EmPO: Theory-Driven Dataset Construction for Empathetic Response Generation through Preference Optimization

Ondrej Sotolar
Faculty of Informatics, Masaryk University, Brno, Czechia
{xsotolar@fi.muni.cz}

Abstract

Empathetic response generation is a desirable aspect of conversational agents, crucial for facilitating engaging and emotionally intelligent multi-turn conversations between humans and machines. Leveraging large language models for this task has shown promising results, yet challenges persist in ensuring both the empathetic quality of the responses and retention of the generalization performance of the models. In this paper, we propose a novel approach where we construct theory-driven preference datasets and use them to align LLMs with preference optimization algorithms to address these challenges. To measure empathetic response generation, we employ the EmpatheticDialogues dataset, assessing empathy with the diff-Epitome and BERTscore metrics, and evaluate the generalization performance on the MMLU benchmark. We make all datasets, source code, and models publicly available.[1]

1 Introduction

Empathetic response generation (ERG) focuses on tuning a conversational agent toward understanding the user's situation, feelings, and experience to generate appropriate, human-like responses. This goal was first attempted by the earliest rule-based chatbots like ELIZA (Weizenbaum, 1966), then more recently with deep learning approaches such as BlenderBot (Roller et al., 2021; Xu et al., 2022; Komeili et al., 2022; Shuster et al., 2022). The introduction of instruction-tuned large language models (LLMs) such as ChatGPT (OpenAI, 2023) has challenged the contemporary approaches to ERG. These models can participate in multi-turn conversations without additional training and are implicitly capable of generating empathetic conversations (Lee et al., 2022a).

LLMs also have shown tremendous generalization capabilities and can draw on world knowledge obtained by training on Internet-size data. These traits were not present in the earlier non-LLM conversational agents. Recent developments focus on approaches that keep the generalization abilities of LLMs while improving empathy in responses, including prompt-engineering methods that do not change the model's weights but rather tune the prompt content to assist the generation by in-context learning (Wang et al., 2024), often assisted by retrieval (Qian et al., 2023).

In the current study, we propose a data-driven solution for ERG with LLMs, based on preference optimization: aligning LLMs via preference optimization algorithms. First, we build a preference dataset using the benchmark EmpatheticDialogues dataset (Rashkin et al., 2019). The dataset contains short multi-turn human-to-human dialogues grounded by emotion labels. We leverage this property to extract responses from the corpus of the polar opposite emotion label using Plutchik's wheel (Plutchik, 2001) such that each prompt is paired with preferred and non-preferred completions. We then fine-tune a foundational LLM using Direct Preference Optimization (Rafailov et al., 2024) to generate responses aligned with the preferred candidate response.

Using diff-Epitome (Lee et al., 2022b), a model-based empathy metric for multi-turn conversations, we show that training LLMs with our preference dataset improves ERG. We also investigate the models' general language understanding using the MMLU benchmark to show how our method impacts the performance on other tasks. Our method is analogous to providing guardrails with helpful/harmful preference datasets (Bai et al., 2022). Contemporary prompt-engineering methods or additional training can be applied to models aligned in this way to adapt them further for any task. We also share novel observations from searching over the hyperparameter configuration space and provide code to apply our method to other datasets and models.

[1] github.com/ondrejsotolar/empo

2 Related Work

Systems specialized for ERG such as KEMP (Li et al., 2022), CEM (Sahand Sabour, 2021), MIME (Majumder et al., 2020), EmpDG (Li et al., 2020), and DIFFUSEMP (Bi et al., 2023) or universal conversational agent like Blenderbot were not built for general instruction-following. Recent approaches such as (Wang et al., 2024; Qian et al., 2023; Li et al., 2024) employ prompt engineering with LLMs to some success. However, their approach differs principally from ours; we restructure the problem of ERG as an LLM alignment problem and produce a method for creating preference datasets and a modeling approach rather than a prompt construction algorithm.

To our knowledge, the current study is the first to explore aligning LLMs for ERG via preference optimization.

Datasets Several dialogue datasets, including IEMOCAP (Busso et al., 2008), MELD (Poria et al., 2019), DailyDialog (Li et al., 2017), Herzig et al., 2016, EmotionLines (Hsu et al., 2018), EmpathicReactions (Buechel et al., 2018), and EmoContext (Chatterjee et al., 2019), contain emotion labels. However, these datasets are either labeled with only a small set of emotions, limited by size, or lack a multi-turn character.

The EmpatheticDialogues (ED) is a dataset with 25K human-to-human dialogues and 32 emotion labels, which are derived from biological responses (Ekman, 1992; Plutchik, 1980) to larger sets of subtle emotions derived from contextual situations (Skerry and Saxe, 2015). It is the benchmark dataset on empathetic conversation as it addresses preceding datasets' limitations.

Earlier deep learning-based conversational agents required large datasets for training from scratch, often at the expense of quality. A subset of ED was analyzed for listener-specific response intents, creating the EDOS dataset (Welivita et al., 2021), which includes 1M movie dialogues annotated with a BERT-based classifier. Recently, synthetic datasets like Chinese SMILE (Qiu et al., 2024) and Korean SoulChat (Chen et al., 2023) have been generated using LLMs.

However, ED and its successors have been criticized for their data annotation and model evaluation approaches (Debnath and Conlan, 2023). In treating ERG as an LLM alignment task, our approach addresses some of these drawbacks.

Evaluating ERG Evaluation of empathy in dialogues has primarily focused on human evaluation and lexical overlap/semantic similarity with reference dialogues. The former is limited by the highly subjective nature of the task, the need to train annotators, and, usually, small sample sizes. We have found the latter to have little relevance for measuring empathy, given the highly creative nature of generative LLM responses, in agreement with the findings of Liu et al., 2016. We instead rely on model-based empathy metrics, which use classification models trained specifically for the task of measuring empathy in multi-turn conversations: diff-EPITOME (Lee et al., 2022b) which is based on the EPITOME (Sharma et al., 2020) classifiers. For measuring the generalization abilities, we use the standard benchmarks such as MMLU (Hendrycks et al., 2020), using the `lm-evaluation-harness` (Bommasani et al., 2023).

LLM Alignment by Preference Optimization Aligning generative models with human feedback has improved their helpfulness, factual accuracy, and ethical behavior, among other aspects (Ouyang et al., 2022). Methods like RLHF (Christiano et al., 2017), including training algorithms such as PPO (Schulman et al., 2017) and DPO (Rafailov et al., 2023) have consistently been more effective than relying solely on supervised fine-tuning (SFT). Human feedback can take many forms: PPO requires human preferences to rank the generations, and DPO requires datasets of prompts and pairs of preferred/rejected completions.

3 Experimental Setup

We conduct our experiments using the Zephyr-7B (Tunstall et al., 2023) model, a variant of the Mistral-7B foundational LLM (Jiang et al., 2023) fine-tuned for multi-turn dialogues using the UltraChat dataset (Ding et al., 2023). The smallest variant with 7 billion parameters already provides good general language understanding ability as it matches LLama-2-70B on many NLP benchmarks (Beeching et al., 2023). For all training steps, we explore the hyperparameter configuration space and optimize for a minimal impact on generalization while improving empathy scores.

Supervised Fine-tuning As Tunstall et al., 2023 showed, SFT is a necessary first step in alignment, which ensures that the preference dataset is indomain for the aligned model. We perform SFT using the standard causal auto-regressive objective using the individual dialogues from EmpatheticDialogues. As per the definition of the dataset, the odd turns are considered the "user prompts" and the even turns the "responses". We limit the SFT training to the even-indexed turns by masking the odd turns to ignore them in the loss-function computation. For computational efficiency, we fine-tune using LoRA adapters (Hu et al., 2021).

Preference Optimization For DPO , we build a preference dataset from the EmpatheticDialogues consisting of preferred/rejected completions. For each dialogue, we target the last even turn (the last "response" to user) as the generation target while including the previous turns as context. This is the standard way of processing the dataset, also done in previous works.

For constructing the completion pairs, we leverage a property of the ED dataset: each dialogue is associated with an emotion label. These were used as grounding during the ideation of each dialogue. For the preferred completion, we take the ground truth - the original response. For its rejected counterpart, we use Plutchik's wheel of emotions (Figure 4), and the derivative emotional dyads (Figure 5, Plutchik, 2001) to find the polar opposite emotion labels resulting in a lookup table (see A.3). For each completion, we randomly select one from the group of completions labeled with this opposite label. Because the stability can suffer from this random selection, we draw a fresh random completion for each new epoch of training and search for the hyperparameter configuration to offset the repetition of the preferred completions.

Empathy Evaluation For measuring empathy, we use the diff-Epitome metric (Lee et al., 2022a): a model-based metric specifically developed for measuring empathy in dialogues. diff-Epitome is an evolution of the Epitome classifiers (Sharma et al., 2020), which measure empathy in dialogue on a scale (0-2) from none (0), through weak (1), to strong (2) on three dimensions: empathetic responses (ER), explanations (EX), and interpretations (IP). The diff-Epitome uses a similar but open-ended continuous scale (0-) on the same dimensions by averaging the scores across the entire dataset, thus providing a measure of difference from the ground truth. Given our data-driven approach, we consider a lower difference from the ground truth better.

In preliminary human-evaluation experiments with lexical-overlap metrics such as BLEU or ROUGE and vector-similarity such as BERTscore, we found them uncorrelated with the perceived quality of the generated responses. This was especially evident with the largest LLMs, such as GPT4 and Claude, which generated high-quality responses yet received low scores from overlap metrics. We attribute this to the highly diverse nature of LLM generations. However, we keep using the semantic similarity BERTscore for sanity and increased metric diversity.

Language Understanding Evaluation We measure the general language of the models with established benchmarks, such as the Massive Multitask Language Understanding (MMLU) (Hendrycks et al., 2020) MMLU is designed to assess how LLMs learn and apply knowledge across a variety of domains by measuring their performance on 57 different multiple-choice style question tasks from STEM, humanities, and world knowledge. This benchmark requires the model to be capable of general instruction following.

4 Results

Table 1 compares the performance of the baseline, unaligned model to the supervised fine-tuned model (SFT) and to the SFT model further aligned with a preference optimization algorithm (DPO).

SFT We observed a clear improvement across the board in empathy and similarity metrics provided by the SFT step. However, the more intensely the model is fine-tuned, the more it over-fits on the ED dataset, and the general performance drops, as shown with language understanding metrics in Figure 1. Mitigating this, but not avoiding it, is possible through careful hyperparameter optimization. First, as Tunstall et al., 2023 suggests, we limited the training to one epoch. The key to reducing over-fitting was using a large `lora_rank` (r) and low `lora_alpha` (α) relative to the rank. The large α supports retention of the original model's abilities, and the rank r controls the impact size of the adapter's fine-tuned weights. Figure 2 shows the empathy metrics saturating with $\alpha \approx \frac{r}{4}$. Furthermore, we observed that increasing the rank above 1024 brings diminishing returns: the computational efficiency of LoRA vanishes while the

Model	**MMLU (5s)** ↑	**diff-ER** ↓	**diff-EX** ↓	**diff-IP** ↓	**FBert** ↑
baseline: Zephyr-7B-sft-full	**.588** ± .00	0.861 ± .00	1.113 ± .00	1.431 ± .00	.804 ± .00
Zephyr-7B + SFT	.580 ± .00	0.888 ± .03	0.657 ± .03	0.750 ± .02	**.867** ± .00
Zephyr-7B + SFT + DPO	.572 ± .03	**0.660** ± .01	**0.627** ± .02	**0.648** ± .07	.728 ± .03

Table 1: Language understanding, empathy, and semantic similarity with EmpatheticDialogues. All presented results are means (± SD) of scores over multiple (4) training runs with the same hyperparameters but different seeds.

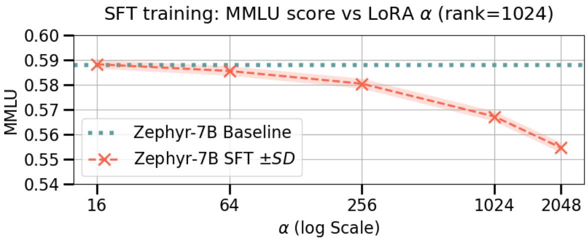

Figure 1: SFT: Impact of LoRA α on the MMLU score. Trained with: `learning_rate=1-e5, batch_size=64`.

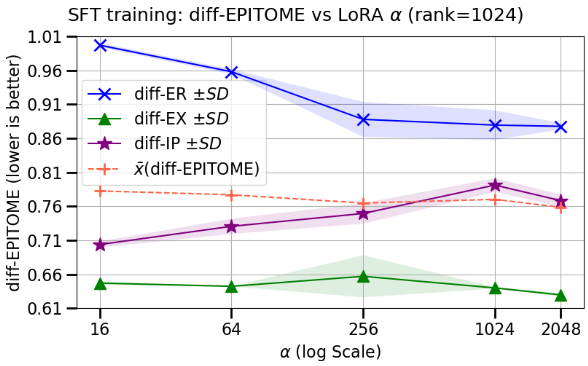

Figure 2: SFT: Impact of LoRA adapter rank on the diff-Epitome score during SFT. Trained with: `learning_rate=1-e5, batch_size=64`.

over-fitting problem remains. We use the best configuration jointly for empathy and generalization ($r = 1024, \alpha = 256, lr = 1e-5$) to report SFT results and to perform DPO experiments.

DPO The preference alignment with DPO improved the empathy metrics over the SFT while retaining the model's general performance. Nevertheless, its hyperparameters need to be set suitably to achieve this. Unsuitable hyperparameter configuration leads to training instability in addition to over-fitting (see Figure 3). Notably, we faced problems with the stability of the training introduced by the dataset construction process: random selection of rejected completions from the group of dialogues labeled with polar opposite emotions. We solved it by training for more epochs: re-drawing new randomly selected rejected completions for the same preferred ones for each epoch. The optimum configuration for high empathy scores while controlling over-fitting was training for three epochs, same as in Tunstall et al., 2023, with low β (0.025) but a relatively high learning rate (1e–5). We did not optimize BERTscore in SFT or DPO training, and results indicate no correlation with the diff-Epitome metric.

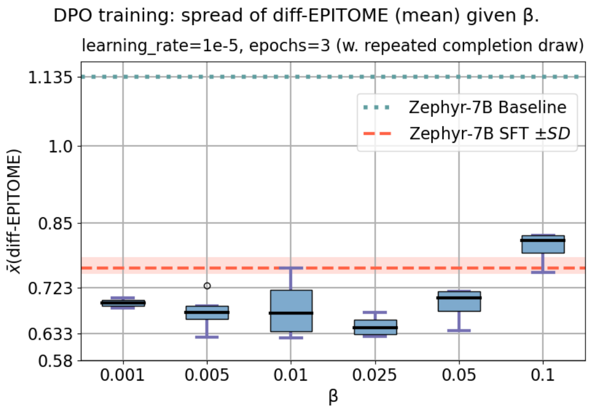

Figure 3: Stability of the DPO training measured by the spread of metric scores between multiple runs with the same hyperparameters. A smaller spread signifies higher training stability. We measure the mean and spread of the diff-Epitome metric, averaged across its three dimensions (ER, EX, IP: lower is better). We observe that setting $\beta >> 0.05$ leads to over-fitting on the preferred completions from the training set.

5 Conclusion

We introduced a new method for constructing preference datasets to align large language models (LLMs) for empathetic response generation while maintaining their overall language understanding capabilities. We validate our findings using the EmpatheticDialogues dataset and evaluate using the diff-Epitome and BERTscore metrics for empathy and standard language understanding benchmarks. Our method achieves the desired alignment of LLMs and provides a robust foundation for further enhancements, such as through prompt engineering. Furthermore, our method is directly applicable to other datasets with emotion labels.

Limitations

Empathetic Responses Lahnala et al., 2022 argue that most NLP research defines *empathy* loosely as understanding and appropriately responding to others' emotions, focusing mainly on detecting sentiment, emotions, or supportive interactions in text as indicators of empathy. This perspective assumes that systems achieving emotional recognition and responding in line with the target's sentiment are empathetic. However, this approach overlooks the critical aspect of cognitive empathy, which involves understanding another person's perspective, a gap highlighted by established human empathy theories (Debnath and Conlan, 2023).

Instead, we use a fully data-driven approach. We utilize the EmpatheticDialogs dataset as a source of the ground-truth empathetic responses. Even though there is criticism (Debnath and Conlan, 2023), dialogues in this dataset are written by human conversation partners and thus align well with real-world empathetic interactions.

References

Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, Nicholas Joseph, Saurav Kadavath, Jackson Kernion, Tom Conerly, Sheer El-Showk, Nelson Elhage, Zac Hatfield-Dodds, Danny Hernandez, Tristan Hume, Scott Johnston, Shauna Kravec, Liane Lovitt, Neel Nanda, Catherine Olsson, Dario Amodei, Tom Brown, Jack Clark, Sam McCandlish, Chris Olah, Ben Mann, and Jared Kaplan. 2022. Training a helpful and harmless assistant with reinforcement learning from human feedback.

Edward Beeching, Clémentine Fourrier, Nathan Habib, Sheon Han, Nathan Lambert, Nazneen Rajani, Omar Sanseviero, Lewis Tunstall, and Thomas Wolf. 2023. Open llm leaderboard. `https://huggingface.co/spaces/open-llm-leaderboard/open_llm_leaderboard`.

Guanqun Bi, Lei Shen, Yanan Cao, Meng Chen, Yuqiang Xie, Zheng Lin, and Xiaodong He. 2023. DiffusEmp: A diffusion model-based framework with multi-grained control for empathetic response generation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2812–2831, Toronto, Canada. Association for Computational Linguistics.

Rishi Bommasani, Percy Liang, and Tony Lee. 2023. Holistic evaluation of language models. *Annals of the New York Academy of Sciences*, 1525(1):140–146.

Sven Buechel, Anneke Buffone, Barry Slaff, Lyle Ungar, and João Sedoc. 2018. Modeling empathy and distress in reaction to news stories. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4758–4765, Brussels, Belgium. Association for Computational Linguistics.

Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower, Samuel Kim, Jeannette N Chang, Sungbok Lee, and Shrikanth S Narayanan. 2008. Iemocap: Interactive emotional dyadic motion capture database. *Language resources and evaluation*, 42:335–359.

Ankush Chatterjee, Kedhar Nath Narahari, Meghana Joshi, and Puneet Agrawal. 2019. SemEval-2019 task 3: EmoContext contextual emotion detection in text. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 39–48, Minneapolis, Minnesota, USA. Association for Computational Linguistics.

Yirong Chen, Xiaofen Xing, Jingkai Lin, Huimin Zheng, Zhenyu Wang, Qi Liu, and Xiangmin Xu. 2023. Soulchat: Improving llms' empathy, listening, and comfort abilities through fine-tuning with multi-turn empathy conversations.

Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. 2017. Deep reinforcement learning from human preferences. *Advances in neural information processing systems*, 30.

Alok Debnath and Owen Conlan. 2023. A critical analysis of empatheticdialogues as a corpus for empathetic engagement. In *Proceedings of the 2nd Empathy-Centric Design Workshop*, EMPATHICH '23, New York, NY, USA. Association for Computing Machinery.

Ning Ding, Yulin Chen, Bokai Xu, Yujia Qin, Zhi Zheng, Shengding Hu, Zhiyuan Liu, Maosong Sun, and Bowen Zhou. 2023. Enhancing chat language models by scaling high-quality instructional conversations.

Paul Ekman. 1992. Are there basic emotions?

Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. 2020. Measuring massive multitask language understanding. *arXiv preprint arXiv:2009.03300*.

Jonathan Herzig, Guy Feigenblat, Michal Shmueli-Scheuer, David Konopnicki, Anat Rafaeli, Daniel Altman, and David Spivak. 2016. Classifying emotions in customer support dialogues in social media. In *Proceedings of the 17th annual meeting of the special interest group on discourse and dialogue*, pages 64–73.

Chao-Chun Hsu, Sheng-Yeh Chen, Chuan-Chun Kuo, Ting-Hao Huang, and Lun-Wei Ku. 2018. Emotion-Lines: An emotion corpus of multi-party conversations. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*

(LREC 2018), Miyazaki, Japan. European Language Resources Association (ELRA).

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models.

Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. 2023. Mistral 7b. *arXiv preprint arXiv:2310.06825*.

Mojtaba Komeili, Kurt Shuster, and Jason Weston. 2022. Internet-augmented dialogue generation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8460–8478.

Allison Lahnala, Charles Welch, David Jurgens, and Lucie Flek. 2022. A critical reflection and forward perspective on empathy and natural language processing. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 2139–2158, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Young-Jun Lee, Chae-Gyun Lim, and Ho-Jin Choi. 2022a. Does GPT-3 generate empathetic dialogues? a novel in-context example selection method and automatic evaluation metric for empathetic dialogue generation. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 669–683, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Young-Jun Lee, Chae-Gyun Lim, and Ho-Jin Choi. 2022b. Does gpt-3 generate empathetic dialogues? a novel in-context example selection method and automatic evaluation metric for empathetic dialogue generation. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 669–683.

Qintong Li, Hongshen Chen, Zhaochun Ren, Pengjie Ren, Zhaopeng Tu, and Zhumin Chen. 2020. EmpDG: Multi-resolution interactive empathetic dialogue generation. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4454–4466, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Qintong Li, Piji Li, Zhaochun Ren, Pengjie Ren, and Zhumin Chen. 2022. Knowledge bridging for empathetic dialogue generation.

Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. 2017. DailyDialog: A manually labelled multi-turn dialogue dataset. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 986–995, Taipei, Taiwan. Asian Federation of Natural Language Processing.

Zaijing Li, Gongwei Chen, Rui Shao, Dongmei Jiang, and Liqiang Nie. 2024. Enhancing emotional generation capability of large language models via emotional chain-of-thought.

Chia-Wei Liu, Ryan Lowe, Iulian V Serban, Michael Noseworthy, Laurent Charlin, and Joelle Pineau. 2016. How not to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation. *arXiv preprint arXiv:1603.08023*.

Navonil Majumder, Pengfei Hong, Shanshan Peng, Jiankun Lu, Deepanway Ghosal, Alexander Gelbukh, Rada Mihalcea, and Soujanya Poria. 2020. MIME: MIMicking emotions for empathetic response generation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8968–8979, Online. Association for Computational Linguistics.

OpenAI. 2023. Chatgpt. Accessed: 2024-06-14.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744.

Robert Plutchik. 1980. A general psychoevolutionary theory of emotion. In *Theories of emotion*, pages 3–33. Elsevier.

Robert Plutchik. 2001. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist*, 89(4):344–350.

Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. 2019. MELD: A multimodal multi-party dataset for emotion recognition in conversations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 527–536, Florence, Italy. Association for Computational Linguistics.

Yushan Qian, Weinan Zhang, and Ting Liu. 2023. Harnessing the power of large language models for empathetic response generation: Empirical investigations and improvements. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 6516–6528, Singapore. Association for Computational Linguistics.

Huachuan Qiu, Hongliang He, Shuai Zhang, Anqi Li, and Zhenzhong Lan. 2024. Smile: Single-turn to multi-turn inclusive language expansion via chatgpt for mental health support.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2024. Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems*, 36.

Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2019. Towards empathetic open-domain conversation models: A new benchmark and dataset. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5370–5381, Florence, Italy. Association for Computational Linguistics.

Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Eric Michael Smith, Y-Lan Boureau, et al. 2021. Recipes for building an open-domain chatbot. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 300–325.

Minlie Huang Sahand Sabour, Chujie Zheng. 2021. Cem: Commonsense-aware empathetic response generation. *arXiv preprint arXiv:2109.05739*.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Ashish Sharma, Adam Miner, David Atkins, and Tim Althoff. 2020. A computational approach to understanding empathy expressed in text-based mental health support. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5263–5276.

Kurt Shuster, Jing Xu, Mojtaba Komeili, Da Ju, Eric Michael Smith, Stephen Roller, Megan Ung, Moya Chen, Kushal Arora, Joshua Lane, et al. 2022. Blenderbot 3: a deployed conversational agent that continually learns to responsibly engage. *arXiv preprint arXiv:2208.03188*.

Amy E Skerry and Rebecca Saxe. 2015. Neural representations of emotion are organized around abstract event features. *Current biology*, 25(15):1945–1954.

Lewis Tunstall, Edward Beeching, Nathan Lambert, Nazneen Rajani, Kashif Rasul, Younes Belkada, Shengyi Huang, Leandro von Werra, Clémentine Fourrier, Nathan Habib, Nathan Sarrazin, Omar Sanseviero, Alexander M. Rush, and Thomas Wolf. 2023. Zephyr: Direct distillation of lm alignment.

Lanrui Wang, Jiangnan Li, Chenxu Yang, Zheng Lin, Hongyin Tang, Huan Liu, Xiaolei Huang, Yanan Cao, Jingang Wang, and Weiping Wang. 2024. Sibyl: Sensible empathetic dialogue generation with visionary commonsense knowledge.

Joseph Weizenbaum. 1966. Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45.

Anuradha Welivita, Yubo Xie, and Pearl Pu. 2021. A large-scale dataset for empathetic response generation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1251–1264.

Jing Xu, Arthur Szlam, and Jason Weston. 2022. Beyond goldfish memory: Long-term open-domain conversation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5180–5197.

A Training Details

A.1 Model Settings

We trained the models using the HuggingFace Transformers library. In the SFT step, we applied standard sequence-to-sequence training with **cross-entropy** loss on tokens from the user prompts (not on the tokens from "replies"). In both the SFT and DPO training steps, we optimized the model with **AdamW** optimizer and effective **batch size** of 64. We used **learning rate** of 1e-5 without **warmup** and a cosine **lr decay**. The models were trained in bf16 **precision**.

A.2 Hardware

To train our models, we used two NVIDIA A100 80GB GPUs. We trained LoRA adapters with ranks ranging from 16 to 2048 for models with 7BM parameters. The total training wall time, including preliminary experiments, was 17 days.

A.3 DPO Preference Dataset

For training with DPO, we introduced a method to construct a preference dataset from the EmpatheticDialogues dataset. With each training run or a single epoch within a multi-epoch run, the dataset is created anew as described in Section 3. This section contains auxiliary material to support the description of the preference dataset creation.

Each dialog in the EmpatheticDialogues dataset is associated with an emotion label. To construct the training example for DPO, we paired the preferred dialog completion (in our case, using the ground truth) with one rejected competition. To find the rejected completion, we proposed using opposite emotion labels. The opposites are based on two theory-based sources: Plutchik's wheel (Figure 4), emotional dyads (Figure 5), and we proposed the rest ourselves. The opposites form the lookup table 2, which is queried each time a preferred/rejected pair is constructed.

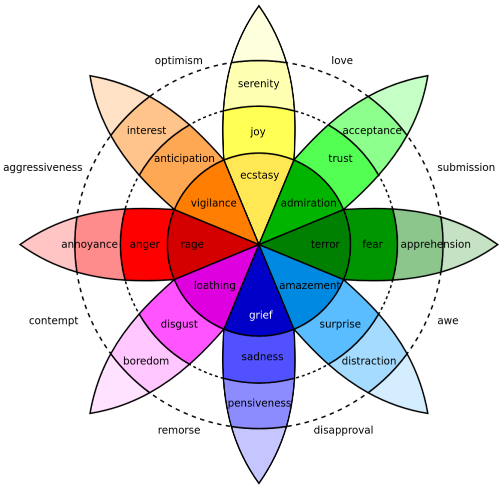

Figure 4: Plutchik's wheel of emotions (Creative Commons). It shows eight basic emotions: joy, trust, fear, surprise, sadness, anticipation, anger, and disgust. The wheel of emotions groups these eight basic emotions based on the physiological purpose of each into polar coordinates reflecting their similarity and intensity.

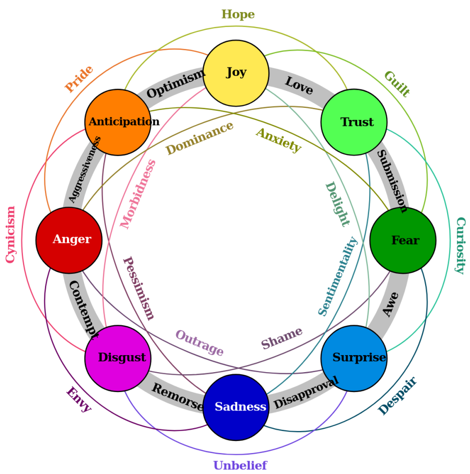

Figure 5: Plutchik's emotion's dyads (Creative Commons). When two emotions are elicited together, they form the primary dyad. If they are one petal apart, it is a secondary dyad; if they are two petals distant from each other, it is a tertiary dyad. Opposite dyads are on the opposite side.

Emotion label	**Opposite label**	**Source**
afraid	angry	wheel
angry	afraid	wheel
sad	joyful	wheel
grateful	disgusted	wheel
surprised	anticipating	wheel
trusting	disgusted	wheel
disgusted	trusting	wheel
anticipating	surprised	wheel
content	anxious	wheel
apprehensive	annoyed	wheel
joyful	sad	wheel
proud	ashamed	dyads
prepared	anxious	dyads
ashamed	proud	dyads
guilty	proud	dyads
nostalgic	hopeful	dyads
anxious	content	dyads
hopeful	nostalgic	dyads
sentimental	apprehensive	
jealous	faithful	
embarrassed	confident	
excited	devastated	
annoyed	apprehensive	
lonely	caring	
faithful	jealous	
terrified	furious	
confident	embarrassed	
furious	terrified	
disappointed	impressed	
caring	lonely	
impressed	disappointed	
devastated	excited	

Table 2: The opposite emotion labels lookup table. The emotion labels are sourced from the EmpatheticDialogues dataset. Each pair of opposites is associated with how the pair was determined: using Plutchiks's wheel, emotion dyads, or our proposal.