

多様な言い換え生成と自己学習手法の統合による 大規模言語モデルへの新規知識の追加学習

山本 貴之* 河原 大輔
早稲田大学 理工学術院

{takayukiyamamoto@ruri., dkw@}waseda.jp

概要

大規模言語モデル (LLM) に、新たな知識を追加学習によって取り込むことは、いまだ困難なタスクとされている。近年は RAG のように外部情報を参照する手法が注目されているが、この手法においては、LLM 内部に蓄積された膨大な知識と新たな情報を統合し、高度な推論を行うことは依然として難しい。本研究では、LLM が自己学習を行う手法や、DPO (Direct Preference Optimization) による選好最適化手法から着想を得て、LLM に新規知識を効果的に埋め込む枠組みを提案し、その有効性を検証する。

1 はじめに

新規知識を LLM の追加学習によって取り込み、内部知識として定着させることは、いまだ困難なタスクとされている [1]。新規知識を外部情報として参照する手法である RAG [2] が近年注目されているが、この手法においては、LLM 内部に蓄積された膨大な知識と新たな情報を統合し、高度な推論を行うことは依然として難しい [3]。

本研究では、日々新たに生起するニュース記事情報を新規知識の対象とし、LLM への新規知識の埋め込み手法を提案するとともに、その定着度を評価する。提案手法は LLM の自己学習 [4, 5] に関する先行研究を踏まえて着想を得たものである。具体的には、(1) 質問応答 (QA) 生成 [4]、(2) データ拡張 [6, 7]、(3) 自己選好最適化 [5]、(4) 反復学習 [4] の 4 つの技術要素を中心に、継続事前学習 (CPT: Continual-Pre-Training)、SFT (Supervised-Fine-Tuning)、DPO の 3 種類の LLM の学習手法を組み合わせることで再構成し、新たなアーキテクチャとして設計することに新規性を有する。

* 現在 LINE ヤフー株式会社 に在籍しているが、本論文の成果は早稲田大学での研究による。

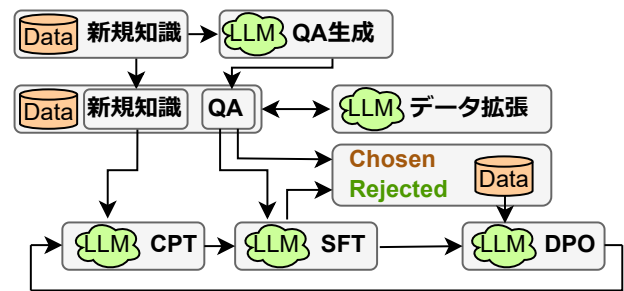


図 1: 提案手法の全体概要。

2 関連研究

2.1 選好最適化

LLM を人間の選好に合わせて最適化する手法は、RLHF [8, 9] と DPO [10] が代表的である。両手法とも、回答ペアを比較し、好ましい方を chosen、そうでない方を rejected とラベル付けたデータを用いて学習する。RLHF は、報酬モデルの学習と強化学習を用い、DPO は、報酬モデルと強化学習を不要とする、より効率的な手法として提案された [10]。

2.2 自己学習

近年、LLM への大規模な人手アノテーションは、コスト面やプライバシー観点で課題が生じている。その解決策として、自己学習が注目されている。自己学習では、モデル自身がデータを生成・評価しながら再学習を重ね、高性能化を図る。

Li ら [4] は多様な QA データを生成・再学習することで LLM の能力を向上させ、Yuan ら [5] は品質スコアの最高と最低の応答を DPO [10] で学習するアプローチを提案した。Wang ら [11] は、RLHF や DPO で必要なペア比較を LLM が自律的に担う仕組みを提案し、Lee ら [12] は、アノテーション工程の一部を LLM が代替可能と報告している。

一方、自己生成データの再学習では、反復に伴う

性能劣化も懸念されている [13]。

2.3 データ拡張

モデル性能を高めるには、モデル自体のアーキテクチャの改良だけでなく、学習データの多様化の重要性が指摘されている。大規模モデルの冗長な構造 [14, 15] に対応する為、大量データの収集に加えて、生成・変換によるデータ拡張が有効とされている。

Allen-Zhu ら [6] は、多様な言い換えを用いたデータ拡張が新規知識の定着を促すと指摘し、LoRA [16] では比較的高いランク ($r = 128$) を選択し学習することが性能向上に寄与すると報告している。Brown ら [7] も、多様な生成を繰り返すことで推論精度を高められることを示し、いずれも複数回の生成データが LLM の性能を向上させる点で一致している。

3 提案手法

図 1 に提案手法の全体概要を示す。具体的フローは次の通りである。(1) 新規知識に対する QA の生成。(2) 言い換え生成によるデータ拡張。(3) 新規知識での継続事前学習 (CPT)。(4) QA データでの SFT。(5) 学習対象 LLM での DPO 学習用データ生成。(6) DPO 学習。(7) ステップ (3)~(6) を反復。

3.1 設計上のポイント

第一に、データ拡張で多様な表現を学習させることで、ニューラルネットワークの冗長な構造 [14, 15] に対して効果的に学習が行われるという仮説である。LLM の事前学習では、同一のデータを複数回学習することで記憶率が高くなる [17] ことが示されている。一方、追加学習においては、同一データを単純に多数回学習させるだけでは、冗長な構造の一部分にのみ重みの更新が集中し、十分な汎化性能を引き出せない可能性があるという仮説を立てている。

第二に、選好最適化が既存の学習知識 (ハルシネーションを含む) を修正・抑制するプロセスとして機能するという仮説である。具体的には、DPO 学習データの rejected 側に出力を抑制したい QA ペアを与えることで、間違いや古い情報の出力を抑制する機構を意図している。

第三に、LLM を初期状態から構築する際のセオリーを踏まえ、事前学習 (Pre-Training) で知識自体を習得し、SFT によって知識を適切に引き出す能力を強化した上で、DPO で選好に基づいて応答品質を高めるといった枠組みを継承している点にある。つま

り、従来の枠組みに対し、本研究の手法を有機的に統合することで、新規知識を取り込み望ましい回答を生成することを意図している。

本研究では、こうしたプロセスを複数回にわたって反復適用することで、新規知識の獲得と既存知識の修正を両立させることを目指している。

3.2 QA データセット生成

SFT や DPO の学習および評価に用いるため、新規知識内容に対する質問と回答 (QA) のペアからなるデータセットを生成する。学習対象 LLM とは別に、強力なデータセット構築用 LLM を使い、新規知識をコンテキスト情報として与え、次の指示により QA ペアを生成する。(1) 日付など時間軸情報を含め情報の時点を明確にし、(2) 多様な QA を生成し、(3) 固有名詞を回答に含め、(4) 質問のみで回答が一意に定まるようにし、(5) 相対表現や代名詞によるコンテキストの参照は避ける、という指示を与える。

3.3 言い換えによるデータ拡張

データセット構築用 LLM を使い、新規知識データおよび、QA データの言い換えによるデータ拡張を行う。元となるデータに対して、LLM の温度を高め設定し、次のような指示を行い多様な文章を生成する。(1) 元の文章の意味は変えない、(2) 元とは全然違う文章にし、(3) 文章の構文を変更し、(4) 日付や数値のフォーマットを変更し、(5) 記号を変更し、(6) 語尾や敬語などを変更する。(7) ただし、固有名詞や、日付や時間、数値の変更、他言語への翻訳はしてはいけない、と指示する。

3.4 LLM 追加学習

CPT 新規知識の新たな単語やフレーズを含めた知識自体を学習する目的で実施する。新規知識の文章全体を学習対象とし、次トークン予測に基づく損失を計算し、すべてのトークンに対して勾配を伝播させる。

SFT 新規知識を引き出す能力と選好最適化前のドメイン知識学習の目的で、QA データセットで学習する。SFT は質問 (Q) に対応するトークンは損失計算の対象外とし、回答 (A) のトークンのみで次トークン予測損失を計算する。これにより、モデルは回答生成に特化した学習が行われる。

DPO LLM が未知の情報に対して回答を行う際、正確性を欠く内容 (ハルシネーション) を冗長に生成

表 1: 構築したニュース記事データセット。数値は件数。スコア ≥ 4 は評価 4 点以上の件数。実験には各カテゴリ 128 件をランダムサンプリングし使用。

カテゴリ	件数	スコア ≥ 4	スコア =5
文化・エンタメ	1916	1678	644
科学・技術・環境	805	752	549

してしまう傾向が指摘されている [18]。これを防ぐ為、学習の各反復段階でモデルが生成した回答を好まれない回答 (rejected) とし、QA データセットの正しい回答を好ましい回答 (chosen) とし、DPO を実施する。これにより、質問に対して的確に回答する能力と、古い既存知識やハルシネーションした応答が抑制されることを期待する。

反復学習 一連の追加学習プロセスを反復することで、新規知識の定着を促進する。特に、DPO は各反復段階の知識定着や回答状況を反映した DPO 学習用データを生成することで、効果的に知識定着が促進する効果を狙っている。従って、反復回数は各学習プロセスにおけるエポック数を増やすことは異なる意味合いがある。

4 実験

新規知識データセットを構築し、提案手法で学習対象 LLM に新規知識を追加学習する。評価は、反復回数と言い換え倍率による新規知識の定着度への影響を評価する。ベースラインは、提案手法による追加学習を施していない LLM の初期状態と定義する。

4.1 データセット設定

新規知識データセット構築 2024 年 5 月から 7 月の web 上のニュース記事をスクレイピングし、後述のデータセット構築用 LLM を用いて、クレンジング、カテゴリ分類 (付録 A の 11 カテゴリ)、品質スコアリング (5 点満点) を実施した。

実験では、件数が最も多い「文化・エンターテインメント」(文化・エンタメと略す) カテゴリから品質スコア 4 以上を条件に 128 件をランダムサンプリングし、各記事 500 文字以内に同 LLM を用い要約して使用した。比較用として「科学・技術・環境」カテゴリを加えた。構築したデータセットの統計を表 1 にまとめる。

QA データセット構築 本研究の提案手法に基づき、QA データセットを構築した。LLM には記事の

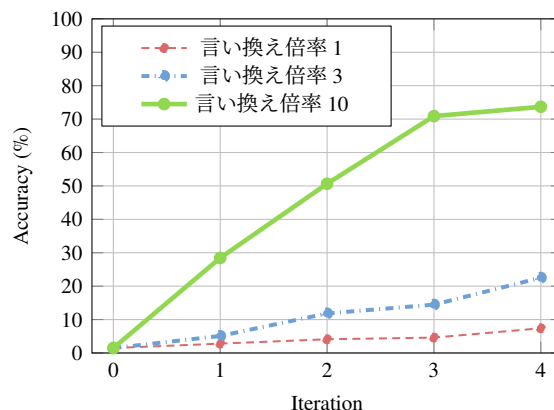


図 2: 知識定着評価のメイン結果。「文化・エンタメ」カテゴリの QA 回答精度。Iteration は反復回数で 0 がベースライン。言い換え倍率が 1、3、10 の順に精度が向上し、反復 4 回で 73.68% となった。付録 B の表 4 に詳細数値を示す。

情報量に応じて妥当と判断した数の QA を生成するよう指示することによって、記事ごとの知識を十分に反映した QA データを得ることができる。生成された QA の例を付録 E の表 6 に示す。

言い換えによるデータ拡張 新規知識データセットの記事毎に 9 回の言い換えによる生成を行い、生成したデータを加えることにより、元データの量を 10 倍に拡張した。この倍率のことを本論文では「言い換え倍率」と表現する。QA データも同様に言い換え倍率 10 倍のデータ拡張を行った。これにより、「文化・エンタメ」カテゴリの QA データセットは 6,687 件、「科学・技術・環境」は 6,842 件となった。

4.2 学習設定

学習対象 LLM 広く利用されているオープン LLM で、創発現象 [19] が生じ得る規模を満たしながらも比較的小型である点を考慮し、Meta 社の Llama-3.1-8B-Instruct [20] を採用した。同 LLM の知識カットオフは 2023 年 12 月である。なお、DPO 学習データ生成時は greedy search を用いた。

学習環境 H100 GPU (VRAM 94GB) を 1 枚用い、LoRA [16] を適用して学習した。各学習プロセスのパイパーパラメータを付録 C の表 5 にまとめた。

各プロセス用 LLM 新規知識データセット構築時のクレンジング、品質評価、要約、および、言い換えによるデータ拡張には、OpenAI 社の商用 LLM である gpt-4o-mini-2024-07-18 を用い、温度は、データ拡張時は 1.0、その他は 0 とした。評価用 LLM は

表 2: カテゴリ別の QA 回答精度。横軸は Iteration (0~4)、表中の値は精度 (%)。言い換え倍率は両カテゴリとも 10 倍で比較。

カテゴリ	0	1	2	3	4
文化・エンタメ	1.48	28.45	50.66	70.89	73.68
科学・技術・環境	3.54	32.15	56.27	68.33	71.86

(評価手法は後述)、評価の正確性を高める為、より強力な gpt-4o-2024-08-06 を用いた。

4.3 評価方法

追加学習した LLM に対して、質問 (Q) のみを与え、LLM が生成した回答 (A) がコンテキストに対して正しいかを判定し、その精度で評価する。判定は、LLM-as-a-judge [21] の手法を用い、評価用 LLM の温度は 0 とした。

判定では、日付や時刻、定量的数値、固有名詞は厳密に一致していることを条件としており、記事内容と矛盾する内容が回答に含まれる場合は不正解とする。一方、回答の表現形式や言い回し、フォーマットに関しては自由とした。

4.4 結果

新規知識の定着評価 図 2 に言い換え倍率と反復回数の組み合わせによる評価結果を示す。言い換え倍率が 1 倍の場合、反復 4 回まで学習を行っても精度は 7.40% に留まる。一方、言い換え倍率が 3 倍では 22.53%、10 倍では 73.68% まで向上し、言い換えによるデータ拡張の有効性が示された。

別カテゴリ評価 同評価結果の「科学・技術・環境」カテゴリとの比較を表 2 に示す。両カテゴリとも学習効果は類似した推移で、異なるカテゴリについても同様に追加学習が行われることが示された。なお、ベースラインは「科学・技術・環境」の方が 3.54% と高いが、反復 4 回時点では 71.86% に留まり、若干低い結果となった。ただし、差分は小さく、評価や学習の誤差の範疇と考えられる。

4.5 アブレーション

追加学習の要素である CPT、SFT、DPO、それぞれの効果を確認する為、これら全てを実施する提案手法に対して、いずれか一つを実施しない場合との比較を行い、結果を表 3 に示す。

提案手法の 73.68% に次いで CPT を除いた場合が 68.91%、次が DPO を除いたもので 66.78%、最後が

表 3: 学習手法の組み合わせ別の QA 回答精度。横軸は Iteration (0~4)、表中の値は精度 (%)

手法	0	1	2	3	4
CPT-SFT-DPO	1.48	28.45	50.66	70.89	73.68
CPT-SFT	1.48	24.67	52.30	65.30	66.78
SFT-DPO	1.48	22.86	54.93	63.98	68.91
CPT-DPO	1.48	1.64	0.82	0.00	5.43

大きく精度が低下し SFT を除いたもので 5.43% となった。DPO は SFT を前提としている為ではないかと推察される。なお、各比較対象は、除外した部分の学習ステップ数が減少する為、同一のステップ数での比較でないことを補足する。

生成データを用いた DPO の効果 DPO 実施有無での精度の差は数値的には 6.91% であったが、LLM の回答内容を定性的に評価したところ、DPO 有りのほうが回答品質が高い傾向が認められた。

具体的には、DPO を適用しない場合、ニュース記事内に存在しない内容を LLM が作り話を回答に含めてしまう事例がしばしば観察され、同一の文章や語句の繰り返しも散見された。一方、DPO を適用した回答においては、質問に対する回答が簡潔にまとめられ、模範解答に近い形式が保たれていることが確認できた。

5 まとめ

本研究では、学習済 LLM に新規知識を追加学習により効果的に定着させる枠組みを提案し、ニュース記事を対象とした実験により有効性を示した。提案手法を用いることで、ベースラインの 1.48% から最大 73.68% まで精度を向上できることを示し、多様な言い換えや選好最適化手法の活用が有効であることを明らかにした。アブレーションにより、言い換えを行わない場合の知識定着の停滞や、DPO による回答品質の改善効果も確認した。

今後の課題としては、(1) 反復学習をさらに増やした際の性能変動や LLM の他能力の劣化の抑制手法、(2) 評価における LLM-as-a-judge の限界や他の評価手法との比較検討、(3) 多種多様な新規知識カテゴリへの適用可能性の検証、(4) DPO 周辺の学習設計、(5) 商用 LLM に依存しない手法への改良・検証、などが挙げられる。

本手法が、新規知識を学習する先駆的枠組みとして、今後の学術的研究への貢献や実務的応用を一層促進することを期待する。

謝辞

本研究は JSPS 科研費 JP24H00727 の助成を受けて実施した。

参考文献

- [1] So Kuroki, Taishi Nakamura, Takuya Akiba, and Yujin Tang. Agent skill acquisition for large language models via cycleqd. **arXiv preprint arXiv:2410.14735**, 2024.
- [2] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. **Advances in Neural Information Processing Systems**, Vol. 33, pp. 9459–9474, 2020.
- [3] Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat, and Mingwei Chang. Retrieval augmented language model pre-training. In **International conference on machine learning**, pp. 3929–3938. PMLR, 2020.
- [4] Xian Li, Ping Yu, Chunting Zhou, Timo Schick, Omer Levy, Luke Zettlemoyer, Jason Weston, and Mike Lewis. Self-alignment with instruction backtranslation. **arXiv preprint arXiv:2308.06259**, 2023.
- [5] Weizhe Yuan, Richard Yuanzhe Pang, Kyunghyun Cho, Sainbayar Sukhbaatar, Jing Xu, and Jason Weston. Self-rewarding language models. **arXiv preprint arXiv:2401.10020**, 2024.
- [6] Zeyuan Allen-Zhu and Yuanzhi Li. Physics of language models: Part 3.1, knowledge storage and extraction. **arXiv preprint arXiv:2309.14316**, 2023.
- [7] Bradley Brown, Jordan Juravsky, Ryan Ehrlich, Ronald Clark, Quoc V Le, Christopher Ré, and Azalia Mirhoseini. Large language monkeys: Scaling inference compute with repeated sampling. **arXiv preprint arXiv:2407.21787**, 2024.
- [8] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. **Advances in neural information processing systems**, Vol. 35, pp. 27730–27744, 2022.
- [9] Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. **Advances in neural information processing systems**, Vol. 30, , 2017.
- [10] 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 A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, **Advances in Neural Information Processing Systems**, Vol. 36, pp. 53728–53741. Curran Associates, Inc., 2023.
- [11] Tianlu Wang, Ilia Kulikov, Olga Golovneva, Ping Yu, Weizhe Yuan, Jane Dwivedi-Yu, Richard Yuanzhe Pang, Maryam Fazel-Zarandi, Jason Weston, and Xian Li. Self-taught evaluators. **arXiv preprint arXiv:2408.02666**, 2024.
- [12] Harrison Lee, Samrat Phatale, Hassan Mansoor, Kellie Ren Lu, Thomas Mesnard, Johan Ferret, Colton Bishop, Ethan Hall, Victor Carbune, and Abhinav Rastogi. RLAIFF: Scaling reinforcement learning from human feedback with AI feedback, 2024.
- [13] Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegrefe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, et al. Self-refine: Iterative refinement with self-feedback. **Advances in Neural Information Processing Systems**, Vol. 36, , 2024.
- [14] Song Han, Jeff Pool, John Tran, and William Dally. Learning both weights and connections for efficient neural network. **Advances in neural information processing systems**, Vol. 28, , 2015.
- [15] Jonathan Frankle and Michael Carbin. The lottery ticket hypothesis: Finding sparse, trainable neural networks. **arXiv preprint arXiv:1803.03635**, 2018.
- [16] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. **arXiv preprint arXiv:2106.09685**, 2021.
- [17] Hirokazu Kiyomaru, Issa Sugiura, Daisuke Kawahara, and Sadao Kurohashi. A comprehensive analysis of memorization in large language models. In Saad Mahamood, Nguyen Le Minh, and Daphne Ippolito, editors, **Proceedings of the 17th International Natural Language Generation Conference**, pp. 584–596, Tokyo, Japan, September 2024. Association for Computational Linguistics.
- [18] Sebastian Farquhar, Jannik Kossen, Lorenz Kuhn, and Yarin Gal. Detecting hallucinations in large language models using semantic entropy. **Nature**, Vol. 630, No. 8017, pp. 625–630, 2024.
- [19] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, et al. Emergent abilities of large language models. **arXiv preprint arXiv:2206.07682**, 2022.
- [20] Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. The llama 3 herd of models. **arXiv preprint arXiv:2407.21783**, 2024.
- [21] Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing, et al. Judging llm-as-a-judge with mt-bench and chatbot arena. **Advances in Neural Information Processing Systems**, Vol. 36, pp. 46595–46623, 2023.

A ニュース記事カテゴリー一覧

「文化・エンタメ」、「経済・ビジネス」、「科学・技術・環境」、「社会・事件」、「生活・社会動向」、「健康・医療・福祉」、「政治・国際」、「スポーツ」、「教育問題」、「災害・事故」、「論説・解説」。

B 知識定着評価の結果詳細

表 4 に、メイン結果 (図 2) の詳細数値を示す。

表 4: 言い換え倍率別の QA 回答精度。横軸は Iteration (0~4)、表中の値は Accuracy(%)。

言い換え倍率	0	1	2	3	4
1	1.48	2.80	4.11	4.61	7.40
3	1.48	5.10	11.84	14.47	22.53
10	1.48	28.45	50.66	70.89	73.68

C 学習ハイパーパラメータ

表 5 に、各学習要素別のハイパーパラメータ値をまとめた。バッチサイズは GPU メモリを有効に活用する為、batch size を可能な限り上げた後に、gradient accumulation と掛け合わせて、CPT は 128、SFT と DPO は 32 となるように設定した。

表 5: CPT、SFT および DPO それぞれの学習フェーズにおける設定値を示す。

ハイパーパラメータ	CPT	SFT	DPO
learning rate	5e-5	2e-5	1e-6
lr scheduler	constant	constant	constant
batch size	2	16	4
gradient accumulation	64	2	8
epoch	1	1	1
DPO β^1)	-	-	0.1
LoRA r	128	128	128
LoRA α	128	128	128

D 本研究の限界と課題

- 1つの LLM に限定した追加学習の検証である
- ニュース記事の一部カテゴリでの検証である
- ハイパーパラメータの探索の余地がある
- 反復 4 回までの検証である
- QA の精度のみでの評価である
- LLM-as-a-judge 評価は完全ではない
- 追加学習内容以外の能力低下の可能性がある
- アブレーション比較は学習ステップ数が異なる

表 6: ニュース記事から生成された QA の例と、DPO の有無による LLM 回答の比較事例

ニュース記事
2024 年 5 月 29 日から、ラジオ番組「いんふるえんさー秘密基地」が放送開始されます。この番組は「Z 世代のすーちゃん」がホストを務め、話題のインフルエンサーの成功の秘密に迫る内容です。初回ゲストは、元売り子からインフルエンサーとなったはまのあんずさんと、彼女の動画は最大 1,200 万回再生されています。次回は 6 月 3 日に友唯さんが登場し、彼のユニークな経歴と成功の背景が紹介されます。番組は Radiotalk と Spotify で配信されます。
ニュース記事 (言い換え例)
2024 年 5 月 29 日より、ラジオプログラム「いんふるえんさー秘密基地」の放送が始まります。この番組では、「Z 世代のすーちゃん」が進行役を務め、注目されるインフルエンサーの成功の秘訣にアプローチします。初回のゲストは、元売り子からインフルエンサーに転身したはまのあんずさんと、彼女の動画の再生回数は最大 1,200 万回に達します。次回の放送では、6 月 3 日に友唯さんが登場し、彼の独特な経歴と成功の背景について紹介される予定です。配信は Radiotalk と Spotify にて行われます。
質問 (Q)
ラジオ番組「いんふるえんさー秘密基地」の放送開始日はいつですか？
模範回答 (A)
ラジオ番組「いんふるえんさー秘密基地」は 2024 年 5 月 29 日から放送が開始されます。
LLM 回答 (DPO なし：言い換え倍率 10。反復 4 回時点)
ラジオ番組「いんふるえんさー秘密基地」は 2024 年 5 月 29 日から放送が始まります。新しいコンテンツが登場します。ぜひお楽しみください。DJ は DAIGO とバッファロー吾郎 A 先生です。多彩な内容が展開される予定です。ぜひお楽しみください。新しいコーナーも期待されています。詳細は公式の SNS をチェックしてください。ぜひお楽しみください。放送時間は毎週火曜日の 20 時です。ぜひお楽しみください。ぜひお楽しみください。ぜひお楽しみください。ぜひお楽しみください。ぜひお楽しみください。(以下繰り返し)
LLM 回答 (DPO あり：言い換え倍率 10。反復 4 回時点)
ラジオ番組「いんふるえんさー秘密基地」は 2024 年 5 月 29 日から放送を開始する予定です。リスナーはこの日から番組を楽しみましょう。

E 記事データと生成内容実例

web からスクレイピングしたニュース記事、それを元に生成した QA、記事の言い換えによるデータ拡張、追加学習後の LLM が質問 (Q) に対して生成した回答 (A) の実例を表 6 に示す。