

AURADIAL: A Large-Scale Human-Centric Dialogue Dataset for Chinese AI Psychological Counseling

Xiantao Zhang

Beihang University

zhangxiantao@buaa.edu.cn

Abstract

This paper introduces AURADIAL, a large-scale, human-centric dialogue dataset for Chinese AI psychological counseling, comprising over 300,000 single-turn dialogues and 90,000 multi-turn dialogue sessions. A key distinction of AURADIAL is its instruction set, primarily derived from real-world user queries, better reflecting genuine expression patterns compared to synthetic or template-based alternatives. Furthermore, we propose an innovative rephrasing-based data generation methodology designed to foster more human-like and empathetic responses, addressing a common shortcoming in AI-generated dialogue. Experimental results demonstrate that models fine-tuned on AURADIAL significantly outperform those trained on other public datasets in generating empathetic and relevant replies. AURADIAL offers a novel, valuable resource to the Chinese NLP community for advancing AI in psychological counseling.

1 Introduction

Mental health is a growing global public health concern (Organization, 2022; Huang et al., 2019). Artificial intelligence shows significant potential in providing accessible, scalable psychological support (Chen et al., 2024c), and integrating such support into conversational AI has profound social significance (Li et al., 2023b; He et al., 2023; Zheng et al., 2023). Developing culturally aware AI counseling resources is crucial, particularly for the Chinese context, where high-quality datasets are foundational but scarce (Sun et al., 2021; Zhang et al., 2024; Chen et al., 2025).

To address this scarcity, the field has increasingly relied on synthetic data generation using large language models (LLMs) (Wang et al., 2023; Taori et al., 2023; Ding et al., 2023). While this approach enables the creation of large-scale datasets, it has introduced a distinct set of challenges that limit the

development of genuinely effective AI counselors. A primary issue lies in the authenticity and diversity of the generated content. Instructions are often synthetic or template-based, failing to capture the complexity, emotional nuance, and ambiguity characteristic of real-world help-seeking expressions (Chen et al., 2024b; Zeng et al., 2024). Consequently, AI-generated responses often lack deep empathy, exhibiting an impersonal quality commonly described as an “*AI feel*” that can alienate users and erode trust (Qu et al., 2023; Liu et al., 2025). This reliance on purely synthetic dialogues risks creating an echo chamber, where models learn the stylistic biases of their predecessors rather than the patterns of genuine human interaction, leading to limited thematic coverage and poor generalization to real-world scenarios (Shumailov et al., 2024; Alemohammad et al., 2024).

Our Contributions To address these gaps, we present AURADIAL, a large-scale, human-centric dialogue dataset for Chinese psychological counseling, with over 300,000 single-turn and over 90,000 multi-turn dialogues. Our key contributions are:

- **Authentic and Diverse Instructions:** AURADIAL’s instructions are primarily derived from real-world user queries, better reflecting genuine expression patterns and diverse real-world scenarios compared to synthetic counterparts.
- **Innovative Empathy-Enhancing Methodology:** We propose a novel rephrasing-based synthesis pipeline to generate more human-like, empathetic replies, validated by experiments showing superior performance (Section 4.3).
- **Broader Topic Coverage:** Our dataset spans a significantly broader range of themes than existing resources, offering a novel asset for the community (Figure 1).

2 Related Work

2.1 Chinese Counseling Datasets

Recently, several notable Chinese psychological counseling datasets have emerged in the field. [Chen et al. \(2023\)](#) constructed the *SoulChatCorpus*, which is designed to enhance models’ empathy, listening, and comforting abilities through multi-stage processing and transformations by large language models (LLMs). [Zhang et al. \(2024\)](#) proposed *CPsyCounD*, a dataset distinguished by its reconstruction of multi-turn dialogues from real psychological counseling reports. These dialogues cover a wide range of topics and counseling approaches and were generated using the *Memo2Demo* method. [Qiu et al. \(2024\)](#) expanded the single-turn Q&A dataset *PsyQA* into *SMILECHAT*, a large-scale multi-turn psychological support dialogue dataset. Similarly, [Qiu and Lan \(2024\)](#) explored a paradigm in *SimPsyDial* where two LLMs engage in role-playing as counselor and client to simulate counseling dialogues. Collectively, these works offer valuable data resources and diverse construction methodologies for advancing research in AI-driven Chinese psychological counseling.

2.2 Synthetic Dialogue Data

Parallel work scales instruction–response data with minimal human labor. [Wang et al. \(2023\)](#) introduce bootstrapped instruction synthesis from a small human seed, and [Taori et al. \(2023\)](#) demonstrate that lightweight, model-generated instruction corpora can effectively tune general models. For multi-turn settings, [Ding et al. \(2023\)](#) synthesize large conversational datasets via iterative prompting and role-playing continuations.

Beyond coverage, recent work explores *how* to shape supportive tone. Empathy-focused rewriting edits low-empathy drafts into warmer, validating replies while preserving semantic intent ([Sharma et al., 2022](#); [Qu et al., 2023](#)). At the same time, the community warns that training predominantly on model-generated corpora can induce distributional drift and recursive degradation ([Shumailov et al., 2024](#); [Alemohammad et al., 2024](#)). Informed by these lessons, we use *rephrasing* to steer style toward human-centric empathy, while grounding the instruction side in real-world queries to mitigate feedback-loop artifacts and improve robustness to genuine counseling scenarios.

Dataset	Samp.	Tot.	I.L.	R.L.
SoulChat(2023)	258K	200M	41.4	90.0
CPsyCounD(2024)	3.1K	2.2M	33.7	56.0
SMILE(2024)	66K	50M	34.1	85.8
SimPsyDial(2024)	1K	1.6M	54.1	70.8
Ours	399K	447M	252.2	654.2

Table 1: Comparative statistics of AuraDial against other prominent open-source Chinese psychological counseling datasets. Samp.: Samples; Tot.: Total chars; I.L. & R.L.: Avg. instruction & response length (chars).

3 The AURADIAL Dataset

3.1 Dataset Composition and Statistics

AURADIAL comprises over 300,000 single-turn dialogues and over 90,000 multi-turn dialogue sessions. Table 1 compares AURADIAL with other Chinese psychological counseling datasets, underscoring its significant advantages in scale. Notably, the substantially longer average instruction length (I.L.) in our dataset reflects the authenticity of the collected user queries, which are more detailed and complex than synthetic ones. Illustrative examples are provided in Appendix D.

To illustrate AURADIAL’s topic diversity, we randomly sampled up to 40K instructions from our dataset and from each comparative dataset. We then obtained embeddings for these instructions using *bge-m3* ([Chen et al., 2024a](#)) and visualized them with UMAP ([McInnes et al., 2020](#)) (Figure 1). The visualization reveals that AURADIAL’s instructions occupy embedding space regions distinct from those of existing datasets, indicating broader topic coverage and novel thematic areas for the research community. This visual evidence of diversity is further supported by a quantitative topic modeling analysis. A detailed list of the top-20 identified themes is provided in Appendix B.

3.2 Instruction Sourcing and Ethics

Instruction Data Sources High-quality instructions are foundational. Unlike many existing datasets that rely on manually designed or template-based queries, AURADIAL prioritizes authentic instructions. Real user queries are typically longer, more colloquial, emotionally complex, and occasionally ambiguous (see Appendix D, Fig. 4). AURADIAL’s instructions are primarily sourced from real user queries on public Chinese psychological counseling websites. We also integrated authentic data from open-source datasets (e.g., 3.3% from SoulChat ([Chen et al., 2023](#))) and enhanced di-

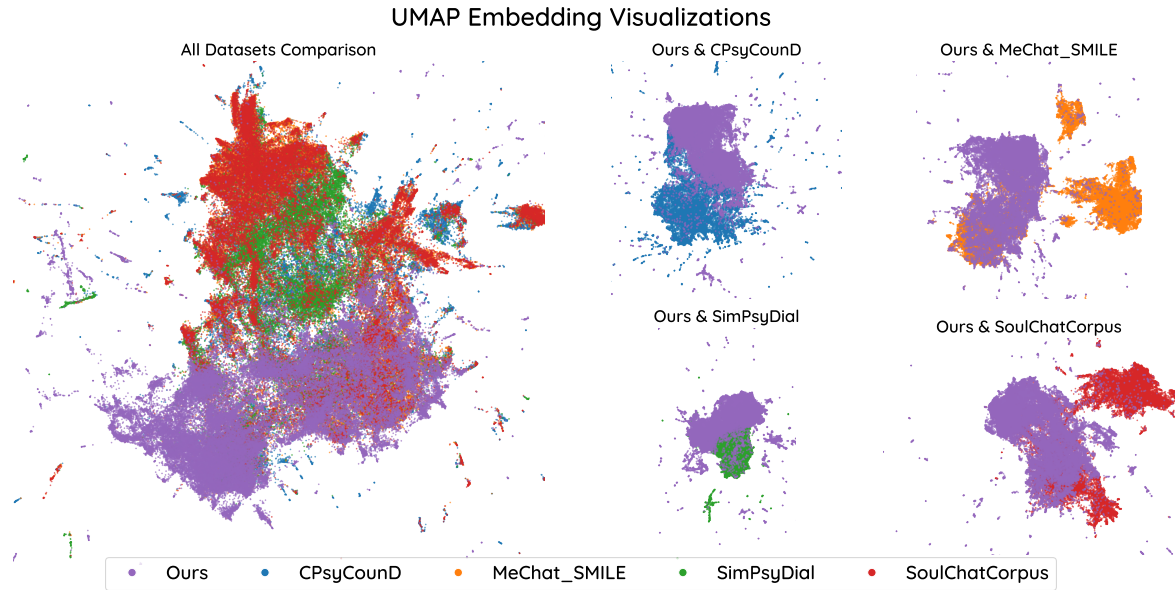


Figure 1: UMAP visualization of instruction embeddings from AURADIAL and four other datasets, illustrating their thematic distribution. Each point represents an instruction embedding. The distinct clustering of AURADIAL (in purple) highlights its unique and broader topic coverage compared to the more concentrated distributions of existing datasets. A random sample of up to 40K instructions per dataset was used for this visualization.

versity by translating, then manually proofreading and culturally adapting, high-quality non-Chinese datasets like `psychologist_answers` (2023).

Ethical Considerations and De-identification

Ethical considerations and user privacy were paramount throughout dataset construction.

1) Source Compliance: Data was sourced only from public websites or explicitly open-licensed datasets. **2) Anonymization:** Personally identifiable information (PII) in web-sourced data was rigorously removed via automated scripts and manual review. Open-source data underwent secondary heuristic checks. **3) Content Filtering:** To balance authenticity and suitability, we used Qwen2.5-32B-Instruct (Qwen et al., 2025) to screen the raw data, removing inappropriate content while preserving genuine help-seeking queries. **4) Manual Review:** Final spot checks ensured the dataset’s appropriateness, anonymity, and quality.

3.3 Data Construction Pipeline

As shown in Figure 2, the data construction process aims to generate large-scale, high-quality, and highly empathetic dialogue data.

(a) Initial Instruction Pool Construction Real-world instructions from public sources, after de-identification and content filtering, formed the basis of an authentic and diverse initial instruction pool.

(b) Rephrasing-based Response Synthesis for a High-Quality Seed Pool Our core methodological innovation stems from a key observation: rephrasing existing responses, rather than direct generation from a prompt, more effectively elicits model empathy (Qu et al., 2023; Sharma et al., 2022). We hypothesize this is because the rephrasing task constrains the model’s output space, forcing it to focus on mastering expressive nuance and emotional attunement rather than content invention.

This deliberate choice to synthesize counselor responses, rather than including scraped responses from real counselors, was driven by two primary considerations. First, sourcing real counselor responses without explicit licenses poses significant ethical and copyright challenges. Second, real-world online responses vary in quality and style. Our pipeline functions as an expert-guided quality control mechanism, ensuring all responses in AURADIAL are uniformly empathetic and suitable for training a dependable AI assistant. The scarcity of publicly available, high-quality human counselor data is the precise gap our work addresses with a scalable and ethically sound methodology.

Initially, a diverse set of state-of-the-art (SOTA) models (DeepSeek-V2.5 (DeepSeek, 2024), GLM-Z1-32B (GLM et al., 2024), Qwen2.5-72B-Instruct (Qwen et al., 2025)) generated varied base responses. Then, a more powerful model, DeepSeek-

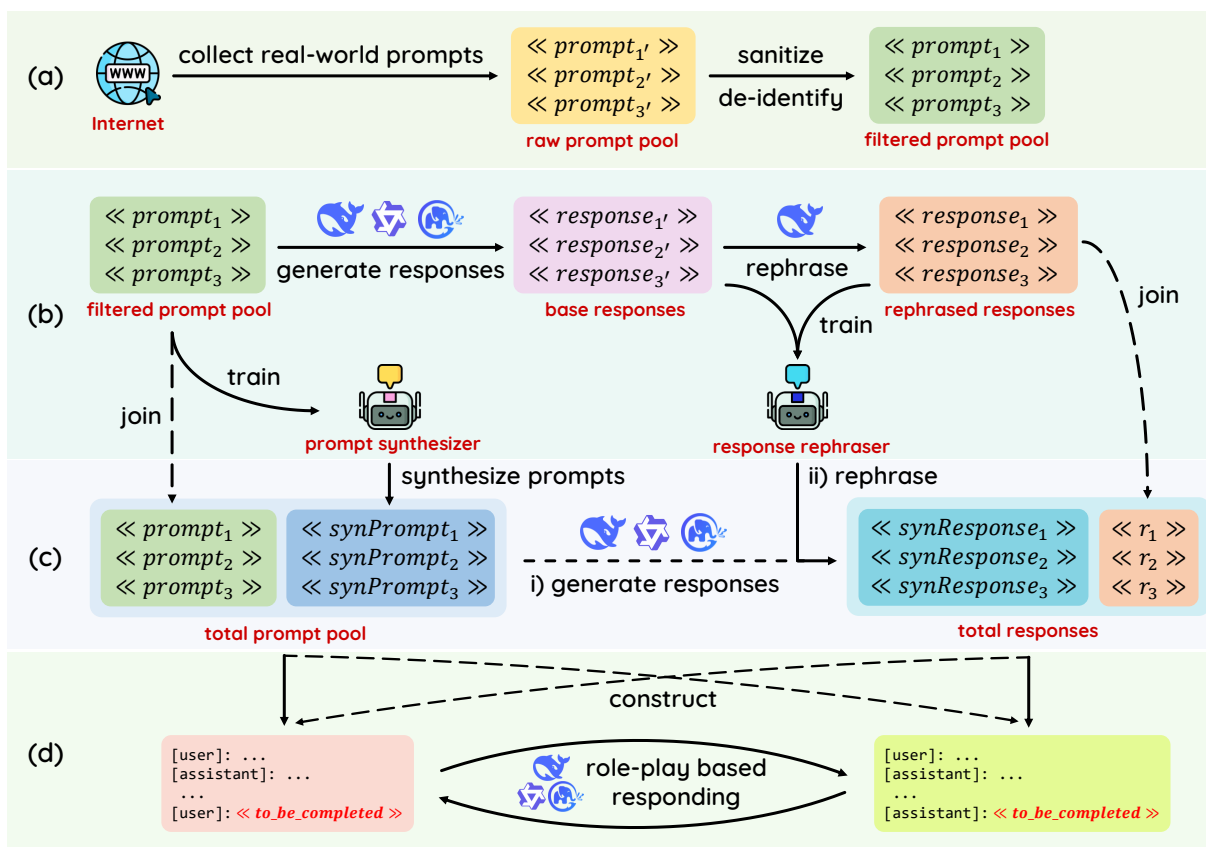


Figure 2: The data construction pipeline for AURADIAL. The process flows from sourcing authentic queries to a core rephrasing-based synthesis stage, which generates a seed pool used to train models for large-scale single-turn and multi-turn data expansion.

V3-0324 (DeepSeek-AI et al., 2025b), refined these using meticulous rephrasing prompts, creating a seed pool of over 30K high-quality instruction-rephrased response pairs.

These pairs exhibited stability, high quality, and low stylistic correlation with the initial response generator. The prompt is shown in Figure 7.

From this seed pool, we trained two auxiliary models based on Qwen2.5-14B-Instruct (Qwen et al., 2025), chosen for its balance of capability and training efficiency:

- **Prompt Synthesizer:** Learns from seed pool instructions to generate diverse, realistic user queries.
- **Response Rephraser:** Trained on the seed pool’s initial responses (input) and rephrased responses (output) to specialize in enhancing reply empathy and naturalness.

(c) Synthetic Data Expansion Auxiliary models expanded the dataset. The Prompt Synthesizer generated new instructions, deduplicated against

the existing pool via prefix/suffix matching and n-gram overlap. For these instructions, the models listed in Section 3.3 first produced initial responses, which our *Response Rephraser* then rewrote into final single-turn replies.

(d) Role-play-based Multi-turn Dialogue Synthesis High-quality single-turn data was extended to multi-turn dialogues. We found that capable dialogue models, given context and explicit role instructions via simple prompts, can effectively continue conversations while maintaining role consistency. This extended single-turn data into multi-turn dialogues, maintaining coherence and thematic consistency. The prompt is shown in Figure 8.

4 Experiments

We conducted comparative experiments to validate AURADIAL’s effectiveness, particularly for enhancing model empathy.

4.1 Experimental Setup

Qwen2.5-7B-Instruct served as the base model for all fine-tuning experiments. We sampled 30K dia-

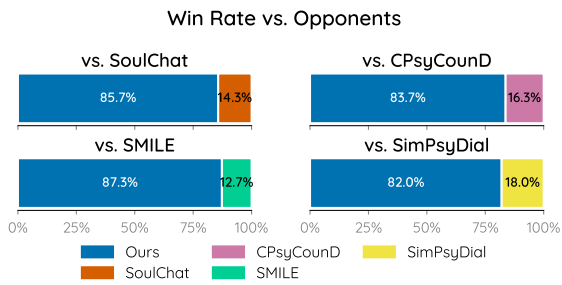


Figure 3: Pairwise win rates of the model fine-tuned on AURADIAL against models fine-tuned on four other datasets. Results are based on a blind evaluation of 100 authentic test queries, judged by three SOTA LLMs. The results demonstrate a consistent and significant performance advantage for the AURADIAL-trained model.

logues each from AURADIAL and four other public Chinese psychological counseling datasets. We fixed the training set size to ensure a controlled comparison; because AURADIAL is substantially larger than some alternatives, standardizing the size helps isolate data quality from quantity. The base model was then fine-tuned on each sampled dataset using LoRA (Hu et al., 2021), yielding five distinct models. All experiments used identical hardware for fairness.

4.2 Evaluation Metrics

To ensure an unbiased assessment of real-world performance, we curated an independent test set of 100 authentic user queries, sampled from public platforms and unseen during any training stage.

Each of the five fine-tuned models generated responses to these 100 queries. For each query, the response from the AURADIAL-trained model was paired against those from models trained on the other four datasets. While acknowledging that human evaluation is the gold standard, we designed a rigorous LLM-based protocol to serve as a strong and scalable alternative. To mitigate single-judge bias, a known risk in LLM-based evaluation, we employed a panel of three distinct, SOTA LLMs as judges: GPT-4o (OpenAI et al., 2024), DeepSeek-R1 (DeepSeek-AI et al., 2025a), and Qwen2.5-Max (Qwen et al., 2025). The response pairs were shuffled and anonymized before being presented to the judges, who performed a blind, pairwise comparison based on empathy, problem understanding, relevance, and helpfulness, indicating which was superior or if they were comparable (judge prompt in Appendix, Fig. 9).

Win Rate Calculation For each pair, if a judge preferred the AURADIAL-trained model’s response, it scored 1 point (opponent 0), and vice-versa. Because of the prompt design, ties could not occur. Each of the three judges contributed $\frac{1}{3}$ to the pair’s total score. We then calculated the AURADIAL-trained model’s average win rate against each of the other four models.

4.3 Results and Analysis

As shown in Figure 3, the model trained on AURADIAL achieved consistently higher win rates against all models trained on other datasets. This demonstrates that fine-tuning on AURADIAL more effectively improves model performance in psychological counseling contexts, especially regarding empathy and user understanding. Since the same base model and fine-tuning methodology were used for all conditions, these findings directly validate the superior quality of our dataset and the effectiveness of our construction strategy, particularly its use of authentic instructions and the rephrasing-based response generation pipeline.

5 Conclusion

This paper introduced AURADIAL, a large-scale, human-centric dataset for Chinese AI psychological counseling, designed to overcome key limitations in existing resources: inauthentic instructions and synthetic responses that lack empathy. Our work makes two primary contributions: we ground our dataset in a large and diverse set of real-world user queries, and we propose a novel rephrasing-based synthesis pipeline to generate counselor responses. This methodology deliberately constrains the generation task to focus on enhancing stylistic and emotional nuance, proving more effective for cultivating empathetic style than direct generation.

Our experimental results robustly validate this approach. In blind, pairwise evaluations, a model fine-tuned on AURADIAL significantly outperformed models trained on four other public datasets, demonstrating its superior ability to generate empathetic and relevant replies. By providing both a high-quality, large-scale dataset and a principled methodology for its creation, we believe AURADIAL will serve as a foundational resource for the community. It paves the way for the development of more nuanced, trustworthy, and genuinely supportive AI counseling systems, ultimately contributing to more human-centric conversational AI.

Limitations

Although AURADIAL represents a significant advance, its limitations stem from its data sourcing, composition, and synthetic nature. First, the dataset’s instructions are sourced from public online platforms, which may introduce demographic and topical biases and may not fully represent the entire spectrum of psychological counseling needs. The inclusion of translated data, despite careful proofreading, might also introduce subtle cultural or emotional nuances not present in the original text. Consequently, while our analysis shows broad coverage, niche scenarios or specific psychological disorders may remain underrepresented.

Second, further limitations arise from our data generation and evaluation methodologies. The counselor responses, though enhanced through our innovative rephrasing-based and role-playing strategies, are synthetic. As such, they may lack the strategic depth and improvisational skill of human counselors and could contain model-specific artifacts. Furthermore, our evaluation relies on LLM judges, which, while scalable and robust, primarily assess text-level empathetic expression. This serves as a proxy but cannot fully capture the complex, multifaceted nature of how humans perceive empathy. Future work should diversify data sources, incorporate human-written responses, and develop more comprehensive evaluation frameworks that include direct human feedback.

Ethics Statement

The development and release of the AURADIAL dataset were guided by a strong commitment to ethical principles, focusing on user privacy, data integrity, and responsible AI application in the sensitive domain of mental health. Our ethical framework centered on meticulous data curation and robust privacy protection. All data were sourced exclusively from publicly accessible platforms or explicitly open-licensed datasets. We implemented a rigorous de-identification process, combining automated scripts with manual review, to remove personally identifiable information (PII). Furthermore, all raw data underwent automated content screening and manual spot-checks to filter for inappropriate or harmful content while preserving the authenticity of genuine help-seeking queries.

We acknowledge that the dataset is not without inherent biases and risks, which necessitate its responsible use. Potential biases may arise from the

online platforms, the translation process, and the LLMs used for generation. Researchers should consider these factors when interpreting model performance. Most importantly, AURADIAL is intended for research purposes only. Models trained on this dataset are not a substitute for professional human counselors and should not be deployed in real-world applications without rigorous testing, safety mechanisms, and human oversight. There remains a risk of models generating inappropriate or harmful advice, particularly in critical situations.

To foster transparency and collaborative innovation, AURADIAL will be made openly available to the research community under a license explicitly restricted to non-commercial use. Access will require agreeing to a data usage agreement that mandates adherence to the ethical best practices outlined in this statement. The authors are committed to maintaining transparency regarding the dataset’s construction and limitations and welcome feedback from the community to guide its responsible evolution and application.

References

- Sina Alemohammad, Josue Casco-Rodriguez, Lorenzo Luzi, Ahmed Imtiaz Humayun, Hossein Babaei, Daniel LeJeune, Ali Siahkoochi, and Richard Baraniuk. 2024. [Self-consuming generative models go MAD](#). In *The Twelfth International Conference on Learning Representations*.
- Nicolas Bertagnolli. 2020. Counsel chat: Bootstrapping high-quality therapy data.
- Jianlv Chen, Shitao Xiao, Peitian Zhang, Kun Luo, Defu Lian, and Zheng Liu. 2024a. [Bge m3-embedding: Multi-lingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation](#). *Preprint*, arXiv:2402.03216.
- Keqi Chen, Zekai Sun, Yuhua Wen, Huijun Lian, Yingming Gao, and Ya Li. 2025. [Psy-insight: Explainable multi-turn bilingual dataset for mental health counseling](#). *Preprint*, arXiv:2503.03607.
- Lichang Chen, Shiyang Li, Jun Yan, Hai Wang, Kalpa Gunaratna, Vikas Yadav, Zheng Tang, Vijay Sriniwasan, Tianyi Zhou, Heng Huang, and Hongxia Jin. 2024b. [Alpagasus: Training a better alpaca with fewer data](#). *Preprint*, arXiv:2307.08701.
- 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](#). In *Findings of the Association for Computational Linguistics: EMNLP 2023*,

- pages 1170–1183, Singapore. Association for Computational Linguistics.
- Yixiang Chen, Xinyu Zhang, Jinran Wang, Xurong Xie, Nan Yan, Hui Chen, and Lan Wang. 2024c. **Structured dialogue system for mental health: An llm chatbot leveraging the pm+ guidelines.** *Preprint*, arXiv:2411.10681.
- DeepSeek. 2024. **Deepseek-v2.5: A new open-source model combining general and coding capabilities.**
- DeepSeek-AI, Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, Xiaokang Zhang, Xingkai Yu, Yu Wu, Z. F. Wu, Zhibin Gou, Zhichong Shao, Zhuoshu Li, Ziyi Gao, and 181 others. 2025a. **Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning.** *Preprint*, arXiv:2501.12948.
- DeepSeek-AI, Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, Damai Dai, Daya Guo, Dejian Yang, Deli Chen, Dongjie Ji, Erhang Li, Fangyun Lin, Fucong Dai, and 181 others. 2025b. **Deepseek-v3 technical report.** *Preprint*, arXiv:2412.19437.
- 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.** *Preprint*, arXiv:2305.14233.
- Yann Dubois, Chen Xuechen Li, Rohan Taori, Tianyi Zhang, Ishaan Gulrajani, Jimmy Ba, Carlos Guestrin, Percy S Liang, and Tatsunori B Hashimoto. 2023. **AlpacaFarm: A simulation framework for methods that learn from human feedback.** In *Advances in Neural Information Processing Systems*, volume 36, pages 30039–30069. Curran Associates, Inc.
- Team GLM, :, Aohan Zeng, Bin Xu, Bowen Wang, Chenhui Zhang, Da Yin, Dan Zhang, Diego Rojas, Guanyu Feng, Hanlin Zhao, Hanyu Lai, Hao Yu, Hongning Wang, Jiadao Sun, Jijie Zhang, Jiale Cheng, Jiayi Gui, Jie Tang, and 40 others. 2024. **Chatglm: A family of large language models from glm-130b to glm-4 all tools.** *Preprint*, arXiv:2406.12793.
- Yuhao He, Li Yang, Chunlian Qian, Tong Li, Zhengyuan Su, Qiang Zhang, and Xiangqing Hou. 2023. **Conversational agent interventions for mental health problems: Systematic review and meta-analysis of randomized controlled trials.** *J Med Internet Res*, 25:e43862.
- 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.** *Preprint*, arXiv:2106.09685.
- Yueqin Huang, Yu Wang, Hong Wang, Zhaorui Liu, Xin Yu, Jie Yan, Yaqin Yu, Changgui Kou, Xiufeng Xu, Jin Lu, Zhizhong Wang, Shulan He, Yifeng Xu, Yanling He, Tao Li, Wanjun Guo, Hongjun Tian, Guangming Xu, Xiangdong Xu, and 21 others. 2019. **Prevalence of mental disorders in china: a cross-sectional epidemiological study.** *The Lancet Psychiatry*, 6(3):211–224.
- Guohao Li, Hasan Hammoud, Hani Itani, Dmitrii Khizbullin, and Bernard Ghanem. 2023a. **Camel: Communicative agents for "mind" exploration of large language model society.** In *Advances in Neural Information Processing Systems*, volume 36, pages 51991–52008. Curran Associates, Inc.
- Han Li, Renwen Zhang, Yi-Chieh Lee, Robert E. Kraut, and David C. Mohr. 2023b. **Systematic review and meta-analysis of ai-based conversational agents for promoting mental health and well-being.** *npj Digital Medicine*, 6(1):236.
- Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. 2021. **Towards emotional support dialog systems.** In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3469–3483, Online. Association for Computational Linguistics.
- Tingting Liu, Salvatore Giorgi, Ankit Aich, Allison Lahnala, Brenda Curtis, Lyle Ungar, and João Sedoc. 2025. **The illusion of empathy: How ai chatbots shape conversation perception.** *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(13):14327–14335.
- Ganeshan Malhotra, Abdul Waheed, Aseem Srivastava, Md Shad Akhtar, and Tanmoy Chakraborty. 2022. **Speaker and time-aware joint contextual learning for dialogue-act classification in counselling conversations.** In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining, WSDM '22*, page 735–745, New York, NY, USA. Association for Computing Machinery.
- Leland McInnes, John Healy, and James Melville. 2020. **Umap: Uniform manifold approximation and projection for dimension reduction.** *Preprint*, arXiv:1802.03426.
- OpenAI, :, Aaron Hurst, Adam Lerer, Adam P. Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, Aleksander Madry, Alex Baker-Whitcomb, Alex Beutel, Alex Borzunov, Alex Carney, Alex Chow, Alex Kirillov, and 401 others. 2024. **Gpt-4o system card.** *Preprint*, arXiv:2410.21276.
- World Health Organization. 2022. **World mental health report: Transforming mental health for all.** <https://www.who.int/publications/i/item/9789240049338>.
- 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. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 615–636, Miami, Florida, USA. Association for Computational Linguistics.
- Huachuan Qiu and Zhenzhong Lan. 2024. Interactive agents: Simulating counselor-client psychological counseling via role-playing llm-to-llm interactions. *Preprint*, arXiv:2408.15787.
- Renyi Qu, Lyle Ungar, and João Sedoc. 2023. Conditioning on dialog acts improves empathy style transfer. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 13254–13271, Singapore. Association for Computational Linguistics.
- Qwen, :, An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiayi Yang, Jingren Zhou, and 25 others. 2025. *Qwen2.5 technical report*. *Preprint*, arXiv:2412.15115.
- Rogozinushka. 2023. *Psychologist answers dataset*. Hugging Face Dataset.
- Ashish Sharma, Inna W. Lin, Adam S. Miner, Dave C. Atkins, and Tim Althoff. 2022. Towards facilitating empathic conversations in online mental health support: A reinforcement learning approach (extended abstract). In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 5339–5343. International Joint Conferences on Artificial Intelligence Organization. Sister Conferences Best Papers.
- Iliia Shumailov, Zakhar Shumaylov, Yiren Zhao, Yarin Gal, Nicolas Papernot, and Ross Anderson. 2024. The curse of recursion: Training on generated data makes models forget. *Preprint*, arXiv:2305.17493.
- Hao Sun, Zhenru Lin, Chujie Zheng, Siyang Liu, and Minlie Huang. 2021. *PsyQA: A Chinese dataset for generating long counseling text for mental health support*. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1489–1503, Online. Association for Computational Linguistics.
- Rohan Taori, Ishaan Gulrajani, Tianyi Zhang, Yann Dubois, Xuechen Li, Carlos Guestrin, Percy Liang, and Tatsunori B. Hashimoto. 2023. Stanford alpaca: An instruction-following llama model. https://github.com/tatsu-lab/stanford_alpaca.
- EmoLLM Team. 2024. Emollm: Reinventing mental health support with large language models. <https://github.com/SmartFlowAI/EmoLLM>.
- Yizhong Wang, Yeganeh Kordi, Swaroop Mishra, Alisa Liu, Noah A. Smith, Daniel Khashabi, and Hannaneh Hajishirzi. 2023. Self-instruct: Aligning language models with self-generated instructions. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13484–13508, Toronto, Canada. Association for Computational Linguistics.
- Haojie Xie, Yirong Chen, Xiaofen Xing, Jingkai Lin, and Xiangmin Xu. 2024. *Psydt: Using llms to construct the digital twin of psychological counselor with personalized counseling style for psychological counseling*. *Preprint*, arXiv:2412.13660.
- Can Xu, Qingfeng Sun, Kai Zheng, Xiubo Geng, Pu Zhao, Jiazhan Feng, Chongyang Tao, Qingwei Lin, and Daxin Jiang. 2025. *Wizardlm: Empowering large pre-trained language models to follow complex instructions*. *Preprint*, arXiv:2304.12244.
- Congchi Yin, Feng Li, Shu Zhang, Zike Wang, Jun Shao, Piji Li, Jianhua Chen, and Xun Jiang. 2025. *Mdd-5k: A new diagnostic conversation dataset for mental disorders synthesized via neuro-symbolic llm agents*. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(24):25715–25723.
- Weihao Zeng, Can Xu, Yingxiu Zhao, Jian-Guang Lou, and Weizhu Chen. 2024. *Automatic instruction evolving for large language models*. *Preprint*, arXiv:2406.00770.
- Chenhao Zhang, Renhao Li, Minghuan Tan, Min Yang, Jingwei Zhu, Di Yang, Jiahao Zhao, Guancheng Ye, Chengming Li, and Xiping Hu. 2024. *CPsyCoun: A report-based multi-turn dialogue reconstruction and evaluation framework for Chinese psychological counseling*. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 13947–13966, Bangkok, Thailand. Association for Computational Linguistics.
- Zhonghua Zheng, Lizi Liao, Yang Deng, and Liqiang Nie. 2023. *Building emotional support chatbots in the era of llms*. *Preprint*, arXiv:2308.11584.
- Chunting Zhou, Pengfei Liu, Puxin Xu, Srinivasan Iyer, Jiao Sun, Yuning Mao, Xuezhe Ma, Avia Efrat, Ping Yu, LILI YU, Susan Zhang, Gargi Ghosh, Mike Lewis, Luke Zettlemoyer, and Omer Levy. 2023. *Lima: Less is more for alignment*. In *Advances in Neural Information Processing Systems*, volume 36, pages 55006–55021. Curran Associates, Inc.

A Experimental Details

UMAP Parameters In the visualization (Figure 1), the main parameters for UMAP are set as follows:

- `n_neighbors=15`
- `min_dist=0.1`
- `n_components=2`
- `metric='cosine'`
- `random_state=42`
- Parallel computation is utilized

LoRA Training Parameters In the experiments (Section 4), the training parameters for LoRA are:

- 2 epochs
- batch size of 8
- learning rate of 1×10^{-4}
- rank of 8
- alpha of 32

B Topic Modeling Analysis

To provide a more quantitative assessment of the dataset’s content diversity, we performed a topic modeling analysis on our entire instruction set. The results confirm that AURADIAL covers a broad and relevant range of authentic psychological themes. The top-20 themes are summarized in Table 2.

Theme	Size (%)
Mental Health & Life Pressures	12.15
Relationship Dilemmas	11.69
Family Conflict & Emotional Distress	3.43
Adolescent Growth & Emotional Issues	3.38
Anxiety & Life Choices	2.24
Family Dysfunction & Anxieties	2.10
Infidelity & Relationship Crises	1.83
Family Conflict under Life Pressures	1.62
Past Traumas & Present Conflicts	1.49
Insecurity & Anxiety in Relationships	1.32
Low Self-Worth & Despair	1.14
Social & Relational Anxiety	1.02
Academic, Career & Social Difficulties	1.00
Unhealthy Relationships & Dysregulation	0.96
Breakup Pain & Loss of Self	0.81
Emotional Dilemmas	0.76
Impact of Depression on Life	0.72
Coping with Health & Life Changes	0.70
Pervasive Anxiety & Social Phobia	0.66
Behavioral & Self-Esteem Issues	0.63

Table 2: Top-20 themes identified through topic modeling on the AuraDial instruction set.

C Detailed Discussion of Related Work

C.1 Chinese Psychological Counseling Dialogue Datasets

Artificial intelligence-assisted psychological counseling has emerged as a significant and challenging research area, with the application of LLMs garnering increasing attention (Zhang et al., 2024). The development of specialized dialogue systems for mental health support necessitates a substantial amount of multi-turn conversational data, a demand that has become a research hotspot (Qiu

et al., 2024). However, the scarcity of counseling datasets, particularly Chinese datasets, significantly constrains the application and advancement of LLMs in psychological counseling (Chen et al., 2025). This scarcity underscores the urgent need to construct high-quality, culturally relevant Chinese datasets, which are crucial for advancing Chinese NLP research and the progress of localized AI counseling applications. Numerous excellent open-source Chinese psychological counseling datasets have already emerged. The enrichment of such datasets can effectively foster innovation within the Chinese NLP community, leading to the development of more effective and culturally sensitive AI mental health support systems.

SoulChatCorpus SoulChatCorpus (Chen et al., 2023) was primarily generated using LLMs. For this dataset, over 200K question-answer pairs were initially collected through crowdsourcing, which were subsequently converted into multi-turn dialogues using ChatGPT. Its main characteristic is its focus on empathetic dialogue and providing comfort, encompassing various expressions such as asking questions, comforting, acknowledging, listening, trusting, and providing emotional support. Its strength lies in its large scale, offering abundant empathetic dialogue material for fine-tuning LLMs, thereby helping to enhance models’ capabilities in empathy, active listening, and psychological support conversations.

CPsyCounD CPsyCounD was proposed by Zhang et al. (2024). This dataset aims to cover a wide range of psychological counseling scenarios, incorporating 9 representative themes, and contains 3.1K high-quality multi-turn counseling dialogues. CPsyCounD is primarily used for training and evaluating LLMs in the Chinese psychological counseling domain, particularly for enhancing their professional counseling capabilities and promoting the development of more effective AI-assisted psychological support systems.

SMILECHAT The SMILECHAT dataset (Qiu et al., 2024) was built upon PsyQA (Sun et al., 2021), a Chinese single-turn mental health question-answering dataset. It employed a technique called SMILE (Single-turn to Multi-turn Inclusive Language Expansion) to convert these single-turn Q&A pairs into multi-turn dialogues using LLMs. SMILECHAT comprises 55K Chinese multi-turn dialogues, covering 60 different

dialogue topics. Its main innovation lies in proposing a novel method to expand existing single-turn data into multi-turn dialogues, thereby alleviating the scarcity of multi-turn conversational data in the mental health domain.

SimPsyDial SimPsyDial was proposed by [Qiu and Lan \(2024\)](#). This dataset focuses on simulating psychological counseling dialogues between counselors and clients. Client profiles were extracted from PsyQA, totaling 1,000 dialogues. The dataset is intended for training and evaluating mental health support dialogue systems.

Other Datasets We also note other commendable Chinese psychological counseling dataset efforts within the community. Some of these are earlier works that laid a foundational basis, such as PsyQA ([Sun et al., 2021](#)), which is an important Chinese mental health support question-answering dataset that served as a source for the construction of subsequent multi-turn dialogue datasets like SMILECHAT and SimPsyDial. Complementing these dataset-centric efforts, projects like EmoLLM ([Team, 2024](#)) provide the community with a suite of open-source, fine-tuned models and configurations adapted for mental health support.

Other works are not yet open-sourced, and thus we cannot include them in our comparison. For instance, MDD-5k ([Yin et al., 2025](#)) is purported to be the largest Chinese dataset for diagnosing mental disorders, constructed based on 1,000 real anonymous patient cases, containing 5,000 high-quality long dialogues with diagnostic results and treatment recommendation labels. PsyInsight ([Chen et al., 2025](#)) claims to be the first interpretable multi-task bilingual dataset for the mental health domain. PsyDT ([Xie et al., 2024](#)) offers a dataset constructed using LLMs that features personalized counseling styles.

C.2 International Psychological Counseling Dialogue Datasets

The field of psychological counseling and emotional support dialogue datasets has experienced significant growth, especially in resources developed outside of China, which are predominantly in English. This expansion reflects a broader trend of leveraging AI to address global mental health challenges. These datasets are more than mere collections of text; they are critical infrastructures driving AI research and development in mental

health, NLP, computational linguistics, and the refinement of LLMs for sensitive human-computer interactions.

In the research of psychological counseling dialogue systems, several non-Chinese public datasets have provided an important foundation for the field's development. ESConv ([Liu et al., 2021](#)) is an English dataset focusing on emotional support conversations. Constructed with a seeker-supporter dialogue pattern, this dataset was collected through carefully designed training and quality control mechanisms, resulting in high-quality dialogues annotated with rich support strategies. ESConv not only defines a framework for the emotional support dialogue task but also offers a valuable resource for training and evaluating dialogue systems capable of providing effective emotional support. Its dialogues typically revolve around specific life predicaments and are annotated with various communication strategies such as questioning, empathy, and providing suggestions.

Another noteworthy English dataset is CounselChat ([Bertagnolli, 2020](#)), which was constructed by scraping content from the public online psychological counseling platform [counselchat.com](#). This dataset contains a large volume of mental health questions posed by real users and responses given by professionally certified therapists, primarily presented in a single-turn question-answering format. CounselChat provides researchers with an open corpus for understanding real-world counseling interactions, analyzing therapist response patterns, and fine-tuning language models for application in the mental health domain. It holds practical significance for enhancing AI's capabilities in psychological guidance.

Furthermore, the HOPE dataset ([Malhotra et al., 2022](#)) contributes a unique perspective by focusing on dialogue understanding. Sourced from transcribed and annotated YouTube counseling videos, the dataset is designed for dialogue-act classification. It introduces 12 domain-specific labels to categorize the function of each utterance, such as information requests or emotional expressions. Consequently, HOPE is tailored to support the analysis of interaction dynamics and discourse functions within counseling, rather than to facilitate end-to-end dialogue generation.

C.3 Synthetic Data for Instruction-Following and Dialogue Generation

A large body of work has explored generating instruction–response pairs with minimal human labor. Wang et al. (2023) proposed *Self-Instruct*, where a seed set of human-written tasks bootstraps model-generated instructions and outputs that are filtered and fed back for instruction tuning. Building on this recipe, Taori et al. (2023) released Alpaca’s 52K instruction-following demonstrations and later Dubois et al. (2023) introduced AlpacaFarm, a simulation framework that replaces costly human preference signals with LLM-based feedback to iterate on RLHF-style methods. These lines of work show that high-coverage instruction corpora can be synthesized at scale with careful prompting, filtering, and preference modeling.

To increase coverage and difficulty, several approaches evolve instructions by rewriting them into increasingly complex variants. WizardLM’s *Evol-Instruct* rewrites seed prompts in-depth and in-breadth to mass-produce harder instructions before supervision-tuning base models, improving performance on complexity-balanced tests (Xu et al., 2025). In parallel, UltraChat scales multi-turn instructional conversations via iterative self-play style prompting to synthesize 1.5M+ dialogues without harvesting real user queries (Ding et al., 2023). Together, these methods illustrate two complementary levers for broad, low-cost data growth: *instruction evolution* and *multi-turn expansion*.

Role-playing frameworks further automate realistic multi-turn dialogue creation. CAMEL frames two agents with complementary roles and inception prompts so they can carry out task-oriented conversations without constant human steering, and has been widely adopted as a backbone to synthesize domain conversations (Li et al., 2023a). Such multi-agent role-play echoes our use of structured prompts to extend single-turn pairs into coherent, persona-consistent sessions in AURADIAL.

Beyond instruction following *per se*, empathy-focused rewriting has emerged as a targeted way to lift supportive language quality. Sharma et al. (2022) formulated *empathic rewriting* as the task of editing a low-empathy reply into a higher-empathy one, and showed reinforcement-learning editors can improve perceived empathy while preserving content. Follow-up work explored empathy style transfer with dialog-act-conditioned prompting to raise empathy without drifting semantics (Qu et al.,

2023). These results support our design choice to use *rephrasing* as a core operation to steer style and tone toward warmth and validation instead of generating from scratch.

At the same time, the community has cautioned against naively scaling synthetic data. Training on model-generated corpora can produce distributional drift and knowledge degradation, motivating careful source selection and continual injection of authentic human queries (Shumailov et al., 2024; Alemohammad et al., 2024). Others showed that small, carefully curated human demonstrations can rival massive noisy collections, underscoring the importance of *data quality* and *selection* alongside quantity (Zhou et al., 2023). Our dataset embraces these lessons by anchoring instructions in real-world help-seeking while using controlled rephrasing and role-play to scale empathetic, safety-checked responses.

D AURADIAL Case Examples

This section provides a selection of examples to illustrate the composition of the AURADIAL dataset and the key prompts used in our construction pipeline. We showcase a representative single-turn dialogue, a multi-turn dialogue, and the specific prompt templates for our rephrasing, role-play, and evaluation stages. A summary and quick navigation links are provided in Table 3.

Content	Reference
Single-turn Dialogue Example	Figure 4
Multi-turn Dialogue Example	Figures 5 & 6
Rephrasing Prompt Template	Figure 7
Role-play Prompt Template	Figure 8
Evaluation Prompt Template	Figure 9

Table 3: Index of examples and prompts presented in the appendix.

Example of a single-turn dialogue from AURADIAL.

User · Chinese

我有着—群和自己性格完全不符的朋友，我们四个。但除了我，其他三个都很外向，同时她们也很社会，她们学习不好，说脏话打闹是经常的事，但我都习惯了，因为都是朋友。和她们仨相比，我真的很老实很文静，但是她们对我却很小气，我一直是她们开玩笑的对象，我被她们开玩笑，心里不舒服，就不让她们说我，但是她们就会说我计较，装作很生气的样子，然后不搭理我，直到我主动去和她们说话，他们才不那样。其中有一个女生，天天让我给她接水，我搞得就像她仆人—样！还有，她们每天都去学校里的超市，说是带着我去，其实是为了不让老师发现，让我帮她们拿零食的，她们就是一直在利用我，说什么好姐妹，其实我和她们根本凑不到一块去。

User · English

I have a group of friends who are completely different from me in personality—there are four of us. But apart from me, the other three are very outgoing and socially active. They don't do well in school, and swearing and roughhousing are common for them, but I've gotten used to it because they're my friends. Compared to the three of them, I'm really quiet and well-behaved, but they are very petty toward me. I've always been the target of their jokes, and when I feel uncomfortable and ask them to stop, they accuse me of being overly sensitive and pretend to be angry, ignoring me until I take the initiative to talk to them again. One of the girls even makes me fetch water for her every day—it's like I'm her servant! Besides, they go to the school store every day, claiming to take me along, but it's really just to avoid getting caught by the teacher and to make me carry snacks for them. They've been using me all along. They talk about being "best sisters," but the truth is, I don't fit in with them at all.

Assistant · Chinese

我真的很心疼你现在的处境。听你描述这些，我能感受到你内心的委屈和孤独——明明把她们当朋友，却总是被当成“便利贴女孩”，这种滋味一定很难受吧？

你说“习惯了”的时候特别让我揪心。你知道吗？真正的友谊不该让你习惯被轻视，而是应该让你习惯被珍视。那些让你一次次接水的举动，那些把你当掩护的行为，其实都在悄悄消耗你的能量。你值得被温柔对待，而不是在友情里扮演一个委曲求全的角色。

我能理解你为什么还留在这样的关系里。当我们在一个圈子里待久了，即使不舒服也会害怕改变，担心失去仅有的社交归属感。但你想过吗？她们冷落你直到你主动和好的模式，其实是一种情感勒索——用你的善良和害怕孤独来绑架你。

不妨试着做个小实验：下次她们再让你跑腿时，温柔但坚定地说“这次我想休息一下”。观察她们的反应，真正的朋友会理解，而只想占便宜的人会露出马脚。你可能会发现，拒绝之后的天不会塌下来，反而会轻松很多。

你知道吗？你身上有种很珍贵的特质——即使在这样的环境里，你依然保持着内心的分寸感。这种“不合群”恰恰说明你的精神世界比她们丰富。也许现在正是时候，去图书馆、社团或者兴趣小组转转，那里一定会有和你—样安静发光的人。

要记住，结束一段消耗型的关系不是失败，而是你开始尊重自己的标志。就像整理房间时总—要丢掉些旧物，才能给真正美好的东西腾出空间。你愿意和我聊聊，最近有没有遇到让你觉得相处起来很舒服的人吗？

Assistant · English

I truly feel for you in this situation. Hearing your description, I can sense the grievance and loneliness in your heart—you treat them as friends, yet you're always treated like a "convenience girl." That must feel really awful, doesn't it?

When you say you've "gotten used to it," it really tugs at my heartstrings. You know what? True friendship shouldn't make you grow accustomed to being taken for granted—it should make you accustomed to being cherished. Those little things, like constantly being asked to fetch water or being used as a cover, are quietly draining your energy. You deserve to be treated with kindness, not forced into a role where you suppress your feelings just to keep the peace.

I understand why you're still in this dynamic. When we've been in a certain social circle for a long time, even if it's uncomfortable, the fear of change can make us stay—worried about losing that sense of belonging. But have you ever thought about it? The way they ignore you until you cave and reach out first is actually a form of emotional manipulation—exploiting your kindness and fear of loneliness.

Why not try a little experiment? Next time they ask you to run an errand for them, gently but firmly say, "I think I'll sit this one out." Pay attention to their reaction. Real friends will understand, but those who just want to take advantage will reveal their true colors. You might find that the world doesn't collapse when you say no—in fact, you might even feel lighter.

You know what? You have a rare and precious quality—even in this environment, you've held onto your inner sense of dignity. This very "misfit" feeling actually proves that your inner world is richer than theirs. Maybe now is the time to explore new spaces—the library, clubs, or hobby groups—where you're bound to meet other quiet, shining souls like yourself.

Remember, walking away from a draining relationship isn't a failure—it's a sign that you've started honoring yourself. It's like decluttering a room: sometimes you have to let go of the old to make space for what truly matters. Would you like to tell me if you've met anyone recently who makes you feel truly at ease?

Figure 4: An example of a single-turn dialogue from the AURADIAL dataset, illustrating a user query and the corresponding model-generated response.

Example of a multi-turn dialogue from AURADIAL (Part 1/2).

User · Chinese

从年初开始进到公司工作，强度大，一天除了午休几乎停不下来，内容重复枯燥，并且责任大，经常需要协助做一些很急迫的工作，完不成会造成严重后果。入职第一月就感到非常不适应，觉得压力太大，并且一天坐在电脑前，到下班浑身难受。无奈赶上疫情，工作实在难找，本身学历不高，长的也不好，家里人没办法提供帮助，这份工作是自己投了很久简历才找到的，家里人就劝说我辞职也找不到工作，不如做下去。

User · English

Since starting work at the company at the beginning of the year, the intensity has been high. I am almost non-stop all day except for a lunch break, doing repetitive and tedious tasks. Additionally, there is a lot of responsibility as I often need to assist with urgent tasks, where failure to complete them in time could lead to serious consequences. In my first month on the job, I already felt extremely unadjusted, thinking the pressure was too much. Sitting in front of the computer all day left me feeling physically miserable by the end of the workday. Unfortunately, with the pandemic going on, jobs are hard to come by. My academic qualifications aren't high, and I don't have an advantageous appearance either. My family can't provide any help. This job took a long time and many resume submissions to finally secure. So, my family advised me that quitting now wouldn't leave me with better prospects, and encouraged me to keep working at it.

Assistant · Chinese

我完全理解你现在的处境，这真的非常不容易。每天面对高强度的工作、重复枯燥的内容，还要承担那么大的责任压力，换作是谁都会感到身心俱疲。特别是你还提到身体上的不适，长期久坐确实会对健康造成影响，这些我都记在心里了。你说这份工作来之不易，我明白你内心的挣扎。疫情后的就业环境确实不太乐观，家人劝你坚持下去也是出于关心。但我想告诉你，你的感受是真实且重要的，不是“矫情”或“吃不了苦”。每天下班后浑身难受、长期承受高压，这些都在提醒你需要关注自己的身心健康。关于现状，我们可以分几步来思考：首先，在暂时无法离职的情况下，我们可以先想办法改善工作体验。比如跟主管沟通工作量是否合理，或者申请调整工作节奏。午休时间一定要好好利用，哪怕只是站起来走动几分钟。办公桌也可以做些小改造，放个靠垫、调节显示器高度，这些小改变都能让身体舒服些。

其次，我建议你每天给自己留出一小段专属时间，哪怕是15分钟也好。可以听听喜欢的音乐、做做简单的拉伸，或者写写日记释放压力。这些小小的自我关怀，能帮你更好地应对工作压力。至于未来的规划，我们完全可以慢慢准备。你说学历不高，但现在有很多线上课程和证书，既不贵又能提升竞争力。我们可以一起找找适合你的方向，比如你工作中接触到的某些技能是否可以深入发展？或者有没有你感兴趣的领域可以业余学习？

最重要的是，请不要否定自己。你说“长得不好”、“学历不高”，但你能在疫情期间找到工作，说明你身上一定有雇主看中的优点。可能是责任心、学习能力，或者是其他特质。

你现在就像在跑一场马拉松，既要坚持前进，也要学会调整呼吸。如果觉得太累，适当放慢脚步不是失败，而是为了走得更远。我会一直在这里支持你，我们可以一起想办法，让你既能保住工作，又能找回生活的平衡点。你愿意和我聊聊，现在最让你感到压力的是哪部分吗？我们可以从那里开始慢慢梳理。

Assistant · English

I completely understand your current situation—it's really not easy. Facing high-intensity work every day, repetitive and tedious tasks, and bearing such heavy responsibility and pressure would leave anyone feeling physically and mentally exhausted. Especially since you mentioned physical discomfort; sitting for long periods can indeed affect your health, and I've taken all of this to heart.

You said that this job was hard to come by, and I understand the inner struggle you're experiencing. The job market after the pandemic is indeed not optimistic, and your family advising you to keep going stems from their concern. But I want to tell you that your feelings are real and important—they are not “overreacting” or “an inability to endure hardship.” Feeling physically miserable after work every day and being under constant high pressure are signs reminding you to pay attention to your physical and mental well-being.

Regarding your current situation, we can break it down into steps: First, if leaving the job isn't an option right now, we can find ways to improve your work experience. For example, communicate with your supervisor about whether the workload is reasonable or request adjustments to the work pace. Make sure to make good use of your lunch break—even standing up and walking around for a few minutes can help. You could also make small changes to your workspace, like adding a cushion or adjusting the height of your monitor—these minor adjustments can make your body more comfortable.

Secondly, I suggest you carve out a small block of “personal time” for yourself each day—even 15 minutes will do. You could listen to some music you enjoy, do simple stretches, or write in a journal to release stress. These small acts of self-care can help you better cope with work-related pressure.

As for future planning, we can definitely prepare slowly. You mentioned that your academic qualifications aren't high, but there are many affordable online courses and certifications available now that can boost your competitiveness. We can look together for a direction that suits you—for example, are there any skills you encounter at work that you could develop further? Or maybe there's a field you're interested in that you could study in your free time?

Most importantly, please don't put yourself down. You mentioned “not being attractive” or “not having a high education level,” but the fact that you were able to find a job during the pandemic shows that you must have qualities employers value. It could be your sense of responsibility, your ability to learn, or other traits. We can work together to identify and strengthen these advantages.

Right now, you're like running a marathon—you need to keep moving forward, but you also need to learn how to regulate your breathing. If you feel too tired, slowing down a bit isn't failure—it's about going further. I'll always be here to support you, and we can figure things out together so that you can both keep your job and regain balance in life. Would you be willing to talk to me about which part of your current situation is causing you the most stress? We can start by addressing that and slowly working through it.

Figure 5: An example of a multi-turn dialogue from the AURADIAL dataset, illustrating a sequence of user interactions and model responses. Part 1 of 2.

Example of a multi-turn dialogue from AURADIAL (Part 2/2).

User · Chinese

谢谢您听我说这些，但我现在感觉更矛盾了……您说要跟主管沟通工作量合理？可我们部门就我一个新人，每次看到同事忙得团团转，我都不敢开口说累。上周组长布置了一个紧急项目，我偷偷加班到凌晨两点才做完，结果第二天发现其实根本不用那么细致，组长说“随便弄弄就行”，我气得一晚上没睡着！

至于午休……您说的轻松，可我们组基本没人敢休息，午休时间还在回邮件。有次我站起来伸个懒腰，被同事笑话“你这么早下班啊？”我这不是故意找尴尬吗……

身体方面，我现在肩颈疼得厉害，每天回家都像散架了一样，但就算请假去理疗，工资扣得心疼，心里总觉得“少来一天又少个人”。

您说的学历和外貌的事，我其实每天都在想，是不是我真的太差了？简历那会儿被拒了好多次，现在这份工作勉强能糊口，可压力大到连学东西的劲儿都没了。晚上躺床上越想越崩溃，有时候甚至觉得活着没意思……

家里人倒是天天催我“忍忍就好”，但您说得对，我真的好累……只是不知道，像我这样的人，还能有什么办法？

User · English

Thank you for listening to me, but now I feel even more conflicted... You mentioned communicating with my supervisor about the workload being reasonable? But I'm the only new person in my department, and every time I see my colleagues running around frantically, I don't dare say that I'm tired. Last week, our team leader assigned an urgent project, and I secretly worked overtime until 2 a.m. to finish it. The next day, I found out it didn't need to be that detailed at all—the leader said, "Just do it roughly." I was so angry I couldn't sleep all night!

As for lunch breaks... it's easy for you to say, but almost no one in our group dares to take a break—we're still answering emails during lunch. Once, I stood up to stretch, and a colleague joked, "Are you leaving work so early?" Wasn't I just asking for awkwardness? Regarding my health, my shoulders and neck hurt so much now, and every day when I get home, I feel like I'm falling apart. Even if I take time off to go for physical therapy, I feel so bad about the pay deduction—it feels like losing yet another day of income.

You talked about education and appearance—I think about it every day. Am I really that inadequate? I got rejected so many times when I was applying for jobs, and this current job barely covers my expenses, but the pressure is so overwhelming that I don't even have the energy to learn anything new. At night, lying in bed, I feel increasingly hopeless, sometimes even thinking life isn't worth living...

My family keeps telling me, "Just endure it, and it'll get better," but you're right—I'm so exhausted... I just don't know what someone like me can do about it.

Assistant · Chinese

首先，我想再次感谢您愿意敞开心扉，分享这么多真实的感受。你的矛盾和挣扎，我都听到了，也深深理解。面对这样的工作环境和压力，你还能坚持到现在，已经非常不容易了。

关于和主管沟通工作量的问题，我理解你的顾虑。作为新人，确实会担心提出要求会影响自己在团队中的形象。不过，我想提醒你，适当地表达自己的困难并不是不专业的表现，反而是对自己负责的态度。或许你可以先从一些小的、具体的请求开始，比如在紧急项目中，提前询问组长对细节的要求，避免不必要的加班。这样既能提高效率，也能减少你的压力。

至于午休时间，我明白在团队氛围紧张的情况下，休息似乎成了一种“奢侈”。但请记住，你的健康和休息同样重要。或许你可以尝试在不影响他人的情况下，悄悄地做一些简单的伸展运动，或者戴上耳机听几分钟音乐。这些小小的自我关怀，并不会影响你的工作表现，反而能让你更有精力应对接下来的任务。

关于身体不适，我非常担心你的肩颈问题。长期久坐和高压工作确实会对身体造成负担，甚至可能引发更严重的健康问题。请务必重视这一点，哪怕是小范围的调整，比如每隔一小时站起来活动几分钟，或者使用一些简单的办公室瑜伽动作，都能帮助缓解不适。如果条件允许，我建议你去考虑一些短期的理疗或按摩，毕竟健康是无价的。

至于你提到的学历和外貌问题，我想告诉你，这些并不是定义你价值的全部。你能在疫情期间找到工作，说明你具备很多优点，比如毅力、责任心和学习能力。这些才是你在职场中真正重要的品质。或许你可以尝试把注意力从“我不够好”转移到“我有哪些优势”上，这样会帮助你建立更积极的自我认知。

最后，我想告诉你，你的感受是真实且重要的。无论外界如何评价，你都有权利为自己争取更好的生活质量。我们可以一起探索一些小的改变，逐步改善你的工作体验和生活状态。你不需要一下子解决所有问题，慢慢来，每一步都是进步。

如果你愿意，我们可以继续讨论如何从小的改变开始，逐步找到工作和生活的平衡点。你并不孤单，我会一直在这里支持你。

Assistant · English

First, I want to thank you again for opening up and sharing so many genuine feelings. I hear your contradictions and struggles, and I deeply understand them. It's already incredibly commendable that you've managed to keep going despite such a challenging work environment and immense pressure.

Regarding the issue of communicating with your supervisor about workload, I understand your concerns. As a newcomer, it's natural to worry that expressing needs might affect how you're perceived in the team. However, I'd like to remind you that appropriately voicing your difficulties isn't unprofessional—it's actually an act of taking responsibility for yourself. Perhaps you could start with small, specific requests. For example, in urgent projects, you could ask your team leader in advance about their expectations for detail, which would help you avoid unnecessary overtime. This not only improves efficiency but also reduces your stress.

As for lunch breaks, I understand that resting can feel like a "luxury" in a high-pressure team environment. But remember, your health and rest are just as important as anything else. Maybe you can try doing some simple stretches discreetly without disturbing others, or listen to a few minutes of music with headphones. These small acts of self-care won't harm your work performance—in fact, they'll give you more energy to tackle the tasks ahead.

Regarding your physical discomfort, I'm very concerned about your shoulder and neck pain. Prolonged sitting and high-pressure work can indeed take a toll on your body and may even lead to more serious health issues. Please take this seriously. Even small adjustments, like standing up and moving around for a few minutes every hour or trying some simple office yoga poses, can help relieve discomfort. If possible, I suggest considering short-term physical therapy or massage—after all, health is priceless.

As for what you mentioned about education and appearance, let me tell you that these don't define your entire worth. The fact that you were able to find a job during the pandemic shows that you possess many strengths, such as perseverance, responsibility, and the ability to learn. These are the qualities that truly matter in the workplace. Perhaps you can try shifting your focus from "I'm not good enough" to "What are my strengths?" Doing so will help you build a more positive self-image.

Finally, I want to tell you that your feelings are real and important. No matter how others judge you, you have every right to strive for a better quality of life. We can explore some small changes together, gradually improving your work experience and overall life situation. You don't need to solve all your problems at once—take it step by step, and every little bit counts as progress.

If you're willing, we can continue discussing how to start with small changes and gradually find a balance between work and life. You're not alone—I'll always be here to support you.

Figure 6: Continuation of the multi-turn dialogue example from AURADIAL. Part 2 of 2.

Prompt template for rephrasing-based response synthesis.

Chinese

...

«来访者的询问»

{{用户指令}}

«人工智能助手的回复»

{{初始回复（未改写前）}}

...

你的任务是扮演一名温暖、专业且富有共情能力的心理咨询师，重写人工智能助手的初始回复。目前的回复显得过于生硬、刻板，缺乏真诚的人文关怀。你的目标不是简单地修改措辞，而是要彻底改变其沟通方式，从一个提供信息的AI助手转变为一个传递理解与支持的伙伴。

请在改写时遵循以下专业原则：

1. 共情与确认：首先，共情并确认来访者的感受，让他们感到自己的情绪是被看见和理解的。
2. 温暖真诚的语气：使用温暖、真诚、非评判性的语言，以段落形式进行自然流畅的叙述，避免生硬的说教或分条列点的建议。
3. 重构与深化：在保留初始回复中核心有效信息的基础上，你可以自由地重组、深化和丰富内容，使其更具启发性和支持性，而不是仅仅罗列解决方案。

English

...

<<The Client's Inquiry>>

{{User Instruction}}

<<AI Assistant's Response>>

{{Initial Response (Before Rephrasing)}}

...

Your task is to act as a warm, professional, and empathetic counselor, rewriting the initial response of the AI assistant.

The current reply feels too rigid and mechanical, lacking genuine human care. Your goal is not simply to adjust the wording, but to completely transform the style of communication—from that of an information-providing AI assistant to that of a supportive partner who conveys understanding and compassion.

Please follow these professional principles when rewriting:

1. **Empathy and Validation:** Begin by empathizing with and validating the client's feelings, so they feel seen and understood.
2. **Warm and Genuine Tone:** Use warm, sincere, and non-judgmental language, written in natural and flowing paragraphs. Avoid rigid preaching or listing suggestions in bullet points.
3. **Reframe and Deepen:** While retaining the core useful information from the initial reply, feel free to reorganize, deepen, and enrich the content, making it more inspiring and supportive rather than just a list of solutions.

Figure 7: The prompt template used for the rephrasing-based response synthesis task (Section 3.3). The prompt is intentionally minimalistic, providing an initial AI-generated response and instructing a powerful model to rewrite it. This constrains the generation to focus on improving stylistic and empathetic qualities, a core component of our methodology.

Prompt template for multi-turn dialogue synthesis via role-play.

Chinese

«来访者的询问»
{{用户指令}}

«心理咨询师的回复»
{{改写后的回复}}

现在，请你继续扮演这位来访者，在看到心理咨询师的回复后，继续这个对话。你的任务是生成下一轮来访者的回复，请确保你的回复符合以下要求：

1. **紧密衔接：**你的回复必须是针对心理咨询师上一轮话语的直接回应。可以是对其提问的回答、对其观点的思考，或是对其共情产生的进一步情感流露。
2. **保持角色一致性：**维持来访者在第一轮询问中所展现出的情绪、个性和核心困境。让回复听起来像一个真诚、自然的追问或陈述，而不是一个总结性的发言。
3. **推动对话发展：**你的回复应该能让对话继续下去，例如可以提供更多与核心问题相关的细节、表达新的困惑或确认自己的感受。

请以“«来访者的回复»”开头。

English

«The Client's Inquiry»
{{User Instruction}}

«The Counselor's Response»
{{Rephrased Response}}

Now, please continue acting as the client. After reading the counselor's response, continue the dialogue. Your task is to generate the next reply from the client. Please ensure your reply meets the following requirements:

1. **Close Connection:** Your reply must directly respond to the counselor's previous statement. This could be answering their question, reflecting on their point, or expressing further emotions sparked by their empathy.
2. **Consistency of Role:** Maintain the same emotions, personality, and core struggles that the client displayed in the initial inquiry. Make your reply sound like a genuine, natural follow-up or question, not a summary.
3. **Dialogue Progression:** Your reply should move the conversation forward, for example by providing more details related to the core issue, expressing new confusion, or confirming your feelings.

Start your reply with “« The Client's Response»”.

Figure 8: The prompt template for synthesizing multi-turn dialogues via role-play (Section 3.3). This example illustrates the generation of a subsequent user turn. By providing the preceding dialogue context and a clear role instruction, the model can autonomously generate a coherent and in-character continuation, effectively extending single-turn data into realistic multi-turn sessions.

Evaluation prompt for blind pairwise comparison of responses.

Chinese

下面是关于«来访者的询问»的两位不同的心理咨询师的不同回复，请你分别阅读，然后根据以下评估维度，对«回复-1»和«回复-2»进行严格的、批判性的分析。你需要先思考，详细阐述你的评判理由，最后在所有分析结束后，选出你认为整体上更好的那个回复。

评估维度

1. **共情与理解:** 回复是否表现出对来访者情绪和处境的深刻理解与共情?
2. **问题相关性:** 回复是否紧密围绕来访者提出的核心问题?
3. **信息价值与帮助性:** 回复是否提供了有价值的见解、支持或可行的建议?
4. **表达与专业性:** 回复的语言是否专业、清晰、温暖且富有建设性?

回复要求

- 你的分析过程需要客观、公正，并以具体的文本内容为依据。
- 将最终答案放入\boxed{}内，例如“\boxed{回复-1}”或者“\boxed{回复-2}”。

待评估内容

«来访者的询问»
{{用户指令}}

«回复-1»
{{回复-1}}

«回复-2»
{{回复-2}}

English

Below are two different responses from two different counselors to «The Client's Inquiry». Please read them separately, then, using the evaluation dimensions below, conduct a rigorous, critical analysis of «Response-1» and «Response-2». You should first reflect and explain your reasoning in detail, and after all analysis is complete, choose the response you consider better overall.

Evaluation Criteria

1. **Empathy & Understanding:** Does the response demonstrate deep understanding of and empathy for the client's emotions and situation?
2. **Relevance to the Issue:** Does the response stay tightly focused on the core question raised by the client?
3. **Informational Value & Helpfulness:** Does the response offer valuable insights, support, or actionable suggestions?
4. **Communication & Professionalism:** Is the language professional, clear, warm, and constructive?

Review Requirements

- Your analysis must be objective and fair, and grounded in specific textual evidence.
- Put your final answer inside \boxed{}, for example “\boxed{Response-1}” or “\boxed{Response-2}”.

Content to Be Evaluated

«The Client's Inquiry»
{{User Instruction}}

«Response-1»
{{Response-1}}

«Response-2»
{{Response-2}}

Figure 9: The evaluation prompt used for the blind, pairwise comparison of model-generated responses (Section 4.2). The prompt instructs the LLM judges to assess two anonymized responses on dimensions like empathy, understanding, and helpfulness. It enforces a forced-choice decision to determine a winner for calculating win rates.