンケート結果を図8に示す。図8の結果から、一部の参加者が「少しストレスを感じた」と回答していることが確認された。

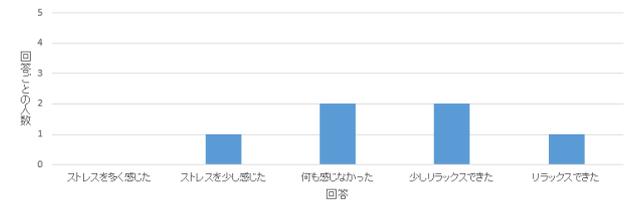


図 8.システムの使用時の感情について

面接官の応答が適切であったかどうかに関するアンケート結果を図9に示す。多くの参加者が「少し適切」または「適切」と回答した一方で、「少し適切ではない」と回答した参加者も見られた。

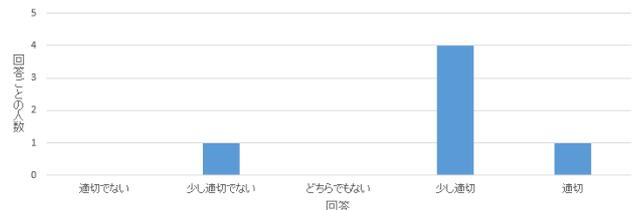


図 9.面接官の回答について

本システムの使用によって自己への気づきが得られたかに関するアンケート結果を、図10に示す。図10の結果から

ら、本システムを使用することで、参加者全員が何らかの気づきを得たことが確認された。

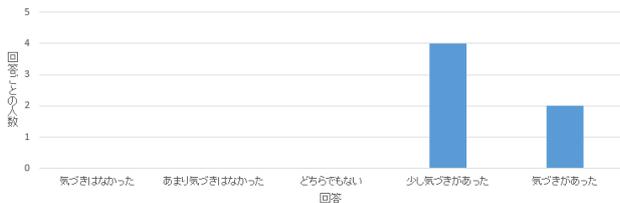


図 10.気づきを得たかについて

自身の改善点を見つけることができたかに関するアンケート結果を、図 11 に示す。図 11 の結果から、参加者全員が何らかの改善点を見出したことが確認された。



図 11.改善すべき点を見つけることができたか

面接に対する自己否定感が軽減されたかに関するアンケート結果を、図 12 の結果から、2名の参加者が「少し軽減できた」と回答した一方で、残りの参加者は軽減を実感していなかったことがわかる。

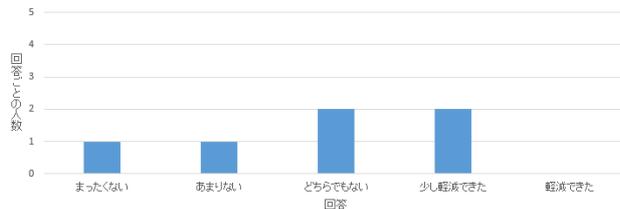


図 12.自己否定感が軽減されたかについて

自由に質問を行うことができたかに関するアンケート結果を、図 13 に示す。図 13 の結果から、自由に質問ができたと回答した参加者がいた一方で、全く質問できなかったと回答した参加者も確認された。

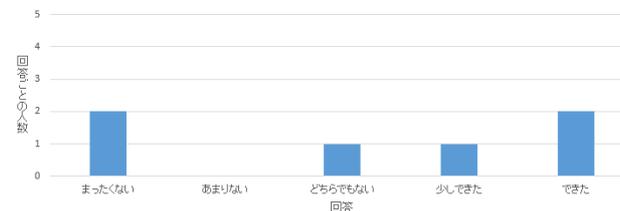


図 13.自由な質問ができたかについて

面接後に行われたフィードバックが適切であったかに関するアンケート結果を、図 14 に示す。図 14 の結果から、参加者全員が「少し適切」または「適切」と回答しており、フィードバックは概ね適切であったと評価されたことがわかる。

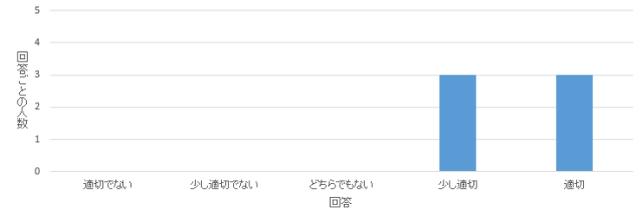


図 14.フィードバックが適切だったかについて

失敗しても大丈夫だと感じたかに関するアンケート結果を、図 15 に示す。図 15 の結果から、「大丈夫だと感じた」と回答した参加者がいた一方で、そう感じなかった参加者も確認された。

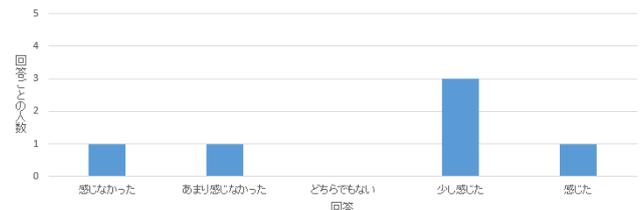


図 15.失敗しても大丈夫だと感じたかについて

本システムを再度使用したいと感じたかに関するアンケート結果を、図 16 に示す。図 16 の結果から、多くの参加者が再度使用したいと回答した一方で、1名はあまりそう感じなかったことが確認された。

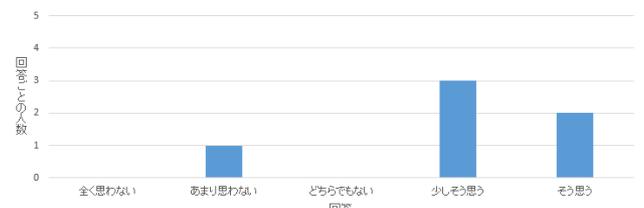


図 16.またシステムを使用したいか

本物の面接に近いと感じたかに関するアンケート結果を、図 17 に示す。図 17 の結果から、多くの参加者が本物の面接に近いとは感じなかったことが確認された。

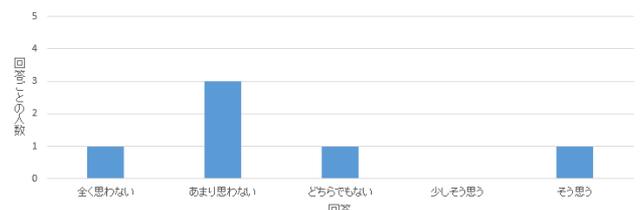


図 17.本物の面接に近いと感じたか

また、自由記述形式による本システムに対する意見も収集した。その結果、面接内容の適切性や応答速度、および自己への気づきに関する点では高い評価が得られた一方で、システム内で表示される会話ログ機能については否定的な意見も見られた。

5. 考察

図 8 のシステムの使用感、および図 9 の面接官の応答に関するアンケート結果から、予備実験で用いた本システム

は、面接試験において一定の実用性を有することが示唆された。特に、多くの参加者が応答を「適切」または「少し適切」と評価しており、参加者に応じた面接が実現されたことがうかがえる。

また、図 10 の気づきを得たか、図 11 の改善すべき点を発見できたか、および図 14 のフィードバックの適切性に関するアンケート結果から、参加者全員が本システムの利用を通じて自己の成長を実感したことが確認された。

これらの結果から、本システムはグループコミュニケーション能力向上の支援を目的としたプロトタイプとして有効に機能することが示唆される。

一方で、図 16 の再度使用したいかという設問では多くの参加者が肯定的に回答したものの、図 8 の使用時の感情、図 12 の自己否定感の軽減、図 13 の自由な質問ができたか、および図 15 の失敗しても大丈夫だと感じたかに関する結果では、一部に否定的な回答も見られた。これらの項目は、Daryanto ら[4]の研究においても指摘されているように、反復的な訓練におけるモチベーションの低下に影響を与える可能性がある。そのため、本番環境では、システムの導入説明を充実させるとともに、対話構成やフィードバックの設計を工夫する必要がある。

特に、LLM は参加者を励ます役割を担うと同時に、過度に干渉しないバランスの取れた応答が求められる。さらに、LLM の出力内容には信頼性の問題も存在する。Nofal ら[5]の研究でも指摘されているように、LLM による回答にはモデルごとのバイアスが含まれる可能性があり、今後本番環境で LLM を用いる際には、出力の妥当性を評価するための指標の構築が必要となる。

また、図 13 の自由な質問ができたかに関するアンケート結果は参加者によって回答が分かれた。これは、実験参加者の主体性の違いによるものである可能性が考えられる。したがって、今後は LLM のプロンプト設計や対話設計において、参加者の主体性を引き出す工夫が求められる。主体性が高まることで、より双方向的な対話を実現され、コミュニケーションスキル向上への効果が期待される。

さらに、図 17 の本物の面接に近いと感じたかに関するアンケートでは、あまりそうは思わなかったと回答した参加者が最も多かった。この要因としては、自由記述にも見られたように、図 5 に示す通り本システムでは会話内容が画面に表示されるため、参加者の注意が文字情報に向かい、面接官への視線が分散したことが考えられる。また、NPC の動作が会話内容と連動していなかったことや、音声の印象がキャラクターと一致していなかったことも、没入感の低下に影響したと考えられる。

したがって、本番環境においては、NPC の動作や音声表現の改善が必要である。さらに、実際の面接環境に近づけるためには、適度な緊張感を再現する演出も重要であり、それにより面接スキルやコミュニケーション能力のさらなる向上が期待できる。

6. おわりに

本研究では、対話生成 AI エージェントによるグループコミュニケーションの自動評価に基づく振り返り支援システムについて検討した。その予備実験として、1 対 1 の面接試験を想定したシステムを開発し、実験参加者に対するアンケート調査を通じてその有効性を評価した。

アンケート結果からは、本システムが参加者の自己認知や改善点の発見に寄与し、グループコミュニケーション能

力向上を目的としたシステムのプロトタイプとして一定の有用性を持つことが示唆された。

一方で、LLM による応答の信頼性や、実際の面接との臨場感の差異など、改善すべき課題も明らかとなった。今後は、LLM のプロンプトエンジニアリングの工夫や、NPC の動作・音声の自然さの向上といった改良を通じて、より現実的で効果的なコミュニケーション支援システムの構築が期待される。

参考文献

- [1] 一般社団法人日本経済団体連合会, 「2018 年度新卒採用に関するアンケート調査結果」, <http://www.keidanren.or.jp/policy/2018/110.html?v=s> (2025 年 7 月 24 日閲覧) .
- [2] 山部温志, 佐野睦夫: グループコミュニケーション能力向上のための VR 空間自動評価トレーニングシステムの実装, インタラクション 2025 論文集, pp. 891-895.
- [3] エン・ジャパン, 9 割が「コミュニケーション能力は現在の仕事に影響する」と回答。『エンバイト』ユーザーアンケート, <https://corp.en-japan.com/newsrelease/2020/21098.html> (2025 年 7 月 24 日閲覧) .
- [4] DARYANTO, Taufiq, et al. Conversate: Supporting reflective learning in interview practice through interactive simulation and dialogic feedback, Proceedings of the ACM on Human-Computer Interaction, Vol.9 No.1, pp.1-32, 2025.
- [5] Nofal, A. B., Ali, H., Hadi, M., Ahmad, A., Qayyum, A., Johri, A., ... & Qadir, J. AI-enhanced interview simulation in the metaverse: Transforming professional skills training through VR and generative conversational AI, Computers and Education: Artificial Intelligence, Vol.8, 100347, 2025.