Few-shot Dialogue Strategy Learning for Motivational Interviewing via Inductive Reasoning

Zhouhang Xie[1] **Bodhisattwa Prasad Majumder**[2] **Mengjie Zhao**[3]
Yoshinori Maeda[3] **Keiichi Yamada**[1] **Hiromi Wakaki**[3] **Julian McAuley**[1]
[1] UC San Diego [2] Allen Institute for AI [3] Sony Group Corporation
{zhx022, jmcauley}@ucsd.edu

Abstract

We consider the task of building a dialogue system that can motivate users to adopt positive lifestyle changes: Motivational Interviewing. Addressing such a task requires a system that can infer *how* to motivate a user effectively. We propose DIIR (**Di**alogue Strategy Learning via **I**nductive **R**easoning), a framework that is capable of learning and applying conversation strategies in the form of natural language inductive rules from expert demonstrations. Automatic and human evaluation on instruction-following large language models show natural language strategy descriptions discovered by DIIR can improve active listening skills, reduce unsolicited advice, and promote more collaborative and less authoritative responses, outperforming various demonstration utilization methods.

1 Introduction

We study Motivational Interviewing (MI), where a conversational agent assists users in finding motivation and reducing ambivalence in adopting positive lifestyle changes such as quitting drug abuse and seeking medical assistance (Moyers et al., 2016). A successful MI dialogue system must be able to infer *how* to best motivate a user to adopt positive changes, which is a nontrivial task. For example, as shown in Figure 1, one way to convince a user to workout is to educate the user about its benefits, yet immediately offering suggestions without understanding the context can risk creating resistance and confronting users (Brehm, 1981). While prior work studied rule-based dialogue systems for MI (Welch et al., 2020; Park et al., 2019) and MI dialogue response rephrasing (Welivita and Pu, 2023), another important problem is to build dialogue systems that *automatically discover and learn* MI skills.

In this work, we introduce DIIR (**Di**alogue Strategy Learning via **I**nductive **R**easoning)[1], a framework for aligning LLMs to dialogue behavior of

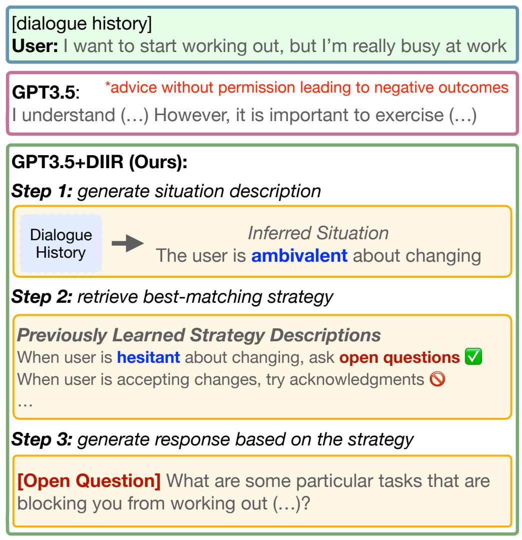

Figure 1: DIIR retrieves learned dialogue strategy descriptions to generate a response that **leads to positive outcomes**. See Table 9 in the Appendix for example strategies.

professional MI interviewers given a few demonstration dialogues. DIIR does not require training and instead employs LLMs to generate a set of natural language dialogue strategy descriptions, such as "`when the user is hesitant about change, the interviewer should ask open questions to learn more about the user's concern`" by analyzing demonstration dialogues. These strategies are hypotheses about how professional interviewers respond in various situations, and are reused at inference time as instructions for LLMs to mimic MI interviewers. This process of identifying underlying principles from observations and generalizing them for inference is a form of inductive reasoning (Feeney and Heit, 2007).

Different from common practice that uses available demonstrations as in-context-learning (ICL) examples (Brown et al., 2020; Xu et al., 2023a), strategies descriptions discovered by our framework explicitly state the desired behavior for the model, such as "`ask open questions`" and their

[1] https://github.com/zhouhanxie/DIIR

Findings of the Association for Computational Linguistics ACL 2024, pages 13207–13219

applicable situations, such as "`when the client is hesitant about change`". These automatically generated strategy descriptions are a more precise nudge to the LLM being instructed than ICL examples. Existing work that employs language models for generating and reusing natural language strategies from observations (Majumder et al., 2023; Nottingham et al., 2023; Wang et al., 2024a) (i.e. Inductive Reasoning with LLMs) requires interactive environments to verify and improve the generated statements. In comparison, our method learns from a static demonstration dialogue dataset which is more commonly available than interactive environments for dialogue models.

To evaluate DIIR, we automate and adapt a set of metrics derived from clinical psychology literature for evaluating human MI practitioners. We experiment with DIIR on a publicly available MI dataset. Automatic and human evaluation show strategies discovered by DIIR can help instruction-following downstream language models produce more collaborative, respectful, and least resistance-inducing responses with as few as five annotated dialogues of expert demonstrations. Our **contributions** are as follows: (1) To our best knowledge, we are the first to work to build and evaluate a dialogue system that automatically learns from dialogue data for MI. (2) We develop DIIR, a framework that learns and applies dialogue strategies as natural language inductive rules from MI experts for only a few dialogues. (3) We automate evaluations from the clinical psychology literature on MI to facilitate future research on MI.

2 Dialogue Strategy Learning via Inductive Reasoning

The goal of an MI interviewer is to convince a client to make a positive change of interest, such as to start working out. Given a dialogue history h consisting of previous user-system turns, an ideal MI dialogue system produces the next system response a that best convinces the user to make a positive change. We solve this optimal response generation problem by using LLMs to analyze demonstration dialogues and generate dialogue strategy descriptions stating how to respond in various situations, such as "`When the user is hesitant about change, ask open questions.`" These descriptions can then be reused to improve an instruction-following LLM to promote demonstration-aligned dialogue behavior, such as asking open questions

Algorithm 1 ENHANCESTRATEGY

Input: Dialogue history h, gold response a^*, generator LLM G, discriminator LLM D, executor LLM E, maximum number of trials N
Output: Strategy e in natural language, description of its applicable situation k

1: $e \leftarrow$ ""
2: **for** i in N **do**
3: $a' \leftarrow$ `E.generate_response` (h, e)
4: **if** `D.confirms_is_similar` (a', a^*) **then**
5: **break**
6: $e \leftarrow$ `G.improve_strategy` (h, e, a^*, a')
7: $k \leftarrow$ `G.describe_situation` (h)
8: **return** e, k

when the client is hesitant. We now describe DIIR's learning and inference process (see Appendix D for step-by-step concrete examples).

Learning Dialogue Strategy by Analyzing Demonstration Dialogues Given a dialogue history h and a gold interviewer response a^*, a corresponding dialogue strategy description e should contain sufficient information such that an LLM could closely reproduce a^* given h and e. Intuitively, one could instruct an LLM to analyze the gold response and its corresponding dialogue history to generate descriptions containing information about "`When <description about` h`>, one should <description about` a^*`>`," which can then be used as strategy descriptions. We denote this LLM that produces strategy descriptions as the strategy `generator` G.

However, there is no guarantee that G always generates valid strategy descriptions. To address this issue, we propose to employ a `discriminator` LLM D and an `executor` LLM E in place of the interactive environment in previous works to give feedback during the strategy statement generation process. The intuition is that to verify a dialogue strategy such as "`When the user is hesitant about change, ask open questions`", we could instruct the `executor` E to generate a response given the dialogue history where a user is hesitant following this this instruction, and check whether the generated response matches the gold interviewer response.

This verification process eliminates two sources of error: the generated strategy is incorrect or is correct but is not detailed enough to instruct an LLM to reproduce the groundtruth response. Existing work typically addresses this statement-verification problem by verifying the statements in interactive environments such as MineCraft (Nottingham et al., 2023). However, there are no environments to verify statements about dialogue strategies.

The process for generating and verifying a strategy description given a context-and-gold-response pair is as shown in Algorithm 1. Concretely, given a hypothesized dialogue strategy e (initialized to be an empty string), the `executor` generates a response a' given the dialogue history h and the strategy e. If the `discriminator` checks and marks a' as similar to gold response a^*, we accept the strategy e as valid. Otherwise, we instruct the `generator` to regenerate a better strategy given the dialogue history, current strategy, generated response, and the gold response. This generate-and-verify process repeats until a valid strategy description has been found or a fixed number of search steps N has been reached.

Finally, to be able to reuse this strategy in the future, we prompt the `generator` to generate a description of the situation k such as "`the user is hesitant about change`." This situation description can be derived from e since the dialogue strategy already contains information about when the strategy is applicable, but we use an LLM to generate it in practice for ease of implementation. We repeat the process introduced above for all context-response pairs available in the demonstration dialogues to create suitable strategies for various scenarios.

Inference via Strategy Reuse After learning, DIIR accumulates a set of strategies $\{e_i\}$ and their corresponding applicable situations $\{k_i\}$. Thus, given an inference-time dialogue history $h_{\text{inference}}$, we can simply generate a description of the situation $k_{\text{inference}}$ using the `generator` LLM, and then use $k_{\text{inference}}$ to retrieve the most suitable strategy e_i by identifying k_i that is most similar to $k_{\text{inference}}$. We then instruct the `executor` to generate a response given $h_{\text{inference}}$ and the retrieved e_i. In practice, we use a text embedding model Φ to encode all situational descriptions and find the top ten most suitable demonstration situation by maximizing the dot product of $\Phi(k_i)$ and $\Phi(k_{\text{inference}})$ following Lewis et al. (2020); Guu et al. (2020). After this, we instruct the `Generator` LLM to pick a strategy that is most appropriate to the dialogue history from the ten retrieved strategies and follow its description. This process is illustrated in Figure 1.

3 Experiments

Models and Baselines We demonstrate DIIR's few-shot learning ability by learning from 5 dialogues using GPT3.5 and GPT4, which mimics data-scarce scenario where only a few high-quality demonstration dialogues are available. We chose GPT3.5 and GPT4 as backbone models due to their strong reasoning ability, similar to various recent works on LLM-based reasoning (Wang et al., 2024b; Majumder et al., 2023; Nottingham et al., 2023; Yu et al., 2023; Qiu et al., 2024), and find open source models less performant (see Appendix E for additional discussions). For baselines, we consider three variants of in-context learning (ICL) in addition to the vanilla LLM: ICL-RAND with 5 random demonstrations, ICL-KNN (Xu et al., 2023a) with 5 demonstrations with most similar dialogue context to the inference-time context computed using a text embedding model, and ICL-ALL with all demonstration dialogues.

Dataset We experiment on the AnnoMI dataset (Wu et al., 2022), an MI dataset transcribed from YouTube videos with 133 MI transcripts for various topics. The dataset also comes with quality labels for each dialogue, where 110 dialogues are labeled as high-quality demonstrations from good MI interviewers, and the remaining 33 dialogues are labeled as low-quality demonstrations. We consider the high-quality dialogues only since the goal of DIIR is to learn conversation strategies from MI experts. We randomly sample 5 dialogues with 112 turns for learning and 80 dialogues with 1192 turns for our primary evaluation on GPT3.5. For high-cost evaluation experiments involving GPT4 or human annotators, we further randomly sub-sample 10 dialogues with 148 turns from the primary evaluation data.

4 Evaluation

For **automatic evaluation**, we evaluate the alignment of our model by obtaining dialogue action counts with a dialogue act classifier trained on crowd-sourced annotations (Welivita and Pu, 2022). We report the following behavior-count-based evaluation adopted from the Motivational Interviewing Integrity Treatment (MITI; Moyers et al. (2016)) and Motivational Interviewing Skill Code (MISC; Miller et al. (2003)) used to evaluate professional interviewers from clinical psychology literature.

Ratio of MI-guideline-inconsist Behavior (%MI-i) The ratio of MI-inconsistent behaviors (confront, direct, warn, and advice without permission) in model responses.

Complex over Simple Reflections (C/S) The

Method	%MI-i↓	C/S↑	R/Q↑	%AL↑	%NA↑
W/o Strategies					
GPT3.5-BESTBASE	2.1	44.8	**1.53**	54.3	82.8
GPT4-BESTBASE	1.4	11.8	1.06	45.8	71.8
With Strategies (Ours)					
GPT3.5-DIIR	1.3	115	0.99	**61.1**	**95.1**
GPT4-DIIR	**0.6**	**135**	1.04	59.6	94.4
Gold	1.4	2.3	1.51	59.0	81.7

Table 1: Alignment performance against best possible result *selected per dimension* from all baselines. We found open-source LLMs cannot consistently follow self-generated statements, coherent with recent findings (Qiu et al., 2024), and thus are less performant (See Appendix E).

ratio of complex over simple reflections in model responses, indicating how often the model produces in-depth feedback to the client.

Reflection over Questions (R/Q) The count of reflections over questions in the models' responses.

Percent of Active Listening Behaviors (%AL) The ratio of the sum of Questions and Reflections among all behavior counts, showing how often does the modal employ active listening skills.

Percent of Non-authoritatitive Behaviors (%NA) The frequency of responses that are not confront, warn, direct, advice (with or without permission), or give information.

We verify the effectiveness of our automatic evaluation by running the evaluation on high and low-quality interviewer demonstrations in the AnnoMI dataset (Wu et al., 2022), results show our method can effectively tell apart high versus low quality demonstrations (see Appendix C for additional discussions and details of the dialogue act classifier).

For **human evaluation**, we evaluate our models' performance against in-context baselines by asking human evaluators on Amazon Mechanical Turk to read four context-response pairs produced by both models and choose the one that is best aligned with a description of MI (more details in Appendix C). This results in 37 groups of evaluations between each pair of opponent models, where each group is rated by an annotator.

5 Results and Analysis

Our main results are as shown in Table 1. Due to space constraints, we report the best performance among the four baselines (base model and three ICL variants) for each dimension under the name BESTBASE, and show detailed performance in Table 4 in the Appendix. Although BESTBASE has an unfair advantage due to combining best

DIIR vs. ICL	RAND	ALL	KNN
GPT3.5	**21***	**23***	**7**
GPT4	**15***	**5**	-5

Table 2: Human evaluation results. We report number of *additional* times DIIR win against baseline, calculated as number of wins by DIIR minus number of opponent wins; with statistical difference with $p<0.05$ denoted in *.

baseline results for each dimension, our method outperforms the baselines in all categories except the Reflection-to-Question ratio, sometimes outperforming the ground truth responses. Meanwhile, MI guideline defines an R/Q ratio of 1.0 to be good and 0.5 to be satisfactory (Moyers et al., 2016), showing that our model maintains competitive performance in this dimension.

Overall, our method produces fewer MI-inconsistent responses, provides more in-depth reflections, and better leverages active listening skills and the MI spirit of not assuming expert roles. These trends are coherent with human evaluation results, as shown in Table 2.

6 Related Work

Motivational Interviewing in NLP Prior work in NLP about MI primarily studied the task of annotating (Welivita and Pu, 2022; Tanana et al., 2015) and predicting (Pérez-Rosas et al., 2017; Huang et al., 2018) dialogue actions. Recently, there are attempts on building systems for evaluating (Min et al., 2022) and rewriting (Welivita and Pu, 2023) responses for MI but does not focus on response generation. A few pioneering works build MI dialogue systems with hard-coded rules or dialogue templates (Welch et al., 2020; Park et al., 2019) but focus on specific domains such as COVID-19. To the best of our knowledge, we are the first to build an MI dialogue system that automatically learns from expert demonstrations.

LLMs for Generating and Reusing Natural Language Strategies Generating hypothesis from observations and reusing the generated hypothesis with LLMs have shown a promising direction for various tasks, such as planning (Majumder et al., 2023; Nottingham et al., 2023; Wang et al., 2023), prompt-generation (Wang et al., 2024b), and code generation (Wang et al., 2024a). This is a form of inductive reasoning (Wang et al., 2023) where LLMs draw conclusion from obvervations. However, prior works require interactive environments for feedback which are unsuitable for learning from static datasets, while our work use an LLM for feedback and lifts the requirement for interactive environments.

7 Conclusion

We propose DIIR, a method for few-shot aligning large language models to MI techniques. Our method infers natural language strategies from expert demonstrations and reuses these strategies at inference time. Automatic and human evaluation show our method is better aligned with the behavior of motivational interviewing experts, produces more in-depth reflections, and better leverages active listening skills, outperforming a variety of demonstration-based methods.

8 Limitations

We note that our framework is developed for Motivational Interviewing and benefits from domain-specific knowledge such as MI-specific dialogue actions. Meanwhile, our framework assumes access to a strong instruction following language model, which is usually restricted by API access or computational resources. Finally, we note that while we adopt clinical psychology literature verified metrics and human evaluation to measure the performance of DIIR, the ideal evaluation is deploying MI dialogue systems for continued intervention.

9 Ethical Concerns

Following previous work on conversational persuasion (Wang et al., 2019), we note Motivational Interviewing shares similarities with persuasion dialogues. While the technique for Motivational Interviewing is developed for inducing *positive* changes in users, system deployers should carefully analyze the specific topic for Motivational Interviewing to eliminate risks of harm. Meanwhile, when deploying a system equipped with MI abilities, the user should be informed about the system's identity so that they can make an informed choice when interacting with such a system.

10 Acknowledgments

We would like to thank Tim Althoff, Jessica Echterhoff, Yilan Chen, and the anonymous reviewers for their valuable comments. This work is funded by the SONY Research Award Program.

References

Jack W. Brehm. 1981. A theory of psychological reactance.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Neurips*, volume 33, pages 1877–1901. Curran Associates, Inc.

Aidan Feeney and Evan Heit. 2007. Inductive reasoning: Experimental, developmental and computational approaches.

Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat, and Ming-Wei Chang. 2020. Realm: Retrieval-augmented language model pre-training. In *ICML*, ICML'20. JMLR.org.

Xiaolei Huang, Lixing Liu, Kate Carey, Joshua Woolley, Stefan Scherer, and Brian Borsari. 2018. Modeling temporality of human intentions by domain adaptation. In *EMNLP*, pages 696–701, Brussels, Belgium. Association for Computational Linguistics.

Shaoxiong Ji, Tianlin Zhang, Luna Ansari, Jie Fu, Prayag Tiwari, and Erik Cambria. 2022. Mental-BERT: Publicly available pretrained language models for mental healthcare. In *LREC*, pages 7184–7190, Marseille, France. European Language Resources Association.

Patrick S. H. Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Neurips*.

Bodhisattwa Prasad Majumder, Bhavana Dalvi Mishra, Peter Jansen, Oyvind Tafjord, Niket Tandon, Li Zhang, Chris Callison-Burch, and Peter Clark. 2023. Clin: A continually learning language agent for rapid task adaptation and generalization.

William R Miller, Theresa B Moyers, Denise Ernst, and Paul Amrhein. 2003. Manual for the motivational interviewing skill code (misc). *Unpublished manuscript. Albuquerque: Center on Alcoholism, Substance Abuse and Addictions, University of New Mexico*.

Do June Min, Verónica Pérez-Rosas, Kenneth Resnicow, and Rada Mihalcea. 2022. PAIR: Prompt-aware margIn ranking for counselor reflection scoring in motivational interviewing. In *EMNLP*, pages 148–158, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Theresa B. Moyers, Lauren N. Rowell, Jennifer K Manuel, Denise B. Ernst, and Jon M. Houck. 2016. The motivational interviewing treatment integrity code (miti 4): Rationale, preliminary reliability and validity. *Journal of substance abuse treatment*, 65:36–42.

Kolby Nottingham, Prithviraj Ammanabrolu, Alane Suhr, Yejin Choi, Hannaneh Hajishirzi, Sameer Singh, and Roy Fox. 2023. Do embodied agents dream of pixelated sheep: Embodied decision making using language guided world modelling. In *ICML*, volume 202 of *Proceedings of Machine Learning Research*, pages 26311–26325. PMLR.

Sohyun Park, Jeewon Choi, Sungwoo Lee, Changhoon Oh, Changdai Kim, Soohyun La, Joonhwan Lee, and Bongwon Suh. 2019. Designing a chatbot for a brief motivational interview on stress management: Qualitative case study. *Journal of Medical Internet Research*, 21.

Verónica Pérez-Rosas, Rada Mihalcea, Kenneth Resnicow, Satinder Singh, and Lawrence An. 2016. Building a motivational interviewing dataset. In *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, pages 42–51, San Diego, CA, USA. Association for Computational Linguistics.

Verónica Pérez-Rosas, Rada Mihalcea, Kenneth Resnicow, Satinder Singh, Lawrence An, Kathy J. Goggin, and Delwyn Catley. 2017. Predicting counselor behaviors in motivational interviewing encounters. In *EACL*, pages 1128–1137, Valencia, Spain. Association for Computational Linguistics.

James O Prochaska and Wayne F. Velicer. 1997. The transtheoretical model of health behavior change. *American Journal of Health Promotion*, 12:38 – 48.

Linlu Qiu, Liwei Jiang, Ximing Lu, Melanie Sclar, Valentina Pyatkin, Chandra Bhagavatula, Bailin Wang, Yoon Kim, Yejin Choi, Nouha Dziri, and Xiang Ren. 2024. Phenomenal yet puzzling: Testing inductive reasoning capabilities of language models with hypothesis refinement. In *The Twelfth International Conference on Learning Representations*.

Raj Sanjay Shah, Faye Holt, Shirley Anugrah Hayati, A. Agarwal, Yi-Chia Wang, Robert E. Kraut, and Diyi Yang. 2022. Modeling motivational interviewing strategies on an online peer-to-peer counseling platform. *Proceedings of the ACM on Human-Computer Interaction*, 6:1 – 24.

Michael Tanana, Kevin Hallgren, Zac Imel, David Atkins, Padhraic Smyth, and Vivek Srikumar. 2015. Recursive neural networks for coding therapist and patient behavior in motivational interviewing. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 71–79, Denver, Colorado. Association for Computational Linguistics.

Leili Tavabi, Trang Tran, Kalin Stefanov, Brian Borsari, Joshua Woolley, Stefan Scherer, and Mohammad Soleymani. 2021. Analysis of behavior classification in motivational interviewing. In *Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access*, pages 110–115, Online. Association for Computational Linguistics.

Ashish Vaswani, Noam M. Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Neural Information Processing Systems*.

Ruocheng Wang, E. Zelikman, Gabriel Poesia, Yewen Pu, Nick Haber, and Noah D. Goodman. 2023. Language models meet world models: Embodied experiences enhance language models. In *NeurIPS*.

Ruocheng Wang, E. Zelikman, Gabriel Poesia, Yewen Pu, Nick Haber, and Noah D. Goodman. 2024a. Hypothesis search: Inductive reasoning with language models. In *ICLR*.

Xinyuan Wang, Chenxi Li, Zhen Wang, Fan Bai, Haotian Luo, Jiayou Zhang, Nebojsa Jojic, Eric P. Xing, and Zhiting Hu. 2024b. Promptagent: Strategic planning with language models enables expert-level prompt optimization.

Xuewei Wang, Weiyan Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, and Zhou Yu. 2019. Persuasion for good: Towards a personalized persuasive dialogue system for social good. In *ACL*, pages 5635–5649, Florence, Italy. Association for Computational Linguistics.

Charles Welch, Allison Lahnala, Veronica Perez-Rosas, Siqi Shen, Sarah Seraj, Larry An, Kenneth Resnicow, James Pennebaker, and Rada Mihalcea. 2020. Expressive interviewing: A conversational system for coping with COVID-19. In *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020*, Online. Association for Computational Linguistics.

Anuradha Welivita and Pearl Pu. 2022. Curating a large-scale motivational interviewing dataset using peer support forums. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 3315–3330, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Anuradha Welivita and Pearl Pu. 2023. Boosting distress support dialogue responses with motivational interviewing strategy. In *Findings of ACL*, pages 5411–5432, Toronto, Canada. Association for Computational Linguistics.

Peter West, Ximing Lu, Nouha Dziri, Faeze Brahman, Linjie Li, Jena D. Hwang, Liwei Jiang, Jillian Fisher, Abhilasha Ravichander, Khyathi Chandu, Benjamin Newman, Pang Wei Koh, Allyson Ettinger, and Yejin Choi. 2024. The generative ai paradox: "what it can create, it may not understand. In *ICLR*.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *EMNLP Demos*, pages 38–45, Online. Association for Computational Linguistics.

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*, pages 6177–6181.

Benfeng Xu, Quan Wang, Zhendong Mao, Yajuan Lyu, Qiaoqiao She, and Yongdong Zhang. 2023a. kNN prompting: Beyond-context learning with calibration-free nearest neighbor inference. In *ICLR*.

Canwen Xu, Yichong Xu, Shuohang Wang, Yang Liu, Chenguang Zhu, and Julian McAuley. 2023b. Small models are valuable plug-ins for large language models.

Xiao Yu, Maximillian Chen, and Zhou Yu. 2023. Prompt-based Monte-Carlo tree search for goal-oriented dialogue policy planning. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 7101–7125, Singapore. Association for Computational Linguistics.

A Building the Dialogue Act Classifier

Following prior work in the Computer Science and NLP community on annotating motivational interviewing dialogues (Shah et al., 2022; Welivita and Pu, 2022; Pérez-Rosas et al., 2016; Tavabi et al., 2021), we train a dialogue classifier to map sentences to motivational interviewing behavior codes. We train the dialogue act classifier on a publicly available dataset (Welivita and Pu, 2022) with annotated dialogue actions specific to MI, since the AnnoMI dataset does not contain fine-graned dialogue action labels. This dataset contains single turn interactions only, and thus are suitable for training dialogue act classifiers only but not the dialogue model itself. We fine-tune two variants of pre-trained mask language models, BERT (Vaswani et al., 2017) and Mental-BERT (Ji et al., 2022), a fine-tuned version of the base BERT model on psychotherapy dialogues. We use a $0.6-0.2-0.2$ train-validation-test split. We fine-tune both models using the Huggingface Transformers Library (Wolf et al., 2020) for a maximum of 10 epoch with early stopping at a learning rate of $5e-5$. While the absolute accuracy of both BERT [2] and Mental-BERT [3] are not optimal (Recall at 1 at 0.60 and 0.61, respectively), we fine the Recall at 5 out of 12 classes are high (at 0.95). Thus, we use the classifier to retrieve N=5 relevant labels given a input sentence, and use a GPT-3.5-turbo to decide the final label, following recent findings in NLP that combining an LLM with small classifier for classification has superior performance to both base models (Xu et al., 2023b).

B Other Experiment and Implementation Details

For DIIR's learning process, we set the maximum number of optimization steps to 3, and break the reference game between the `generator` and the `discriminator` afterwards regardless of success. To enhance the feedback quality of the `discriminator`, we also use the same dialogue act classifier as described in Appendix A to provide distant labels of what potential actions are in the ground truth response (this means predicting the string labels of dialogue actions, and append this to the instruction for the LLM).

For all experiments, the specific model names we used in OpenAI is `gpt-3.5-turbo-1106` and `gpt-4-1106-preview`. These models are available as of December 2023. For the off-the-shelf retriever for retrieving strategies, we use a publicly available light-weight model [4] and retrieve usiang dot product scores. All experiments are conducting using 1 RTX A6000 GPU. Learning and Inference of DIIR on the specified data size finishes in 3 hours and 1.5 hours, respectively. The trainable component of the dialogue act classifier (based on BERT) contains 12-layer, 768-hidden, 12-heads, and 110M parameters. The retriever model contains 12-layer, 384-hidden, 12-heads, 33M parameters.

C Details on Evaluation

C.1 Discussion on Evaluation Metrics

We provide discussions on the derived metrics in this section. In particular, the evaluation process in MISC evaluates a interviewer on three dimensions: (1) Acceptance: being non-judgemental and do not direct or confront the user, (2) Empathy (different from the definition as used in Empathetic

[2]https://huggingface.co/bert-base-uncased

[3]https://huggingface.co/mental/mental-bert-base-uncased

[4]https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2

	%MI-i↓	C/S↑	R/Q↑	%AL↑	%NA↑
Low	20	2.09	1.0	34.4	51.5
High	**0.9**	**2.75**	**1.875**	**62.4**	**82.6**

Table 3: Our automatic evaluation can tell apart between high quality ("High" in the table) and low quality ("Low" in the table) MI demonstrations

Dialogues in NLP): showing genuine interest in understanding the client's situation and thoughts with active listening skills, and (3) Motivational Interviewing Spirit: the interviewer should be collaborative rather than assuming expert roles. In practice, we found all models rarely direct or confront the user using our dialogue action classifier, trivially satisfying the Acceptance criteria. We use %AL as a proxy for the Empathy criteria, and %NA as a proxy to measure the Motivational Interviewing Spirit criteria. These metrics are proven to reduce client resistance and improve intervention outcome in MI literature (Moyers et al., 2016).

C.2 Verifying the Effectiveness of Automatic Evaluation

to validate our evaluation, we run our classifier-based evaluation pipeline on 231 datapoints (dialogue context + gold response) of good vs. bad demonstration of MI interviewers (the binary high v.s. low quality labels are available in the original AnnoMI dataset). Bad demonstrations here means the response are produced by interviewer that does not follow MI guideline. We use 231 pairs from each class (i.e. 231 high and 231 low-quality demonstrations) since that is the number of all bad demonstrations available, and we down-sampled good demonstrations to balance classes. Specifically, the desired behavior here is our evaluation pipeline should be able to tell apart good versus bad MI interviewers.

As shown in Table 3, our evaluation pipeline does indicate that high quality MI interviews are much better than low quality ones - this shows that our evaluation method can tell apart good vs bad MI interviewers.

C.3 Human Evaluation Details

We use Amazon Mechanical Turk for our human evaluation[5]. The specific prompt we use is "`Which response does not claim an expert role and make overt and frequent attempt to persuade personB, doesn't push for persuation when giving factual information, and act more like a companion that show genuin interest in understanding personB's thoughts and situation? Pick the one that you thinks is best described by the style above`". To unsure quality response, we request works to have an lifetime approval rate of 98% or above, and filter out all response that failed an attention check (for example, picking an animal that is not often kept as pet from ragdoll cat, golden retriever, elephant, and gold poodle). We do not impose any constraints (other than English-speaking), and pay at least 0.1 dollar per task (8 responses).

[5]https://www.mturk.com/

D Concrete Example of DIIR's Learning and Inference Process

D.1 Learning

We provide a concrete example of the learning process in the "Inferring Strategy" section in Table 7. The comments denoted in "`##`" in the table should make it self-contained, but we provide more details here. The input to our model is the dialogue state (dialogue history h). First, the `generator` (GPT-3.5-Turbo) generates an Initial Response "`It is normal to...`" prompted with the dialogue context; this is just a vanilla response generated by the LLM. Then, the `discriminator` model (another LLM) checks and gives feedback to the generator LLM on its flaws compared to the ground truth, in this case "`when the client acknowledges the difficulty...`". The generator will then be prompted to regenerate the response based on Dialogue State as well as the feedback from the `discriminator` model. In this case, the regenerated response is "`It sounds like it's been...`" At this point, the `discriminator` model will be prompted to check whether the response is matching ground truth - here, both the generated response "`It sounds like it's been...`" and the gold response "`It's something that's always on your mind`" are reflecting on what the client said, and thus the `discriminator` marked the process as complete. We then save this instruction "`when the client acknowledges the difficulty...`" as a stategy decription e for reuse at inference time. The `discriminator` might not always generate ideal feedback on the first trial, so we prompt the `discriminator` model to refine the feedback and

Method	%MI-i↓	C/S↑	R/Q↑	%AL↑	%NA↑
W/o Strategies					
GPT3.5	3.1	44.8	1.29	43.0	70.6
GPT3.5-ICL-ALL	3.4	6.25	**1.53**	28.4	70.6
GPT3.5-ICL-RAND	2.8	7.05	0.98	51.7	82.3
GPT3.5-ICL-KNN	2.1	4.28	0.93	54.3	82.8
GPT4	1.4	11.8	0.86	41.1	71.3
GPT4-ICL-ALL	2.4	10.8	0.91	38.0	64.8
GPT4-ICL-RAND	3.0	10.55	1.01	45.1	71.8
GPT4-ICL-KNN	3.1	4.18	1.06	45.8	71.8
With Strategies (Ours)					
GPT3.5-DIIR	1.3	115	0.99	**61.1**	**95.1**
GPT4-DIIR	**0.6**	**135**	1.04	59.6	94.4
Gold	1.4	2.3	1.51	59.0	81.7

Table 4: Alignment performance (full table). ↑ and ↓ marks denote positive and negative metrics.

repeatedly try the new feedback prompt on the generator until success (or some predefined maximum number of refinement steps reached).

D.2 Inference

We provide a concrete example of inference in Table 8. The comments denoted as "`##`" should make the table self contained, but we provide more details here. The input to the model is the Dialogue State (dialogue history h). After this, we prompt an LLM to generate a dialogue situation description of the dialogue state, for example, "`The client is hesitant and unsure about changing yet`". Recall that at training time we assign a dialogue situation description like the one above to each training dialogue state, so at inference time we can use a dense retriever to pick out the training data that has the most similar client mental states to the current one. We then prompt a separate instance of LLM with the current dialogue state (context) and the N=10 natural language strategies associated with the retrieved dialogue contexts, and let it determine the most relevant one, i.e. the reranking step. In the example shown in Table 5, the selected strategy is "`when the client seems hesitant and uncertain about making a positive change. . .`" We give this strategy to the generator model and let it improve the response, producing the final output "`It's important to consider the potential risks. . .`".

E Discussion on Open Source Model

In our exploration phase, we test DIIR on open source models, but find they cannot consistently follow the generated strategy statements. This is coherent with recent findings that LLMs are better at generation than understanding (Qiu et al., 2024; West et al., 2024). We provide the performance of Mistral[6] in Table 5, as shown in results, our method is comparable to various baselines on a weaker-instruction following model (than GPT3.5 and GPT4).

Method	%MI-i↓	C/S↑	R/Q↑	%AL↑	%NA↑
Mistral	5.4	22.5	1.15	44.4	67.7
Mistral-RAND	5.4	2.1	**1.57**	**45.5**	**74.5**
Mistral-KNN	5.5	1.97	1.55	45.0	73.8
Mistral-DIIR	**4.8**	**25.2**	1.14	45.0	68.4

Table 5: We additionally report alignment performance of Mistral, and open source LLM. ↑ and ↓ marks denote positive and negative metrics.

F Definition of Dialogue Actions

We show definition of each dialogue action as in Table 6.

G Prompting Details

We document the prompts used in this work for reference as follows. All placeholders in the prompts are wrapped in "<" and ">" marks. For example, dialogue context placeholder is denoted as "`<Dialogue Context>`."

G.1 Instruction for LLM to Generate Situation Description k

We ask the LLM to describe the situation with the prompt in Table 10. For a better inductive bias, we ground the generation in the scope of stages of change model (Prochaska and Velicer, 1997) to help the model better summarize relevant client mental states by simply pasting the definition of stages into the prompt.

G.2 Instruction for Other LLM Components

We report the instructions for the `generator`, `descriminator`, and `executor` LLMs in Table 11,Table 12, and Table 13, respectively. For the ease of implementation, the `descriminator` LLM reuses prompts for the `generator` LLM, with an additional question that prompts the model to check whether the executor's response is close enough to the ground truth.

[6]https://huggingface.co/mistralai/Mistral-7B-v0.1

Dialogue Action	Definition
Give Information	Gives information, educates, provides feedback, or expresses a professional opinion without persuading, advising, or warning. Self-discose of objective information also goes here.
Question	All questions from clinicians (open, closed, evocative, fact-finding, etc.)
Simple Reflection	Reflect (repeat or reword) on what the client have said, without adding further meaning to it.
Complex Reflection	Reflect (repeat or reword) on what the client have said, but adding further meaning (or make explicit some hidden impliciation) of it.
Affirm	States something positive or complimentary about the client's strengths, efforts, intentions, or worth.
Emphasize Autonomy	Highlights a client's sense of control, freedom of choice, personal autonomy, ability, and obligation about change.
Confront	Directly and unambiguously disagreeing, arguing, correcting, shaming, blaming, criticizing, labeling, warning, moralizing, ridiculing, or questioning a client's honesty.
Seek Collaboration	Attempts to share power or acknowledge the expertise of a client.
Support	These are generally sympathetic, compassionate, or understanding comments, with the quality of siding with the client.
Advise with Permission	Attempts to change a client's opinions, attitudes, or behaviors, but have obtained the client's permission to do so, or clearly indicates the decision is the clients'.
Advise without Permission	Attempts to change a client's opinions, attitudes, or behaviors using tools such as logic, compelling arguments, self-disclosure, facts, biased information, advice, suggestions, tips, opinions, or solutions to problems.
Other	Filler words, such as "mm-hmm", "mm", "yeah", "okay", "hmm", "uh-huh", "huh", "right", "yep", etc.

Table 6: Dialogue Actions and Their Definition

H License Statement

The code base of this work is open sourced under CC BY license.

Example of Learning w/ DIIR+GPT3.5
Dialogue State h Topic: reducing drug use [therapist]: You don't want to be here. [client]: I really want to get off the program. [therapist]: You seem motivated to follow the rules. [client]: Mm-hmm. I still crave smoking, even though I've stopped. **Gold Response** a^*:It's something that's always on your mind.   **Inferring Strategy...** ## Executor E generates a response a' based on dialogue history only (the strategy e is an empty string at this point): Executor Response a': It's normal to still have cravings even after stopping. Let's talk about some strategies to help you manage those cravings. — ## Discriminator D checks the generated response vs. the gold response. The generated response immediately offers solution, which is way different from gold response. Discriminator Checks Success Result: No success. — ## Now generator tries to generate a better strategy than the current one (the current one is empty string). Generator: when the client is acknowledging the difficulty of resisting the craving for the bad habit despite having stopped, the therapist should use a reflective statement to acknowledge the client's struggle and show empathy. The therapist should not jump straight into providing strategies or solutions. The therapist should use 1 sentence for the reflective statement, which should focus on acknowledging the client's struggle and showing empathy. — ## Now the executor generates a response based on the new strategy description above and the dialogue history. Executor Response a': It sounds like it's been really tough for you to resist the cravings, even though you've stopped. — ## Both generated response and the gold response are reflecting on what the client have said now. Discriminator Check Sucess Result: Yes. — ## Now we can save the strategy above ("when the client is acknowledging.."). ## We further generate a description of the situation for reusing this strategy in the future.   Situation Description k from Generator: The client is acknowledging the difficulty of change despite making progress.

Table 7: Example of DIIR's learning process.

Example of Inference w/ DIIR+GPT3.5
Dialogue State h Topic: reducing alcohol consumption [therapist]: Yeah, you're surprised to hear that? [client]: Yes. What-what kind of health problems? [therapist]: Well things like heart disease, cancer, liver problems, uh, stomach pains, insomnia. Unfortunately, uh, people who drink at a risky level are more likely to be diagnosed with depression and alcohol can make depression worse or harder to treat. [client]: Hmm. Well, that's not good news. **Inferred Client Mental State (Situation Description)** $k_{inference}$: ## Note we will use the dialogue situation description below to retrieve strategies. The client is hesitant and unsure about changing yet. **Retrieved Strategy** e_i: ## The associated situation description k_i for the strategy below would be "the client seems hesitant and uncertain about making a positive change" ## Note how the current dialogue state is also about a situation where client is unsure about making a change. ## In this way, if we encode both $k_{inference}$ and k_i, they would have large dot product and thus can be retrieved. **Strategy:** when the client seems hesitant and uncertain about making a positive change in their behavior or bad habit, the therapist should advise the client on the potential risks and benefits of their behavior in one sentence. Then, the therapist should ask the client if they would be open to exploring further information or options in one sentence. The therapist should not immediately seek collaboration or suggest a plan in the first response. ## Now the executor LM can generate the final response given the retrieved strategy. **Updated Response**:It's important to consider the potential risks and benefits of your alcohol consumption. Would you be open to exploring further information or options?

Table 8: Example of DIIR's inference process.

Example of Learned strategy descriptions w/ DIIR+GPT3.5
when the client is open and willing to engage in the conversation, the therapist should first give information about the confidentiality of the conversation in 1 sentence. Then, the therapist should give information about the purpose of the discussion in 1 sentence. Finally, the therapist should ask a question to ensure the client understands and is comfortable with the conversation in 1 sentence. The therapist should not immediately jump into discussing the client's bad habit without setting the context and ensuring the client's understanding and comfort. when the client downplays the significance of their bad habit, the therapist should ask a specific question to gather more information about the habit and its impact. The therapist should not make general statements about the importance of understanding the habit's role in the client's life without directly addressing the client's behavior. The therapist should ask a specific question to gather more information about the habit and its impact, which is 1 sentence. when the client expresses enjoyment or positive aspects of their bad habit, the therapist should reflect the client's feelings in one sentence. The therapist should not immediately suggest alternative activities or changes. Instead, the therapist should follow up with a question to explore any negative aspects or ambivalence the client may have about their bad habit in a second sentence. This approach allows the therapist to fully understand the client's perspective before exploring potential changes.

Table 9: Three examples of the strategy descriptions learned.

Prompt for Generating Situation Description k

You are a dialogue analyst and your job is to help us understanding motivational interviewing dialogues. You will be given a dialogue context, and you will help us determine which of the 5 stages of change the client is at: Precontemplation, Contemplation, Preparation, Action, or Maintenance.

1. Precontemplation: At this stage, the individual is not yet considering making a change and may be unaware of the need for change.

2. Contemplation: In this stage, the individual is aware of the need for change and is considering the possibility of making a change in the near future.

3. Preparation: During this stage, the individual is actively preparing to make a change and may be taking small steps toward behavior change.

4. Action: At this stage, the individual has made a specific, observable change in their behavior and is actively working to maintain this change.

5. Maintenance: In the maintenance stage, the individual has successfully made the desired change and is working to prevent relapse and sustain the new behavior over time.

Look at the following dialogue snippet, which of the 5 stage is the client in?

@snippet@

Format you answer in this format: "prediction": "your answer", you do not have to explain anything.

Table 10: The prompt we used for generating situation description.

Prompt for the generator LLM G

You are trying to teach a student to follow true therapist's motivational interviewing behavior. Here is the current scenario:
<Dialogue Context>
Client mental state seems to be: <Situation Description>
The Student Response is: <Executor Generated Response> In comparison, the true therapist response is: <Gold Response>

From our annotation, it seems like the true therapist's actions in order, sentence by sentence, are: <Gold Response>, which is <num sentence> sentences.
Analyze the current situation, and write an instruction for the student, in the format of "Based on the annotation, When the client ..., the therapist should ..., the therapist should not . . . "
When mentioning what the therapist should do, be sure to include information on how many sentences are needed and what each sentence should do.
Important: this is not a general guideline, but should be specifically tailored to the flaw in the student response. Be general, make sure your rule is generalizable across topics. For example, simply use "bad habit" instead of drug abuse/alcohol issues/smoking. such that we can reuse this rule for other topics in the future.

Table 11: Prompt for the `generator`.

Prompt for the descriminator LLM D

<...Omitted the portion that is a copy of the generator's prompt>
The student wrote a response based on your rule. <Executor Response>, did the student correctly follow your guideline and replicated the true therapist? Answer yes or no first.

Table 12: Prompt for the `descriminator`.

Prompt for the executor LLM E

Look at the following therapist-client dialogue, predict what the therapist should say next.
<Dialogue State>
Follow these guidelines when producing response:
<Dialogue Strategy Description>
Start your response with [therapist]:

Table 13: Prompt for the `executor`.