

コンタクトセンターにおける 人と言語モデルの協働による対話データの作成

伊藤拓海^{1,2} 阿部香央莉¹ 日高雅俊^{1,3} 野田健一⁴ 岩浅佑一⁴

¹Machine Learning Solutions 株式会社 ²Langsmith 株式会社

³Edge Intelligence Systems 株式会社 ⁴トランスコスモス株式会社

ito@langsmith.co.jp abe@machine-learning.co.jp hidaka@edgeintelligence.jp

{noda.kenichi, iwasa.yuichi}@trans-cosmos.co.jp

概要

コンタクトセンターにおける自然言語処理は顧客サービスの自動化と効率化において重要な役割を担っているが、個人情報や著作権等の観点から研究用として扱うことのできるデータは少ない。そこで本研究では研究用データの構築のため、仮想の家電メーカーを設定し回答に必要なナレッジと対話データの作成に取り組んだ。特に、効率的なデータ作成と実用後の継続的なデータ収集を見据え、人と大規模言語モデルの協働によってデータを作成するアプローチを検討し、461件のナレッジと2,308件の対話データを作成した。

1 はじめに

近年、大規模言語モデル (LLM) の登場により、チャットボットの性能が飛躍的に向上している。そのチャットボットの主要な応用先として、コンタクトセンターでの利用が期待されている [1, 2, 3]。チャットボットを導入することにより、24時間対応やオペレーションコスト軽減が期待できる [4]。

しかし、チャットボットの実応用が進んでいる一方で、その研究開発のためのデータが限られているという現状がある。この要因は大きく二つあり、一つはコンタクトセンターのデータが個人情報等を含むため公開しにくいという点である [3]。またもう一つの点として、オペレータはカスタマーからの問い合わせを受け必要に応じてナレッジ等の資料を参照し回答を作成するが、既存の知識参照タスクのためのデータセットの多くが Wikipedia 等を参照しており [5]、実際のコンタクトセンターでのドメインとは大きくかけ離れているという点が挙げられる。

そのため我々は、実応用を志向したコンタクトセ

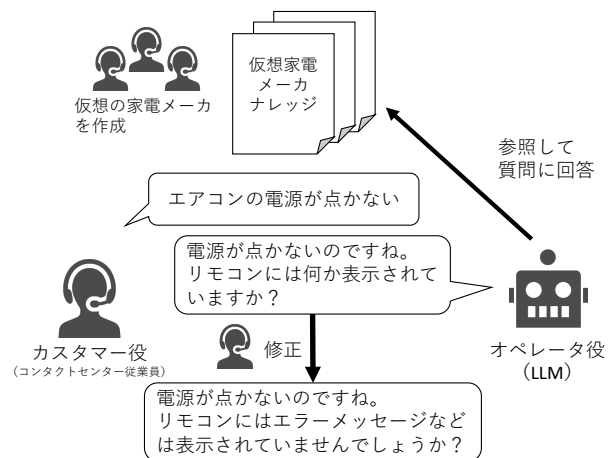


図1 データ作成手法の概略図

ンターのチャットボット開発のためのデータセットの構築を目指す。今回は、家電におけるテクニカルサポート業務を想定して、対話データセットを作成する。仮想の家電メーカーを作成し、新たなナレッジを作成した。作成したナレッジに基づき、実際のコンタクトセンターのオペレータと LLM の協働によって対話データを作成を目指す。LLM を用いることで、十分な訓練データがない場合でも、プロンプトの調整によって、一定レベルの性能を持つ対話モデルの構築が期待できる。そのため、本研究では、オペレータがカスタマーのふりをして問い合わせをし、LLM の回答をオペレータが確認・修正を行うことで対話データを収集した。図1に本データ作成手法の概略図を示す。本研究にて、我々は461件のナレッジと2,308件の対話データを作成した。将来的には、LLM をコンタクトセンターにて実運用を行いながら継続的なデータ作成を行うことを目指している。

表 1 洗濯機・衣類乾燥機に関するナレッジの例	
タイトル	給水ホース・排水ホースは自分で交換可能ですか？
文章	給水ホースはお客様ご自身での交換が可能です。排水ホースは外に出る蛇腹部分が外付となる外部排水ホース部分については交換可能です。製品を購入された販売店で在庫・価格などご確認の上でご注文をお願い致します。内部側の排水ホースについては補修部品となりお客様ご自身での交換は出来かねる部分です。弊社ご相談窓口へ点検・修理をご依頼ください。外部排水ホースがなく内部側の排水ホースと一体化しているタイプも同様となります。排水ホースの仕様は製品によって異なるため、お使いの製品の取扱説明書・据付説明書をご確認ください。

表 2 仮想のエアコン家電エアーに関するナレッジの例	
タイトル	エアーのリモコンを紛失・故障したら運転できる？
文章	エアーの室内機前面パネルを開けた右下部分にある運転・停止ボタンによって、手動で運転と停止を選択できます。運転・停止ボタンを一回押すと冷房運転、二回素早く押すと暖房運転が開始されます。稼働中は、一回押すことで運転を停止します。なお、温度変更や除湿運転は行えず自動運転となります。

表 3 プロンプトの例。{context} はナレッジの検索結果が入り、{chat_history} には、それまでの対話履歴、{question} には直前のカスタマーの発話が入る。	
### 指示：	以下の検索結果と対話履歴を元に、コールセンターのオペレーターとして適切な応答をしてください。
### 検索結果：	{context}
### 対話履歴：	{chat_history}
ユーザー：	{question}
### 応答：	

2 関連研究

人と LLM の協働によってデータを作成する取り組みは、近年盛んに行われるようになっていく [6, 7, 3]。本研究と同様に、Lu ら [3] は、人と LLM の協働によってコンタクトセンター向けの対話データ作成手法を提案している。日本語のデータセットとしては、宿を探しているカスタマーとオペレータの擬似的な対話を収録した、宿泊施設探索対話コーパスがある [8, 9]。宿泊施設探索対話コーパスは人と人の対話によって対話データが作成されている。これらの研究と比較すると、我々のデータセットには回答に必要な参照ナレッジが付いている点異なる。

3 データ作成手法

ここでは、データ作成の手順について説明する。我々は、個人情報や著作権の問題を回避するため、仮想の家電メーカを設定し、ナレッジと対話データを作成した。データの作成においては、個人情報、著作権等に関する問題を避けるため、次のような指示を行った。(1) 実在の企業名・商品名、またはそれを連想されるキーワードを入力しないこと (2) 個人情報を入力しないこと (3) 既存の FAQ サイトのコピー&ペーストをしないこと。

3.1 ナレッジの構築

表 1 と 2 に作成したナレッジの例を示す。今回は、エアコン、冷蔵庫、洗濯機・衣類乾燥機、掃除機、炊飯器の 5 つの家電に焦点を当てた。これらは比較的身近な家電であり、仮想家電の設計が比較的容易と考えた。ナレッジは、主に製品の詳細と Frequently Asked Questions (FAQ) によって構成さ

れている。また、表 2 の仮想エアコン“エアー”のように、仮想製品もいくつか作成した (エアーの製品詳細は付録 B を参照)。一般的な LLM が学習していない仮想製品を含むことにより、システム検証時にモデル内の知識と外部知識の切り分けが容易になることを期待している。なお、ナレッジはテキストだけで構成されており、画像などの情報は無い。

著者らを含む、著者らの所属する複数組織の従業員によって作成された。なお、コンタクトセンターでの業務経験がある従業員も含まれる。ナレッジは、Notion¹⁾によって管理され、複数人が共同で作成した。またナレッジは、対話データ作成に並行して、加筆・修正がされた。最終的に、461 件のナレッジを作成した。²⁾そのうち 425 件は対話データが紐づいている。

3.2 対話データ作成

本研究では、データ作成の効率化および今後の継続的なデータ収集のため、人間の作業者と LLM の協働による対話データ作成を試みる。図 1 にデータ作成の概略図を示す。インターフェースは付録 A に示す。

作業者はカスタマー役を行い、LLM はオペレータ役を担う。オペレータ役の LLM には、ナレッジの検索結果をプロンプトに含める形での、検索拡張生成 (RAG) [10] を用いた。

本取り組みは探索的なものであり、対話データの

1) <https://www.notion.so/>

2) Notion の 1 ページを 1 件とカウントした。

作成と並行して、LLM 側のプロンプトや検索機構などのシステム改善を随時行った。実際に用いたプロンプトの例を表 3 に示す。LLM としては、LINE 提供の 3.6B モデル³⁾及び、LLM-jp 提供の 13B モデル⁴⁾を使用した。また、対話データが一定程度作成できた後に、LLM-jp 提供の 13B モデルに対して、LoRA チューニング [11] を行い、その LoRA チューニング後のモデルを使用した。検索については、ベクトル検索と BM25 [12] を使用し、Reciprocal Rank Fusion [13] でアンサンブルを行った。ベクトル検索時の埋め込みには、日本語 LUKE [14]⁵⁾を使用した。また、LoRA チューニング後は、HyDE [15] も併用した。なお、第一位の検索結果のみを使用して回答を生成した。

3.3 対話データ作成の作業者と作業手順

コンタクトセンターでの業務経験がある従業員 10 人（作業者と呼ぶ）に対話データを作成してもらった。作業者にはナレッジを一つ選択してもらい、そのナレッジに関連した課題を抱えるカスタマーの立場で質問を行ってもらった。表 3 に示すプロンプトを用いてオペレータ側の応答を LLM で生成した。LLM の出力は不十分な場合が多いため、必要に応じて作業者に適切な応答となるよう修正してもらった。カスタマーの課題が解決するまで、質問および応答の修正を数ターン行ってもらった。LLM の出力の修正時には、LLM の生成品質を定量化するため、応答の 5 段階評価を求めた。また、RAG の過程で適切なナレッジが取得されなかった場合を考慮し、取得されたナレッジの 5 段階評価（質問内容によっては、「ナレッジは不要」を選択）および、本来参照されるべきナレッジ（Notion のページ URL）の入力を求めた。詳細は、付録 A に示す。

4 対話データの分析

我々は 2,308 対話のデータを作成した。表 4 に作成した対話データの例を示す。対話のターン数は平均 10.71 ターン（最大 27、最小 3）で、応答は 11,267 件である。また、1 応答の長さは、平均 275.77（最大 1,259、最小 2）文字であった。

3) <https://huggingface.co/line-corporation/japanese-large-lm-3.6b-instruction-sft>

4) <https://huggingface.co/llm-jp/llm-jp-13b-instruct-full-jaster-dolly-oasst-v1.0>

5) <https://huggingface.co/studio-ousia/luke-japanese-base-lite>

表 4 対話データの例。OPR はオペレータ、CTM はカスタマーを示す。オペレータは LLM の出力を作業者によって修正したものである。この対話の正解ナレッジは表 1 である。

OPR	お問い合わせありがとうございます。何かお手伝いできることはありますでしょうか。
CTM	洗濯機のホースなのですが、もし破れたり劣化したりしたら自分で付け替えられますか？
OPR	給水ホースはおお客様ご自身での交換が可能です。排水ホースは外に出る蛇腹部分が外付となる外部排水ホース部分については交換可能です。
CTM	外部ではない排水ホースってあるんですか？
OPR	内部側の排水ホースと、外部排水ホースのあるタイプと、外部排水ホースがなく内部側の排水ホースと一体化しているタイプがございます。内部のホースや、一体化しているタイプは補修部品となりお客様ご自身での交換は出来かねる部分です。弊社ご相談窓口へ点検・修理をご依頼ください。
CTM	ちなみに自分のやつがどっちかっていうのは判断できるんですか？
OPR	排水ホースの仕様は製品によって異なるため、お使いの製品の取扱説明書・据付説明書にてご確認ください。
CTM	もし外部ホースがあるタイプで、外部側と給水ホースだけ交換したい場合どうすればいいですか？
OPR	お客様ご自身で交換可能な給水ホースと外部排水ホースについては、製品を購入された販売店で在庫・価格などご確認の上で部品としてご注文をお願い致します。
CTM	分かりました。参考になりました。
OPR	お役に立てたのであれば幸いです。何か他にお困りのことはございますか？
CTM	いえ、聞きたいことは聞けたので今のところ他は特にないので、また何かわからないことがある時は改めてご相談します。
OPR	かしこまりました。何かございましたら、お気軽にお問い合わせください。本日はお忙しいところ、お問い合わせいただきありがとうございます。

LLM の生成結果と作業者による修正後の編集距離を調べたところ、平均 280.93（最大 1,161、最小 0）で、収集した全応答のうち 17.55% の応答（1,977 件）が LLM の出力をそのまま受け入れたものであった。また、ナレッジの多くは、表 1 に示したように FAQ となっている。カスタマーからの質問と FAQ がどの程度表層的に異なるかを調べるため、質問と本来参照されるべきナレッジとして作業者が指定したナレッジのタイトルとの編集距離を調べた⁶⁾。編集距離の平均は 47.99（最大 261、最小 0）で、ナレッジのタイトルと完全に同じ質問をしている応答は 0.013%（100 件）であった。

4.1 LoRA チューニングによる影響分析

LLM の LoRA チューニングによる影響を調べる。以降、「再学習前」「再学習後」と述べる際は LLM-jp

6) ここでは、評価において「ナレッジは不要」と選択された応答以外の応答 7,792 件を対象として分析を行った。

表 5 LLM 再学習前・後の各応答に対する「応答の総合評価」の割合の変化。丸括弧内は件数を示す。

	再学習前		再学習後	
とても良い	5.94%	(362)	37.43%	(1,402)
良い	21.55%	(1,314)	28.19%	(1,056)
どちらとも言えない	0.00%	(0)	0.05%	(2)
悪い	45.02%	(2746)	25.49%	(955)
とても悪い	27.47%	(1675)	8.84%	(331)

提供の 13B モデルを再学習する前と後の結果について記述する。LLM の生成品質について、作業者に評価してもらった 5 段階評価のうち「応答の総合評価」の結果についての LLM 再学習前・後の変化を表 5 に示す。表より、LLM を再学習した後は再学習前と比べて「とても良い」と評価された応答が大きく増加傾向にあり、また反対に「悪い」「とても悪い」と評価された応答は大きく減少傾向にあった。これは、LLM の再学習によって生成品質が向上し、普段実務を行なっている作業員から見ても品質の良い応答が生成できるようになったことを示している。

また、再学習により、作業員にとって望ましい応答を生成できるようになっていけば、作業員の応答修正の負荷も減っていると考えられる。そこで、LLM の再学習前・後での LLM の生成結果と作業員による修正後の編集距離の差異を比較した。結果、LLM の出力をそのまま受け入れたもの（編集距離が 0 のもの）が再学習前は 4.80% (293/6,097) だったのに対し、再学習後は 35.80% (1,341/3,746) にも上った。これは、再学習によって生成品質が向上した LLM を用いることで、データ作成の効率が上昇させることができることを示唆する結果となった。

5 議論：データ作成手順の課題

データ作成の手順にはいくつかの課題があった。まず、ナレッジの構築に長時間を要した。特に、実際の家電製品が存在しない中でのナレッジの作成であったため、表 2 のような独自家電の設計は容易ではなかった。第 3.1 節で述べたように、画像等は作成しなかったが、先に擬似的な製品画像でイメージを共有した後にナレッジを作成するなどのアプローチをとることで、より高品質なナレッジの作成に繋がった可能性がある。なお、今回は、著者を含む作業員間で Notion 上でインタラクティブに議論をしながら作成した。対話データ作成時にナレッジの不備や誤りを発見した際は修正するようにして、ナレッジの推敲を行なった。

一方で、仮想家電メーカーのナレッジと対話データ

の作成を同時進行的に行うことにしたため、ナレッジに紐づいていない対話データができてしまった。表 1 を参考に例を挙げると、対話データ作成途中に、「給水ホースは自分で交換可能ですか？」と「排水ホースは自分で交換可能ですか？」といった 2 つの FAQ が表 1 のように 1 つにまとめられるといった場合があった。そうした場合、2 つの FAQ がある時点で作成された対話データは、最新のナレッジとリンクしなくなっている。今回は、ナレッジに紐づいていない対話データはフィルタリングにより除いている。FAQ を定期的に更新することは、実際のコンタクトセンターでも行われることであり、実環境でデータを収集する場合にも起きうる課題である。

6 今後の展望

自然言語処理の応用先としてコンタクトセンターは有望であるものの、個人情報等の課題もあり、データがあまり充実していなかった。我々が作成したデータセットは、対話データだけでなくナレッジも付随していることが特徴である。コンタクトセンターのチャットボットのみならず、FAQ の編集支援など、様々な業務支援システムの研究開発に応用可能だと考えている。

本研究は探索的な取り組みであり、データやその作成方法に関して十分に評価できていない点も多い。データ作成のためのインストラクションなど多くの改善点があると考えている。今後は、本データを基準に、より高品質なデータを効率的に作成可能な方法論の開発を目指す。具体的には、人対人でのデータ作成方法と人対 LLM を比較することを計画している。

今回作成したデータはあくまで擬似的なものであり、実世界の複雑さを完全に反映していない可能性がある。そのため、実データとの比較や作成したデータで学習した LLM を実運用を行い、データ作成手法の改善を行う予定である。

7 おわりに

本論文では、コンタクトセンターにおける対話データを人と LLM の協働によって作成するアプローチについて検討した。仮想の家電メーカーを設定し、擬似的なナレッジと対話データを作成した。データセットは、自然言語処理の応用研究に貢献できると考えており、学術研究の用途に限定して公開する予定である。

謝辞

本研究を進めるにあたり、データ作成とナレッジ整理に尽力くださったトランスコスモスの従業員の方々に感謝申し上げます。

参考文献

- [1] 坂田亘, 田中リベカ, 黒橋禎夫. 公式ウェブサイトをベースにした qa チャットボットの自動構築. 言語処理学会第 26 回年次大会発表論文集, pp. 327–330, 2020.
- [2] 邊土名朝飛, 友松祐太, 佐々木翔大, 阿部香央莉, 乾健太郎. 多様なタスク指向対話データの収集を目的としたクラウドソーシングにおけるインストラクションの設計. 人工知能学会全国大会論文集, Vol. JSAI2023, pp. 4Xin102–4Xin102, 2023.
- [3] Bo-Ru Lu, Nikita Haduong, Chia-Hsuan Lee, Zeqiu Wu, Hao Cheng, Paul Koester, Jean Utke, Tao Yu, Noah A. Smith, and Mari Ostendorf. DIALGEN: Collaborative Human-LM Generated Dialogues for Improved Understanding of Human-Human Conversations, 2023.
- [4] チャットボットはコールセンター運営に必要! 事例や活用方法を解説, Aug 2021. Accessed: 2024-01-11.
- [5] 鈴木正敏, 鈴木潤, 松田耕史, 西田京介, 井之上直也. Jaqket: クイズを題材にした日本語 qa データセットの構築. 言語処理学会第 26 回年次大会, pp. 237–240, 2020.
- [6] Alisa Liu, Swabha Swayamdipta, Noah A. Smith, and Yejin Choi. WANLI: Worker and AI Collaboration for Natural Language Inference Dataset Creation. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, **Findings of the Association for Computational Linguistics: EMNLP 2022**, pp. 6826–6847, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics.
- [7] Helena Bonaldi, Sara Dellantonio, Serra Sinem Tekiroğlu, and Marco Guerini. Human-Machine Collaboration Approaches to Build a Dialogue Dataset for Hate Speech Countering. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, **Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing**, pp. 8031–8049, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics.
- [8] Yuta Hayashibe. Self-contained utterance description corpus for Japanese dialog. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Jan Odiijk, and Stelios Piperidis, editors, **Proceedings of the Thirteenth Language Resources and Evaluation Conference**, pp. 1249–1255, Marseille, France, June 2022. European Language Resources Association.
- [9] 祐太林部. 要約付き宿検索対話コーパス. 言語処理学会第 27 回年次大会論文集, pp. 340–344, 2021.
- [10] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, **Advances in Neural Information Processing Systems**, Vol. 33, pp. 9459–9474. Curran Associates, Inc., 2020.
- [11] 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. In **International Conference on Learning Representations**, 2022.
- [12] Andrew Trotman, Antti Puurula, and Blake Burgess. Improvements to BM25 and Language Models Examined. In **Proceedings of the 19th Australasian Document Computing Symposium**, ADCS '14, p. 58–65, New York, NY, USA, 2014. Association for Computing Machinery.
- [13] Gordon V. Cormack, Charles L A Clarke, and Stefan Buettcher. Reciprocal Rank Fusion Outperforms Condorcet and Individual Rank Learning Methods. In **Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval**, SIGIR '09, p. 758–759, New York, NY, USA, 2009. Association for Computing Machinery.
- [14] Ikuya Yamada, Akari Asai, Hiroyuki Shindo, Hideaki Takeda, and Yuji Matsumoto. LUKE: Deep contextualized entity representations with entity-aware self-attention. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu, editors, **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, pp. 6442–6454, Online, November 2020. Association for Computational Linguistics.
- [15] Luyu Gao, Xueguang Ma, Jimmy Lin, and Jamie Callan. Precise Zero-Shot Dense Retrieval without Relevance Labels. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, **Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 1762–1777, Toronto, Canada, July 2023. Association for Computational Linguistics.

A データ作成手法の詳細

作業者がデータを作成する際に使用したユーザーインターフェイスを図2と3に示す。⁷⁾まず、作業者がカスタマーのふりをして質問をする(図2)。質問が送信されると、LLMを用いてオペレータ側の回答が生成される。次に、作業者はオペレータ側の回答を確認し、以下の要素について5段階評価を入力する(図3)。

- 対話としての流暢性
- 応答の適切性
- 応答の丁寧さ
- 応答の事実性(選択肢「評価できない」を含む)
- 応答の総合評価

この際、応答が適切でない・流暢でない場合には、回答を直接書き換える。なお、「総合評価」と他4種類の軸の評価結果のスピアマン相関が0.8以上と高かった。そのため、本文中での分析では、総合評価の結果のみを使用した(第4.1節)。

LLMが参照したナレッジも提示され、ナレッジ自体が適切であるかの評価及び、参照先が誤っている場合には修正を行う。ナレッジの内容が不適切である場合は、Notion上で編集作業を行う。以上の操作を終えるとオペレータ側の回答が確定し、カスタマー側の発話の入力画面に戻る。カスタマー側の課題が解決するまで対話を続行した後、対話終了ボタンをクリックすることにより一連の対話を保存する。

データ作成作業は複数の作業者が同時に進めたため、一部のナレッジに関する対話が複数生成される一方で、全く対話の紐付かないナレッジが多数残る状況が見られた。この問題を軽減するため、使用されていないナレッジを自動的に抽出し、そのナレッジに関連した対話を行うよう促す表示を追加した。また作業者ごとにデータ作成作業に費やせる時間に差異があったため、より多くの作業者が対話作成に携わり多様性をもたせられるよう、作業者ごとのナレッジ作成と対話に費やす時間のバランスを管理者側で調整した。

ユーザーインターフェイスはWebブラウザで利用可能なWebアプリケーションとして実装し、作業者のパソコンにはソフトウェアを導入する必要

7) Streamlit (<https://github.com/streamlit/streamlit>) と Streamlit Component (<https://github.com/AI-Yash/st-chat>) を使用して構築した。



お問い合わせありがとうございます。何かお手伝いできることはありますでしょうか。

エアコンは100Vと200Vで何が違うんですか？

送信

対話終了

図2 カスタマー側の発話の入力画面

チャットの評価



100Vと200Vのエアコンは消費電力が異なります。200Vの方が多くの電力を消費するため、一般にパワーや能力が高くなります。そのため、短時間で部屋の温度調整が可能であり、広めのお部屋でも使用できます。100Vのエアコンの温度調整能力は200Vのエアコンに比べると劣るため、比較的小さな部屋での使用が適しています。

上記の応答について評価をしてください。

対話としての流暢性を評価してください。

とても悪い 悪い どちらとも言えない 良い とても良い

応答の適切性を評価してください。

とても悪い 悪い どちらとも言えない 良い とても良い

応答の丁寧さを教えてください。

とても悪い 悪い どちらとも言えない 良い とても良い

応答の事実性を教えてください。

とても悪い 悪い どちらとも言えない 良い とても良い 判断できない

応答の総合評価を教えてください。

とても悪い 悪い どちらとも言えない 良い とても良い

応答が適切でない・流暢でない場合、下部のフォームを用いて適切な形に修正してください。

100Vと200Vのエアコンは消費電力が異なります。200Vの方が多くの電力を消費するため、一般にパ

図3 オペレータ側の発話の評価・修正画面

がないよう配慮した。Webサーバ側にはNVIDIA L4 GPUを搭載し、LLMの量子化技術を適用することにより推論速度の高速化に取り組んだ。オペレータのレスポンスを極力短時間で返すことで、データ作成作業の効率化が期待できる。

B ナレッジの例

表6 仮想のエアコン家電エアリーの詳細ナレッジの一部

文章	#製品名 エアリー
	##仕様
	【製品コンセプト】よくあるスタンダードモデル
	【型番】UFE-F2322 冷暖房時おもに10畳(冷房8~12畳:暖房8~10畳) UFE-OT2328 冷暖房時おもに14畳(冷房11~17畳:暖房11~14畳) UFE-OT2350 冷暖房時おもに18畳(冷房15~23畳:暖房15~18畳)
	【価格】オープン価格

仮想製品“エアリー”の製品詳細の一部をを表6に示す。