

LLM を対象としたプロンプトのテキスト分析

和田 縁¹ 木村 泰知¹ 谷辺 哲史² 中川 裕志³ 安野 智子⁴ 寺田 麻佑⁵
¹小樽商科大学 ²早稲田大学 ³理化学研究所革新知能統合研究センター
⁴中央大学 ⁵一橋大学
kimura@res.otaru-uc.ac.jp

概要

本研究の目的は、ユーザが大規模言語モデルに入力するプロンプトをユーザが実際に入力しているテキストとして捉え、その特徴と生成結果に対する信頼のあり方を明らかにすることである。ChatGPTのリンク共有機能を用いて実際の対話ログを収集し、プロンプトの文字数、入力回数、記述パターンといった形式的特徴を分析するとともに、生成結果に対するユーザの評価および信頼度を調査した。分析の結果、プロンプトの量的特徴にかかわらず高い満足度と信頼が示される一方で、課題内容によって評価傾向に差が生じることが明らかとなった。

1 はじめに

近年、大規模言語モデル (Large Language Models: LLMs) の発展に伴い、ChatGPTをはじめとする生成 AI の活用が急速に拡大している [1, 2]。キーワード検索中心の利用から自然言語による対話へと移行するなかで、ユーザが入力する文章 (プロンプト) は、生成される回答の品質を左右する重要な要素として位置づけられている [3]。ユーザは対話的なプロンプトのやり取りを通じて、要約文の生成、英文校正、提案作成などを行い、得られた出力をさまざまな場面で利用している [4]。このような状況では、自由記述式アンケートやレポート等の提出文が LLM により生成・編集された可能性を含み、提出文が必ずしもユーザ自身の直接的な記述のみから構成されるとは限らない。したがって、ユーザの実質的な記述行為を捉えるには、最終生成文だけでなく、生成に至るまでのプロンプトと出力の相互作用を分析対象として位置づける必要がある。

そこで、本研究では、ユーザが LLM に入力するプロンプトに焦点を当て、「ChatGPT のリンク共有」機能を活用して、プロンプトおよび出力結果を分析する。具体的には、プロンプトの特徴 (文字数、入

力回数、記述方法など) を定量的に整理する。さらに、アンケート調査によって ChatGPT の出力に対するユーザの信頼を測定し、プロンプトの特徴と信頼との関連を分析する。

2 関連研究

2.1 自由記述のアンケート調査

教育現場においても、LLM による文章生成の活用とその影響について問題提起がなされている。佐藤 (2023) は、LLM が受講生のレポートを高品質かつ即時に生成できる一方で、その生成物が人間による執筆と区別困難である点を指摘している。このような状況は、レポート提出を評価手段として用いることのリスクや意義の再検討を、教育機関および教員に求めるものであると論じている [5]。

2.2 プロンプトエンジニアリング

プロンプトエンジニアリングの技能と、LLM から得られる回答の品質との関係についても検討が進められている。鎌田 (2025) は、LLM による自動採点結果に基づいて分析を行ったところ、レポート生成プロンプトに具体例の提示や明確な目標設定を含めた場合、生成物の内容は明瞭かつ一貫性を持ち、自動採点において高評価を得やすいことが実験データから確認された。一方で、指示が曖昧、あるいはテーマが過度に広範な場合には、内容が散漫となり、評価が低下する傾向が認められた。今後は、人手採点結果との相関分析を行うとともに、評価基準の細分化や多様なテーマ・文体に対応可能なプロンプト設計の最適化が求められる、と述べた [6]。

3 プロンプト分析の方法

本研究では、文章生成 AI (ChatGPT) の活用実態および利用過程に関する特徴を明らかにすることを目的として、Google フォームによるアンケート調査

- 課題① 「新型コロナウイルスの流行は、ワクチンを販売したい製薬会社の利権が絡んでいる」という主張について真偽を調べてください。
 課題② 「大学の英語の授業において、学生がより主体的に授業に取り組めるようなアイデア」を挙げてください。
 課題③ 「情報収集時におけるファクトチェックをどの程度行うべきか」というテーマのレポートを1,000字で書いてください。

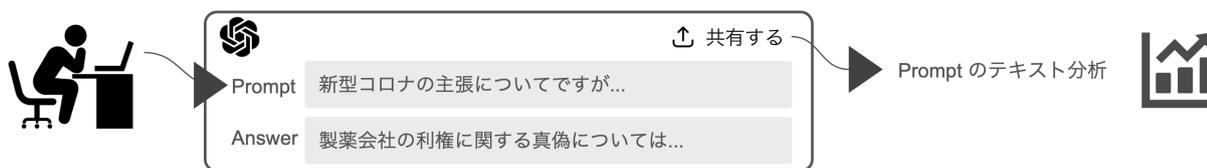


図1 調査概要：3つの課題について ChatGPT で生成した回答の共有リンクを提出させ、テキスト分析を行う。

を実施した。本調査では、図1に示すように、3つの課題に対して ChatGPT を用いて取り組む設問を設け「共有リンク」を提出してもらうことで、回答者の実際の利用行動とプロンプト生成過程を取得することとした。アンケート調査の対象者は、大学のゼミナールまたは研究室に所属する学生である。アンケートの回答所要時間は約30分である。調査期間は2025年11月1日から同年11月30日までの1か月間とした。下記に3つの課題を示す。

課題①

「新型コロナウイルスの流行は、ワクチンを販売したい製薬会社の利権が絡んでいる」という主張について、正しいのか、間違っているのか、ChatGPT を用いて真偽を調べてください。

課題②

「大学の英語の授業において、学生がより主体的に授業に取り組めるようなアイデア」を、あなたが納得するアイデアが出てくるまで ChatGPT に挙げさせてください。

課題③

ChatGPT を用いて、「情報収集時におけるファクトチェックをどの程度行うべきか」というテーマのレポートを1,000字で書いてください。

これらの3つの課題に対して、ChatGPT の共有リンクを提出してもらうとともに、その後、得られた回答に対する評価を求めた。評価項目は、ChatGPT の回答に対する主観的評価を測定するため、次の2つの観点から評価してもらうこととした。①「課題に対して適切な答えを得られた」という課題適合性の観点と、②「ChatGPT を使用することで、自分1人で考えるよりも良い答えを得られた」という付加価値の観点である。各項目は、回答者の主観的評

価を詳細に把握するため、「そう思わない」「あまりそう思わない」「どちらともいえない」「ややそう思う」「そう思う」の5段階で評価を求めた。また、本調査では、上記の ChatGPT との対話が陰謀論的信念に及ぼす影響を検証するための質問も加えた。具体的には、課題①の「新型コロナウイルス」に関連する内容であり、ChatGPT を用いた課題遂行の前後において同一の質問項目を提示し、回答者の認知的態度の変化を測定した。この反復測定法により、文章生成 AI との相互作用が個人の信念体系に与える効果を定量的に評価することとした。評価項目は、新型コロナウイルスに関する陰謀論的信念の程度を測定するため、3つの具体的な陰謀論的主張を用いて構成した。具体的には、①「新型コロナウイルスの流行には、ワクチンを販売したい製薬会社の利権が絡んでいる」、②「新型コロナウイルスの流行や、ワクチンの副作用による被害が生じたことの背後には、何者かの陰謀があった」、③「新型コロナウイルスに関するマスコミの報道は、ウイルスへの恐怖をおおる内容に偏っていた」の3項目である。これらの項目に対しては、「きっと正しい」「おそらく正しい」「どちらともいえない」「おそらく正しくない」「きっと正しくない」の5段階で測定した。

4 アンケート調査の結果

本アンケートの回答数は78件であり、そのうち有効回答は68件、無効回答は10件であった。無効回答には、設問1~3のいずれかの回答リンクが欠落しているもの、プロンプトが共有されていないもの、または ChatGPT が「思考中」の状態に停止し、調査者が会話内容を確認できないものが含まれる。調査対象は、一橋大学、早稲田大学、小樽商科大学、福岡大学、中央大学、宇都宮共和大学、ならびに国立情報学研究所 (NII) に所属する学生である。

5 プロンプト分析の考察

5.1 LLM への信頼度について

LLM の出力結果をどの程度信頼しているのかについて分析した。図 4 に示すとおり、課題①では「お題に対して適切な答えを得られた」と回答した参加者が 65 名 (95.6%)、「ChatGPT を使用することで、自分一人で考えるよりも良い答えを得られた」と回答した参加者が 63 名 (92.6%) であり、課題①から課題③の中で最も高い満足度であった。課題②では、「お題に対して適切な答えを得られた」および「ChatGPT を使用することで、自分一人で考えるよりも良い答えを得られた」のいずれについても 61 名 (89.7%) が肯定的に評価した。課題③では、「お題に対して適切な答えを得られた」が 53 名 (77.9%)、「ChatGPT を使用することで、自分一人で考えるよりも良い答えを得られた」が 56 名 (82.4%) となり、課題①および課題②と比べて満足度はやや低下したものの、全体としては高い水準を維持していた。

陰謀論的信念への影響

LLM との対話が陰謀論的信念に影響を及ぼすのかについて分析した結果を示す。新型コロナウイルスの流行やワクチンをめぐる複数の設問において、**LLM 利用後には陰謀論を信じる参加者が減少する傾向が確認された。**

例えば、図 3 「新型コロナウイルスの流行について、ワクチンの副作用や流行そのものについて、何者かの陰謀によって世の中が悪い方向に動かされていたと思いますか」という質問では、LLM 利用前後で、陰謀論を信じる側の参加者数は 15 名 (22.0%) から 6 名 (8.8%) に減少した。また、「新型コロナウイルスの流行には、ワクチンを販売したい製薬会社の利権が絡んでいる」という質問では、陰謀論を信じる側の参加者数は 4 名 (5.8%) から 3 名 (4.4%) へ減少した。

一方で、「きっと正しい」と陰謀論を信じる人が増えた例も確認された。この理由としては、【アンケートの対象者である大学生らは、オールドメディアと呼ばれるマスコミの報道に触れておらず、マスコミの悪い面だけを切り取った SNS の投稿などに触れ、信頼度が下がっているため】だと考えられる。

表 1 プロンプト回数と満足度の相関係数

| 課題 | 適切な答え | | 一人で考えるよりも良い | |
|----|--------|------|-------------|------|
| | 相関係数 | p 値 | 相関係数 | p 値 |
| ① | 0.034 | 0.78 | 0.057 | 0.64 |
| ② | -0.183 | 0.13 | 0.030 | 0.81 |
| ③ | -0.113 | 0.36 | -0.182 | 0.14 |

* $p < 0.05$ (両側検定)

表 2 AI に関する知識と満足度の相関係数

| 課題 | 適切な答え | | 一人で考えるよりも良い | |
|----|--------|------|-------------|------|
| | 相関係数 | p 値 | 相関係数 | p 値 |
| ① | 0.237* | 0.05 | 0.076 | 0.54 |
| ② | 0.045 | 0.71 | 0.055 | 0.65 |
| ③ | 0.013 | 0.91 | -0.209 | 0.09 |

* $p < 0.05$ (両側検定)

5.2 LLM への満足度について

まず、**プロンプト回数が多いほど満足度が高まるかを検証した。**表 1 のプロンプト回数と満足度の相関係数からわかるように、「適切な答えが得られた」という評価項目では、課題②が 0.045、課題③が 0.013 であり、いずれも相関は認められなかった。また、「一人で考えるよりも良い答えが得られた」という評価項目においても、課題①は 0.076、課題②は 0.055 と、明確な相関は見られなかった。一方で、「適切な答えが得られた」に関する課題①では 0.237 と弱い正の相関が認められたのに対し、「一人で考えるよりも良い答えが得られた」に関する課題③では -0.209 と弱い負の相関が確認された。これらの結果から、課題①から課題③のいずれにおいても、**プロンプト回数と満足度との相関係数は小さく、明確な関連は認められず、プロンプト回数に関係なく満足度が高い傾向があることが確認された。**図 4 に課題①から 3 へ対する満足度が高い傾向がある結果を示す。

次に、**AI に関する知識の程度が ChatGPT の回答に対する満足度に影響を与えるかを検討した。**表 2 に示すとおり、「適切な答えが得られた」および「一人で考えるよりも良い答えが得られた」という評価項目における満足度との相関係数はいずれも小さく、有意な相関は確認されなかった。一方、AI 関連用語の理解状況には差が見られた。図 2 が示すように、「ディープフェイク」については内容を理解している参加者が 77.9% と多数を占めたのに対し、「エコーチェンバー」「フィルターバブル」「ハルシ

ネーション」については、いずれも内容を理解している参加者は半数未満にとどまっていた。これらの結果から、課題①から課題③のいずれにおいても、AIに関する知識量と満足度との相関係数は小さく、明確な関連は認められなかった。

5.3 LLM の操作方法について

本研究では、AIに関する知識の程度が、プロンプトの入力回数、文字数、ならびに記述方法に差異をもたらすかを検討した。表3および表4に示すとおり、AIに関する知識とプロンプト入力回数との相関係数は、課題①で0.126、課題②で0.252、課題③で0.223であり、課題②および課題③において弱い正の相関が認められた。また、AIに関する知識とプロンプト文字数との相関係数は、課題①で0.231、課題②で0.325、課題③で0.214となり、すべての課題において弱い相関が確認された。一方で、プロンプト入力回数に関する課題①では相関が認められなかった。これは、当該課題が比較的定型的であり、少ない試行で十分な回答が得られる内容であったため、参加者全体のプロンプト入力回数が低水準に収束したことによるものと考えられる。以上の結果から、AIに関する知識が高いほど、より多く、かつ詳細なプロンプトを入力する傾向があることが示唆される。

表3 AIに関する知識とプロンプト回数の相関

| 課題 | 相関係数 | p 値 |
|----|--------|------|
| ① | 0.126 | 0.30 |
| ② | 0.252* | 0.04 |
| ③ | 0.223 | 0.07 |

* $p < 0.05$ (両側検定)

表4 AIに関する知識とプロンプト文字数の相関

| 課題 | 相関係数 | p 値 |
|----|--------|------|
| ① | 0.231 | 0.06 |
| ② | 0.325* | 0.01 |
| ③ | 0.214 | 0.08 |

* $p < 0.05$ (両側検定)

5.4 アンケート調査の課題

本調査では、参加者全員に一律の報酬を付与した結果、提示した設問をそのまま LLM に入力する事例が多く見られ、十分な質を有するサンプルが限定的であった。今後の課題として、回答の質や取り組み姿勢に応じた報酬設計を導入する必要性が示唆さ

れる。

6 おわりに

本研究では、LLM に入力されるプロンプトに着目し、「ChatGPT のリンク共有」機能を用いて、ユーザのプロンプトおよび生成結果を分析した。その結果、本調査においては、AIに関する知識とプロンプト数、ならびに LLM の回答に対する満足度との間に有意な相関は確認されなかった。一方、AIに関する知識とプロンプト数およびプロンプト文字数との間には、弱い相関が認められた。今後は、調査方法を改善し、より精度の高いデータ収集を行う必要がある。

謝辞

本研究は、JST、RISTEX、JPMJRS24L3 および JPMJRS25L2 の支援を受けたものです。

参考文献

- [1] OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. GPT-4 Technical Report, 2023.
- [2] Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, Shyamal Buch, Dallas Card, Rodrigo Castellon, Niladri Chatterji, Annie Chen, Kathleen Creel, Jared Quincy Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren Gillespie, Noah Goodman, Tatsunori Hashimoto, Peter Henderson, Daniel E. Ho, Dan Jurafsky, Pratyusha Kalluri, Omar Khattab, Pang Wei Koh, Ranjay Krishna, Percy Liang, and many others. On the opportunities and risks of foundation models, 2021.
- [3] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, Vol. 55, No. 9, pp. 195:1–195:35, 2023.
- [4] Aaron Chatterji, Tom Cunningham, David Deming, Zoë Hitzig, Christopher Ong, Carl Shan, and Kevin Wadman. How people use chatgpt, 2025.
- [5] 佐藤修. 高等教育での大規模言語モデル利用についての一考察. *情報経営*, Vol. 85, pp. 37–40, 2023.
- [6] 鎌田光宣. 生成 ai におけるプロンプトエンジニアリングの効果—レポート生成と自動採点を用いた分析—. *千葉商大紀要*, Vol. 40, p. 38, 2025.

A 付録

あなたは、以下にお示しする言葉の意味を知っていますか。
[エコーチェンバー]

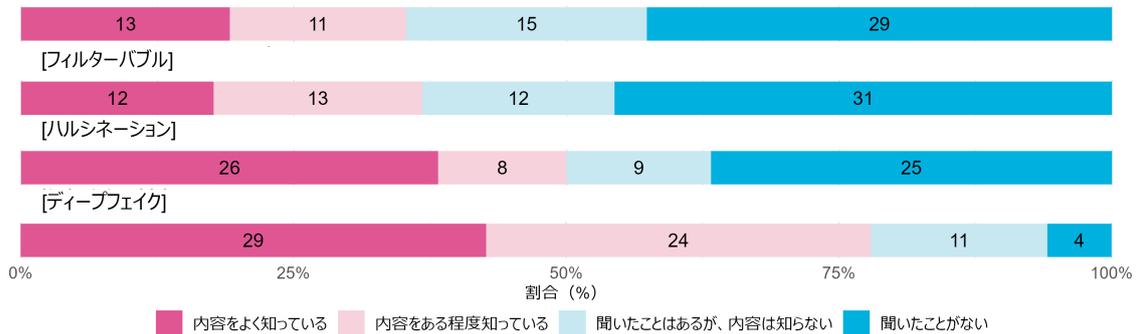


図2 AIに関する知識を問う質問への回答

新型コロナウイルスの流行について、ワクチンの副作用や流行そのものについて、何者かの陰謀によって世の中が悪い方向に動かされていたと思いますか。



あなたは、以下にお示しする文章の内容についてどう考えますか。
[新型コロナウイルスの流行には、ワクチンを販売したい製薬会社の利権が絡んでいる]

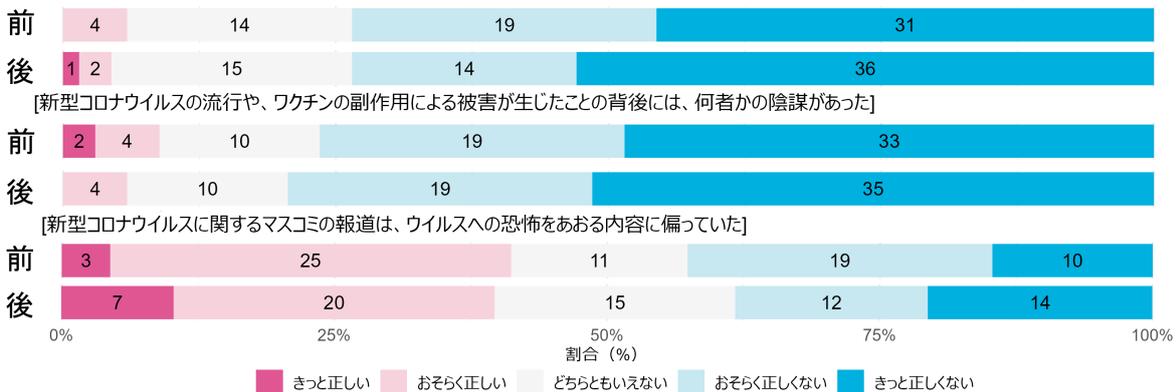


図3 陰謀論に関する質問への回答

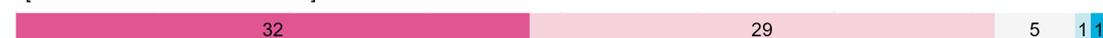
お題①へのChatGPTの回答について、あなたはどのように思いますか。
[お題に対して適切な答えを得られた]



[ChatGPTを使用することで、自分1人で考えるよりも良い答えを得られた]



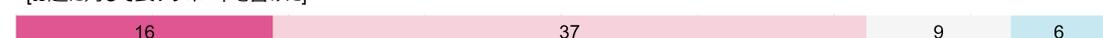
お題②へのChatGPTの回答について、あなたはどのように思いますか。
[お題に対して適切な答えを得られた]



[ChatGPTを使用することで、自分1人で考えるよりも良い答えを得られた]



お題③へのChatGPTの回答について、あなたはどのように思いますか。
[お題に対して良いレポートを書けた]



[ChatGPTを使用することで、自分1人で考えるよりも良いレポートを書けた]



図4 課題①～③への満足度