

Mind-Inference for Multi-Turn Emotional Support: Distinguishing Personal and Factual with Adaptive Memory

Extended Abstract

Peilin Li*
Northwestern Polytechnical
University
Xi'an, China
lipeilin@mail.nwpu.edu.cn

Qiuyun Zhang*
Northwestern Polytechnical
University
Xi'an, China
qiuyunzhang@mail.nwpu.edu.cn

Yukun Lian*
Tsinghua University
Beijing, China
lianyk22@mails.tsinghua.edu.cn

Chengxuan Li*
East China University of Science and
Technology
Shanghai, China
22013185@mail.ecust.edu.cn

Bin Guo†
Northwestern Polytechnical
University
Xi'an, China
guob@nwpu.edu.cn

Mengqi Chen
Northwestern Polytechnical
University
Xi'an, China
chenmengqi@mail.nwpu.edu.cn

Han Wang
Northwestern Polytechnical
University
Xi'an, China
1099255210@mail.nwpu.edu.cn

Yuyang Fang
Northwestern Polytechnical
University
Xi'an, China
fangyuyang@mail.nwpu.edu.cn

Geyang Song
Northwestern Polytechnical
University
Xi'an, China
768904277@qq.com

Yuan Gao
Northwestern Polytechnical
University
Xi'an, China
rudygao@mail.nwpu.edu.cn

ABSTRACT

Emotional Support Conversation (ESC) requires models to track a user's evolving psychological state and produce supportive responses across multiple turns. Existing ESC methods improve empathy with emotion clues, strategy planning, or commonsense reasoning, but they often treat personal and factual information uniformly and therefore struggle with stale memories in long interactions. We present MIA, a compact framework for multi-turn ESC with three coordinated modules: EToM for hierarchical mind inference, PFD for adaptive Personal-Factual weighting, and ORM for removing obsolete psychological inferences during memory updates. Evaluations on ESConv and CPsyCounD show consistent gains over strong baselines, while ablations confirm that both adaptive factor weighting and obsolescence-aware memory are necessary. These results indicate that explicit Personal/Factual separation is a practical design principle for robust multi-turn emotional support.

*These authors contributed equally to this work.

†Corresponding author.



This work is licensed under a Creative Commons Attribution International 4.0 License.

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). <https://doi.org/10.65109/CIMN4023>

KEYWORDS

Emotional Support Conversation; Theory of Mind; Multi-turn Dialogue; Adaptive Memory; Large Language Models

ACM Reference Format:

Peilin Li, Qiuyun Zhang, Yukun Lian, Chengxuan Li, Bin Guo, Mengqi Chen, Han Wang, Yuyang Fang, Geyang Song, and Yuan Gao. 2026. Mind-Inference for Multi-Turn Emotional Support: Distinguishing Personal and Factual with Adaptive Memory: Extended Abstract. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026)*, Paphos, Cyprus, May 25 – 29, 2026, IFAAMAS, 3 pages. <https://doi.org/10.65109/CIMN4023>

1 INTRODUCTION

Emotional Support Conversation (ESC) has become a core task for socially aware dialogue systems [16, 26], with recent LLM-based methods improving empathy and strategy quality through planning and instruction tuning [5, 10, 32, 33]. However, real ESC is inherently multi-turn: user affect and intention evolve across turns, as documented by prior multi-turn analyses [4, 12, 17, 27]. This makes robust state tracking harder than single-shot generation [24, 30].

Existing work has advanced strategy-aware generation [5, 32] and richer context modeling [10, 33], yet two limitations remain. First, many systems still compress user state into a single latent representation, which blurs *Personal* clues and *Factual* clues; this harms interpretability and controllable planning, especially when

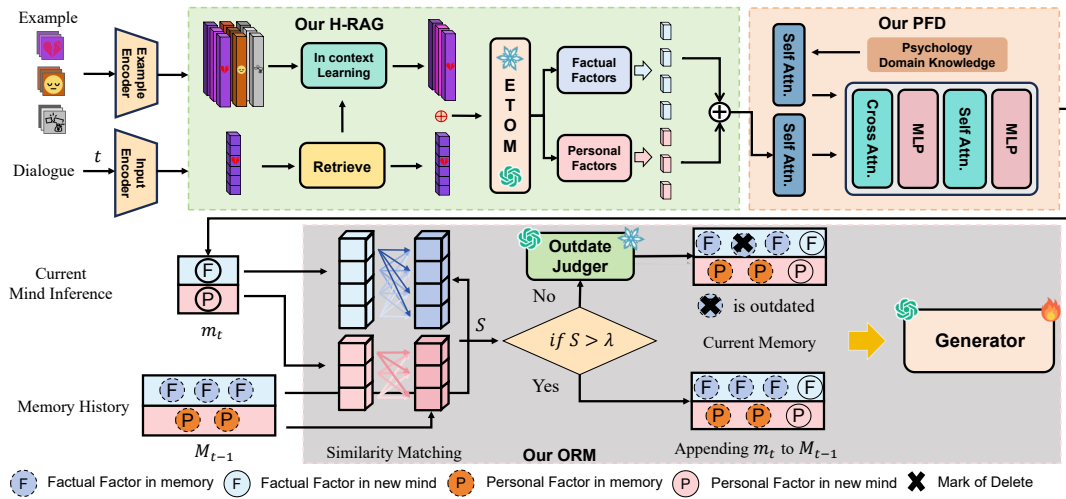


Figure 1: Overview of MIA for multi-turn ESC. At each turn, EToM derives structured Personal/Factual states, PFD estimates turn-level emphasis, and ORM updates memory by preserving stable context while removing obsolete personal states before response generation.

support goals shift within a dialogue [13, 18, 23]. Second, memory often updates uniformly, so stale psychological assumptions may persist and bias later responses [1, 32, 33]. This concern is consistent with evidence from ESC data [16, 26] and longitudinal interaction studies [15], where emotional states change faster than events.

Related cognitive literature also motivates differentiated processing: mental-state inference and reflective regulation are dynamic [2, 6, 7, 22], while decision under uncertainty often combines fast affective signals and slower deliberative signals [8, 9, 11]. Inspired by this, we propose **MIA (Mind-Inference with Adaptive memory)**, a compact framework for multi-turn ESC that separates factorized inference from temporal memory maintenance.

2 MIA FRAMEWORK

As shown in Figure 1, MIA first performs structured mind inference, then applies turn-adaptive factor fusion, and finally updates memory before response generation. This decomposition follows practical ESC observations from prior strategy-centric systems [5, 10, 32]: explicit state interfaces simplify control, and selective memory prevents drift in long conversations [4, 25, 33].

EToM: hierarchical mind inference. EToM factorizes user information into *Personal* (emotion, belief, intention, desire) and *Factual* (fact, cause, result) components. Unlike undifferentiated latent summaries [29], this hierarchy yields interpretable intermediate states that can be explicitly inspected or weighted at each turn [23, 28]. The separation also aligns with cognitive accounts of mentalization and appraisal [2, 7, 22].

PFD: adaptive factor weighting. Because turn-level needs vary, PFD dynamically reweights Personal/Factual streams before generation. In affectively intense turns, stronger Personal grounding tends to improve emotional attunement; in clarification or action-planning turns, stronger Factual grounding improves specificity

and usefulness. This adaptive mechanism complements strategy-conditioned ESC methods [5, 32, 33] and mitigates one-style reasoning collapse observed in purely static fusion [3, 14, 20, 21, 24].

ORM: obsolescence-aware memory update. Instead of accumulating all inferred states, ORM removes obsolete Personal inferences while retaining stable Factual context. This reduces stale-state propagation across long dialogues, a known risk in memory-heavy pipelines [1, 17, 27]. The design is compatible with representation practices in dialogue systems [19, 31], but differs by explicitly modeling psychological obsolescence at turn granularity.

3 RESULTS AND DISCUSSION

We evaluate MIA on ESConv [16] and CPsyCounD [26], comparing against representative ESC baselines including PAL [5] and SUPPORTER [32]. Following common ESC practice [10, 33], we report perplexity and overlap metrics (BLEU/ROUGE). MIA improves consistently on both datasets: on ESConv, it achieves PPL 7.39, BLEU-4 22.25, and ROUGE-L 22.44; on CPsyCounD, it reaches PPL 3.87, BLEU-4 25.99, and ROUGE-L 42.22 (strong baseline references: 9.97/16.42 on ESConv and 4.10/20.96/39.38 on CPsyCounD).

Ablation further supports the modular design. Removing **PFD** weakens turn-level adaptation (e.g., ESConv BLEU-4 drops to 20.13), while removing **ORM** harms multi-turn consistency (BLEU-4 drops to 19.72). These patterns align with prior findings that ESC quality depends jointly on strategy selection [5, 32] and context management over long horizons [4, 14, 25, 28].

4 CONCLUSION

We presented **MIA**, a compact multi-turn ESC framework combining EToM, PFD, and ORM. By separating dynamic mental cues from stable factual context, MIA improves response quality and robustness on ESConv and CPsyCounD, enabling more reliable long-horizon emotional support.

REFERENCES

- [1] Sanghwan Bae, Donghyun Kwak, Soyoun Kang, Min Young Lee, Sungdong Kim, Yuni Jeong, Hyeri Kim, Sang-Woo Lee, Woomyoung Park, and Nako Sung. 2022. Keep Me Updated! Memory Management in Long-term Conversations. In *Findings of the Association for Computational Linguistics: EMNLP 2022*. 3769–3787.
- [2] Aaron T. Beck. 1987. Cognitive therapy and the emotional disorders: A. T. Beck. *British Journal of Psychiatry* 150 (1987), 870 – 871.
- [3] Zhuang Chen, Yaru Cao, Guanqun Bi, Jincenzi Wu, Jinfeng Zhou, Xiyao Xiao, Si Chen, Hongning Wang, and Minlie Huang. 2025. SocialSim: Towards Socialized Simulation of Emotional Support Conversation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 1274–1282.
- [4] Zhiliang Chen, Xinyuan Niu, Chuan-Sheng Foo, and Bryan Kian Hsiang Low. 2025. Broaden your SCOPE! Efficient Multi-turn Conversation Planning for LLMs with Semantic Space. In *ICLR*.
- [5] Jiale Cheng, Sahand Sabour, Hao Sun, Zhuang Chen, and Minlie Huang. 2023. PAL: Persona-Augmented Emotional Support Conversation Generation. In *Findings of the Association for Computational Linguistics: ACL 2023*. 535–554.
- [6] Albert Ellis. 1962. Reason and emotion in psychotherapy.
- [7] Peter Fonagy, György Gergely, Elliot L Jurist, and Mary Target. 2002. Affect regulation, mentalization, and the development of the self. (2002).
- [8] Keith Frankish. 2010. Dual-Process and Dual-System Theories of Reasoning. *Philosophy Compass* 5 (2010), 914–926.
- [9] Clara E Hill and Karen M O'Brien. 1999. Helping skills: Facilitating exploration, insight, and action. (1999).
- [10] Mengzhao Jia, Qianglong Chen, Liqiang Jing, Dawei Fu, and Renyu Li. 2023. Knowledge-enhanced Memory Model for Emotional Support Conversation. arXiv:2310.07700 [cs.CL]
- [11] Daniel Kahneman and Amos Tversky. 2013. Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I*. World Scientific, 99–127.
- [12] Haitao Li, Yifan Chen, Hu YiRan, Qingyao Ai, Junjie Chen, Xiaoyu Yang, Jianhui Yang, Yueyue Wu, Zeyang Liu, and Yiqun Liu. 2025. LexRAG: Benchmarking Retrieval-Augmented Generation in Multi-Turn Legal Consultation Conversation. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 3606–3615.
- [13] Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2016. A Diversity-Promoting Objective Function for Neural Conversation Models. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics*. 110–119.
- [14] Junlin Li, Bo Peng, Yu-Yin Hsu, and Chu-Ren Huang. 2024. Be Helpful but Don't Talk Too Much - Enhancing Helpfulness in Conversations through Relevance in Multi-Turn Emotional Support. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*. 1976–1988.
- [15] Shuo Liu, Kaining Ying, Hao Zhang, Yue Yang, Yuqi Lin, Tianle Zhang, Chuanhao Li, Yu Qiao, Ping Luo, Wenqi Shao, et al. 2024. Convbench: A multi-turn conversation evaluation benchmark with hierarchical ablation capability for large vision-language models. *Advances in Neural Information Processing Systems* 37 (2024), 100734–100782.
- [16] 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)*. 3469–3483.
- [17] Shichao Ma, Xinfeng Zhang, Zeng Zhao, Bai Liu, Changjie Fan, and Zhipeng Hu. 2025. DialogDraw: Image Generation and Editing System Based on Multi-Turn Dialogue. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 24795–24803.
- [18] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. 311–318.
- [19] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*.
- [20] Sahand Sabour, Chujie Zheng, and Minlie Huang. 2021. CEM: Commonsense-aware Empathetic Response Generation. *ArXiv abs/2109.05739* (2021).
- [21] Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (San Francisco, California, USA) (AAAI'17)*. 4444–4451.
- [22] Kathy A. Stepien and A. Baernstein. 2006. Educating for empathy. A review. *Journal of general internal medicine* 21 5 (2006), 524–30.
- [23] Quan Tu, Yanran Li, Jianwei Cui, Bin Wang, Ji-Rong Wen, and Rui Yan. 2022. MISC: A Mixed Strategy-Aware Model integrating COMET for Emotional Support Conversation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*. 308–319.
- [24] Jincenzi Wu, Zhuang Chen, Jiawen Deng, Sahand Sabour, Helen Meng, and Minlie Huang. 2024. COKE: A Cognitive Knowledge Graph for Machine Theory of Mind. arXiv:2305.05390 [cs.CL]
- [25] Dingbo Yuan, Yipeng Chen, Guodong Liu, Chenchen Li, Chengfu Tang, Dongxu Zhang, Zhenkui Wang, Xudong Wang, and Song Liu. 2025. DMT-RoleBench: A Dynamic Multi-Turn Dialogue Based Benchmark for Role-Playing Evaluation of Large Language Model and Agent. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 25760–25768.
- [26] 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. *ACL*.
- [27] Lu Zhang, Chen Li, Yu Lei, Zhu Sun, and Guanfang Liu. 2024. An empirical analysis on multi-turn conversational recommender systems. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 841–851.
- [28] Tengan Zhang, Xinjie Zhang, Jinning Zhao, Li Zhou, and Qin Jin. 2024. ESCoT: Towards Interpretable Emotional Support Dialogue Systems. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*. 13395–13412.
- [29] Zhuosheng Zhang, Aston Zhang, Mu Li, Hai Zhao, George Karypis, and Alex Smola. 2024. Multimodal Chain-of-Thought Reasoning in Language Models. arXiv:2302.00923 [cs.CL] <https://arxiv.org/abs/2302.00923>
- [30] Yaowei Zheng, Richong Zhang, Junhao Zhang, Yanhan Ye, Zheyuan Luo, Zhangchi Feng, and Yongqiang Ma. 2024. LlamaFactory: Unified Efficient Fine-Tuning of 100+ Language Models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*.
- [31] Zhonghua Zheng, Lizi Liao, Yang Deng, Libo Qin, and Liqiang Nie. 2024. Self-Chats from Large Language Models Make Small Emotional Support Chatbot Better. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*. 11325–11345.
- [32] Jinfeng Zhou, Zhuang Chen, Bo Wang, and Minlie Huang. 2023. Facilitating Multi-turn Emotional Support Conversation with Positive Emotion Elicitation: A Reinforcement Learning Approach. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*. 1714–1729.
- [33] Jinfeng Zhou, Chujie Zheng, Bo Wang, Zheng Zhang, and Minlie Huang. 2023. CASE: Aligning Coarse-to-Fine Cognition and Affection for Empathetic Response Generation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*. 8223–8237.