Adaptive Data Generation and Direct Preference Optimization for Medical Forum Summarization

Anonymous EMNLP submission

Abstract

Summarizing medical forums requires clinical precision. However, traditional supervised fine-tuning (SFT) methods rely on static datasets that cannot adapt to evolving clinical language and diverse user queries. This study aims to overcome these limitations by combining SFT with synthetic data generation and direct preference optimization (DPO). Using PerAnsSumm dataset, we trained a Mistral 7B model to generate synthetic preference data labeled as "rejected" and mark expert summaries as "chosen". KeyBERT-extracted keywords augmented the inputs to enhance contextual relevance. Our DPO-adapted model significantly outperformed the baselines, achieving scores of 0.458 (ROUGE-Lsum), 0.511 (SacreBLEU), and 0.880 (METEOR). Keyword integration prevented performance degradation when adapting to new summary types, increasing the METEOR score by 13.4% in originally excluded categories. This study confirms that using synthetic data and preference optimization reduces the need for costly annotations and enables flexible, clinically precise summarization.

1 Introduction

Large language models (LLMs) have advanced rapidly, enabling their use in clinical summarization tasks like radiology reports and patient queries. However, achieving high-quality summarization in a medical forum context remains challenging due to the diversity of user queries and the need for clinical precision. Traditional supervised fine-tuning (SFT) methods require extensive annotated data, and even when available, the static nature of such datasets can limit the model's ability to adapt to evolving language and clinical practices.

To mitigate these problems, we describe a two-stage method, shown in Figure 1. First, we use SFT to train a large summarization model with 80% of the data from the publicly available medical questions dataset. Secondly, we generate synthetic data where generated summaries are labeled as "rejected" and original expert summaries as "chosen", providing a preference dataset for DPO that enables the model to learn fine-grained quality distinctions.

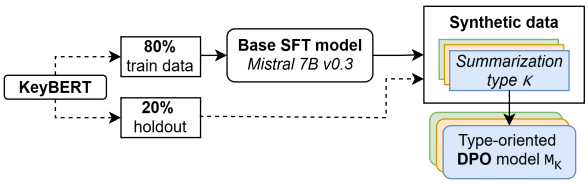

Figure 1: Proposed medical forum summarization pipeline involves KeyBERT extracting top-5 keywords and a Mistral 7B SFT model fine-tuning 80% of the data. The remaining 20% are held out to train synthetic preference data, employed to train individual DPO models for each type of summary.

Our innovation pipeline uses KeyBERT keyword extraction to enhance contextual understanding, improving accuracy and preventing performance degradation when adapting to new summaries. Our framework addresses data scarcity in medical NLP by combining synthetic data generation with optimization, reducing the reliance on costly annotations while maintaining clinical precision.

2 Related Work

Recent studies in medical summarization have demonstrated the effectiveness of transforming intricate medical dialogues into concise and accurate summaries (Liu et al., 2024; Fraile Navarro et al., 2025). Conventional SFT methodologies are predominantly dependent on substantial labeled datasets. However, in medical forums and analogous environments, the challenge of acquiring high-quality human-labeled data has led researchers towards synthetic data generation and preference-guided training methods (Ouyang et al., 2022).

Synthetic data has proven to be effective in addressing medical scarcity and privacy issues. Re-

search shows its uses in simulation, algorithm testing, and training; an evaluation mechanism is in development (Rujas et al., 2025). For instance, Medically Aware GPT-3 has been used as a medical dialogue summarization training data generator, with synthetic training datasets that demonstrate superiority over much larger datasets with human labels (Chintagunta et al., 2021).

Continually with advances in synthetic data, direct preference optimization (DPO) has emerged as a compelling alternative to traditional reinforcement learning from human feedback (RLHF) to align language model output with human preferences. In contrast to RLHF, where a reward model must first be trained, DPO tunes the base model directly with human preference data, thus creating a less cumbersome and often more reliable training pipeline (Rafailov et al., 2023). Concurrent medical work has utilized DPO in combination with parameter-efficient fine-tuning (PEFT) to generate discharge summaries for patients, reducing clinician workload and ensuring that summaries convey high-quality, relevant information for clinicians (Ahn et al., 2024; Xiao et al., 2024).

The integration of synthetic data generation and DPO holds considerable potential for medical forum summarization. UltraMedical presents a preference-annotated collection of synthetic and manually harvested biomedical data, demonstrating the utility of high-quality synthetic datasets for fine-tuning domain-specific models (Zhang et al., 2024). Furthermore, recent work in generating medical reports with DPO has shown that spurious output, such as hallucinated prior examination information, can be effectively suppressed while maintaining clinical accuracy (Banerjee et al., 2024).

Unlike previous studies that used synthetic data generation and preference-based optimization separately in medical applications, our work combines both, adaptively generating training data while applying DPO to refine multifaceted medical forum discussion summaries.

3 Methods

3.1 Data Preparation and Fine-tuning

In order to develop a robust summarization model tailored for medical forum contexts, a two-stage training pipeline was constructed. We employ an open-source dataset derived from the *PerAnsSumm* task (Naik et al., 2024), which comprises medical forum questions and their corresponding expert-written summaries in different styles. A subset of 200 samples was reserved, maintaining the same proportional distribution of summarization types as the original dataset. This part of the data serves as a reference for evaluating model performance across different summarization types.

The remaining data was divided into two parts: 80% for training and 20% for adaptation. Subsequently, the 80% of the training data were used for the first fine-tuning of our base model. SFT pipeline for our base model leverages *Mistral 7B-v0.3* (Jiang et al., 2023), a 7 billion-parameter Mistral AI developed transformer model with an extended vocabulary and v3 Tokenizer compatibility. In this stage, the model learns to generate coherent summaries from actual medical queries, developing a baseline capability for future refinement processes. Comparison by Glazkov and Makarov's (2024) demonstrated strong ability of Mistral 7B-v0.3 to produce to produce high-quality dialogue summary. We applied Low-Rank Adaptation (LoRA) (Hu et al., 2022) to the base model to enable parameter-efficient fine-tuning.

To enhance the model's understanding of significant terms and improve accuracy in the input information, a keyword extraction feature with KeyBERT is utilized (Grootendorst, 2020). KeyBERT is a lightweight and simple keyword extraction algorithm utilizing BERT embeddings for keywords and key phrases extraction most closely related to the input text. By excluding less significant information, the model can then specifically target key sections of the input, improving the accuracy and relevance of the generated summaries. For any query, a top-5 keyword list is extracted, which is then incorporated into the preprocessing pipeline. These keywords serve as additional context at both the SFT and DPO stages, improving contextual relevance and making the model comprehend key information in the input.

3.2 Synthetic Data Generation for DPO

Following the SFT stage, we generate synthetic samples in preparation for fine-tuning the model for individual types of summarization via DPO. For that purpose, we utilized 20% of the training dataset, whose inputs have been partitioned in terms of respective types of summarization (e.g. INFORMATION, CAUSE, SUGGESTION, EXPERIENCE, QUESTION). Inference for individual input is performed via a base model trained on the SFT stage. These generated summaries were

Table 1: Overall Evaluation Metrics Across Summarization Models for all Summary Types. "*KW*" denotes models enhanced with keyword-augmented input through KeyBERT extraction. The values are shown as mean standard error estimated using a bootstrapping method over per-example metric scores.

Model	ROUGE-1	ROUGE-Lsum	BERTScore-F1	SacreBLEU	METEOR
			Baselines		
DistilBART	0.454±*0.010*	0.353±*0.010*	0.909±*0.002*	0.135±*0.008*	0.545±*0.010*
DistilBART + KW	0.452±*0.010*	0.355±*0.010*	0.910±*0.002*	0.112±*0.009*	0.443±*0.011*
t5-Large	0.516±*0.014*	0.432±*0.015*	0.918±*0.002*	0.409±*0.014*	0.647±*0.015*
t5-Large + KW	0.516±*0.013*	0.429±*0.015*	0.918±*0.002*	0.409±*0.014*	0.647±*0.015*
t5 Base	0.248±*0.011*	0.208±*0.011*	0.870±*0.002*	0.045±*0.007*	0.136±*0.012*
t5-Base + KW	0.247±*0.011*	0.209±*0.010*	0.871±*0.002*	0.045±*0.007*	0.136±*0.012*
t5-v1.1-Large	0.414±*0.013*	0.342±*0.013*	0.875±*0.011*	0.193±*0.010*	0.593±*0.013*
t5-v1.1-Large + KW	0.395±*0.013*	0.322±*0.013*	0.893±*0.002*	0.026±*0.008*	0.085±*0.013*
t5-v1.1-Small	0.206±*0.011*	0.180±*0.011*	0.865±*0.002*	0.032±*0.008*	0.096±*0.010*
t5-v1.1-Small + KW	0.205±*0.011*	0.175±*0.011*	0.864±*0.002*	0.032±*0.008*	0.096±*0.010*
Mistral 7B Baseline	0.533±*0.013*	0.447±*0.014*	0.922±*0.002*	0.409±*0.013*	0.647±*0.014*
Mistral 7B Baseline + KW	0.537±*0.013*	0.441±*0.014*	**0.926**±*0.002*	0.409±*0.014*	0.647±*0.014*
			Trained models		
Mistral 7B - adaption	0.551±*0.012*	0.450±*0.013*	0.923±*0.002*	0.501±*0.013*	0.864±*0.014*
Mistral 7B - adaption + KW	**0.557**±*0.012*	**0.458**±*0.013*	**0.926**±*0.002*	**0.511**±*0.014*	**0.880**±*0.014*

labeled as "rejected", while the original (gold) summaries were marked as "chosen".

This labeling scheme aligns with state-of-the-art breakthroughs in preference and synthetic feedback and enables us to develop a preference corpus with rich variation in summary quality and factuality. To enable efficient processing, a batch expansion scheme is utilized, in which one input comes with a group of several rejected candidates for careful examination of potential errors and variations.

3.3 Preference Optimization Adaptation

Our approach incorporates DPO for secondary refinement of the base model. DPO directly optimizes the preference margin between the "chosen" text and the "rejected" text (Rafailov et al., 2023). By leveraging the synthetic preference dataset generated on-policy by the model, DPO ensures that the model learns about its weaknesses and iteratively builds its summarization capabilities.

For each summary type, a distinct DPO model is trained with an augmented synthetic corpus. To achieve this objective, the *DPOTrainer* is initialized with hyperparameters such as beta (set to 0.1), and the model is optimized over a sequence of 2048 for prompts and 1024 for responses.

3.4 Quantitative Evaluation

To assess the effectiveness of our approach, we use a suite of evaluation metrics for general summarization and requirements in a medical field. For individual tasks, ROUGE (Lin, 2004), BLEU (Papineni et al., 2002), and METEOR (Banerjee and Lavie, 2005) assess lexical and structural overlaps between generated and referent summaries. Additionally, we utilize BERTScore (Zhang et al., 2020), which leverages contextualized BERT representations to assess semantic similarity— an advantage in medical domains where paraphrasing and synonym use are common.

4 Results

The validation configuration utilizes the previously reserved 200 sample subset, with balanced distribution over all types of summarization. The evaluation metrics are demonstrated in Table 1. Each score is reported as the mean with standard error, where standard error is estimated using a bootstrapping method with 1,000 resamples over the per-example metric scores. This approach provides a robust measure of variability, which is particularly important given the relatively small size of the validation set. Besides, individual performance for each type of summarization and overall aggregated scores for a complete view of model performance were shown in Appendix A and Appendix B.

The results demonstrate statistically significant improvements in lexical metrics (ROUGE, METEOR, and SacreBLEU) and semantic metrics (BERTScore-F1). Notably, adapting models with keyword-augmented input improves performance consistently across most evaluation criteria. The proposed adaptation of the Mistral 7B model achieves the best overall performance, especially

when enhanced with keyword extraction, outperforming both strong baselines and previously fine-tuned models.

5 Discussion

To assess the generalizability of our two-stage training pipeline, we conducted an additional experiment in which the EXPERIENCE summary type was omitted from the initial SFT stage and then re-introduced during DPO adaptation. As shown in Table 2, excluding EXPERIENCE from the base model and subsequently adapting with our pipeline on that category leads to significant gains in all evaluation metrics for EXPERIENCE summaries. In particular, ROUGE-1, ROUGE-L and METEOR scores demonstrate a consistent upward trend, regardless of whether keywords are used, confirming that new summarization types can be seamlessly added post hoc via our DPO procedure.

Table 2: Comparison of EXPERIENCE Metrics with and without DPO

Metric	Category	Performance	
		Base	DPO
w\o Keywords	ROUGE-1	0.390 ± *0.023*	0.394 ± *0.027* ↑
	ROUGE-2	0.155 ± *0.020*	0.171 ± *0.022* ↑
	ROUGE-L	0.289 ± *0.019*	0.303 ± *0.022* ↑
	ROUGE-Lsum	0.289 ± *0.019*	0.303 ± *0.022* ↑
	BERTScore F1	0.895 ± *0.003*	0.890 ± *0.005* ↓
	SacreBLEU	0.210 ± *0.015*	0.258 ± *0.018* ↑
	METEOR	0.477 ± *0.025*	0.603 ± *0.028* ↑
with Keywords	ROUGE-1	0.399 ± *0.027*	0.403 ± *0.024* ↑
	ROUGE-2	0.172 ± *0.023*	0.178 ± *0.023* ↑
	ROUGE-L	0.308 ± *0.023*	0.316 ± *0.023* ↑
	ROUGE-Lsum	0.308 ± *0.023*	0.316 ± *0.023* ↑
	BERTScore F1	0.896 ± *0.004*	0.898 ± *0.004* ↑
	SacreBLEU	0.217 ± *0.019*	0.223 ± *0.018* ↑
	METEOR	0.469 ± *0.026*	0.539 ± *0.026* ↑

However, Table 3 reveals a clear benefit in incorporating keyword-augmented inputs when aiming to preserve performance across all summary types. When keywords are omitted, DPO adaptation in the withheld EXPERIENCE category yields improvements on that specific class but comes at the expense of degraded performance in other categories —most notably SacreBLEU and METEOR, which drop markedly. In contrast, the keyword-augmented variant not only improves EXPERIENCE metrics but also maintains or slightly enhances aggregated metrics across all summary types, demonstrating that keywords act as effective anchors that prevent catastrophic drift during category-specific adaptation.

Taken together, these findings validate the flexibility and robustness of our pipeline. We can extend the model to new summarization types with minimal retraining, and by leveraging keyword guidance, we safeguard overall model quality while still capturing the nuances of each individual summary type. This characteristic is especially valuable in clinical settings, where new reporting requirements or annotation schemes may emerge over time.

Table 3: Comparison of Overall Metrics with and without DPO

Metric	Category	Performance	
		Base	DPO
w\o Keywords	ROUGE-1	0.525 ± *0.012*	0.513 ± *0.012* ↓
	ROUGE-2	0.286 ± *0.013*	0.279 ± *0.013* ↓
	ROUGE-L	0.426 ± *0.013*	0.414 ± *0.012* ↓
	ROUGE-Lsum	0.426 ± *0.013*	0.414 ± *0.012* ↓
	BERTScore F1	0.919 ± *0.002*	0.916 ± *0.002* ↓
	SacreBLEU	0.501 ± *0.012*	0.235 ± *0.011* ↓
	METEOR	0.864 ± *0.014*	0.572 ± *0.013* ↓
with Keywords	ROUGE-1	0.519 ± *0.013*	0.531 ± *0.013* ↑
	ROUGE-2	0.285 ± *0.014*	0.299 ± *0.014* ↑
	ROUGE-L	0.423 ± *0.013*	0.437 ± *0.014* ↑
	ROUGE-Lsum	0.423 ± *0.013*	0.437 ± *0.014* ↑
	BERTScore F1	0.919 ± *0.002*	0.922 ± *0.002* ↑
	SacreBLEU	0.460 ± *0.012*	0.501 ± *0.013* ↑
	METEOR	0.857 ± *0.014*	0.864 ± *0.014* ↑

6 Conclusion

This study aims to address the challenge of generating high-quality medical forum summaries. To this end, it proposes a methodology that overcomes the limitations of traditional SFT methods. The latter rely on static annotated datasets and struggle to adapt to evolving clinical language and diverse user queries. Our two-stage approach, combining SFT with synthetic data generation and DPO, demonstrates significant improvements in both lexical and semantic summary quality while enabling flexible adaptation to new summarization types.

These results hold significant promise for clinical applications, where the ability to adapt to new summarization requirements - such as evolving medical guidelines or emerging terminology - is paramount. By leveraging synthetic data and preference optimization, our approach reduces reliance on costly human annotations while maintaining clinical precision. In addition, the modular design facilitates the seamless integration of new summary types, offering practical utility in real-world settings such as telemedicine platforms or automated clinical reporting.

Limitations

Our two-stage pipeline has shown promising results, but it is important to note that it also has several limitations. First, the PerAnsSumm dataset is publicly available and well-structured, but it is small, homogeneous, and underrepresents clinical language, rare or specialized topics, and multi-turn forum dynamics. Additionally, summary-type labels occasionally overlap and, in some cases, they are incorrect or incomplete. This can introduce bias into training and evaluation processes and limit generalizability to settings with divergent terminology or dialogue structures.

Secondly, our evaluation is based solely on automated metrics (ROUGE, BERTScore, SacreBLEU, METEOR), which capture lexical and semantic overlap. However, these metrics cannot guarantee clinical correctness, factual consistency, or readability, nor can they detect subtle medical errors or missing contraindications. Human evaluation by domain experts is therefore essential to validate safety and practicality.

Finally, the DPO adaptation's modularity comes with computational overhead and sensitivity to hyperparameters like DPO beta, learning rate, and bootstrap resamples. Future work should explore automated hyperparameter tuning or adaptive DPO scheduling for cost reduction and robustness.

References

Imjin Ahn, Hansle Gwon, Young-Hak Kim, Tae Joon Jun, and Sanghyun Park. 2024. NOTE: notable generation of patient text summaries through efficient approach based on direct preference optimization. *CoRR*, abs/2402.11882.

Oishi Banerjee, Hong-Yu Zhou, Subathra Adithan, Stephen Kwak, Kay Wu, and Pranav Rajpurkar. 2024. Direct preference optimization for suppressing hallucinated prior exams in radiology report generation. *CoRR*, abs/2406.06496.

Satanjeev Banerjee and Alon Lavie. 2005. METEOR: an automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization@ACL 2005, Ann Arbor, Michigan, USA, June 29, 2005*, pages 65–72. Association for Computational Linguistics.

Bharath Chintagunta, Namit Katariya, Xavier Amatriain, and Anitha Kannan. 2021. Medically aware GPT-3 as a data generator for medical dialogue summarization. In *Proceedings of the Machine Learning for Healthcare Conference, MLHC 2021, 6-7 August 2021, Virtual Event*, volume 149 of *Proceedings of Machine Learning Research*, pages 354–372. PMLR.

David Fraile Navarro, Enrico Coiera, Thomas W Hambly, Zoe Triplett, Nahyan Asif, Anindya Susanto, Anamika Chowdhury, Amaya Azcoaga Lorenzo, Mark Dras, and Shlomo Berkovsky. 2025. Expert evaluation of large language models for clinical dialogue summarization. *Scientific Reports*, 15(1):1195.

Nikita Glazkov and Ilya Makarov. 2024. Utterance-aware adaptive data labeling and summarization: Exploiting large language models for unbiased dialog annotation. *IEEE Access*, 12:150793–150806.

Maarten Grootendorst. 2020. Keybert: Minimal keyword extraction with bert.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. Lora: Low-rank adaptation of large language models. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de Las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b. *CoRR*, abs/2310.06825.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81.

Yong Liu, Shenggen Ju, and Junfeng Wang. 2024. Exploring the potential of chatgpt in medical dialogue summarization: a study on consistency with human preferences. *BMC Medical Informatics Decis. Mak.*, 24(1):75.

Gauri Naik, Sharad Chandakacherla, Shweta Yadav, and Md Shad Akhtar. 2024. No perspective, no perception!! perspective-aware healthcare answer summarization. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 15919–15932, Bangkok, Thailand. Association for Computational Linguistics.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F. Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, July 6-12, 2002, Philadelphia, PA, USA*, pages 311–318. ACL.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D. Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*.

Miguel Rujas, Rodrigo Martín Gómez Del Moral Herranz, Giuseppe Fico, and Beatriz Merino-Barbancho. 2025. Synthetic data generation in healthcare: A scoping review of reviews on domains, motivations, and future applications. *Int. J. Medical Informatics*, 195:105763.

Wenyi Xiao, Zechuan Wang, Leilei Gan, Shuai Zhao, Wanggui He, Luu Anh Tuan, Long Chen, Hao Jiang, Zhou Zhao, and Fei Wu. 2024. A comprehensive survey of datasets, theories, variants, and applications in direct preference optimization. *CoRR*, abs/2410.15595.

Kaiyan Zhang, Sihang Zeng, Ermo Hua, Ning Ding, Zhang-Ren Chen, Zhiyuan Ma, Haoxin Li, Ganqu Cui, Biqing Qi, Xuekai Zhu, Xingtai Lv, Jinfang Hu, Zhiyuan Liu, and Bowen Zhou. 2024. Ultramedical: Building specialized generalists in biomedicine. *CoRR*, abs/2406.03949.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. Bertscore: Evaluating text generation with BERT. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net.

A Mistral 7B - Base adaption without keywords

Table 4: Comparison of DPO Adaptations by Task Type for Mistral 7B - Base adaption

Task Type	Metric	Solution					
		Base SFT	CAU. DPO	EXP. DPO	INF. DPO	QUEST. DPO	SUG. DPO
CAUSE	ROUGE-1	**0.535**±*0.051*	0.533±*0.051*	0.532±*0.051*	0.533±*0.051*	**0.535**±*0.051*	**0.535**±*0.051*
	ROUGE-2	0.347±*0.060*	0.342±*0.061*	0.347±*0.060*	**0.351**±*0.061*	0.347±*0.060*	0.348±*0.060*
	ROUGE-L	0.472±*0.056*	0.477±*0.056*	0.483±*0.055*	0.484±*0.055*	0.472±*0.056*	**0.485**±*0.055*
	ROUGE-Lsum	0.472±*0.056*	0.477±*0.056*	0.483±*0.055*	0.484±*0.055*	0.472±*0.056*	**0.485**±*0.055*
	BERTScore F1	**0.928**±*0.007*	0.924±*0.007*	0.924±*0.006*	0.925±*0.007*	0.926±*0.007*	0.924±*0.007*
	SacreBLEU	**0.078**±*0.060*	**0.078**±*0.051*	**0.078**±*0.049*	**0.078**±*0.060*	**0.078**±*0.053*	**0.078**±*0.057*
	METEOR	**0.295**±*0.057*	**0.295**±*0.057*	**0.295**±*0.057*	**0.295**±*0.058*	**0.295**±*0.057*	**0.295**±*0.057*
EXPERIENCE	ROUGE-1	**0.508**±*0.031*	0.506±*0.031*	0.505±*0.032*	0.501±*0.029*	0.505±*0.031*	0.503±*0.030*
	ROUGE-2	0.256±*0.036*	0.256±*0.037*	**0.258**±*0.038*	0.249±*0.033*	0.256±*0.037*	0.252±*0.035*
	ROUGE-L	**0.413**±*0.033*	0.411±*0.034*	0.411±*0.035*	0.405±*0.032*	0.411±*0.034*	0.408±*0.033*
	ROUGE-Lsum	**0.413**±*0.033*	0.411±*0.034*	0.411±*0.035*	0.405±*0.032*	0.411±*0.034*	0.408±*0.033*
	BERTScore F1	**0.916**±*0.005*	0.915±*0.005*	**0.916**±*0.005*	0.914±*0.005*	**0.916**±*0.005*	0.914±*0.005*
	SacreBLEU	0.256±*0.033*	0.256±*0.034*	0.246±*0.035*	0.246±*0.027*	**0.267**±*0.034*	0.256±*0.029*
	METEOR	0.517±*0.035*	0.517±*0.035*	0.427±*0.036*	0.427±*0.034*	**0.518**±*0.035*	0.517±*0.035*
INFORMATION	ROUGE-1	**0.576**±*0.014*	0.574±*0.014*	0.569±*0.014*	0.574±*0.014*	0.572±*0.015*	0.574±*0.014*
	ROUGE-2	**0.322**±*0.018*	0.318±*0.018*	0.314±*0.018*	0.320±*0.018*	0.319±*0.018*	0.320±*0.018*
	ROUGE-L	0.453±*0.018*	0.451±*0.018*	0.451±*0.017*	**0.455**±*0.018*	0.449±*0.018*	**0.455**±*0.018*
	ROUGE-Lsum	0.453±*0.018*	0.451±*0.018*	0.451±*0.017*	**0.455**±*0.018*	0.449±*0.018*	**0.455**±*0.018*
	BERTScore F1	**0.923**±*0.003*	0.922±*0.003*	0.922±*0.003*	**0.923**±*0.003*	0.922±*0.003*	**0.923**±*0.003*
	SacreBLEU	**0.198**±*0.017*	0.130±*0.017*	0.170±*0.017*	0.145±*0.019*	0.130±*0.017*	0.130±*0.018*
	METEOR	**0.412**±*0.019*	0.400±*0.019*	0.390±*0.019*	0.399±*0.018*	0.400±*0.019*	0.400±*0.018*
QUESTION	ROUGE-1	**0.652**±*0.047*	**0.652**±*0.047*	0.600±*0.065*	0.597±*0.065*	0.626±*0.058*	0.595±*0.066*
	ROUGE-2	**0.463**±*0.061*	**0.463**±*0.061*	0.430±*0.072*	0.439±*0.072*	0.444±*0.069*	0.422±*0.075*
	ROUGE-L	**0.632**±*0.049*	**0.632**±*0.049*	0.585±*0.065*	0.590±*0.065*	0.605±*0.059*	0.578±*0.067*
	ROUGE-Lsum	**0.632**±*0.049*	**0.632**±*0.049*	0.587±*0.064*	0.592±*0.064*	0.605±*0.059*	0.580±*0.066*
	BERTScore F1	**0.948**±*0.007*	**0.948**±*0.007*	0.937±*0.011*	0.936±*0.011*	0.942±*0.009*	0.937±*0.011*
	SacreBLEU	**0.501**±*0.063*	**0.501**±*0.063*	**0.501**±*0.068*	**0.501**±*0.068*	**0.501**±*0.066*	**0.501**±*0.070*
	METEOR	**0.864**±*0.059*	**0.864**±*0.059*	**0.864**±*0.079*	**0.864**±*0.079*	**0.864**±*0.073*	**0.864**±*0.081*
SUGGESTION	ROUGE-1	0.527±*0.024*	0.529±*0.024*	0.528±*0.025*	0.537±*0.023*	**0.540**±*0.023*	0.527±*0.024*
	ROUGE-2	0.291±*0.027*	0.292±*0.027*	0.291±*0.027*	0.300±*0.027*	**0.303**±*0.026*	0.291±*0.027*
	ROUGE-L	0.423±*0.024*	0.425±*0.024*	0.421±*0.025*	0.432±*0.024*	**0.435**±*0.024*	0.425±*0.025*
	ROUGE-Lsum	0.423±*0.024*	0.425±*0.024*	0.421±*0.025*	0.432±*0.024*	**0.435**±*0.024*	0.425±*0.025*
	BERTScore F1	0.923±*0.004*	0.923±*0.003*	0.922±*0.004*	0.923±*0.003*	**0.924**±*0.003*	0.922±*0.004*
	SacreBLEU	**0.409**±*0.024*	**0.409**±*0.024*	**0.409**±*0.024*	**0.409**±*0.025*	**0.409**±*0.024*	**0.409**±*0.024*
	METEOR	**0.647**±*0.026*	**0.647**±*0.026*	**0.647**±*0.026*	**0.647**±*0.026*	**0.647**±*0.026*	**0.647**±*0.026*

B Mistral 7B - Base adaption with keywords

Table 5: Comparison of DPO Adaptations by Task Type for Mistral 7B - Base adaption + KW

Task Type	Metric	Solution					
		Base SFT	CAU. DPO	EXP. DPO	INF. DPO	QUEST. DPO	SUG. DPO
CAUSE	ROUGE-1	**0.535**±*0.052*	0.529±*0.052*	0.529±*0.052*	0.522±*0.052*	0.531±*0.052*	0.526±*0.052*
	ROUGE-2	**0.331**±*0.060*	0.325±*0.060*	0.324±*0.060*	0.317±*0.061*	0.326±*0.060*	0.322±*0.060*
	ROUGE-L	**0.471**±*0.057*	0.462±*0.057*	0.462±*0.058*	0.454±*0.059*	0.463±*0.058*	0.460±*0.058*
	ROUGE-Lsum	**0.471**±*0.057*	0.462±*0.057*	0.462±*0.058*	0.454±*0.059*	0.463±*0.058*	0.460±*0.058*
	BERTScore F1	**0.927**±*0.008*	**0.927**±*0.008*	**0.927**±*0.008*	0.926±*0.008*	**0.927**±*0.008*	0.926±*0.008*
	SacreBLEU	**0.078**±*0.061*	**0.078**±*0.061*	**0.078**±*0.061*	**0.078**±*0.061*	**0.078**±*0.061*	**0.078**±*0.061*
	METEOR	**0.295**±*0.057*	**0.295**±*0.057*	**0.295**±*0.057*	**0.295**±*0.057*	**0.295**±*0.057*	**0.295**±*0.057*
EXPERIENCE	ROUGE-1	0.496±*0.032*	0.493±*0.032*	0.495±*0.032*	0.495±*0.032*	0.493±*0.032*	**0.498**±*0.032*
	ROUGE-2	**0.260**±*0.038*	0.257±*0.038*	0.257±*0.038*	0.256±*0.038*	0.258±*0.038*	0.258±*0.038*
	ROUGE-L	0.408±*0.034*	0.405±*0.035*	0.407±*0.035*	0.407±*0.034*	0.405±*0.035*	**0.409**±*0.034*
	ROUGE-Lsum	0.408±*0.034*	0.405±*0.035*	0.407±*0.035*	0.407±*0.034*	0.405±*0.035*	**0.409**±*0.034*
	BERTScore F1	0.914±*0.005*	0.914±*0.005*	**0.915**±*0.005*	**0.915**±*0.005*	0.914±*0.005*	**0.915**±*0.005*
	SacreBLEU	**0.262**±*0.035*	**0.262**±*0.035*	0.246±*0.035*	**0.262**±*0.036*	**0.262**±*0.035*	0.246±*0.035*
	METEOR	**0.546**±*0.038*	**0.546**±*0.038*	0.427±*0.037*	**0.546**±*0.038*	**0.546**±*0.038*	0.427±*0.037*
INFORMATION	ROUGE-1	0.576±*0.015*	0.576±*0.016*	0.575±*0.016*	**0.580**±*0.016*	0.578±*0.016*	0.574±*0.016*
	ROUGE-2	0.320±*0.018*	0.320±*0.019*	0.319±*0.019*	**0.323**±*0.019*	**0.323**±*0.020*	0.321±*0.020*
	ROUGE-L	0.453±*0.018*	0.455±*0.019*	0.452±*0.019*	**0.458**±*0.019*	0.456±*0.019*	0.454±*0.019*
	ROUGE-Lsum	0.453±*0.018*	0.455±*0.019*	0.452±*0.019*	**0.458**±*0.019*	0.456±*0.019*	0.454±*0.019*
	BERTScore F1	**0.927**±*0.003*	0.926±*0.003*	0.926±*0.003*	**0.927**±*0.003*	0.926±*0.003*	0.926±*0.003*
	SacreBLEU	0.125±*0.019*	**0.126**±*0.020*	0.117±*0.020*	0.117±*0.019*	0.125±*0.020*	0.125±*0.020*
	METEOR	**0.377**±*0.018*	**0.377**±*0.019*	0.348±*0.020*	0.348±*0.019*	**0.377**±*0.019*	**0.377**±*0.020*
QUESTION	ROUGE-1	**0.661**±*0.047*	**0.661**±*0.047*	**0.661**±*0.047*	0.647±*0.045*	**0.661**±*0.047*	**0.661**±*0.047*
	ROUGE-2	**0.472**±*0.056*	**0.472**±*0.056*	**0.472**±*0.056*	0.455±*0.056*	**0.472**±*0.056*	**0.472**±*0.056*
	ROUGE-L	**0.640**±*0.047*	**0.640**±*0.047*	**0.640**±*0.047*	0.625±*0.048*	**0.640**±*0.047*	**0.640**±*0.047*
	ROUGE-Lsum	**0.640**±*0.047*	**0.640**±*0.047*	**0.640**±*0.047*	0.625±*0.048*	**0.640**±*0.047*	**0.640**±*0.047*
	BERTScore F1	**0.943**±*0.007*	**0.943**±*0.007*	**0.943**±*0.007*	0.941±*0.007*	**0.943**±*0.007*	**0.943**±*0.007*
	SacreBLEU	**0.511**±*0.055*	**0.511**±*0.055*	**0.511**±*0.055*	0.273±*0.053*	**0.511**±*0.055*	**0.511**±*0.055*
	METEOR	**0.880**±*0.059*	**0.880**±*0.059*	**0.880**±*0.059*	0.688±*0.056*	**0.880**±*0.059*	**0.880**±*0.059*
SUGGESTION	ROUGE-1	**0.552**±*0.024*	0.548±*0.025*	0.545±*0.025*	0.543±*0.025*	0.546±*0.025*	0.546±*0.025*
	ROUGE-2	**0.315**±*0.027*	0.309±*0.027*	0.304±*0.027*	0.303±*0.027*	0.306±*0.027*	0.308±*0.027*
	ROUGE-L	**0.453**±*0.025*	0.447±*0.026*	0.443±*0.026*	0.440±*0.026*	0.444±*0.026*	0.448±*0.026*
	ROUGE-Lsum	**0.453**±*0.025*	0.447±*0.026*	0.443±*0.026*	0.440±*0.026*	0.444±*0.026*	0.448±*0.026*
	BERTScore F1	**0.927**±*0.004*	0.926±*0.004*	0.926±*0.004*	0.926±*0.004*	0.926±*0.004*	**0.927**±*0.004*
	SacreBLEU	**0.409**±*0.026*	**0.409**±*0.025*	**0.409**±*0.026*	**0.409**±*0.025*	**0.409**±*0.025*	**0.409**±*0.025*
	METEOR	**0.647**±*0.027*	**0.647**±*0.027*	**0.647**±*0.028*	**0.647**±*0.028*	**0.647**±*0.028*	**0.647**±*0.028*