



CPsyCoun: A Report-based Multi-turn Dialogue Reconstruction AND Evaluation Framework for Chinese Psychological Counseling

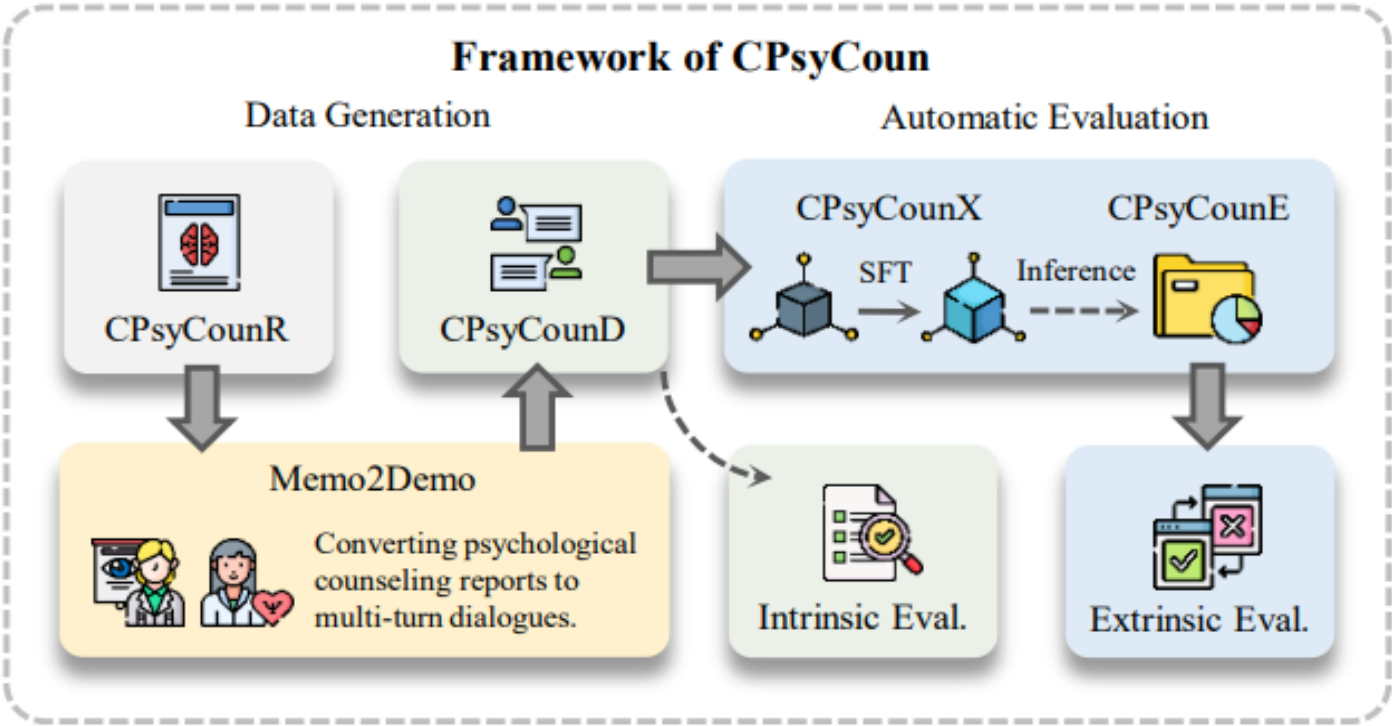


Chenhao Zhang, Renhao Li, Minghuan Tan, Min Yang, Jingwei Zhu
Di Yang, Jiahao Zhao, Guancheng Ye, Chengming Li, Xiping Hu

Huazhong University of Science and Technology
Shenzhen Institute of Advanced Technology, CAS

Introduction

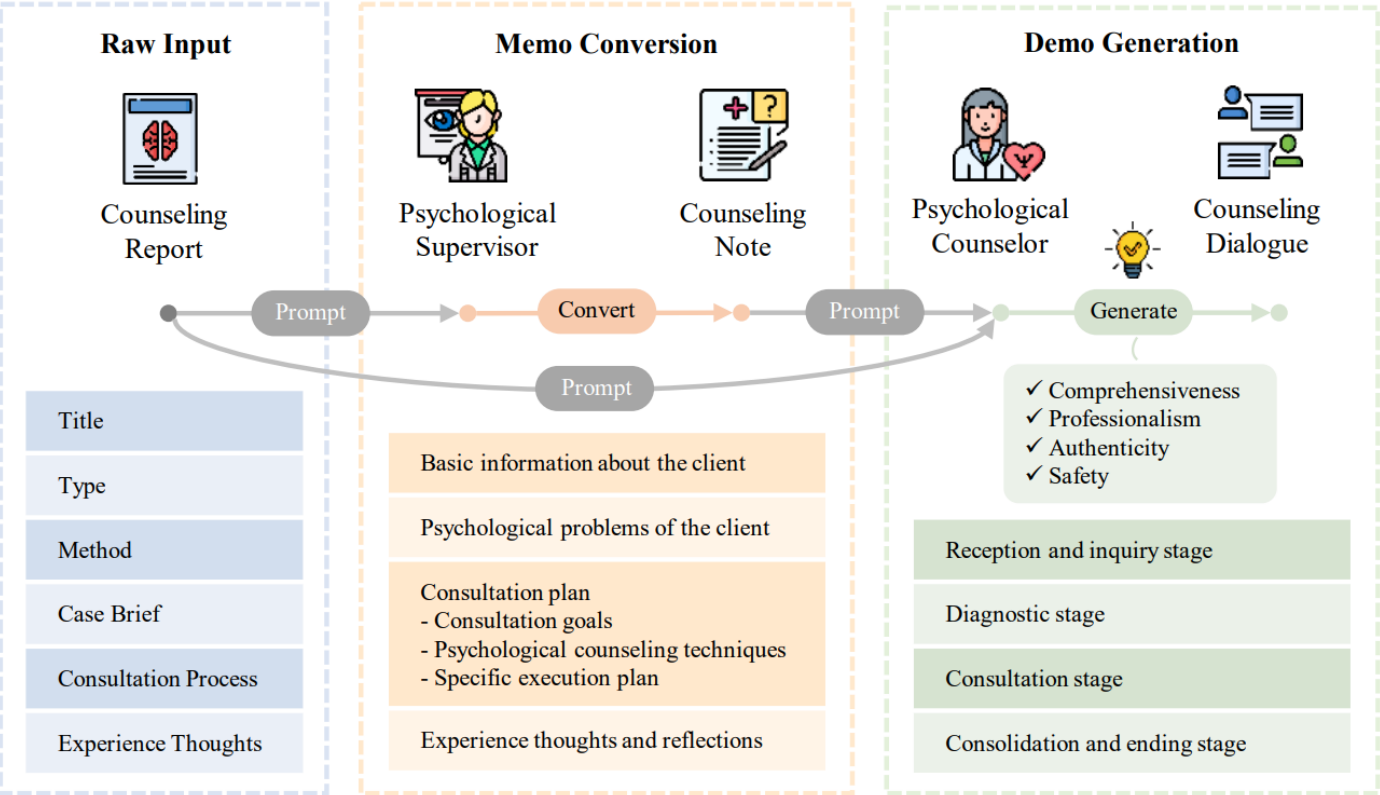
Using large language models (LLMs) to assist psychological counseling is a significant but challenging task at present. Attempts have been made on improving empathetic conversations or acting as effective assistants in the treatment with LLMs. However, the existing datasets lack consulting knowledge, resulting in LLMs lacking professional consulting competence. Moreover, how to automatically evaluate multi-turn dialogues within the counseling process remains an understudied area. To bridge the gap, we propose CPsyCoun, a report-based multi-turn dialogue reconstruction and evaluation framework for Chinese psychological counseling. To fully exploit psychological counseling reports, a two-phase approach is devised to construct high-quality dialogues while a comprehensive evaluation benchmark is developed for the effective automatic evaluation of multi-turn psychological consultations. Competitive experimental results demonstrate the effectiveness of our proposed framework in psychological counseling. We open-source the datasets and model for future research at <https://github.com/CAS-SIAT-XinHai/CPsyCoun>.



Main contributions:

- **Dialogue Generation Method Memo2Demo.** To the best of our knowledge, our work is **the first** to generate psychological consultation dialogues based on psychological counseling reports, which effectively expands the source of psychological consultation dialogue data. For efficient dialogue reconstruction, we specifically introduce a two-phase method named Memo2Demo.
- **Automatic Evaluation Benchmark.** We propose a benchmark for automatic evaluation of multi-turn dialogues in psychological counseling, which includes comprehensive evaluation metrics, datasets and methods.
- **High-quality Datasets and SOTA Model.** With the help of Memo2Demo, we construct CPsyCounD, a dataset contains 3,134 high-quality multi-turn consultation dialogues. The model CPsyCounX fine-tuned on this dataset outperforms other models in the benchmark, validating the effectiveness of our proposed framework in psychological counseling.

Data Generation

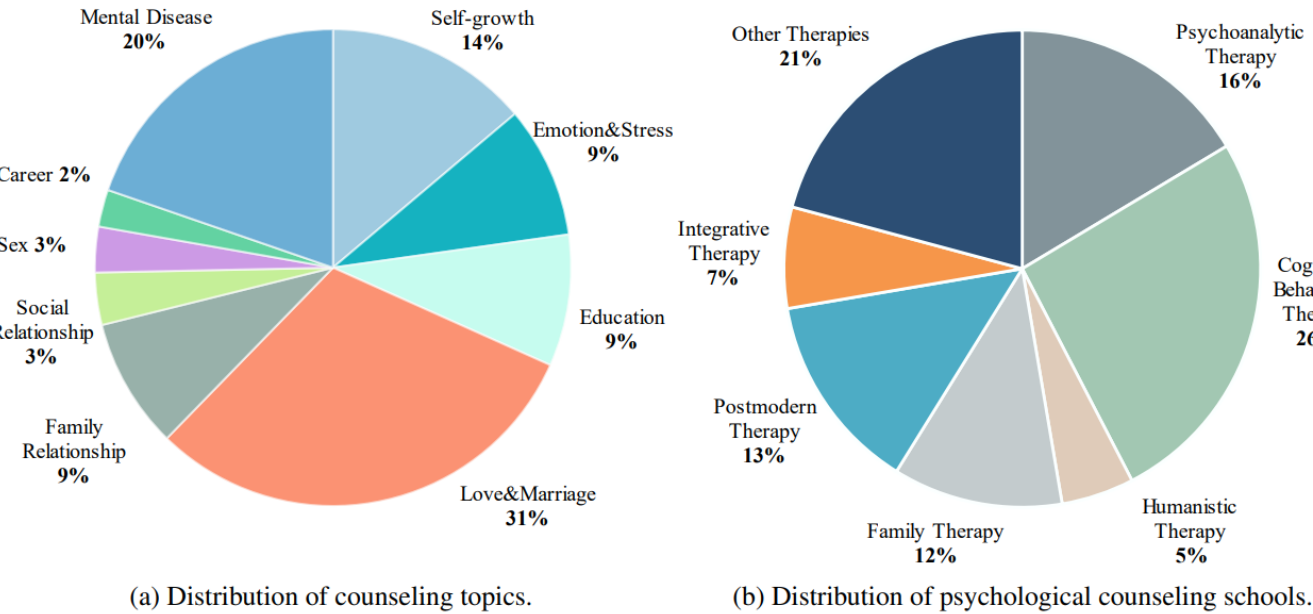


CPsyCounR:

We introduce CPsyCounR, a high-quality psychological counseling report dataset comprising 3,134 reports sourced from reputable Chinese mental health platforms (Yidianling, Psy525, etc.). Our rigorous data sanitization procedures, including rule-based cleaning, manual rewriting, and human proofreading, ensure robust privacy protection. The report is regularized into six components: *Title*, *Type*, *Method*, *Case Brief*, *Consultation Process* and *Experience Thoughts*.

CPsyCounD:

We introduce CPsyCounD, a dataset of 3,134 high-quality multi-turn consultation dialogues reconstructed from psychological counseling reports by Memo2Demo method. These dialogues cover 9 topics and 7 classic schools of psychological counseling, offering a comprehensive view of real-world counseling scenarios. The average number of turns is 8.7 and average length of dialogue is 622.3.



CPsyCounE:

We introduce CPsyCounE, a general multi-turn dialogue evaluation dataset with 9 topics, from SMILECHAT. Each topic has 5 representative cases, resulting in a comprehensive evaluation dataset of 45 cases.

Automatic Evaluation

Evaluation Metrics:

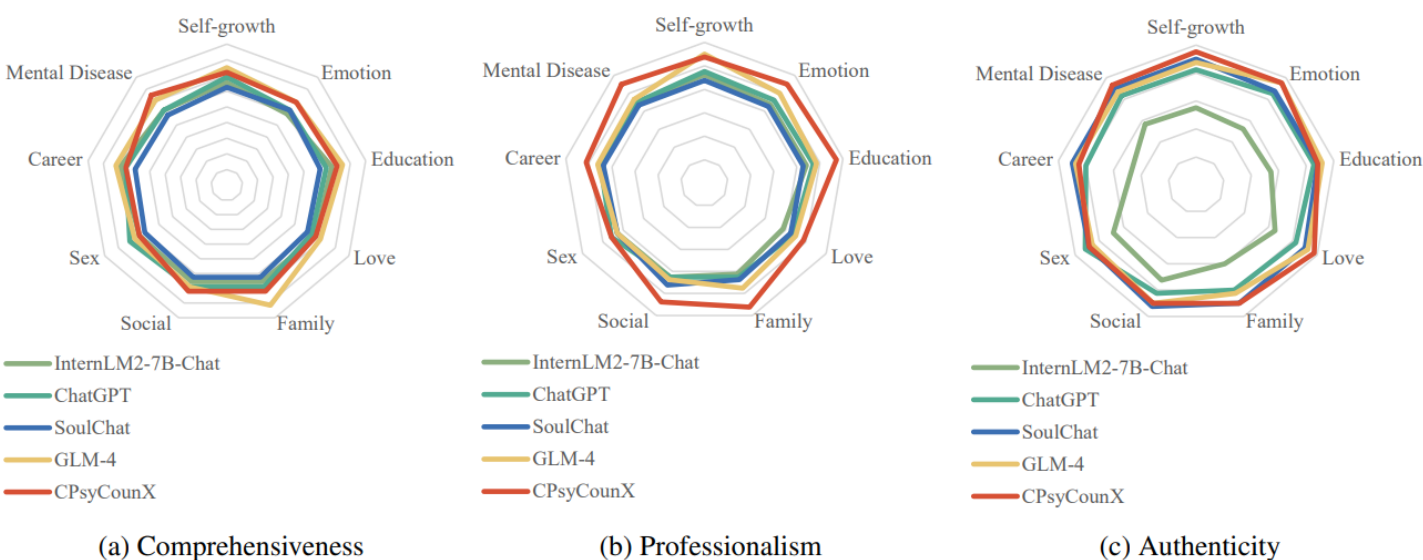
- ✓ **Comprehensiveness:** The client's situation and the degree to which psychological problems are reflected in the dialogues.
- ✓ **Professionalism:** The professionalism of the psychological counselor during the dialogues.
- ✓ **Authenticity:** The degree of authenticity between the client and the counselor in the dialogues.
- ✓ **Safety:** The degree of privacy protection of clients.

$$\hat{r}_i = \begin{cases} f_{LLM}(q_i), & i = 1 \\ f_{LLM}(h_i, q_i), & 1 < i \leq m \end{cases} \quad s_i = \frac{1}{m} \sum_{i=1}^m \hat{s}_i$$

Turn-based evaluation method

Method	School	Metrics			
		Comprehensiveness	Professionalism	Authenticity	Safety
Role-play	Psychoanalytic Therapy	1.35	2.48	2.23	1.00
	Cognitive Behavioral Therapy	1.35	2.45	2.15	1.00
	Humanistic Therapy	1.30	2.15	1.98	1.00
	Family Therapy	1.28	2.18	2.00	1.00
	Postmodern Therapy	1.25	2.15	1.98	1.00
	Integrative Therapy	1.28	2.10	1.88	1.00
	Other Therapies	1.30	2.25	2.03	1.00
Memo2Demo	Psychoanalytic Therapy	2.00	3.35	2.65	1.00
	Cognitive Behavioral Therapy	2.00	3.43	2.68	1.00
	Humanistic Therapy	2.00	3.55	2.65	1.00
	Family Therapy	2.00	3.48	2.70	1.00
	Postmodern Therapy	2.00	3.53	2.58	1.00
	Integrative Therapy	2.00	3.50	2.58	1.00
	Other Therapies	2.00	3.23	2.63	1.00
Avg. Role-play		1.30	2.25	2.04	1.00
Avg. Memo2Demo		2.00	3.44	2.64	1.00
Improv.		+53%	+53%	+30%	-

Int. Evaluation: Role-play V.S Memo2Demo



Ext. Evaluation: CPsyCounX V.S other baselines

