When Less Is More: An Empirical Study of Minimal Responses in Counseling Dialogues and the Behavior of LLMs

Anonymous ACL submission

Abstract

In psychological counseling, effective support is not always delivered through long, information-rich responses. Minimal responses, such as backchannel cues and concise empathic statements, help convey attentive listening, express empathy, and encourage clients to continue expressing themselves. However, existing counseling dialogue systems and evaluation frameworks often favor explicit, content-rich replies, overlooking the interactional value of brief counselor utterances. This paper systematically analyzes minimal responses across multiple counseling dialogue datasets. Using a two-stage filtering method based on utterance length and content, we show that minimal responses are common in human-collected datasets but substantially underrepresented in LLM-generated ones. We further evaluate current LLMs on manually curated dialogue contexts requiring minimal responses. The results show that strong general-purpose models can generate such responses when explicitly instructed, but their performance depends heavily on the model and prompting strategy. Meanwhile, existing counseling dialogue models still do not produce minimal responses reliably, often favoring longer, information-rich replies. These findings suggest that minimal responses are not yet robustly captured in counseling dialogue modeling and should be explicitly incorporated into training and evaluation.

1 Introduction

Mental health issues have attracted increasing attention in recent years (WHO, 2022; Arias et al., 2022), yet many individuals still lack timely access to professional support due to the shortage of counselors and the high cost of services (SAMHSA, 2015). This has motivated growing interest in large language models (LLMs) as virtual counselors (Liu et al., 2023; Qiu et al., 2024; Chen et al., 2023), making high-quality counseling dialogue data essential for both training and evaluation.

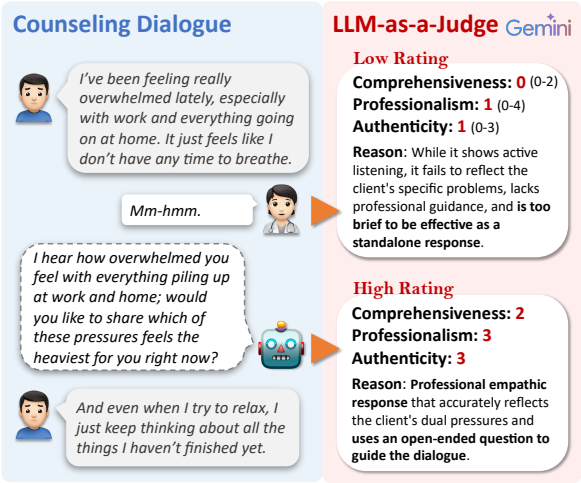

Figure 1: Minimal responses are a key technique in real counseling. However, current LLM-based evaluation methods (here using Zhang et al. (2024)'s prompt with Gemini) tend to undervalue such responses due to their brevity, favoring more elaborate replies.

Collecting real counseling dialogues, however, is challenging because they contain sensitive and private information. Existing public datasets therefore rely on indirect collection strategies, such as transcribing public counseling videos (AnnoMI (Wu et al., 2022)), collecting role-play dialogues with trained counselors (KokoroChat (Qi et al., 2025)), or anonymizing online counseling conversations by rewriting client utterances (PsyDial (Qiu and Lan, 2025)). Since such human-collected resources remain limited in scale, recent work has increasingly turned to LLM-synthesized counseling datasets, including SoulChat (Chen et al., 2023), SmileChat (Qiu et al., 2024), and CACTUS (Lee et al., 2024). In parallel, several evaluation frameworks have been proposed for counseling response generation, typically assessing dimensions such as informativeness, professionalism, realism, and safety (Zhang et al., 2024; Zhao et al., 2024; Li et al., 2026).

While these datasets and evaluation methods

have advanced counseling dialogue research, they often emphasize informative and structurally complete responses. This focus risks overlooking **minimal responses**, brief but interactionally important counselor utterances that play a central role in counseling practice. Examples such as "uh-huh" and "I see" are considered essential micro-skills: although they convey little propositional content, they express empathy, acceptance, and attentiveness without interrupting the client's narrative, thereby encouraging further self-disclosure (Ivey et al., 2018; Hill, 2019). Rather than providing information or direction, minimal responses function as continuers (Sacks et al., 1974; Schegloff, 1982a), facilitating client expression and supporting the therapeutic alliance in client-centered counseling (Rogers, 1957).

In contrast, we observe that counselor utterances in some LLM-generated counseling datasets often follow fixed, information-dense patterns, such as expressing empathy followed by a question[1], as illustrated in Figure 1. Although such responses can help advance the conversation, their structured and directive nature may increase clients' cognitive burden (Hill et al., 1988) and diverge from the more natural rhythm of human counseling interactions. This bias may also be inherited by models fine-tuned on synthetic data, leading them to generate verbose or templated responses. Moreover, utterance-level LLM-based evaluation may further reinforce this tendency by favoring informative and well-structured responses, thereby undervaluing brief but meaningful minimal responses.

Motivated by these observations, this paper investigates minimal responses in counseling dialogues. Our contributions are twofold. First, we propose a rule-based two-stage filtering method based on utterance length and content, and use it to analyze the distribution of minimal responses across multiple human-collected and LLM-generated counseling dialogue datasets. Second, using manually curated dialogue contexts, we evaluate whether current LLMs can generate minimal responses in appropriate situations. Our results show that strong general-purpose models can produce minimal responses when explicitly instructed, but this ability depends heavily on the model and prompting strategy. Existing counseling dialogue models, especially those fine-tuned on synthetic data, still do not reliably produce minimal responses, suggesting that this important counseling skill should be explicitly considered in future data construction, model training, and evaluation.

[1]For example, according to our analysis, 43.9% of the 24,220 counselor utterances across 3,084 dialogues in CPsyCounD (Zhang et al., 2024) end with a question mark. In PsyDTCorpus (Xie et al., 2025), the proportion is 67.4% among 86,054 counselor utterances across 4,760 dialogues.

2 Dataset Analysis

Since minimal responses in existing counseling dialogue datasets have not been systematically examined, this section analyzes seven datasets across Chinese, Japanese, and English. They include four LLM-generated datasets, CACTUS (Lee et al., 2024), CPsyCounD (Zhang et al., 2024), PsyDT-Corpus (Xie et al., 2025), and SmileChat (Qiu et al., 2024), and three human-collected datasets, PsyDial-D4 (Qiu and Lan, 2025), AnnoMI (Wu et al., 2022), and KokoroChat (Qi et al., 2025).

Table 1 reports the statistics of each dataset, including the number of dialogues and the average number of utterances. Because some datasets contain consecutive utterances from the same speaker, we merge adjacent utterances by the same speaker before conducting the analysis.

Minimal responses are not limited to simple backchannels such as "uh-huh" or "I see". Short expressions such as "this must be hard" or "you must feel very sad" can serve a similar function. To identify them more comprehensively, we use a rule-based two-stage filtering method. We first screen counselor utterances by length using thresholds, and then apply a content-based filter with predefined keyword lists. This filter removes short questions, conversation-advancing prompts, directives or suggestions, and short utterances such as thanks, apologies, greetings, polite closings, and simple encouragements, whose primary purpose is not to facilitate client expression. Since minimal responses often encourage clients to continue speaking and appear within longer client narratives, we also require the following client utterance to exceed a predefined length threshold. The keyword lists (Figure 2) and threshold settings are provided in the Appendix A. Table 1 presents the results after two-stage filtering. Minimal responses are almost absent from LLM-generated datasets but appear in all three human-collected datasets. AnnoMI shows the highest proportion, possibly because it is transcribed from face-to-face counseling, where such interactive feedback is more frequent. By contrast, PsyDial-D4 and KokoroChat are text-based online dialogues and contain fewer minimal responses.

Dataset	Lang.	#Dial.	Avg. Utter.	# Counselor Utter.	After Length Filtering (# / %)	After Content Filtering (# / %)	Human Validation (%)
LLM-generated datasets							
CACTUS (Lee et al., 2024)	EN	31,577	30.53	491,304	25 (0.01%)	0 (0.00%)	–
CPsyCounD (Zhang et al., 2024)	ZH	3,084	15.94	24,220	247 (1.02%)	0 (0.00%)	–
PsyDTCorpus (Xie et al., 2025)	ZH	4,760	36.16	86,054	431 (0.50%)	0 (0.00%)	–
SmileChat (Qiu et al., 2024)	ZH	55,165	11.38	309,698	351 (0.11%)	6 (0.00%)	–
Human-collected datasets							
PsyDial-D4 (Qiu and Lan, 2025)	ZH	2,382	75.59	90,031	5,518 (6.13%)	2,092 (2.32%)	73%
AnnoMI (Wu et al., 2022)	EN	133	72.64	4,859	1,501 (30.89%)	1,019 (20.97%)	82%
KokoroChat (Qi et al., 2025)	JA	6,589	71.63	237,735	12,773 (5.37%)	9,730 (4.09%)	89%

Table 1: Statistics of minimal responses in LLM-generated and human-collected counseling dialogue datasets. Minimal responses are identified using a two-stage filtering pipeline based on utterance length and content, with 100 sampled instances per dataset manually inspected to check the reliability of the filtering process.

To validate the filtering method, we randomly sample 100 filtered responses from each dataset and manually annotate them. Among the sampled responses, 73%, 82%, and 89% are manually confirmed as minimal responses, respectively, indicating that the filtering method is reasonably effective. We also examine where the filtered responses appear within dialogues, as detailed in Appendix A.1, which further supports the effectiveness and plausibility of the proposed rule-based filtering method.

3 Experiment

Experimental Setup. To examine whether current models can use minimal responses appropriately, we conduct response generation experiments across multiple languages. From the rule-filtered results of three human-collected datasets in Chinese, Japanese, and English, we manually extract 288, 300, and 300 dialogue history-response pairs, respectively, and verify with `gemini-3.1-flash-lite-preview` that the responses are minimal responses. Figure 6 in Appendix B shows concrete examples of the extracted minimal responses. We then use each dialogue history as input and ask each model to generate a counselor response. We compare open-source models such as Qwen and Llama, the commercial model GPT-5.4, counseling-domain fine-tuned models including SoulChat2.0, MindChat, KokoroChat-Full, and PsyDial-Pi4, and human counselor responses. Since general models may favor longer responses, we design an instructional prompt that encourages concise responses when the client is still expressing emotions or experiences. Details of the models and prompts are provided in the Appendix B.1.

Evaluation. We evaluate generated responses by response length and LLM-based judgment. Specifically, we use `gemini-3.1-flash-lite-preview` to assess whether each response is a minimal response and to rate its likelihood of interrupting the client's expression on a 0–5 scale, where lower is better. The evaluation prompt is provided in the Appendix B.2.

Results. As shown in Table 2, Llama-3 performs well on the English dataset AnnoMI under the instructional prompt, generating mostly simple minimal responses such as "Mm-hmm" and "I see". This may be because AnnoMI is transcribed from speech, where incomplete utterances often contain cues such as trailing hyphens (e.g., "And then-"), making it easier for the model to detect that the client has not finished speaking. We therefore do not include additional models or conduct LLM-based evaluation for English. For the Japanese and Chinese datasets, minimal-response generation depends strongly on both the model and the prompting strategy. The instructional prompt consistently improves performance, and GPT-5.4 achieves the best results, with minimal response rates of 87% and 97%, respectively, showing that strong general-purpose models can generate minimal responses when explicitly instructed. In contrast, counseling dialogue models do not produce minimal responses reliably. Models fine-tuned on synthetic counseling data, namely SoulChat2.0 and MindChat, rarely generate such responses, with rates of 0% and 1.76%, respectively, and instead tend to produce longer, more information-rich replies. Models fine-tuned on human-collected data show better but still limited performance: after translating the extracted

Model	Length-based statistics			LLM-based evaluation	
	Avg. Len.	< 10 (%)	< 5 (%)	MR (%)	Interrupt. ↓
Chinese dataset: PsyDial-D4					
Qwen-8B (general prompt)	76.15	0.00	0.00	0.00	3.48
Qwen-8B (instructional prompt)	24.24	22.54	18.31	24.65	1.73
GPT-5.4 (general prompt)	132.05	0.00	0.00	0.00	3.51
GPT-5.4 (instructional prompt)	8.52	96.48	55.99	97.18	0.05
SoulChat2.0	52.10	0.00	0.00	0.00	2.86
MindChat	58.11	0.00	0.00	1.76	3.64
KokoroChat-Full (input translated into Japanese)	46.81	7.39	2.46	13.73	3.48
Human	5.30	93.66	44.37	100.00	0.04
Japanese dataset: KokoroChat					
Llama-3.1-Swallow (general prompt)	41.29	0.00	0.00	5.33	2.62
Llama-3.1-Swallow (instructional prompt)	20.32	11.33	4.67	40.33	1.69
GPT-5.4 (general prompt)	66.41	0.00	0.00	0.00	2.59
GPT-5.4 (instructional prompt)	10.05	77.67	7.00	87.33	0.26
PsyDial-Pi4 (input translated into Chinese)	12.99	45.00	4.67	54.00	1.08
Human	3.80	96.00	69.33	100.00	0.02
English dataset: AnnoMI					
Llama-3 (general prompt)	49.01	0.33	0.00	–	–
Llama-3 (instructional prompt)	2.13	98.33	95.33	–	–
Human	1.08	100.00	100.00	–	–

Table 2: Results of minimal-response generation. Avg. Len. denotes average response length; MR denotes the percentage judged as minimal responses; Interrupt. denotes the average interruption score, where lower is better. Row colors indicate model categories: blue for base open-source models, orange for synthetic-data fine-tuned models, and purple for human-data fine-tuned models.

dialogues between Japanese and Chinese to avoid possible training-data overlap, KokoroChat-Full and PsyDial-Pi4 achieve minimal response rates of 13.73% and 54.00%, respectively. Overall, these results suggest that minimal responses are not yet robustly captured in counseling dialogue modeling. While strong general-purpose models can generate them with explicit guidance, counseling-domain models, especially those fine-tuned on synthetic data, still struggle to use them appropriately, highlighting the need to explicitly incorporate minimal responses into training and evaluation.

4 Related Work

Backchannels have been widely studied in conversation analysis and spoken dialogue systems that signal attention, understanding, or agreement without taking the conversational floor (Yngve, 1970; Sacks et al., 1974; Schegloff, 1982b). Prior work has explored automatic backchannel prediction and generation, including when to produce a backchannel and which form to use (Ward and Tsukahara, 2000; Gravano and Hirschberg, 2011; Kawahara et al., 2016; Inoue et al., 2025), highlighting their role in making human-human communication smoother and more natural.

Minimal responses in counseling are related to backchannels, but serve more specific therapeutic functions. Brief acknowledgments and encouragers help counselors express empathy, acceptance, and attentiveness while allowing clients to continue exploring their experiences (Ivey et al., 2018; Hill, 2019), which is consistent with client-centered counseling (Rogers, 1957). In text-based online counseling, where nonverbal cues are limited, short responses such as “I see” or “that sounds difficult” can act as lightweight signals of presence and support. However, such responses remain underexplored in counseling dialogue datasets and LLM-based systems. This work addresses this gap by analyzing their distribution and evaluating whether current LLMs can generate them appropriately.

5 Conclusion

This paper investigated minimal responses in psychological counseling dialogues. We found that minimal responses are common in human-collected datasets but largely absent from LLM-generated ones, highlighting a key gap in synthetic counseling data. Generation experiments showed that strong general-purpose models can produce minimal responses with explicit instruction, whereas existing counseling dialogue models, especially those fine-tuned on synthetic data, still fail to use them reliably. These findings suggest that minimal responses should be explicitly considered in future data construction, model training, and evaluation.

Limitations

This study has several limitations. First, our analysis of minimal-response distribution is based on rule-based filtering. Although the rules are designed to be transparent and are supported by manual validation, they may still miss some context-dependent minimal responses. In future work, we plan to involve counseling experts to annotate actual dataset samples, enabling a more systematic analysis of minimal responses.

Second, we use translation to reduce the possibility that manually collected dialogues overlap with the training data of the evaluated models. However, translation quality may introduce bias and affect model behavior, particularly because minimal responses often rely on subtle contextual cues.

Finally, counseling is a long-term dialogue process, while our experiments evaluate response generation only in specific local contexts. Future work should explore suitable methods for assessing the impact of minimal responses on long-term counseling outcomes and further validate them with real human clients.

Ethical Considerations

This study uses existing publicly available counseling dialogue datasets in accordance with the terms of their respective licenses. To the best of our knowledge, these datasets had undergone anonymization or privacy screening before public release. We do not collect new data from real clients or deploy any system for actual counseling. Since counseling dialogues may contain sensitive information, we analyze the data only at the aggregate level and do not attempt to identify or interpret individual clients.

Our findings should not be interpreted as suggesting that LLMs can provide professional counseling. Minimal responses are only one type of counseling-related interactional behavior, and their appropriateness depends on the broader therapeutic context. This work aims to highlight an overlooked aspect of counseling dialogue modeling and to support the development of safer and more realistic counseling dialogue systems.

References

Daniel Arias, Shekhar Saxena, and Stéphane Verguet. 2022. Quantifying the global burden of mental disorders and their economic value. *eClinicalMedicine*, 54:101675.

Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, and Fei Huang et al. 2023. Qwen technical report.

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.

Kazuki Fujii, Taishi Nakamura, Mengsay Loem, Hiroki Iida, Masanari Ohi, Kakeru Hattori, Hirai Shota, Sakae Mizuki, Rio Yokota, and Naoaki Okazaki. 2024. Continual pre-training for cross-lingual llm adaptation: Enhancing japanese language capabilities. In *Proceedings of the First Conference on Language Modeling*, COLM, University of Pennsylvania, USA.

Agustín Gravano and Julia Hirschberg. 2011. Turn-taking cues in task-oriented dialogue. *Computer Speech & Language*, 25(3):601–634.

Clara Hill, Janet Helms, Victoria Tichenor, Sharon Spiegel, Kevin O'Grady, and Elgin Perry. 1988. Effects of therapist response modes in brief psychotherapy. *Journal of Counseling Psychology*, 35:222–233.

Clara E. Hill. 2019. *Helping Skills: Facilitating Exploration, Insight, and Action*, 5th edition. American Psychological Association, Washington, DC.

Koji Inoue, Divesh Lala, Gabriel Skantze, and Tatsuya Kawahara. 2025. Yeah, un, oh: Continuous and real-time backchannel prediction with fine-tuning of voice activity projection. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1: Long Papers*, pages 7171–7181. Association for Computational Linguistics.

Allen E. Ivey, Mary Bradford Ivey, and Carlos P. Zalaquett. 2018. *Intentional Interviewing and Counseling: Facilitating Client Development in a Multicultural Society*, 9th edition. Cengage Learning, Boston, MA.

Tatsuya Kawahara, Takashi Yamaguchi, Koji Inoue, Katsuya Takanashi, and Nigel Ward. 2016. Prediction and generation of backchannel form for attentive listening systems. In *Proceedings of Interspeech 2016*, pages 2890–2894.

Suyeon Lee, Sunghwan Kim, Minju Kim, Dongjin Kang, Dongil Yang, Harim Kim, Minseok Kang, Dayi Jung, Min Hee Kim, Seungbeen Lee, Kyong-Mee Chung, Youngjae Yu, Dongha Lee, and Jinyoung Yeo. 2024. Cactus: Towards psychological counseling conversations using cognitive behavioral theory. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 14245–14274, Miami, Florida, USA. Association for Computational Linguistics.

Yahan Li, Jifan Yao, John Bosco S. Bunyi, Adam C Frank, Angel Hsing-Chi Hwang, and Ruishan Liu. 2026. Counselbench: A large-scale expert evaluation and adversarial benchmarking of large language models in mental health question answering. In *The Fourteenth International Conference on Learning Representations*.

June M. Liu, Donghao Li, He Cao, Tianhe Ren, Zeyi Liao, and Jiamin Wu. 2023. Chatcounselor: A large language model for mental health support. *Preprint*, arXiv:2309.15461.

Naoaki Okazaki, Kakeru Hattori, Hirai Shota, Hiroki Iida, Masanari Ohi, Kazuki Fujii, Taishi Nakamura, Mengsay Loem, Rio Yokota, and Sakae Mizuki. 2024. Building a large japanese web corpus for large language models. In *Proceedings of the First Conference on Language Modeling*, COLM, University of Pennsylvania, USA.

Zhiyang Qi, Takumasa Kaneko, Keiko Takamizo, Mariko Ukiyo, and Michimasa Inaba. 2025. KokoroChat: A Japanese psychological counseling dialogue dataset collected via role-playing by trained counselors. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12424–12443, Vienna, Austria. 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. 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. 2025. PsyDial: A large-scale long-term conversational dataset for mental health support. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 21624–21655, Vienna, Austria. Association for Computational Linguistics.

Carl R. Rogers. 1957. The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology*, 21(2):95–103.

Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735.

SAMHSA. 2015. Behavioral health trends in the united states: Results from the 2014 national survey on drug use and health. Annual report, Substance Abuse and Mental Health Services Administration.

Emanuel A. Schegloff. 1982a. Discourse as an interactional achievement: Some uses of "uh huh" and other things. *Analyzing Discourse: Text and Talk*, pages 71–93.

Emanuel A. Schegloff. 1982b. Discourse as an interactional achievement: Some uses of "uh huh" and other things that come between sentences. In Deborah Tannen, editor, *Analyzing Discourse: Text and Talk*, pages 71–93. Georgetown University Press, Washington, DC.

Nigel Ward and Wataru Tsukahara. 2000. Prosodic features which cue back-channel responses in english and japanese. *Journal of Pragmatics*, 32(8):1177–1207.

WHO. 2022. World mental health report: Transforming mental health for all.

Zixiu Wu, Simone Balloccu, Vivek Kumar, Rim Helaoui, Ehud Reiter, Diego Reforgiato Recupero, and Daniele Riboni. 2022. Anno-mi: A dataset of expert-annotated counselling dialogues. In *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6177–6181.

Haojie Xie, Yirong Chen, Xiaofen Xing, Jingkai Lin, and Xiangmin Xu. 2025. PsyDT: Using LLMs to construct the digital twin of psychological counselor with personalized counseling style for psychological counseling. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1081–1115, Vienna, Austria. Association for Computational Linguistics.

Victor H. Yngve. 1970. On getting a word in edgewise. In *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*, pages 567–578, Chicago, IL. Chicago Linguistic Society.

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.

Haiquan Zhao, Lingyu Li, Shisong Chen, Shuqi Kong, Jiaan Wang, Kexin Huang, Tianle Gu, Yixu Wang, Jian Wang, Liang Dandan, Zhixu Li, Yan Teng, Yanghua Xiao, and Yingchun Wang. 2024. ESC-eval: Evaluating emotion support conversations in large language models. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 15785–15810, Miami, Florida, USA. Association for Computational Linguistics.

A Details of Rule-Based Filtering

Before filtering, adjacent utterances produced by the same speaker are merged into a single turn. We then apply a two-stage filtering pipeline. First, counselor utterances are filtered by length using

Chinese

Category	Excluded patterns
Questions	?, 为什么, 怎么, 怎样, 如何, 什么, 哪个, 哪种, 哪里, 谁, 何时, 几点, 后来呢, 然后呢, 那之后呢, 比如, 例如, 具体, 详细说
Directives and suggestions	请, 可以试试, 你可以, 你要不要, 建议你, 不如, 最好, 应该, 试着, 说说看, 再说说
Thanks and non-listening short expressions	谢谢, 感谢, 下周, 这周, 在呢, 需要, 早, 晚安, 准时, 时间, 周, 今天
Apologies	对不起, 抱歉, 不好意思
Greetings and closings	你好, 您好, 早上好, 晚上好, 再见, 拜拜, 回头见, 辛苦了
Service expressions	不客气, 很高兴帮, 乐意帮, 很开心帮, 希望帮到, 开心
Encouragement	加油, 坚持, 祝福, 祝你, 祝您, 一切都会好, 会好起来, 没事的, 别担心

Japanese

Category	Excluded patterns
Questions	?, ですか, ますか, でしょうか, ませんか, どう, なぜ, なんで, 何, いつ, どこ, 誰, どんな, どのよう, 具体的, それで, その後, 例えば, ちなみに
Directives and suggestions	ください, ましょう, したほうが, するといい, してみて, 話してください, 教えてください, 大丈夫ですか, 平気ですか
Thanks and apologies	ありがとう, ありがとうございます, 有難う, 感謝, すみません, 申し訳, ごめんなさい, 失礼
Greetings and closings	こんにちは, こんばんは, おはよう, よろしく, お願いいたします, お願いします, お疲れ様, 失礼します, 始めます, 終了, 終わり, ではまた, またお話, またね
Service expressions and encouragement	了解, 承知, こちらこそ, 応援, 頑張, だいじょうぶ, 大丈夫ですよ, よかったです

English

Category	Excluded patterns
Questions	Question-final expressions; what, why, how, when, where, who; auxiliary-led questions beginning with do, did, does, are, is, was, were, can, could, would, will, have, has, had, shall, and should
Conversation-advancing prompts	and then, what else, for example, tell me, go on, say more, anything else, what happened, then what
Directives and suggestions	please, try, maybe, you should, you could, let's, consider, think about, remember, just, are you okay, is that okay, are you alright
Thanks and apologies	thank you, thanks, appreciate it, sorry, I apologize, apologies, excuse me
Greetings and closings	hello, hi, hey, good morning, good afternoon, good evening, take care, see you, goodbye, bye, you're welcome, have a nice day, have a good day
Service expressions and encouragement	glad to help, happy to help, here for you, feel free, let me know, good luck, you can do it, hang in there, stay strong, proud of you, hope things get better

Figure 2: Language-specific exclusion patterns used in content-based filtering. The three panels show the original exclusion lists for Chinese, Japanese, and English, respectively.

language-specific thresholds. Second, short utterances whose main function is not to facilitate client expression are removed using predefined exclusion patterns. As shown in Table 3, we also require the immediately following client utterance to exceed a language-specific minimum length, because minimal responses are expected to encourage continued client narration.

Lang.	Counselor utt.	Next client utt.
Chinese	$\leq$ 10 chars.	$\geq$ 10 chars.
Japanese	$\leq$ 10 chars.	$\geq$ 5 chars.
English	$\leq$ 20 chars.	$\geq$ 20 chars.

Table 3: Language-specific length thresholds used in rule-based filtering. Counselor utt. denotes the maximum length of a candidate counselor utterance, and Next client utt. denotes the minimum length required for the immediately following client utterance.

Figure 2 shows the language-specific exclusion patterns used in the content-based filtering stage. These patterns remove short utterances that are questions, conversation-advancing prompts, directives, suggestions, thanks, apologies, greetings, closings, service expressions, or simple encouragements. Although such utterances can be short, their primary function differs from that of minimal responses, which aim to acknowledge the client and facilitate further expression without interrupting the narrative.

A.1 Stage-wise Analysis of Filtered Responses

We further analyze the stage-wise changes of the filtered counselor responses in the three human-collected datasets. Each dialogue is divided into five equal stages, and we calculate the proportion of filtered responses among counselor utterances in each stage. As shown in Figure 3, KokoroChat and AnnoMI show a similar pattern: the filtered responses are frequent in the early and early-middle stages, increase at first, and then gradually decrease as the dialogue progresses. This pattern is consistent with the expected function of minimal responses, as counselors often use them in earlier phases to signal listening and encourage clients to continue speaking, before shifting toward more informative responses in later phases.

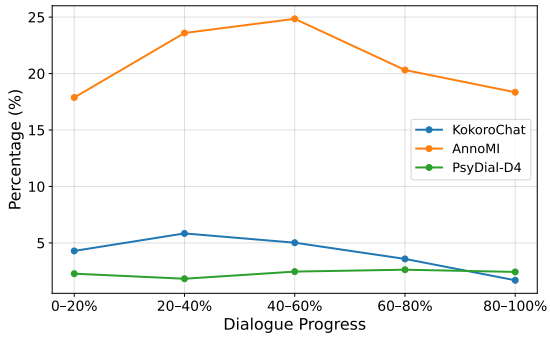

Figure 3: Stage-wise proportions of filtered counselor responses in human-collected datasets. Each dialogue is divided into five equal stages, and the proportion is calculated over counselor utterances in each stage.

By contrast, PsyDial-D4 does not show the same stage-wise trend. A closer inspection indicates that it consists of multiple counseling dialogues with the same client at different time points, rather than single continuous counseling sessions. Its overall structure therefore does not fully follow the typical

progression of a single session, from relationship building and problem exploration to deeper discussion and feedback. This cross-time, multi-session structure weakens the validity of stage-based analysis and may explain why PsyDial-D4 differs from KokoroChat and AnnoMI.

Overall, the stage-wise patterns provide additional support for the effectiveness and plausibility of the proposed two-stage rule-based filtering method.

B Experimental Details

B.1 Models and Inference Prompts

Figure 6 shows concrete examples of the extracted minimal responses.

Models. We evaluate the model variants reported in Table 2. The compared systems include general open-source models, an advanced commercial model, counseling-domain fine-tuned models, and human counselor responses. All model-based experiments are conducted in an inference-only setting; we do not update model parameters or perform additional fine-tuning. Unless otherwise specified, we use the original or default inference configurations provided for each model, including decoding and generation parameters.

- **Qwen-8B** (Bai et al., 2023) We use Qwen3-8B[2] as the general open-source baseline for Chinese. This model is used to examine whether a general-purpose instruction-following model can generate minimal responses without counseling-specific fine-tuning.
- **GPT-5.4.** We use GPT-5.4[3] as one of the strongest commercial models. It serves as an upper-bound reference for evaluating whether advanced proprietary LLMs can follow the minimal-response instruction.
- **SoulChat2.0** (Xie et al., 2025) We use SoulChat2.0-Llama-3.1-8B[4], a counseling-oriented model from the SoulChat2.0 project. This model is included to examine whether models fine-tuned on synthetic counseling data can capture minimal-response behavior.
- **MindChat** We use MindChat-Qwen-7B-v2[5], a Chinese mental-health dialogue model based on Qwen. It is also used as a representative model fine-tuned for psychological support conversations.
- **KokoroChat-Full** (Qi et al., 2025) We use Llama-3.1-KokoroChat-Full[6], which is fine-tuned on Japanese human-collected counseling dialogues. Since the model is trained on Japanese data, we translate the Chinese input dialogues into Japanese before inference.
- **Llama 3.1 Swallow** (Fujii et al., 2024; Okazaki et al., 2024) We use Llama-3.1-Swallow-8B-Instruct-v0.3[7] as the general open-source baseline for Japanese. This model is adapted for Japanese instruction-following and is used to test minimal-response generation without counseling-specific fine-tuning.
- **PsyDial-Pi4** (Qiu and Lan, 2025) We use PsyDial-Pi4[8], which is fine-tuned from Qwen2.5-7B-Instruct on PsyDial-D4. Since this model is trained on Chinese counseling data, we translate the Japanese input dialogues into Chinese before inference.
- **Llama-3** We use Meta-Llama-3-8B-Instruct[9] as the English open-source baseline. This model is included to test whether a general English instruction-following model can generate minimal responses in contexts extracted from AnnoMI.

Inference prompts. We use two prompt settings for the general open-source models and GPT-5.4. The general prompt asks the model to generate the next counselor response based only on the dialogue history. The instructional prompt further encourages the model to use concise minimal responses when the client is still expressing emotions or experiences. Figure 4 shows the prompt templates used for response generation.

B.2 LLM-based Evaluation Prompt

For LLM-based evaluation, we use `gemini-3.1-flash-lite-preview` to assess each generated counselor response from two

[2] https://huggingface.co/Qwen/Qwen3-8B
[3] https://platform.openai.com/docs/models
[4] https://modelscope.cn/models/YIRONGCHEN/SoulChat2.0-Llama-3.1-8B
[5] https://huggingface.co/X-D-Lab/MindChat-Qwen-7B-v2
[6] https://huggingface.co/UEC-InabaLab/Llama-3.1-KokoroChat-Full
[7] https://huggingface.co/tokyotech-llm/Llama-3.1-Swallow-8B-Instruct-v0.3
[8] https://huggingface.co/qiuhuachuan/PsyDial-Pi4
[9] https://huggingface.co/meta-llama/Meta-Llama-3-8B-Instruct

Chinese	Japanese	English
General prompt	**General prompt**	**General prompt**
你是一名专业心理咨询师。 请根据对话历史生成咨询师回复。 只输出回复，不要解释，不要添加额外内容。	あなたはプロの心理カウンセラーです。 対話履歴に基づいて、次のカウンセラーの応答を生成してください。 応答のみを出力し、説明や余計な内容は加えないでください。	You are a professional psychological counselor. Based on the dialogue history, generate the next counselor response. Output only the response, without explanation or extra text.
Instructional prompt	**Instructional prompt**	**Instructional prompt**
你是一名专业心理咨询师。 请根据对话历史生成下一句咨询师回复。 当来访者正在表达情绪或经历时，应优先仅使用简短回应（如"嗯嗯"，"我明白"，"我理解"等）来鼓励其继续表达，并且不做多余回复，把话语权交给来访者。 仅在确实需要时才提供较完整的回应。 只输出一句回复，不要解释。	あなたはプロの心理カウンセラーです。 対話履歴に基づいて、次のカウンセラーの応答を生成してください。 来談者が感情や体験を語っているときは、相手がさらに話し続けられるように、「うんうん」「はい、」「ええ、」などの短い応答のみを優先して用い、余計な返答はせず、話す主導権を来談者に委ねてください。 本当に必要な場合にのみ、やや長めの応答をしてください。 出力は一文の応答のみとし、説明は書かないでください。	You are a professional psychological counselor. Based on the dialogue history, generate the next counselor response. When the client is expressing emotions or describing experiences, you should primarily use a short minimal response (e.g., 'Mm-hmm.', 'Yeah.', 'I see.') to encourage them to continue speaking. Do not add extra content in such cases and let the client keep the floor. Only provide a more complete response when it is truly necessary. Output only ONE response, without explanation.

Figure 4: Prompt templates used for response generation.

perspectives. First, the model judges whether the response is a minimal response. Second, it assigns an interruption risk score from 0 to 5, where lower scores indicate a lower likelihood of interrupting the client's expression. The model is instructed to return only a JSON object containing these two fields. Figure 5 shows the evaluation prompts.

B.3 Translation Procedure for Cross-Lingual Evaluation

To avoid possible overlap between the evaluation samples and the training data of counseling-domain fine-tuned models, we translate the extracted dialogue contexts between Japanese and Chinese before evaluation. Specifically, for KokoroChat-Full, which is fine-tuned on Japanese counseling dialogues, we translate the Chinese samples from PsyDial-D4 into Japanese. Conversely, for PsyDial-Pi4, which is fine-tuned on Chinese counseling dialogues, we translate the Japanese samples from KokoroChat into Chinese. This setting allows each model to receive inputs in its expected language while reducing the risk that the evaluation instances directly overlap with its training data.

We use `gemini-3.1-flash-lite-preview` for translation with a simple instruction prompt. The prompt is shown below:

Prompt for Dialogue Translation

 # Task

Please translate the following counseling dialogue into `{target language}`.

 # Requirements

1. Translate the dialogue sentence by sentence, without omitting any utterance.
2. Preserve the dialogue structure and speaker labels (`counselor` / `client`).
3. Do not summarize, rewrite, or explain the dialogue.
4. Output only the complete translation.
5. Carefully translate the last 10 utterances without omission.
6. Even short responses such as "yes", "uh-huh", and "I see" must be translated.

 # Dialogue

`{dialogue}`

Chinese

```
你是一名心理咨询对话分析专家。

任务：
根据给定的【对话历史】和【候选回复】，判断该回复在当前语境中的性质。

你需要完成两个判断：

（1）Minimal Response 判定
该回复是否属于 minimal response（简短回应）？

判定标准：
- minimal response 通常是简短、低信息量的回应
- 主要作用是鼓励来访者继续表达
- 不引入新信息、不进行分析、不提出新问题
- 示例：
  - “嗯嗯”
  - “我明白”
  - “是的”
  - “听起来很不容易”

输出：
- yes（是 minimal response）
- no（不是 minimal response）

（2）打断风险评分（0–5）
判断该回复在当前对话中“打断来访者表达”的可能性

评分标准：
- 0：完全不会打断（非常自然地鼓励继续表达）
- 1：几乎不会打断
- 2：轻微可能打断
- 3：中等可能打断（开始引导或转移）
- 4：较高可能打断（明显改变话题或节奏）
- 5：很可能打断（直接主导对话、给建议、连续提问等）

重点考虑：
- 来访者是否还在表达情绪或经历
- 回复是否抢夺话语权
- 回复是否过长、分析过多或提问过多

————————————————
【对话历史】
{dialogue_history}

【候选回复】
{response}
————————————————

请严格按照以下 JSON 格式输出，不要添加任何解释：

{
  "minimal": "yes 或 no",
  "interrupt_score": 0-5
}
```

Japanese

```
あなたは心理カウンセリング対話の分析専門家です。

【タスク】
与えられた【対話履歴】と【候補応答】に基づき、その応答の性質を評価してください。

以下の2点について判断してください：

（1）Minimal Response（最小応答）の判定
この応答は minimal response に該当しますか？

判定基準：
- minimal response とは、簡潔で情報量が少ない応答である
- 主な目的は、クライアントの発話を促し、継続させること
- 新しい情報を導入しない
- 分析や解釈を行わない
- 新たな質問をしない

例：
-「うんうん」
-「分かります」
-「そうですね」
-「それは大変でしたね」

出力：
- yes（minimal response である）
- no（minimal response ではない）

（2）発話遮断リスク（0〜5）
この応答がクライアントの発話を「遮断する可能性」を評価してください。

評価基準：
- 0：全く遮断しない（自然に発話継続を促す）
- 1：ほぼ遮断しない
- 2：わずかに遮断の可能性あり
- 3：中程度の遮断可能性（話題の誘導や転換が見られる）
- 4：高い遮断可能性（明確に話題や流れを変える）
- 5：非常に高い遮断可能性（主導的、助言中心、連続した質問など）

特に以下を考慮してください：
- クライアントがまだ感情や経験を語っている途中か
- 応答が発話の主導権を奪っていないか
- 応答が長すぎないか、分析しすぎていないか、質問が多すぎないか

————————————————
【対話履歴】
{dialogue_history}

【候補応答】
{response}
————————————————

以下のJSON形式で厳密に出力してください。説明は不要です：

{
  "minimal": "yes または no",
  "interrupt_score": 0-5
}
```

English

```
You are an expert in psychological counseling dialogue analysis.

Task:
Given the [dialogue history] and a [candidate response], determine the nature of the response in the current context.

You need to make two judgments:

(1) Minimal Response Judgment
Determine whether the response is a minimal response.

Criteria:
- A minimal response is usually brief and low in information content.
- Its main function is to encourage the client to continue expressing themselves.
- It does not introduce new information, provide analysis, or ask a new question.
- Examples:
  - “Mm-hmm”
  - “I see”
  - “Yes”
  - “That sounds very difficult”

Output:
- yes (the response is a minimal response)
- no (the response is not a minimal response)

(2) Interruption Risk Score (0-5)
Judge how likely the response is to interrupt the client’s expression in the current dialogue.

Scoring criteria:
- 0: Does not interrupt at all; naturally encourages continued expression.
- 1: Almost does not interrupt.
- 2: Slightly likely to interrupt.
- 3: Moderately likely to interrupt; begins to guide or shift the dialogue.
- 4: Highly likely to interrupt; clearly changes the topic or pace.
- 5: Very likely to interrupt; directly takes control of the dialogue, gives advice, or asks multiple questions.

Key factors to consider:
- Whether the client is still expressing emotions or experiences.
- Whether the response takes the conversational floor away from the client.
- Whether the response is too long, overly analytical, or contains too many questions.

————————————————
[Dialogue History]
{dialogue_history}

[Candidate Response]
{response}
————————————————

Strictly output the result in the following JSON format. Do not add any explanation:

{
  "minimal": "yes or no",
  "interrupt_score": 0-5
}
```

Figure 5: LLM-based evaluation prompts for judging minimal responses and interruption risk. The Chinese and Japanese prompts are used in the actual experiments, while the English version is translated by ChatGPT and is not used in the experiments.

Chinese (PsyDial-D4)

Client: 是的，特别是有一个朋友，他真的很活泼，总是能够带动大家的气氛。
Counselor: 有这样活泼的朋友在身边一定很棒，和他在一起你有什么特别的感受吗？
Client: 感觉很好，他总是很有正能量，让我觉得世界其实没那么糟。
Counselor: 嗯嗯。
Client: 他还经常带我们去参加一些活动，感觉生活一下子丰富了很多。

Client: 上周我做了一些曲奇，按照你的建议尝试了一些新的事情。
Counselor: 你做了曲奇啊。整个过程怎么样？当时是什么样的心情呢？
Client: 一开始有点紧张，因为我之前很少做这些烘焙的东西。但后来渐渐地就放松了，觉得挺有意思的。
Counselor: 嗯嗯~
Client: 而且做完后和朋友们分享，他们也很喜欢，感觉自己获得了一些认可和成就感。

Client: 嗯，第一点很有共鸣~当觉察到跟另一个生命产生共鸣和联结的时候，会自然催生出一股疗愈的力量。
Counselor: 嗯，这种共鸣和联结确实很重要，让我们感觉到不再孤单。你觉得这些体验对你来说意味着什么？
Client: 是的，不过，最近我有点厌倦这种互相激励的方式。因为我发现自己在帮别人时，内心的压力和焦虑越来越大。
Counselor: 嗯，你说。
Client: 就是那种消耗感越来越明显。我觉得我内心的平衡被打破了，开始感到厌倦和无力。

Client: 是的，我认为自己性格比较踏实，愿意学习，但也比较容易焦虑和过度担心。
Counselor: 看起来这些特质在工作中对你既是一种优势，也可能会让你觉得不太适应某些环境。
Client: 嗯，对，有时候会觉得自己的这些特质在面对高压力和快速节奏的工作时显得不够适合。
Counselor: 嗯嗯
Client: 我常常感觉到自己的进度慢，害怕自己拖团队的后腿。

Client: 有时候在实验室倒是还能正常交流，但一回寝室，她又陷入了那种低迷的状态。
Counselor: 在实验室你们还能交流，但回到寝室她又变得很低落了，对吗？
Client: 是的，必要的交流不算，她大多数时间还是闷闷不乐。
Counselor: 嗯嗯，我明白。
Client: 我们也试过找辅导员想调寝室，但没成功。

Client: 是的，我现在有些不知道该怎么办，一方面我很喜欢他，另一方面又觉得他不理解我。
Counselor: 你说他很自私，除了这件事情，还有其他事情也让你觉得他自私吗？
Client: 是的，他有时候只顾工作，很少会主动花时间陪我。
Counselor: 明白了，
Client: 我觉得我在这段关系中得不到应有的关心和重视。

Client: 当然也有其他的想聊，比如我和朋友们的关系。
Counselor: 明白了，朋友关系对你来说很重要，是吗？
Client: 我真的很在意朋友们。
Counselor: 嗯嗯，可以理解。
Client: 想知道他们有没有建一个没有我的群。

Japanese (KokoroChat)

Client: はいこちらは震度5弱でしたその日は休日で家にいました
Counselor: ええ！？それは怖かったですよね…大丈夫だったのですか？？
Client: ものが倒れたとかは職場も家もなかったですが、
Counselor: ええ
Client: 揺れが大きかったことで、不安が募りました

Client: はい。コロナで事業が危うくなりまして、
Counselor: （お聞きしていますよ）
Client: 資金のことを役所に相談しているんですが、お調べしますと言われ、忙しいんでしょうね、はっきりとした返事がもらえないまま
Counselor: ええ。
Client: 3週間も経ちました。もう、待つ余裕も正直なくて、

Client: そうですね
Counselor: 上司の方はどんな方ですか？
Client: 上司は、仕事の目的はおしえてくれるのですが
Counselor: うんうん
Client: やり方はさっぱり教えてくれず。まるで分からないので、とりあえずこちらから提案しようと、色々と考えながら進めてきたんですが

Client: 養育費を払ってて、それが結構高いんです、だから彼だけの収入じゃきついんですよね
Counselor: ええ
Client: 今の仕事の環境も、たとえどんなに悪くても絶対将来自分のためになるし
Counselor: はい、
Client: やりたくもない仕事における嫌なことを乗り越えてきたなら、やりたいことにおける嫌なことはもっと乗り越えられるってプラスにずっととらえてきたんですでもちょっと疲れてきちゃったなって

Client: はい、主人は収入は少ないですが、ちゃんと真面目に仕事してくれていますのでこれ以上どうしようもないので、私が仕事しないと主人が悪い訳でありませんので
Counselor: ご主人のことも、大切に思っておられるのですね。
Client: ですが、私が仕事する事に反対するので…借金なんて
Counselor: はい
Client: 多分、両親から借りると思うけど、今までも借りてきたので、もう借りてほしくないと思います返せなくなってしまいます

Client: そうですね、昨年の夏休み明けからです。半年ほど経ちました家族以外の誰とも話すこともないまま毎日家にいます。
Counselor: そうでしたか、半年もそんな生活をしていたらしんどくなるもの無理ないですよね・・・
Client: ええ、この先どうしたらいいか分からず
Counselor: はい
Client: 自分だけ社会から置いてけぼりな気もしています。

Client: がまんするしかないんでしょうか
Counselor: 我慢することはないと思います。どうしていったらいいのか一緒に考えていきたいとおもいます。さっきめぐみさんの言葉の中で、はっきり文句みたいに言って「わかっているじゃない」って感じになりたくないというお言葉があったのですが、今ひとつ私がそのあたりの意味がうまく捉えられていないので教えていただけますか？
Client: 上司が好意を持っているってわかっていないようなふりで、にげているので
Counselor: はい
Client: 曖昧な言い方をされているのに、「髪の毛キレイ」とか言わないでくださいとか、はっきり返すと、僕の好意はわかっているんだよねみたいな流れになりそうで

English (AnnoMI)

Client: No. It's just overwhelming when you see them everything they're, you know?
Counselor: Right. Well, so where would you-- I mean, if you-you had to pick somewhere, one or two kind of high priorities or-or things to talk about in our time this morning, where would you wanna start?
Client: I think- I think the thoughts about the dog has been definitely bothering me, that's been one.
Counselor: Okay.
Client: And, um, my-my asthma is out of control too. So maybe those two.

Client: Um, a little bit less junk food. Um, like at times, when I'm like I feel like, "Oh, I think I'll just go have a candy bar," or some worthless thing like that I'll actually have second thoughts before I do it.
Counselor: So you're at the second thought part. And what do you think would-- might take you, uh, over to actually doing it if you wanted to?
Client: Well, I think, uh, when I really start seeing some results at the gym, it's kinda, uh, I think the way it works in my mind is if I've worked out really hard and I've- and I've-- say I've lost a pound that week-
Counselor: Mm-hmm.
Client: -you know, and then as I'm looking at that double or triple fudge four-decker of ice cream Sunday thing, you know, that somebody is eating, I would say, "Wow, that would be about two hours on the treadmill." Then-then that would be probably-- the-- that would keep me from actually ordering one of those.

Client: Yeah. And, um—
Counselor: Do you have any idea why that is?
Client: I do have an idea, I think, why that is. Um, I think that there's been a few things that have happened recently and something that really came to my awareness, when I visited with my family, is that I have consistently through my whole life, probably, put other people first. And I have consistently, uh, almost not even considered myself in the equation. It was, uh, kind of sad in a way, at the time that I realized it. Uh, I didn't realize how severe it actually was, but I was kind of glad that I realized it because I feel like it's never too late to change-
Counselor: True.
Client: -and I feel like I can- I can, uh, respect and value myself just as much as I have other people. I know that's important. And I feel like when I do that, I'm a better person for other people as well.

Client: I-I just hear it every day.
Counselor: Yeah. And what's been your past experience with smoking? Tell me-- Can you tell me a little bit about kind of how the role it plays in your life, how long you've smoked?
Client: I've smoked for 40 years, since I was 14-
Counselor: Mm-hmm.
Client: -and quit a couple of times, off and on in my 20s, but, you know, it's been consistent. And I can't see myself changing now.

Figure 6: Examples of extracted minimal responses in Chinese, Japanese, and English. The bold red utterances are the minimal responses selected by our filtering method. Each example shows a three-turn dialogue history, the counselor's minimal response, and the following client utterance.