

対人スキルと対 AI スキルの関係

西岡 竜生[†] 若宮 翔子[†] 清水 伸幸^{††} 藤田 澄男^{††} 荒牧 英治[†]

[†] 奈良先端科学技術大学院大学 〒 630-0192 奈良県生駒市高山町 8916-5

^{††} LINE ヤフー株式会社 〒 102-8282 東京都千代田区紀尾井町 1-3

E-mail: [†]{nishioka.ryuki,np4,wakamiya,aramaki}@is.naist.jp, ^{††}{nobushim,sufujita}@lycorp.co.jp

あらまし 人と AI が協働することで、AI 単体を上回るパフォーマンスが発揮されることが注目されている。近年においては、大規模言語モデル (Large Language Model; LLM) やチャットボットなどの対話型 AI が急速に発展し、教育分野等への応用が検討されている。人と LLM の協働において、LLM の性能のみならず LLM を利用するユーザの能力も考慮する必要があると考えられる。しかしながら、ユーザのスキルと LLM のパフォーマンスとの関係は明らかにされていない。そこで、本研究では対人スキルと対 AI スキルが相関するかどうかを分析した。具体的には、ユーザのコミュニケーションスキルの高さと ChatGPT を用いたタスクの成果の関係を分析した。その結果、一部の対人スキルと一部のタスクにおける対 AI スキルに関係が見られた。

キーワード 大規模言語モデル (LLM), ChatGPT, HCI, 対人スキル, 対 AI スキル

1 はじめに

人と AI が協働することで、AI 単体を上回るパフォーマンスが発揮されることが注目されている。このような現象はしばしば、フリースタイルチェスにおいて「ケンタウロス」という人と AI の混成チームが AI 単体に勝利したことから「ケンタウロス現象」と呼ばれる¹²。このことから、AI の効果的な活用には人間の介入が不可欠であると考えられる。

近年においては、大規模言語モデル (Large Language Model; LLM) を用いた ChatGPT のようなチャットサービスが普及し、教育分野等への応用が検討されている。このような LLM の応用に関して、LLM の性能を評価する研究や、LLM の利用による可能性を探求した研究がなされている。Bubeck らは様々なタスクを通して GPT-4 の性能を調査している [1]。また、Big Five [2] や Myers-Briggs Type Indicator (MBTI) [3], [4] といった心理測定尺度を用いて LLM を評価し、LLM が疑似的なパーソナリティを持つ可能性や、LLM の挙動を理解するためにはこのような疑似的なパーソナリティの理解が必要であることが示されている [5], [6]。このように、LLM の性能に着目した研究は多いが、LLM を利用する人間の能力を考慮した研究は少ない。

そこで、本研究では人間の能力 (**対人スキル**) と AI との協働によって高いパフォーマンスを発揮する能力 (**対 AI スキル**) の関係について分析を行う。ここで、ChatGPT は対話形式で利用するモデルであることから、人の能力として対人スキルに

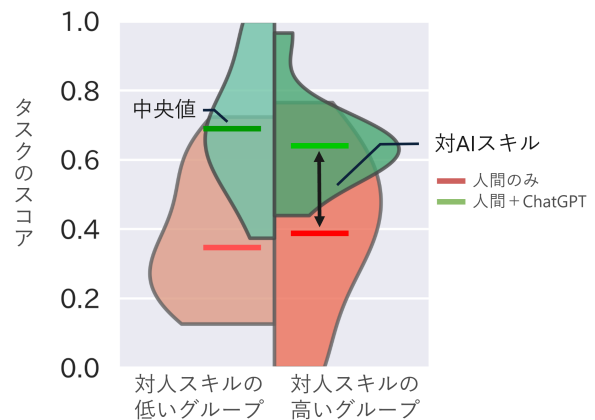


図 1 対人スキルと対 AI スキルの関係分析

着目する。具体的には以下の手順で分析を行う。はじめに、対 AI スキルを測定するための、ChatGPT を利用して取り組む知識タスク、言語タスク、計算タスク、問題解決タスクの 4 種類のタスクを設計する。実験では、心理測定尺度を用いて 6 つの対人スキルを測定した後、タスクに取り組む。この際、設問ごとに ChatGPT を利用可能か否かを指定する。その後、タスクごとに人間が独力で取り組んだ場合と ChatGPT を利用して取り組んだ場合の解答スコアに基づき、ChatGPT を利用した際のスコアの向上度合いを求め、対 AI スキルとする。これらの結果から、6 つの対人スキルと 4 つの対 AI スキルの関係を分析する。本研究のコンセプトと可視化のイメージを図 1 に示す。

2 関連研究

人と LLM の協働によるパフォーマンスが教育分野などで探求されている。Caballé らはソフトウェア教育における対話エージェントの介入がオンライン学習に与える影響を調査している [7]。その結果、対話エージェントによって受講者のオンラ

1: "AI (人口知能) が引き起こす「ケンタウロス現象」に備えていますか?", 株式会社ラーニングデザインセンター, 清宮普美代, <https://ldcjp.com/event-report/ai> (人口知能) が引き起こす「ケンタウロス現象」(2024/2/16 参照)

2: "AI との協働が最適解を生む マルチエージェントシステムが示唆する人間と AI の関係", 立命館大学, <https://shiruto.jp/technology/2389/> (2024/2/16 参照)

イン学習への関与が向上する同時に脱落率が低下する可能性を示している。Ding らはニュースの見出し文の生成を通してライティングタスクにおける人と AI の協働を分析している [8]。人間が手動で生成した場合、LLM が生成した場合、人間と LLM が協働した場合を比較した結果、LLM 単体でも十分な見出しは生成可能だが、出力の修正等における人間の介入が必要であることを示している。Zhao らは物語生成を通して人と LLM の協働を探索している [9]。LLM のみによって生成された物語と、LLM によって生成された文と人間が生成した文を交互に挿入した物語を比較した結果、LLM のみによって生成された物語の方が理解しやすく好まれる傾向が示されている。

AI の性能に着目した人との協働パフォーマンスの向上が検討されている。Li らは LLM が人間のような推論能力や専門スキルを備えるためのフレームワークを提案し、就職フェアのシミュレーションによって評価している [10]。その結果、タスクワークフローの提案等において有望な性能を示している。Rastogi らは人による LLM の動作の監査・評価をサポートする LLM を提案している。実験の結果、ツールの有効性とより良い監査結果に関連する人間の行動が明らかにされている。

以上のように、人と LLM が協働することによる可能性や、人と協働する LLM の性能に関する研究は多数ある。しかしながら、LLM と協働する人の特徴を考慮した研究は少ない。本研究は LLM ではなく LLM を利用する人に着目し、人と AI の協働による可能性を探求している点において以上の研究と異なる。

3 手法

3.1 対人スキル

対人スキルの自己評定尺度として、包括的なコミュニケーションスキルを測定する ENDCOREs [11] を用いる。本尺度はコミュニケーションスキルを対人スキルと基本スキルの階層構造で定義しており、国内外の複数の心理測定尺度を基にコミュニケーションスキルを多面的に評価する。回答には「とても苦手」を 1 点、「とても得意」を 7 点とするリッカート尺度を使用する。使用する質問項目を表 1 に示す。なお、本研究では ENDCOREs の 6 つのメインスキルを対人スキルとし、それぞれ自己統制スキル、表現スキル、解読スキル、自己主張スキル、他者受容スキル、関係調整スキルと呼称する。各対人スキルは、ENDCOREs の各スキルごとの平均によって評価する。

3.2 対 AI スキル

対 AI スキルを測定するために、知識タスク、言語タスク、計算タスク、問題解決タスクの 4 種類のタスクを設計した (表 2)。なお、解答はすべて自由記述にて行われるものとする。

知識タスク 一般的な知識を問うものであり、解答数により評価する。

言語タスク 単語の意味や概念の理解や言語による説明能力を問うものであり、知能検査に用いられる設問を参考に設計した。表 2 の設問例 1 のような場合は、解答数により評価し、例 2 のような単語の意味を問う設問は、解答と広辞苑に記載さ

表 1 ENDCOREs [11] の質問項目。回答は「とても苦手」、「苦手」、「やや苦手」、「普通」、「やや得意」、「得意」、「とても得意」の 7 つから当てはまるものを選択する。

メインスキル	サブスキル	項目文
自己統制	欲求抑制	1. 自分の衝動や欲求を抑える
	感情統制	2. 自分の感情をうまくコントロールする
	道徳観念	3. 善悪の判断に基づいて正しい行動を選択する
	期待応諾	4. まわりの期待に応じた振る舞いをする
表現力	言語表現	5. 自分の考えを言葉でうまく表現する
	身体表現	6. 自分の気持ちをしぐさでうまく表現する
	表情表現	7. 自分の気持ちを表情でうまく表現する
	情緒伝達	8. 自分の感情や心理状態を正しく察してもらう
解読力	言語理解	9. 相手の考えを発言から正しく読み取る
	身体理解	10. 相手の気持ちをしぐさから正しく読み取る
	表情理解	11. 相手の気持ちを表情から正しく読み取る
	情緒感受	12. 相手の感情や心理状態を敏感に感じ取る
自己主張	支配性	13. 会話の主導権を握って話を進める
	独立性	14. まわりとは関係なく自分の意見や立場を明らかにする
	柔軟性	15. 納得させるために相手に柔軟に対応して話を進める
	論理性	16. 自分の主張を論理的に筋道を立てて説明する
他者受容	共感性	17. 相手の意見や立場に共感する
	有効性	18. 友好的な態度で相手に接する
	譲歩	19. 相手の意見をできるかぎり受け入れる
	他者尊重	20. 相手の意見や立場を尊重する
関係調整	関係重視	21. 人間関係を第一に考えて行動する
	関係維持	22. 人間関係を良好な状態に維持するように心がける
	意見対立対処	23. 意見の対立による不和に適切に対処する
	感情対立対処	24. 感情的な対立による不和に適切に対処する

れた該当単語の意味の類似度に基づき評価する。具体的には、Sentence BERT を用いてベクトル化して求めたコサイン類似度を用いる。

計算タスク 論理的思考力や計算能力を問うものであり、中学

		グループA	グループB
タスク	問1	人間のみ	人間+ChatGPT
	問2	人間のみ	人間+ChatGPT
	問3	人間+ChatGPT	人間のみ
	問4	人間+ChatGPT	人間のみ

図2 実験設定

入試における文章題を想定して設計し、1問1点として評価する。

問題解決タスク 明確な正答のない設問を用いて、創造的な思考力や発想力を問うものであり、書籍[12]を参考に設計し、解答数により評価する。

これら4種類のタスクに人間が独力で取り組んだ場合とChatGPTを利用した場合の解答スコアに基づき、ChatGPTの利用によるスコアの向上度合いを求め、対AIスキルとする。具体的には、独力で取り組んだ設問の平均とChatGPTを利用して取り組んだ設問の平均の差を対AIスキルとして評価する。

3.3 対人スキルと対AIスキルの関係

対人スキルと対AIスキルの関係は、スピアマンの順位相関係数を用いて評価する。相関係数は、対人スキルが高いと対AIスキルも高いのか（AIをうまく使えるかどうか）を意味している。なお、前述したとおり、対人スキルは6つのスキルからなり、対AIスキルは4つのタスクごとに定義されるため、24(=4×6)の相関係数が得られる。考察においては、これは個別に議論していくことにする。

4 実験と考察

4.1 実験設定

対人スキルと対AIスキルの関係を分析するために実験を行った。本実験には、奈良先端科学技術大学院大学ソーシャル・コンピューティング研究室に所属する成人12名（男性3名、女性9名）が参加した。実験協力者は、心理測定尺度 ENDCOREs (3.1節)に回答し、4種類のタスク(3.2節)として、著者が選定した合計16の設問（各タスクにつき4問ずつ）に解答した。このとき、設問ごとにChatGPTを利用せず独力で解くか、必要に応じてChatGPTを利用可能かを指定する。今回は、設問の難易度による差の影響を軽減するため、図2のように実験協力者を6名ずつ2グループに分け、グループAは各タスクの設問1と2を解く際は独力(w/o AI)、設問3と4を解く際はChatGPTを利用可(w AI)とし、グループBはこれと反対の設定とした。なお、ChatGPTの利用知識を統一するため、実験協力者はプロンプトガイドラインを確認してからタスクに取り組んだ。このプロンプトガイドラインは、OpenAIの提供す

プロンプトの種類

- **Instruction**

指示を与える。

例:

- **Completion**

入力された文の続きを補完する。

例:

テクニック

- **Instructionの際には具体的に明確な指示を行う。役割や条件を付与する。**

例 1:

例 2:

例 3:

知識問題

問1・問2は一人で、問3・問4はChatGPTを使用して解答してください。解答は一行につき一つとなるよう、改行してください。

知識について、自身とChatGPTのどちらが優秀だと思いますか。*

自身

問1

問1. アメリカの歴代大統領をできるだけ多く挙げてください。*
(※一人で解答してください。)

回答を入力

問1の終了時刻*

時刻

— : —

問1の解答にどの程度自信がありますか。*

自信がある

どちらかといえば自信がある

どちらかといえば自信がない

自信がない

図3 プロンプトガイドラインと解答フォームとして用いた Google Form のスクリーンショット。

るドキュメント³を参考に著者らが作成したものであり、プロンプトの種類や工夫の仕方を記載している。プロンプトガイドラインと解答フォームの例を図3に示す。

これらに加え、実験協力者はChatGPTの利用に関するアンケートに回答した。このアンケートでは、ChatGPTの利用頻度、各カテゴリのタスクにおいて自身とChatGPTのどちらが優れていると思うか、各設問の解答に要した時間、各設問の解答に対する自信、各設問に利用したChatGPTとの対話ログを収集する。

4.2 相関分析

ENDCOREsにおける6つの対人スキルと各タスクにおける対AIスキルとの相関係数をスピアマンの順位相関係数によって求めた。対人スキルの自己評定の結果を表3に、各タスクの結果を表4に、相関分析の結果を表5に示す。タスクの結果は、

3: <https://github.com/openai/openai-cookbook/> (2023/12/20 参照)

表 2 タスクの意図と設問例

カテゴリ	意図	設問例	評価方法
知識タスク	一般的な知識を問う	哺乳類の動物をできるだけ多く挙げてください。	解答数
言語タスク	単語の意味や概念の理解、言語による表現力を問う	例 1. 「机」と「椅子」はどのような点で似ていますか。 例 2. 「従う」とはどういう意味ですか。	解答数 広辞苑に記載されている意味との類似度
計算タスク	論理的思考力、計算能力を問う	A は部屋の掃除に 3 時間かかり、B は同じ部屋を掃除するのに 2 時間かかる。A と B が一緒に掃除すると、どれだけの時間で終わらせることができますか。	1 問 1 点
問題解決タスク	思考力、発想力を問う	10 個のミカンを 3 人に平等に分配する方法をできるだけ多く挙げてください。	解答数

表 3 ENDCOREs による対人スキルの自己評定の結果。値はすべて小数第 3 位で四捨五入している。対人スキルは 1 点から 7 点で評価され、4 点が「普通」であることを示すことから、4 点以下の値を“Low”，4 点より大きい値を“High”とした。“High”の値を太字で示す。

実験協力者	自己統制スキル	表現スキル	解読スキル	自己主張スキル	他者受容スキル	関係調整スキル
H1	4.50	5.50	5.50	5.00	5.75	6.00
H2	5.00	3.00	4.75	2.00	4.75	4.50
H3	5.25	2.25	4.25	3.75	6.00	4.75
H4	4.50	3.75	3.75	3.75	4.50	4.00
H5	4.00	4.75	3.25	4.25	2.00	3.50
H6	2.50	2.50	2.25	3.25	3.50	3.25
H7	4.00	4.00	4.00	4.00	4.00	4.00
H8	4.00	3.00	4.00	3.75	4.25	4.00
H9	3.50	1.75	1.25	3.50	4.00	2.50
H10	5.00	3.50	3.50	2.00	6.00	4.00
H11	5.25	4.00	4.25	4.25	5.25	4.75
H12	3.75	3.50	5.50	4.00	4.00	4.00
平均	4.27	3.46	3.85	3.63	4.50	4.10

解答を表 2 に従って評価した後、各設問のスコアを Min-Max Normalization によって [0, 1] の範囲にスケールしている。なお、表 4 の 2 名の実験協力者 (H4 と H6) は、計算タスクにおいて ChatGPT を利用しなかった。そのため、計算タスクにおける対 AI スキルと対人スキルとの相関係数を算出する際は除外した。表 5 より、言語タスクと自己統制スキル ($r=0.61$, $p < 0.05$)、言語タスクと他者受容スキル ($r=0.77$, $p < 0.01$)、言語タスクと関係調整スキル ($r=0.63$, $p < 0.05$) の間に有意な正の相関が見られた。

4.3 対人スキルと対 AI スキルの関係

対人スキルと各タスクの結果の関係を図 4 に、ChatGPT を利用した際の対話のステップ数と各ステップにおけるプロンプトの平均の長さを図 5 に示す。結果として、タスクによって対人スキルとの関係に違いが見られた。言語タスクは、すべてのメインスキルにおいて、対人スキルの高い人の方が大きく向上している。ChatGPT との対話の特徴として、設問をそのまま投げかける場合が多くみられたが、知識タスクのように追加の

出力を促す場合や、例 2 のような単語の意味を解答する設問では「簡単に説明して」や「辞書のように説明して」といったような指示を追加する場合も見られた。しかしながら、対人スキルによる大きな差は見られなかった。他者受容スキルと強い正の相関が見られた理由として、他者受容スキルの高い人の方が ChatGPT の生成した解答を受け入れやすい傾向にある可能性が考えられるが、今回の結果においては確かでないため、さらなる調査が必要である。一方、知識タスクは、すべてのメインスキルにおいて、対人スキルの低い人の方が大きく向上している。図 5 より、対話の特徴としてはじめに設問をそのまま投げかけたのち「他には？」といった追加の出力を促すことや、「重複のないように」といった指示を追加する傾向にあった。関係調整スキルなどの対人スキルが低い人の方が対話のステップ数が大きいことから、対人スキルの低い人の方が遠慮なく指示を出せたため、より多くの解答を得られたのではないかと考えられる。問題解決タスクは、表現スキルなどの対人スキルの低い人の方がより大きく向上する傾向にあるが、対人スキルの高い人でも他のタスクと比較して大きく向上している。対話の特徴とし

表 4 タスクごとの対 AI スキルの結果。値はすべて小数第 3 位で四捨五入している。“w AI” は ChatGPT を利用した場合，“w/o AI” は ChatGPT を利用せず独力で解答した場合，“対 AI スキル” は “w AI” と “w/o AI” の差を示す。H4 および H6 の実験協力者は計算タスクにおいて ChatGPT を利用しなかったため，“w AI” および “対 AI スキル” の値は空欄となっている。分析対象である対 AI スキルを太字で示す。

実験協力者	知識タスク		言語タスク		計算タスク		問題解決タスク	
	対 AI スキル	(w AI - w/o AI)	対 AI スキル	(w AI - w/o AI)	対 AI スキル	(w AI - w/o AI)	対 AI スキル	(w AI - w/o AI)
H1	0.15	(0.15 - 0.00)	0.32	(0.77 - 0.44)	0.75	(1.00 - 0.25)	0.19	(0.25 - 0.06)
H2	0.51	(0.68 - 0.18)	0.49	(0.54 - 0.06)	-0.75	(0.25 - 1.00)	0.68	(0.78 - 0.10)
H3	0.08	(0.15 - 0.07)	0.43	(0.88 - 0.44)	0.83	(1.00 - 0.17)	0.40	(0.55 - 0.15)
H4	0.87	(0.97 - 0.10)	0.10	(0.63 - 0.53)	-	(- - 0.50)	0.41	(0.71 - 0.30)
H5	0.74	(0.91 - 0.16)	0.04	(0.75 - 0.71)	-0.50	(0.50 - 1.00)	0.31	(0.50 - 0.19)
H6	0.71	(0.94 - 0.23)	-0.18	(0.40 - 0.58)	-	(- - 1.00)	0.59	(0.79 - 0.20)
H7	0.85	(0.91 - 0.06)	-0.03	(0.50 - 0.53)	0.00	(0.00 - 0.00)	1.00	(1.00 - 0.00)
H8	0.62	(0.62 - 0.00)	0.13	(0.44 - 0.32)	-0.50	(0.00 - 0.50)	0.69	(0.79 - 0.10)
H9	0.51	(0.58 - 0.07)	0.09	(0.60 - 0.51)	0.00	(0.00 - 0.00)	0.30	(0.65 - 0.35)
H10	0.02	(0.11 - 0.10)	0.23	(0.73 - 0.50)	0.50	(0.50 - 0.00)	0.39	(0.50 - 0.11)
H11	-0.01	(0.06 - 0.07)	0.05	(0.45 - 0.40)	-0.75	(0.00 - 0.75)	0.55	(0.80 - 0.25)
H12	0.01	(0.05 - 0.03)	0.10	(0.37 - 0.27)	0.00	(0.00 - 0.00)	0.24	(0.24 - 0.00)
平均	0.42	(0.51 - 0.09)	0.15	(0.59 - 0.44)	-0.04	(0.33 - 0.43)	0.48	(0.63 - 0.15)

表 5 ENDCOREs の各メインスキルと成果の向上度合いの相関。*は $p < 0.05$ ，**は $p < 0.01$ となった数値を表す。

タスクカテゴリ	ENDCOREs におけるメインスキル					
	自己統制スキル	表現スキル	読解スキル	自己主張スキル	他者受容スキル	関係調整スキル
知識タスク	-0.44	0.07	-0.52	-0.06	-0.56	-0.50
言語タスク	0.61*	-0.17	0.54	-0.25	0.77**	0.63*
計算タスク	0.03	-0.05	0.08	0.01	0.44	0.19
問題解決タスク	0.09	-0.16	-0.11	-0.32	-0.10	-0.02

て、はじめに設問をそのまま投げかけたのち、必要に応じて指示を追加する場面が見られたが、対人スキルによる差はあまり見られなかった。計算タスクにおいては対人スキルの高い人の方が向上する場合もあるが、より大きく低下する場合もある。このことから、対人スキルに関係なく ChatGPT の利用が適していないと考えられる。

解答時間については、全タスクにおいて、対人スキルの低い人の方が ChatGPT の利用によって短縮する傾向がみられた。対人スキルと各タスクの平均解答時間の関係は付録にて示す。

全体として、ChatGPT を利用した場合の方が成果は向上したが、計算タスクにおいては独力で解答した場合の方が評価が高くなる場合があった。計算タスクにおいて、自身の方が ChatGPT よりも優れていると認識しているにも関わらず、ChatGPT の誤った解答を信用する場面があった。例えば、「5m の木を 1m ずつに切り分けたい。1 回切るのに 5 分かかり、1 回切るごとに 1 分休憩すると、何分で切れますか。」という設問の正しい解は「23 分」であるが、ChatGPT の出力した「26 分」という解を解答している場面があった。また、問題解決課題においては、表 2 の「10 個のミカンをも 3 人に平等に分配する方法をできるだけ多く挙げてください。」という設問に対し、ChatGPT が「まず、各人に 2 個ずつ配り、残り 2 個を 3 人で交互に 1 個ずつ取り合う。」という解答を生成した例が見られ

た。この場合、合計 8 個のミカンにのみ言及されており、残り 2 個に関しては考慮されていない。明確な正答もないため誤りであるという判定も困難であるが、評価の際に注意が必要である。

5 おわりに

本研究では、LLM を利用する人の対人スキルに着目し、人と ChatGPT の協働によるパフォーマンスについて調査を行った。具体的には、ChatGPT を利用して行うタスクとして 4 つのカテゴリのタスクを設計し、実験を実施した。その結果、次の事実が明らかになった。

- 対人スキルの低い人は、知識タスクにおいて ChatGPT の恩恵を受けやすい可能性がある。
- 自己統制スキル、他者受容スキル、関係調整スキルの高い人は言語タスクにおいて ChatGPT の恩恵を受けやすい傾向にある。
- 対人スキルの高さに関わらず、計算タスクに ChatGPT を用いることは有効でない。
- 対人スキルの高さに関わらず、問題解決タスクに ChatGPT を用いることは有効である。

以上のように、一部のタスクと対人スキルには関係があり、人と協働する LLM の開発を行う際は人の能力を考慮した評価

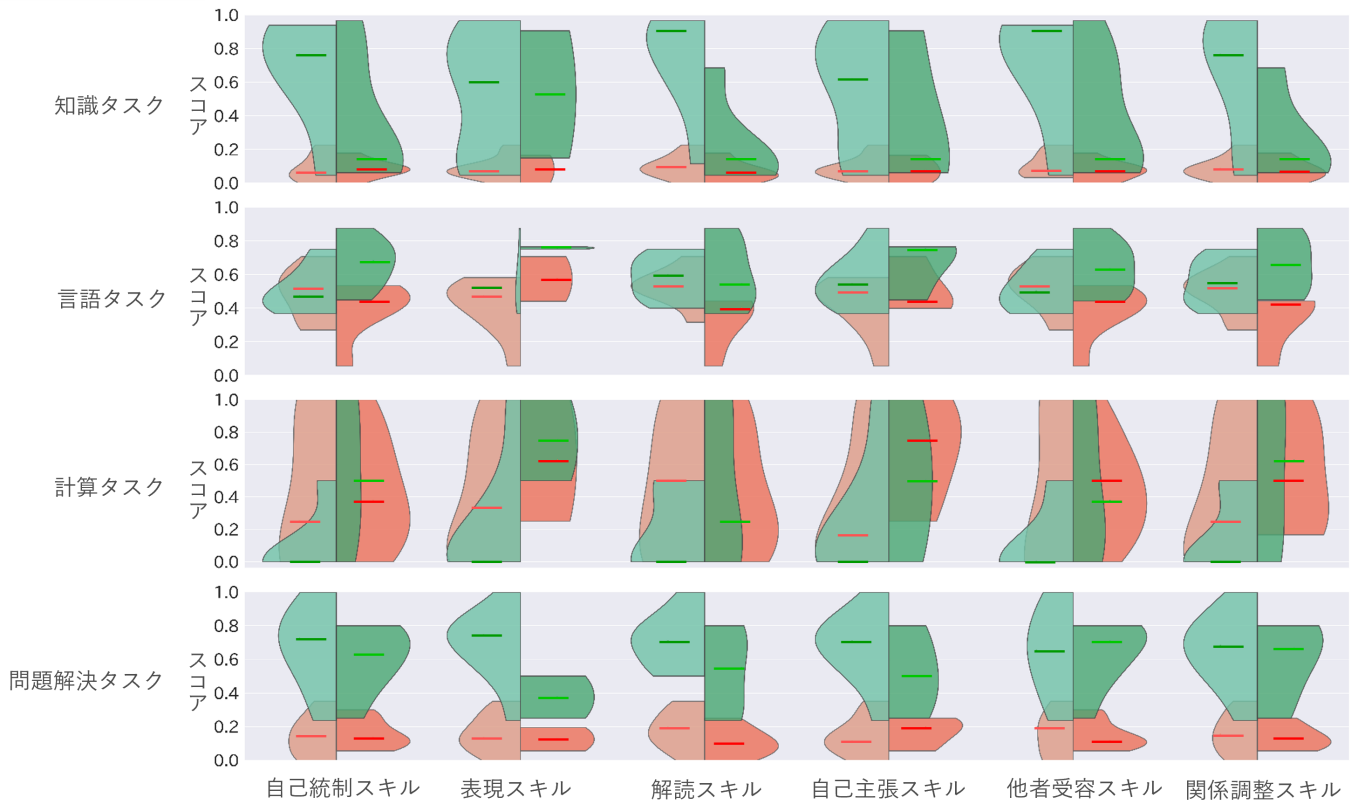


図4 対人スキルとタスクの成果の関係. 横軸は ENDCOREs のメインスキル, 縦軸は各タスクのスコアを示す. バイオリンプロット左側に対人スキルの低い場合, 右側に対人スキルの高い場合を示す. それぞれ赤色のグラフは ChatGPT を利用せず独力で取り組んだ場合, 緑色のグラフは ChatGPT を利用して取り組んだ場合を示す. 図中の線は中央値を示す.

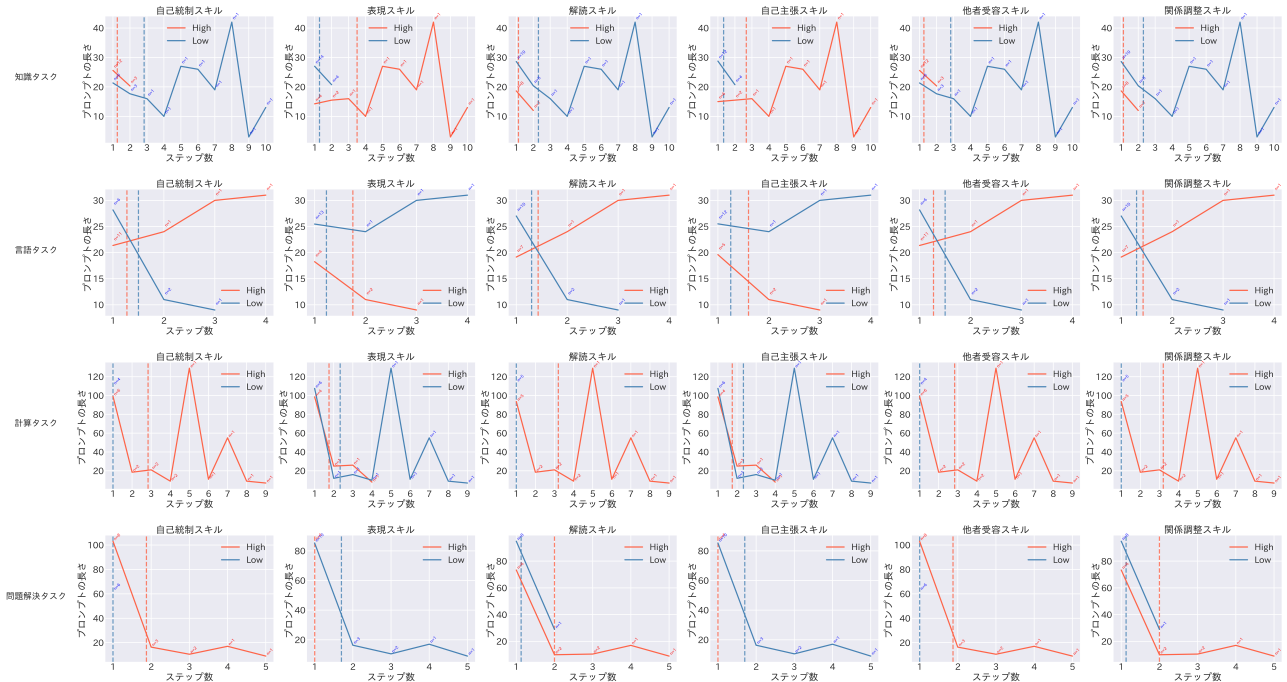


図5 ChatGPT を利用した際の対話のステップ数と各ステップにおけるプロンプトの平均の長さ. 各列に ENDCOREs のメインスキル毎, 各行にタスクカテゴリ毎の結果を示す. 破線は平均ステップ数を示す. 各点における数字は該当するデータ数を示す.

が有用である可能性が示された。しかしながら、タスクに対する評価は十分でなく、タスクの設計や評価方法にさらなる検討が必要である。例えば、問題解決タスクにおいては明確な答えが存在しないため、すべての解答を認めているが、中には類似した解答が複数存在する場合や、問題の解決に繋がるか不明な解答も存在している。そのため、単に解答数のみで評価するのではなく、解答の質を考慮する必要がある。また、本研究ではコミュニケーションスキルを対人スキルとして定義したが、設計したタスクでは ChatGPT とのコミュニケーションが見られない場合があった。そのため、ディベートや意思決定といった、よりコミュニケーションの必要性が高いと考えられるタスクの設計が必要である。さらに、実験データも小規模であるため、追加で実験を行う必要がある。

今後の展望として、クラウドソーシングを用いて大規模な実験を実施する。これにより、偏差値等を用いて6つの対人スキル、4つのタスクにおける対AIスキルをそれぞれ1つの対人スキル、対AIスキルとして総合して評価する。加えて、アンケート結果を考慮した分析を行う。

謝 辞

本研究は、LINE ヤフー株式会社共同研究費の支援を受けたものである。

文 献

- [1] Sébastien Bubeck, et al. Sparks of artificial general intelligence: Early experiments with GPT-4. *arXiv preprint arXiv:2303.12712*, 2023.
- [2] John M Digman. Personality structure: Emergence of the five-factor model. *Annual review of psychology*, Vol. 41, No. 3, pp. 417–440, 1990.
- [3] Isabel Briggs Myers. *The Myers-Briggs Type Indicator: Manual*. Consulting Psychologists Press, 1962.

- [4] Isabel Briggs Myers and Mary H. McCauley. *Manual: A guide to the development and use of the Myers-Briggs Type Indicator*. Consulting Psychologists Press, 1985.
- [5] Guangyuan Jiang, Manjie Xu, Song-Chun Zhu, Wenjuan Han, Chi Zhang, and Yixin Zhu. MPI: Evaluating and inducing personality in pre-trained language models. *arXiv preprint arXiv:2206.07550*, 2022.
- [6] Haocong Rao, Cyril Leung, and Chunyan Miao. Can Chat-GPT Assess Human Personalities? A General Evaluation Framework. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pp. 1184–1194, 2023.
- [7] Santi Caballé and Jordi Conesa. Conversational Agents in Support for Collaborative Learning in MOOCs: An Analytical Review. In *In Proceedings of the 10th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2018)*, pp. 384–394. Springer International Publishing, 2019.
- [8] Zijian Ding, Alison Smith-Renner, Wenjuan Zhang, Joel Tetreault, and Alejandro Jaimes. Harnessing the power of LLMs: Evaluating human-AI text co-creation through the lens of news headline generation. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pp. 3321–3339, 2023.
- [9] Zoie Zhao, et al. More human than human: LLM-generated narratives outperform human-LLM interleaved narratives. In *Proceedings of the 15th Conference on Creativity and Cognition (C&C '23)*, pp. 368–370. Association for Computing Machinery, 2023.
- [10] Yuan Li, Yixuan Zhang, and Lichao Sun. MetaAgents: Simulating Interactions of Human Behaviors for LLM-based Task-oriented Coordination via Collaborative Generative Agents. *arXiv preprint arXiv:2310.06500*, 2023.
- [11] 藤本学, 大坊郁夫. コミュニケーション・スキルに関する諸因子の階層構造への統合の試み. *パーソナリティ研究*, Vol. 15, No. 3, pp. 347–361, 2007.
- [12] 木村尚義. *ずるい考え方 ゼロから始めるラテラルシンキング入門*. あさ出版, 2011.

付 録

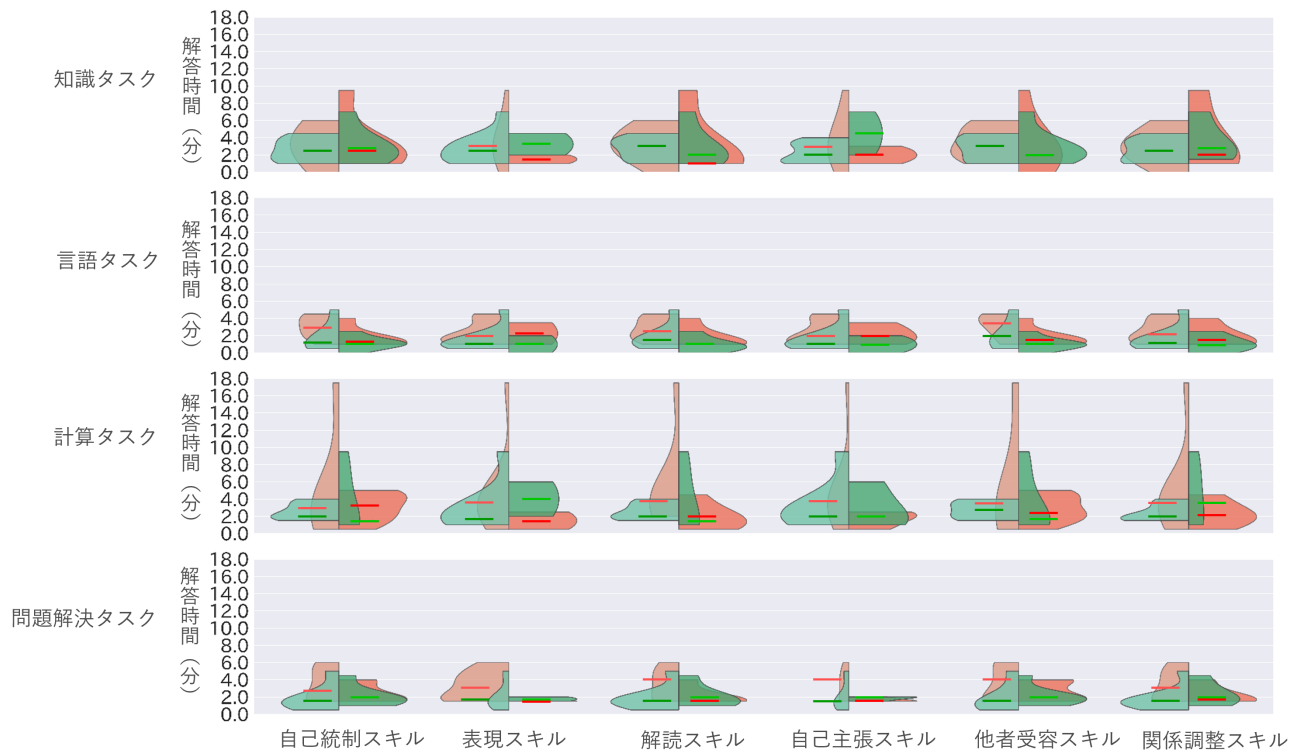


図 6 対人スキルとタスクの平均解答時間 (分) の関係。横軸は ENDCOREs のメインスキルを示し、縦軸は各タスクのスコアを示す。バイオリンプロット左側に対人スキルの低い場合、右側に対人スキルの高い場合を示す。それぞれ赤色のグラフは ChatGPT を利用せず独力で取り組んだ場合、緑色のグラフは ChatGPT を利用して取り組んだ場合を示す。図中の線は中央値を示す。