LLM の日本語指示追従性向上のための人工データセットの構築

守山 慧 [1,2]　児玉 貴志 [2]　中山 功太 [2]
[1] 東京大学　[2]NII LLMC
kei-moriyama@g.ecc.u-tokyo.ac.jp

概要

大規模言語モデル (LLM) が、人間の指示に沿った応答を生成する能力を評価する指標の一つに、指示追従性がある。本研究では、LLM の指示追従性の向上を目的とした人工データ生成手法を提案する。生成対象のデータは、指示データと選好データである。指示データについては、人手で作成された小規模なデータを LLM を用いて拡張する。選好データについては、正例として指示データの応答を用い、負例を LLM で生成する。生成されたデータの質を確保するために、ROUGE-L と LLM-as-a-Judge によるフィルタリングを適用した。実験により、提案手法により生成されたデータを用いて LLM を学習することで、指示追従性の向上が確認できた。

1　はじめに

大規模言語モデル (LLM) が人間の指示に従った応答を生成する能力を評価する指標の 1 つに、指示追従性がある。指示追従性の向上のために、教師あり学習や Direct Preference Optimization(DPO) [1] などの学習方法が用いられる。これらの学習方法には、大規模かつ高品質なデータが必要である。特に、指示追従性においては、指示にさまざまな制約が含まれるため、応答のアノテーションに時間とコストが必要である。

この課題を解決するために、LLM を用いたデータ拡張が提案されている [2]。人間が作成した小規模なシードデータをもとに、LLM が新たな指示や応答を生成することで、低コストで大規模なデータの構築を可能にした。指示追従性を対象とした LLM によるデータ拡張は、主に英語において検証されている [3]。一方で、日本語では指示追従性以外での有効性は示されている [4, 5] が、指示追従性における効果は十分に検証されていない。

本研究では、日本語の指示追従性能の向上を目的とした、LLM によるデータ拡張の効果を検証する。構築するデータは、指示と応答のペアからなる指示データと、指示に対する正例と負例を含む選好データを対象とする。LLM が生成する指示は、制約を持つ必要がある。そこで、指示が持つ制約を指示制約カテゴリとして分類し、その説明を作成した。LLM は、シードデータの指示を指示制約カテゴリとその説明を用いて、新しい指示を生成する。生成された指示に対して、類似の指示を除外する ROUGE-L [6] と、低品質の指示を除外する LLM-as-a-Judge [7] による 2 つのフィルタリングを適用した。フィルタリング後の指示に対して LLM で応答を生成し、別途 LLM-as-a-Judge により評価することで、指示データを構築する。選好データの拡張では、拡張した指示データの指示と応答をそれぞれ選好データの指示と正例とし、負例のみ新たに LLM で生成する。表 1 に、データセットから除外された例と、除外されなかった例を示す。

構築した人工データセットを用いて LLM を学習し、その指示追従性を評価した。評価には、多言語指示追従ベンチマークである M-IFEval [8] の英語および日本語データセットを用いた。その結果、指示データと選好データを用いて LLM を学習することで、指示追従性の向上が確認できた。

2　関連研究

LLM によるデータ拡張は、アノテーションに必要な費用や労力を減らすために提案された。人手により作成された小規模データセットを LLM に与え、書き換えなどの操作をすることで、新しいデータを作成する [2, 3]。教師データを拡張する手法の 1 つに、self-instruct [9] がある。Self-instruct は、複数の指示をプロンプトに含めて指示と応答を生成し、ROUGE-L を用いてフィルタリングする手法である。Self-instruct により生成されたデータを用いて学習されたモデルは、zero-shot 設定における SUPERNI [10] ベンチマークにおける性能が向上した。教師データだけではなく、選好データセットにおけるデータ拡

指示制約カテゴリ	シードデータの指示	生成された指示	生成された応答
文字>句読点	ムーミン一家の家族のメンバーを箇条書きで答えて下さい。	ムーミン一家の家族のメンバーを、句読点（カンマ）を使って列挙してください。	ムーミン、 ムーミンママ、 ムーミンパパ、 ス nufunfa
フォーマット>表>csv	火星の周りには衛星はありますか?あれば名前と英語表記を教えて。回答は 400 字以内で。	火星の衛星の情報を CSV 形式で教えてください。衛星の名前と英語表記を含めてください。	"衛星名","英語表記" "フォボス","Phobos" "デイモス","Deimos"

表 1　生成された指示と応答のペアの具体例。赤い文字は指示制約を表し、青い文字は日本語として崩壊している部分を表す。一行目の例は応答に日本語として成立していない文章を含むため除外され、二行目の例は指示と応答が適切であるためデータに含まれた。

張においても有効性が示されている [11, 12]。

3　提案手法

3.1　指示データセットの生成

最初に、指示データを生成し、そのデータをもとに選好データを生成する。提案手法の概要図を図 1 に示す。LLM が生成する指示は、応答が満たすべき制約を含む必要がある。そこで、指示に含まれる制約を分類した指示制約カテゴリを定義した。例えば、「アクションをこなせる俳優を箇条書きで挙げて下さい。回答以外何も書かないでください。」という指示は、「形式>リスト>順序なし>マークダウン」や「頻度」といったカテゴリに分類される。これらの指示制約カテゴリに対して、説明を作成した。作成した指示制約カテゴリの説明の例を、付録 A に示す。指示の生成に使用するプロンプトには、シードデータの指示、指示制約カテゴリ、および指示制約カテゴリの説明を含めた。図 2 に、指示の生成に用いたプロンプトの例を示す。指示の生成は、以下の 2 つの方法のそれぞれに基づいて行う。

1. LLM は、与えられた指示に、与えられた指示制約カテゴリを追加する。
2. LLM は、与えられた指示を、与えられた指示制約カテゴリを持つ指示に書きかえる。

生成された指示の品質を向上させるため、ROUGE-L [6] と LLM-as-a-Judge [7] を用いたフィルタリングを行う。各指標にはあらかじめ閾値を設定し、その値に基づいて生成された指示や応答をデータセットに含めるかどうかを決定する。ROUGE-L は、データセットに多様な指示が含まれるようにすることを目的として用いる。新しく生成された指示について、次の 2 つの参照文のそれぞれとの ROUGE-L スコアを計算する。

- 与えられたシードデータの指示
- 既に生成され、フィルタリング後に残っている指示

いずれかの参照文との ROUGE-L のスコアが閾値よりも大きい場合、生成された指示はデータセットに含まない。

LLM-as-a-Judge は、以下の三種類の指標について、指示を評価する。

関係性　与えられた指示制約カテゴリと生成された指示の指示制約カテゴリの一致度を評価する。
流暢性　生成された指示が、文法的かつ自然に記述されているか評価する。
冗長性　生成された指示の簡潔さを評価する。

評価は五段階のリッカート尺度（0 が最も低く、5 が最も良い）で行い、どれか 1 つの指標でもスコアが閾値を下回る時、その指示を含まない。評価に用いるプロンプトには、評価の具体例を含めた。図 3 に評価用のプロンプトの具体例を示す。

フィルタリングを通過した指示に対して、LLM は応答を生成する。LLM-as-a-Judge を用いて、生成された応答を評価する。評価には以下の四種類の指標を用いる。

追従性　生成された応答が、指示内の制約に従っているか評価する。
流暢性　生成された指示が、文法的かつ自然に記述されているか評価する。
冗長性　生成された応答の簡潔さを評価する。

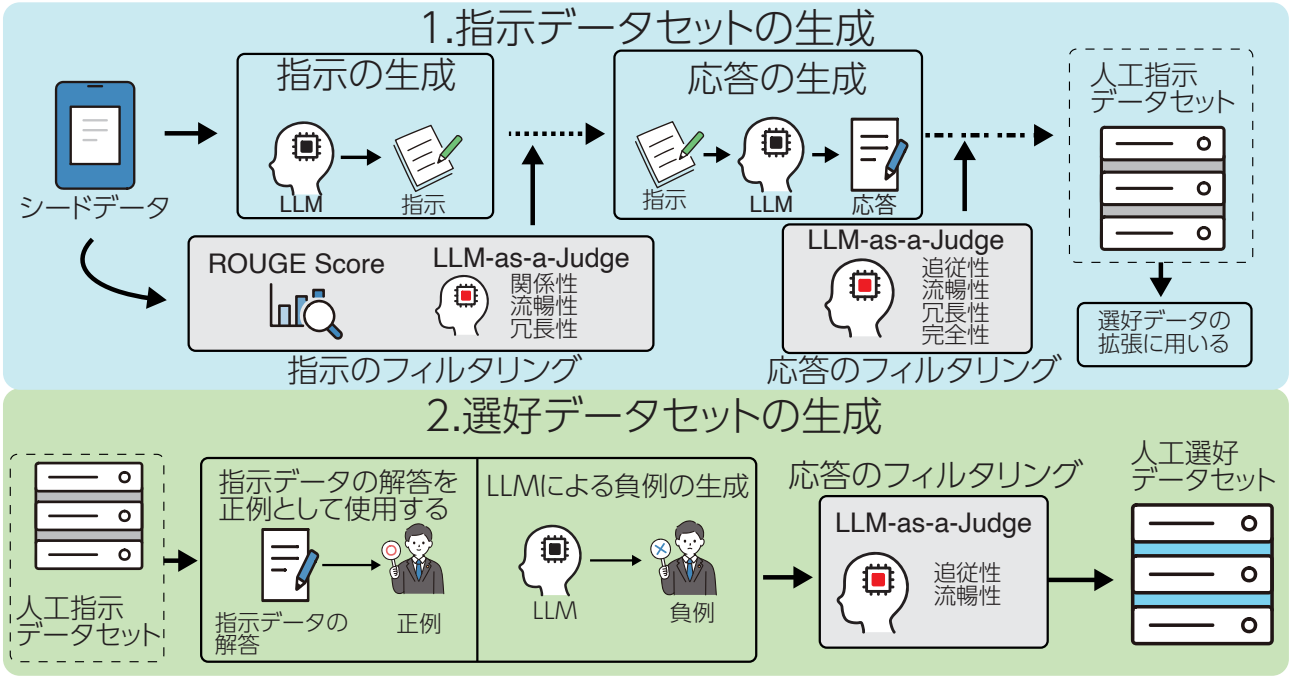

図 1　提案手法の概要図。

完全性　生成された応答が、指示の内容を十分に満たしているか評価する。

指示のフィルタリングと同様の基準にもとづき、応答のフィルタリングを行う。フィルター後の指示と応答のペアを、指示データセットとする。

3.2　選好データの生成

選好データセットの構築には、3.1 節で生成する指示データセットを用いる。指示データセットの指示と応答を、それぞれ選好データの指示と正例として用いる。LLM は負例のみ生成する。

指示追従タスクにおける選好データセットの負例として、以下の 2 つを定義した。

1. 指示制約を従わないが、内容は指示と関係している応答
2. 指示制約に従うが、内容は関係していない応答

各定義ごとに専用のプロンプトを作成し、それぞれの定義に従って LLM は負例を生成する。プロンプトには、具体的な負例の例と、それが負例と判断される理由を含めた。図 4 に負例の生成に用いたプロンプトの例を示す。

生成された負例の質を上げるために、LLM-as-a-Judge を用いて評価する。評価には、以下の二種類の指標を用いる。

追従性　生成された不例が、不例の定義に従うかどうかを評価する。

流暢性　生成された不例が、文法的かつ自然に記述されているか評価する。

指示のフィルタリングと同様の基準に基づいて負例のフィルタリングを行い、選好データを構築した。

4　実験

4.1　実験設定

シードデータには、ichikara-instruction [13] に人手で指示制約を追加した ichikara-instruction-format を用いる。データセットの生成と LLM-as-a-Judge には、Qwen2.5-32B-Instruct[1] を用いる。LLM-as-a-Judge と ROUGE-L によるフィルタリングに用いる閾値は、それぞれ 3 と 0.7 とする。

生成した指示データの効果検証には教師あり学習を、選好データの効果検証には DPO を用いる。学習には、Qwen2.5-1.5B[2]、Llama-3.2-1B[3]、llm-jp-3 の 1.8B[4] と 13B[5] を用いる。教師あり学習は全てのモデルを対象にし、DPO は llm-jp-3 シリーズのモデルにのみ適用する。教師あり学習には、提案手法により生成されたデータに加え、以下のデータセットを

1) https://huggingface.co/Qwen/Qwen2.5-32B-Instruct
2) https://huggingface.co/Qwen/Qwen2.5-1.5B
3) https://huggingface.co/meta-llama/Llama-3.2-1B
4) https://huggingface.co/llm-jp/llm-jp-3-1.8b
5) https://huggingface.co/llm-jp/llm-jp-3-13b

追加で使用する。

- AnswerCarefully [14]
- ichikara-instruction [13]
- ichikara-instruction-format
- ramdom-to-fixed-multiturn-Calm3 [15]
- wizardlm8x22b-logical-math-coding-sft-ja[6)]
- Daring-Anteater [16]
- AutoMultiTurnByCalm3-22B [17]

上記のデータセットと提案手法により構築されたデータを用いて学習されたモデルを `SFT (w/ Synth)` と呼ぶ。また、提案手法により生成されたデータを用いて DPO を適用したモデルを `SFT (w/ Synth)+DPO` と呼ぶ。

ベースラインとして、公開されている指示学習済みの同規模のモデル (`Instruct`) と学習前のモデル (`Base`) を用いる。加えて、提案手法により生成されたデータを除いて、教師あり学習を適用したモデル (`SFT (w/o Synth)`) を用いる。

評価には、多言語対応の指示追従性評価ベンチマークである M-IFEval [8] を用いる。評価対象の言語は、英語と日本語を対象とする。評価指標には、LLM の応答を厳格に評価する Strict 評価を用いる。

4.2 実験結果

表 2 に M-IFEval による評価結果を示す。教師あり学習では、提案手法により生成されたデータを用いることで、Llama3.2 以外のモデルにおいて Base や Instruct モデルよりも良い指示追従性を獲得した。また、`SFT (w/o Synth)` と `SFT (w/ Synth)` を比較すると、提案手法により生成されたデータを用いた方が指示追従性の向上幅が大きい。この傾向は、日本語と英語の両言語において確認された。英語における指示追従性の向上の理由として、指示制約カテゴリに「言語>英語」という英語に関するものがあることが挙げられる。この制約カテゴリは、応答文章に英単語の説明や英語の文章を含めるような制約である。その結果、提案手法により生成された指示や応答に英語が含まれ、英語においても性能が向上していると考えられる。以上より、教師あり学習において、提案手法により生成されたデータは指示追従性の向上において有効であると言える。

DPO においても、提案手法により生成された選好データセットは有効であった。DPO の適用前後を比較すると、1.8B モデルと 13B モデルの両方において、指示追従性の向上が見られた。llm-jp-3-1.8B では、日本語の性能を維持しつつ、英語におけるスコアが向上した。また、llm-jp-3-13B は、評価対象のすべてのモデルの中で最も高い指示追従性を示した。教師あり学習のみの場合、公開されている instruct3 モデルと同等であるが、DPO を適用することで instruct3 モデルよりも良い性能を獲得した。

モデル	モデルの種類	英語	日本語
Llama3.2-1B	Base	18.0	15.9
	Instruct	<u>30.9</u>	<u>21.7</u>
	SFT (w/o Synth)	16.2	15.2
	SFT (w/ Synth)	20.3	15.2
Qwen2.5-1.5B	Base	26.6	14.5
	Instruct	29.6	16.7
	SFT (w/o Synth)	29.3	18.8
	SFT (w/ Synth)	<u>37.6</u>	<u>23.9</u>
llm-jp-3-1.8B	Base	18.0	21.7
	Instruct3	27.0	16.7
	SFT (w/o Synth)	19.9	21.7
	SFT (w/ Synth)	26.6	<u>22.5</u>
	SFT (w/ Synth)+DPO	<u>31.2</u>	<u>22.5</u>
llm-jp-3-13B	Base	37.6	18.1
	Instruct3	36.7	33.3
	SFT (w/o Synth)	34.2	26.1
	SFT (w/ Synth)	38.6	33.3
	SFT (w/ Synth)+DPO	**<u>40.9</u>**	**<u>36.2</u>**

表 2 M-IFEval による評価結果。**太字**は全てのモデルの中での最高値、<u>下線</u>は同一モデル内での最高値を示す。

5 結論

本研究では、指示追従性向上を目的とした人工データセットを構築する手法を提案した。人手で作成されたシードデータをもとに、教師データと選好データを LLM が生成する。高品質なデータセットを構築するために、ROUGE-L と LLM-as-a-Judge を用いて生成された指示と応答をフィルタリングする。提案手法により生成されたデータセットを用いて、LLM を学習、評価した。評価した結果、教師あり学習と DPO の両方において、指示追従性の向上が確認できた。

6) https://huggingface.co/datasets/kanhatakeyama/wizardlm8x22b-logical-math-coding-sft-ja

謝辞

本研究成果の一部は、データ活用社会創成プラットフォーム mdx を利用して得られたものです。

参考文献

[1] Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. In **Thirty-seventh Conference on Neural Information Processing Systems**, 2023.

[2] Renze Lou, Kai Zhang, Jian Xie, Yuxuan Sun, Janice Ahn, Hanzi Xu, Yu Su, and Wenpeng Yin. MUFFIN: Curating multi-faceted instructions for improving instruction following. In **The Twelfth International Conference on Learning Representations**, October 2023.

[3] Can Xu, Qingfeng Sun, Kai Zheng, Xiubo Geng, Pu Zhao, Jiazhan Feng, Chongyang Tao, Qingwei Lin, and Daxin Jiang. Wizardlm: Empowering large pre-trained language models to follow complex instructions, 2025.

[4] Yikun Sun, Zhen Wan, Nobuhiro Ueda, Sakiko Yahata, Fei Cheng, Chenhui Chu, and Sadao Kurohashi. Rapidly developing high-quality instruction data and evaluation benchmark for large language models with minimal human effort: A case study on japanese. In **Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)**, pp. 13537–13547, 2024.

[5] Kazumasa Omura, Fei Cheng, and Sadao Kurohashi. An empirical study of synthetic data generation for implicit discourse relation recognition. In **Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)**, pp. 1073–1085, 2024.

[6] Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In **Text Summarization Branches Out**, pp. 74–81, July 2004.

[7] Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric P Xing, Hao Zhang, Joseph E Gonzalez, and Ion Stoica. Judging LLM-as-a-judge with MT-bench and chatbot arena. **arXiv [cs.CL]**, June 2023.

[8] Antoine Dussolle, Andrea Cardeña Díaz, Shota Sato, and Peter Devine. M-IFEval: Multilingual instruction-following evaluation. **arXiv [cs.CL]**, February 2025.

[9] Yizhong Wang, Yeganeh Kordi, Swaroop Mishra, Alisa Liu, Noah A Smith, Daniel Khashabi, and Hannaneh Hajishirzi. Self-instruct: Aligning language models with self-generated instructions. In **Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, 2023.

[10] Yizhong Wang, Swaroop Mishra, Pegah Alipoormolabashi, Yeganeh Kordi, Amirreza Mirzaei, Atharva Naik, Arjun Ashok, Arut Selvan Dhanasekaran, Anjana Arunkumar, David Stap, Eshaan Pathak, Giannis Karamanolakis, Haizhi Lai, Ishan Purohit, Ishani Mondal, Jacob Anderson, Kirby Kuznia, Krima Doshi, Kuntal Kumar Pal, Maitreya Patel, Mehrad Moradshahi, Mihir Parmar, Mirali Purohit, Neeraj Varshney, Phani Rohitha Kaza, Pulkit Verma, Ravsehaj Singh Puri, Rushang Karia, Savan Doshi, Shailaja Keyur Sampat, Siddhartha Mishra, Sujan Reddy A, Sumanta Patro, Tanay Dixit, and Xudong Shen. Super-NaturalInstructions: Generalization via declarative instructions on 1600+ NLP tasks. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, **Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing**, pp. 5085–5109, December 2022.

[11] Shijia Huang, Jianqiao Zhao, Yanyang Li, and Liwei Wang. Learning preference model for LLMs via automatic preference data generation. In **Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing**, pp. 9187–9199, December 2023.

[12] Suhyun Lee and Changheon Han. Sentimatic: Sentiment-guided automatic generation of preference datasets for customer support dialogue system. In **Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 4: Student Research Workshop)**, pp. 120–128, 2025.

[13] 関根聡, 安藤まや, 後藤美知子, 鈴木久美, 河原大輔, 井之上直也, 乾健太郎. ichikara-isntruction LLM のための日本語イストラクションデータの作成. 言語処理学会第 30 年次大会 (NLP2024), 2024.

[14] Hisami Suzuki, Satoru Katsumata, Takashi Kodama, Tetsuro Takahashi, Kouta Nakayama, and Satoshi Sekine. Answercarefully: A dataset for improving the safety of japanese llm output, 2025.

[15] Kan Hatakeyama. kanhatakeyama/ramdom-to-fixed-multiturn-calm3, 2024.

[16] Zhilin Wang, Yi Dong, Olivier Delalleau, Jiaqi Zeng, Gerald Shen, Daniel Egert, Jimmy J. Zhang, Makesh Narsimhan Sreedhar, and Oleksii Kuchaiev. Helpsteer2: Open-source dataset for training top-performing reward models, 2024.

[17] Kan Hatakeyama. kanhatakeyama/automultiturnbycalm3-22b, 2024.

A 作成した指示制約カテゴリの説明

作成した指示制約カテゴリの具体例を表 3 に示す。

指示制約カテゴリ	作成した指示制約カテゴリの説明
長さ>文	応答に使用する文章の数が指定されている
形式>表>csv	応答文章が、csv 形式の表である制約を含む

表 3 作成した指示制約カテゴリの説明の例。

B 指示の生成などに使用したプロンプト

教師データの生成に使用したプロンプトの例を図 2、LLM-as-a-Judge に使用したプロンプトの例を図 3、選好データの負例の生成に用いたプロンプトの例を図 4 に示す。${category}は指示制約カテゴリ、${category_instruction}は指示制約カテゴリの説明、${seed_instruction}はシードデータの指示に置きかえられる。

あなたは、プロンプトの設計者です。
あなたの目的は、LLM の学習に使用する指示制約データセットのためのプロンプトの作成です。
以下の指示に「${category}」カテゴリに属する指示制約を 20 字程度で追加してください。
「${category}」カテゴリの指示は${category_instruction}である必要があります。
指示は [質問開始] で始まり、[質問終了] で終わるようにしてください。
[質問開始]
${seed_instruction}
[質問終了]

図 2 指示の生成に用いたプロンプトの例。

与えられた AI アシスタントへの指示を次の項目に従い、評価してください。評価指標は次の通りです。
- 関係性 : 与えられた指示が指示カテゴリについてどれくらい適切であるかどうかを評価してください。
- 流暢性 : 指示の日本語表現が自然であるか、文法的な誤りが無いかどうかを評価してください。
- 冗長性 : 指示に無駄な表現が含まれていないか、複数の指示が含まれていないかを評価してください。
評価は短い説明から始め、その後に「評価:[[関係性:1-5、流暢性:1-5、冗長性:1-5]]」という形式で評価を行って下さい。
応答は評価のみを行い、中国語の漢字を含まないように注意してください。
評価の例を以下に示します。
例 1:
(省略)
指示が属するカテゴリは「${category}」で、このカテゴリの内容は「${category_instruction}」です。
AI アシスタントへの指示は次の通りです。
${instruction}
[評価]

図 3 生成された指示を LLM-as-a-Judge により評価するために用いたプロンプトの例。

次の指示に対する応答を生成しください。
ただし、応答は制約に従わないが、内容は関係するようにしてください。
応答は [応答開始] で始まり、[応答終了] で終わるようにしてください。
例 1：
(省略)
以下に、指示を示します。
応答は制約に従わないが、内容は関係する回答を生成してください。
[指示開始]
${instruction}
[指示終了]

図 4 負例を生成するために用いたプロンプトの例。