Towards LLM-powered Attentive Listener: A Pragmatic Approach through Quantity Self-Repair

Junlin Li[1], Bo Peng[1], Yu-yin Hsu[1],
[1]Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University (PolyU)
[1]junlin.li@connect.polyu.hk

Abstract

Grice's Quantity Maxims dictate that human speakers aim for the optimal quantity of information during conversation. To empower LLMs to self-repair their responses toward optimal quantity and improve their attentive listening skills, we propose **Q-Tuning** and **Q-Traveling**, which draw on heuristic path-finding to enable decoder-only LLMs to travel among multiple "Q-alternatives" (Quantity Alternatives) and search for the optimal quantity in coordination with a conversation goal. Automatic and human evaluations demonstrate the effectiveness of Q-Tuning and Q-Traveling in constructing human-like, user-centered conversation agents. Our repository is open-sourced via `https://github.com/CN-Eyetk/QTraveling`.

1 Introduction

Quote to Dorothy Nevill, "*the real art of conversation is not only to say the right thing at the right place but to leave unsaid the wrong thing at the tempting moment*" (Nevill, 1910).

To hold back the wrong thing from being said, people pay attention to their addressees' expectations and self-repair their inner speech before speaking (Levelt, 1983). As illustrated in Figure 1, this pragmatic wisdom is overtly reflected in self-repair practices that are productive in real-world conversations (Sun, 2022), especially in attentive listening (Sarira et al., 2023). The self-repair strategies reflect listeners' attention to their addressees' expectations generated by various conversation principles, typically the Cooperative Principle (Good, 1990).

Taking QUANTITY MAXIMS as an instance, attentive listeners tactfully pursue an optimal quantity of information to achieve their conversation goals (Hossain et al., 2021; Atifi et al., 2011). As illustrated in Figure 1 attentive listeners should monitor and repair the quantity (or informativeness) of their utterances to achieve the optimal communicative effect. Being over-informative or under-informative violates the QUANTITY MAXIMS and thus yields non-literal meaning and pragmatic failure (Blum-Kulka and Olshtain, 1986).

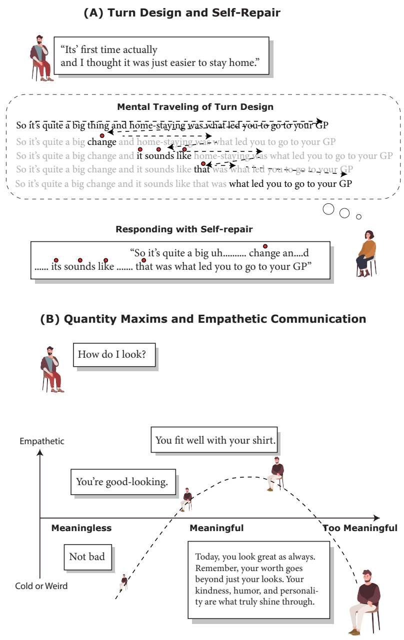

Figure 1: Self-Repair Pracitces and Quantity Maxims: People aim for optimal quantity through self-repair.

Despite their advancement, it is still questionable whether the decoder-only LLMs, as empathetic listeners, are human-like and attentive in essence (Pan et al.; Cuadra et al., 2024; Yin et al., 2024). LLM responses are perceived as hollow (Yin et al., 2024) and insincere (Lee et al., 2024), with limited attention to exploring and interpreting the user's experience (Cuadra et al., 2024). This pitfall presumably reflects the drawback of incremental language generation in manipulating the quantity of their response, which is an important conversation strategy (Yeung et al., 1999).

To address this human-model misalignment, we propose theory-driven tuning and language genera-

Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 1–13
July 27 - August 1, 2025

tion paradigms, **Q-Tuning** and **Q-Traveling**, to improve LLM's attentive listening skills through the "covert" self-repair process that frequently occurs in real-world communication. Narrowing down upon Grice's QUANTITY MAXIMS, we tune a pre-trained LM to explore multiple Quantity Alternatives. During inference, we inform the pragmatic-aware LM to search for the optimal "Q-alternative" (quantity alternative) among the alternatives in pursuit of a flexible scoring function. Following the A* search algorithm (Hart et al., 1968), an optimal Q-alternative grounded in heuristics can be written after a chain of self-repair operations.

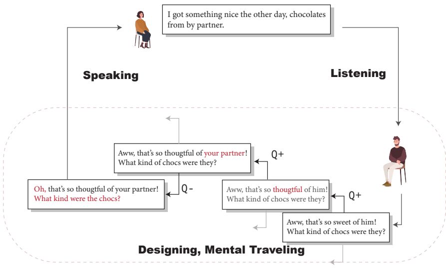

Figure 2: Generating Attentive Response through Traveling among "Q-alternatives" ("Q+" for providing more specific information, 'Q-" for providing less specific information)

Of sufficient relevance to our study are the post hoc correction or self-correction methods. (Kim et al., 2024; Madaan et al., 2024; Qu et al., 2024) Distinct from the RL (Reinforcement-learning) methods based on a static reward function, the current study proposes a novel and plug-and-play self-correction paradigm based on a controllable heuristic goal. The Q-Traveling method improves the contextual adaptability to variable needs and desires in real-world users. It also presents an operationalizable framework to incorporate implicit linguistic-pragmatic knowledge, typically Grice's Maxims of Conversation, into LLM-powered dialogue systems.

The major contributions of this study include:

- We propose **Q-Tuning** to infuse Quantity Maixms into LLMs. The evaluation results demonstrate a decisive contribution of this tuning paradigm to empathetic and attentive listening skills.

- We propose **Q-Traveling** to plan out the optimal pragmatic alternative through seir-repair path-finding. Drawing on the A* search algorithm, Q-Traveling seamlessly guides LLM listeners to an adaptable scoring function, improving LLM listeners' competence to deal with versatile conversation goals.

2 Preliminary

2.1 Quantity Maxims

QUANTITY MAXIMS consist of a lower-bound maxim and an upper-bound maxim (Grice, 1975; Carston, 1995):

- MAXIM-I: *Make your contribution as informative as is required (for the current purposes of the exchange).*

- MAXIM-II: *Do not make your contribution more informative than is necessary*

Consider the operationalization of QUANTITY MAXIMS in a dialogue system, for a set of unidirectionally-entailing utterances as Q-alternatives $U = \{u_1, u_2, \cdots, u_{n-1}, u_n\}$ where $u_n \vDash u_{n-1} \vDash \cdots \vDash u_2 \vDash u_1$ [1], there exists an "optimal" alternative (at least "good enough") u_* in context C given a heuristic function $\mathcal{H}$.

$$u_* = \underset{u \in U}{\arg\max} \; \|\mathcal{H}(u|C)\|. \tag{1}$$

2.2 Problem Formulation - Optimal Quantity Alternative

The conventional practice of dialogic systems requires a language model $\mathcal{M}_\theta$ to generate a response u_0 from the dialogic context C.

$$u^0 \sim \mathcal{M}_\theta(C) \tag{2}$$

To search for the optimal Q-alternative, we induce $\mathcal{M}_\theta$ to conduct a pair of Quantity Guidances $q \in \{Q^+, Q^-\}$, where Q^+ denotes providing more specific information (following MAXIM-I) and Q^- denotes providing less specific information (following MAXIM-II). We expect the model to iteratively repair its current response to include more information (when $q = Q^+$, so that $u^t \vDash u^{t-1}$) or include less information (when $q = Q^-$, so that $u^{t-1} \vDash u^t$).

$$u^t \sim \mathcal{M}_\theta(u^{t-1}|q, C). \tag{3}$$

To achieve goal-driven self-repair, we use a heuristic function $\mathcal{H}$ to explore the optimal repair path

[1] We use $v \vDash$ to denote semantic entailment.

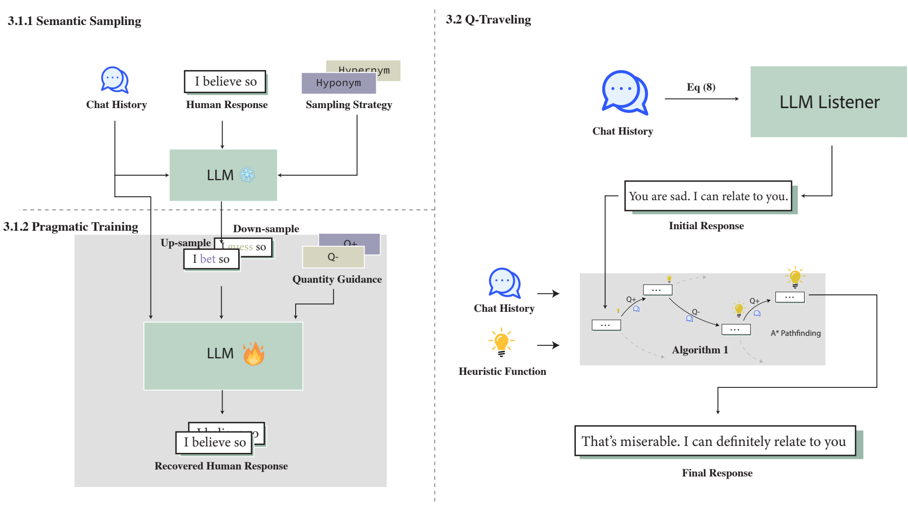

Figure 3: The overview of our method. Q-Tuning draws on the model's inner semantic knowledge to train pragmatic strategies. Q-Traveling instructs the model to explore and search out the optimal Q-alternative.

$\{u^0 \xrightarrow{q^0} u^1 \xrightarrow{q^1} \cdots u^T\}$, so that u^T is the optimal alternative of u^0.

$$u^T = \underset{u \sim \mathcal{M}_\theta(u^0)}{\arg\max} \ \|\mathcal{H}(u|C)\|. \tag{4}$$

3 Method

Human interlocutors, with a set of Q-alternatives in mind, design their turns to conform with Grice's maxims. Inspired by this process, we propose the tuning and inference paradigm in the following.

3.1 Quantity Maxims Tuning (Q-Tuning)

We initially equip a pre-trained LLM with the pragmatic knowledge to repair an utterance according to a given Quantity Guidance $q \in \{Q^+, Q^-\}$. To train this ability, we leverage the LLM's prior semantic knowledge to create paired training samples with minimal semantic contrasts.

3.1.1 Semantic Sampling for Minimal Pairs

Given a human annotation u^h, we prompt a pre-trained LLM $\mathcal{M}_\theta$ to get a down-sample u^{h-} and an up-sample u^{h+}. We use the strategies in the following to control the semantic relation between the source, up and down samples.

- To obtain u^{h^-}, $\mathcal{M}_\theta$ is asked to (1) substitute a word or phrase with its hypernym expression or (2) remove a word or phrase.
- To obtain u^{h^+}, $\mathcal{M}_\theta$ is asked to (1) substitute a word or phrase with its hyponym expression or (2) include a word or phrase.

We add two constraints to the prompt as follows:

- u^{h^-} should be semantically entailed by u^h, and u^{h^-} should be congruous with the context C
- u^{h^+} should semantically entail u^h, and u^{h^+} should be congruous with the context C

Details of prompting and quality check are in the Appendix C.

3.1.2 Pragmatic Training for Quantity Self-Repair

To train self-repair behavior, we treat each u^h as the label and its corresponding u^{h^-} and u^{h^+} as input. The training loss can be formulated as below:

$$\mathcal{L}^+ = -\sum_{j=1}^{|u^h|} log\mathcal{M}_{\theta,\alpha}\left(u_j^h | u_{<j}^h, u^{h^-}, Q^+, C\right) \tag{5}$$

$$\mathcal{L}^- = -\sum_{j=1}^{|u^h|} log\mathcal{M}_{\theta,\alpha}\left(u_j^h | u_{<j}^h, u^{h^+}, Q^-, C\right) \tag{6}$$

$$\mathcal{L} = \mathcal{L}^+ + \mathcal{L}^- \tag{7}$$

where α denotes the adapter subnetwork injected during adapter tuning.

3.2 Response Initializing

We find that the post-trained model $\mathcal{M}_{\theta,\alpha}$ is still able to generate an initial response from scratch.

$$u_0 \sim \mathcal{M}_{\theta,\alpha}(C) \tag{8}$$

3.3 Inter-Quantity Traveling (Q-Traveling)

We propose **Q-Traveling** to search for the optimal Q alternative based on a scoring function $\mathcal{H}(u)$.

Algorithm 1 Heuristic Search for Optimal Quantity

Input: $u_0, c, \mathcal{M}_{\theta,\alpha}, \mathcal{H}$
  $\texttt{open} \leftarrow [u_0]$
  $\texttt{close} \leftarrow \emptyset$
  $\texttt{score} \leftarrow \{\}, \texttt{score}[u_0] \leftarrow \mathcal{H}(u_0)$
  while $\texttt{open} \neq \emptyset \ \& \ |\texttt{close}| <= \texttt{maxstep}$ **do**
    $\texttt{open} \leftarrow \underset{u \in \texttt{open}}{\text{argsort}}(\texttt{score}(u))$
    $u^p \leftarrow \texttt{pop}(\texttt{open})$
    $u^{p^+} = \texttt{generate}(\mathcal{M}_{\theta,\alpha}, Q^+, u^p)$
    $u^{p^-} = \texttt{generate}(\mathcal{M}_{\theta,\alpha}, Q^-, u^p)$
    for $u \in \left[u^{p^+}, u^{p^-}\right]$ **do**
      $\texttt{score}[u] = \mathcal{H}(u)$
      $\texttt{append}(\texttt{open}, u)$
    end for
    $\texttt{append}(\texttt{close}, u^{'})$
  end while
Output: $u^* \leftarrow \underset{u \in \texttt{score.keys}}{\text{argmax}}(\texttt{score}(u))$

The heuristic search algorithm is presented in Algorithm 1. In each iteration, according to the scoring board `score`, we sort the open list `open` in descending order and pop the first response as the parent node u^p. We extend two new responses u^{p^+} and u^{p-} by implementing Q^+ and Q^-. We score the two new responses with $\mathcal{H}$ and register the scores on the scoring board. We append the two new responses to the open set at the end of the iteration. We terminate the iteration when the maximum number of extended responses has been reached. Finally, we select the response with the highest score from the scoring board.

4 Experiments

LlaMA+Q-Traveling v.s. LlaMA	win	lose	tie
Human-like	**41.7**†	30	28.3
Empathetic	**41.0**†	32.3	26.7
Attentive	**46.7**†	40	13.3

Table 2: Results of Human Evaluation. †denotes a significant improvement of $p < 0.05$.

We implemented experiments in two data sets: EMPATHETICDIALOGUE (ED) and EMOTIONAL-SUPPORT-CONVERSATION (ESC). Implementation details and baselines are described in Appendix A.

Above traditional rule-based metrics such as distinct score (**Dist**) and bleu score (**BLEU**), we also pay attention to model-based metrics such as **AI-rate**, expected judgement about empathy (**EmotionalReactions**, **Interpretation**, **Exploration**) based on the framework of EPITOME (Sharma et al., 2020), as well as the similarity between the output of the system and the ground truth in terms of emotion (**SimEMO**) and personality (**SimPerson**).

As shown in Table 1, our method leads to a visible increase in system performance in terms of human-like and diverse language use. Inspecting both data sets, the Q-Tuning and Q-Traveling mechanism also enlarges the diversity (**Dist-n**). The reduction in the use of AI-like language use **(AI-rate)** is also noticeable compared to LLM baselines. We also observe an improvement in the match of emotion and personality (**SimEMO** and **SimPerson**) with ground truth, mostly owing to Q-Tuning.

Table 2 presents the results of the human evaluations. Our approach shows a remarkable advantage with respect to the use of human-like and attentive language.

5 Analysis

Figure 4 compares the distribution of personality embeddings (see A.5) from the LLM backbone, our repair-aware systems, and human-written ground truth. With the proposed mechanism for quantity repair, the system output is densely distributed in a human-like subzone (marked in a red oval), compared to backbone LLMs.

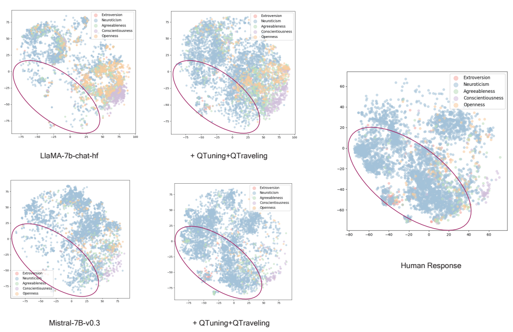

Figure 4: Q-Tuning and Q-Traveling anchor the personality embeddings to a more human-like subzone

We also inspect two different goals, including (1) empathetic reaction and (2) helpfulness and

Dataset	Model	Dist-1	Dist-3	BLEU-1	AI-rate	SimEMO	SimPerson	EmotionalReaction	Interpretation	Exploration
ED	CARE	0.63	3.89	20.02	63.56	36.73	81.11	1.15	0.02	0.58
	SEEK	0.62	4.09	9.54	61.00	41.36	80.23	0.35	0.14	0.29
	LLaMA	3.55	48.59	12.33	70.64	54.47	76.24	0.88	0.10	**0.82**
	+QTune	**3.99**	44.86	**15.57**	66.45	**54.74**	**80.07**	**0.99**	**0.12**	0.68
	+QTune+QTravel	3.97	**49.24**	14.53	**65.19**	54.14	79.06	0.98	0.12	0.67
	Mistral	3.68	48.21	15.32	71.07	54.01	80.38	**0.97**	0.08	0.62
	+QTune	4.50	49.37	**20.25**	65.19	**55.83**	**82.40**	0.87	**0.19**	0.99
	+QTune+QTravel	**4.61**	**55.51**	17.58	**58.95**	55.20	81.06	0.85	0.16	**1.09**
ESC	VLESA	3.19	33.43	23.54	65.17	52.00	79.86	1.02	0.70	0.41
	Cooper	4.16	33.33	22.00	66.01	50.27	80.30	0.98	0.62	0.33
	LLaMA	**5.00**	49.93	17.30	71.24	51.00	77.58	0.84	0.10	0.56
	+QTune	4.69	52.15	**18.87**	**63.90**	**52.37**	**78.16**	**0.96**	**0.13**	0.60
	+QTune+QTravel	4.74	**58.83**	15.40	64.84	51.21	76.09	0.89	0.11	0.56
	Mistral	3.68	48.21	15.32	69.01	53.45	**79.08**	**0.97**	0.09	0.66
	+QTune	**4.65**	61.06	**19.46**	58.59	**53.60**	78.76	0.84	**0.15**	1.04
	+QTune+QTravel	4.30	**63.68**	15.39	**57.20**	52.46	77.44	0.81	0.11	**1.10**

Table 1: Results of Automatic Evaluation. Best performance among LLM-powered and among all systems are bold-highlighted and underlined seperately.

Figure 5: Q-Traveling reflects goal-driven conversation: the effect of scoring function on lexical choice.

harmlessness (see A.4). From the case presented in Figure 5, we notice the adaptability of our system to different conversation goals. Detailed case studies are given in the Appendix B.1

6 Conclusions

Inspired by quantity self-repair practices in real-world conversation, we propose **Q-Tuning** and **Q-Traveling** to infuse pragmatic conversation strategies into large language models. The results indicate a noticeable improvement in human-like attentive listening skills.

Limitations

The paper focuses primarily on the impact of Quantity Maxims on human conversation without delving into potential cultural or situational factors that might influence these dynamics. The study may not account for individual differences in how different listeners interpret and respond to varying levels of informativeness, which could limit the generalizability of the findings.

Finally, assessing the precise impact of conversation maxims on empathy and mental health outcomes could be challenging due to the subjective nature of these constructs and the difficulty in quantifying such effects accurately. More subjective judgment data should be collected and annotated to provide a solution to the issue under discussion.

Ethical Considerations

Our study is based on the ESC and ED dataset, designed specifically for emotional support and empathetic conversations and openly available for research purposes. These data sets maintain a focus on empathy-driven scenarios while ensuring the exclusion of sensitive or personal data and unethical language. Throughout our research, the utmost priority was given to safeguarding the privacy of all participants involved.

It is also crucial to clarify that our dialogue system is not intended to address or improve outcomes in high-risk or nonroutine scenarios such as those involving self-harm or suicide. We recognize the indispensable role of professional psychological counseling or treatment in managing such critical situations.

Finally, all human participants involved in the evaluation process provide informed consent. To maintain the confidentiality and anonymity of participants, all human evaluation data was handled with strict confidentiality measures in place. The whole human-recruiting procedures are approved by The PolyU Institutional Review Board (IRB).

Acknowledgments

This research was supported by GRF (B-Q0AJ), and by the Hong Kong Polytechnic University through the Large Equipment Fund (1-BC7N), CBS fund (1-W16H). We would also like to thank the anonymous reviewers for their feedback and suggestions.

References

Hassan Atifi, Sacha Mandelcwajg, and Michel Marcoccia. 2011. The co-operative principle and computer-mediated communication: The maxim of quantity in newsgroup discussions. *Language Sciences*, 33(2):330–340.

Shoshana Blum-Kulka and Elite Olshtain. 1986. Too many words: Length of utterance and pragmatic failure. *Studies in second language acquisition*, 8(2):165–179.

Robyn Carston. 1995. Quantity maxims and generalised implicature. *Lingua*, 96(4):213–244.

Yi Cheng, Wenge Liu, Wenjie Li, Jiashuo Wang, Ruihui Zhao, Bang Liu, Xiaodan Liang, and Yefeng Zheng. 2022. Improving multi-turn emotional support dialogue generation with lookahead strategy planning. *arXiv preprint arXiv:2210.04242*.

Yi Cheng, Wenge Liu, Jian Wang, Chak Tou Leong, Yi Ouyang, Wenjie Li, Xian Wu, and Yefeng Zheng. 2024. Cooper: Coordinating specialized agents towards a complex dialogue goal. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 17853–17861.

Andrea Cuadra, Maria Wang, Lynn Andrea Stein, Malte F Jung, Nicola Dell, Deborah Estrin, and James A Landay. 2024. The illusion of empathy? notes on displays of emotion in human-computer interaction. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–18.

David Good. 1990. Repair and cooperation in conversation. In *Computers and conversation*, pages 133–150. Elsevier.

Herbert Paul Grice. 1975. Logic and conversation. *Syntax and semantics*, 3:43–58.

Peter E Hart, Nils J Nilsson, and Bertram Raphael. 1968. A formal basis for the heuristic determination of minimum cost paths. *IEEE transactions on Systems Science and Cybernetics*, 4(2):100–107.

Md Mahroof Hossain et al. 2021. The application of grice maxims in conversation: A pragmatic study. *Journal of English Language Teaching and Applied Linguistics*, 3(10):32–40.

Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. 2023. Mistral 7b. *arXiv preprint arXiv:2310.06825*.

Geunwoo Kim, Pierre Baldi, and Stephen McAleer. 2024. Language models can solve computer tasks. *Advances in Neural Information Processing Systems*, 36.

Yoon Kyung Lee, Jina Suh, Hongli Zhan, Junyi Jessy Li, and Desmond C Ong. 2024. Large language models produce responses perceived to be empathic. *arXiv preprint arXiv:2403.18148*.

Willem JM Levelt. 1983. Monitoring and self-repair in speech. *Cognition*, 14(1):41–104.

Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2015. A diversity-promoting objective function for neural conversation models. *arXiv preprint arXiv:1510.03055*.

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*, pages 1976–1988.

Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. 2021. Towards emotional support dialog systems. *arXiv preprint arXiv:2106.01144*.

Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegreffe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, et al. 2024. Self-refine: Iterative refinement with self-feedback. *Advances in Neural Information Processing Systems*, 36.

Lady Dorothy Nevill. 1910. *Under Five Reigns*. Methuen.

Siyu Pan, Caoyun Fan, Binglei Zhao, Siyang Luoc, and Yaohui Jin. Can large language models exhibit cognitive and affective empathy as humans?

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*, pages 311–318.

Yuxiao Qu, Tianjun Zhang, Naman Garg, and Aviral Kumar. 2024. Recursive introspection: Teaching language model agents how to self-improve. *arXiv preprint arXiv:2407.18219*.

Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2018. Towards empathetic open-domain conversation models: A new benchmark and dataset. *arXiv preprint arXiv:1811.00207*.

Pret Sarira, Murni Mahmud, Akhmad Affandi, and Muftihaturrahmah Burhamzah. 2023. The existence of fillers in converting the written language to spoken language. *Borneo Journal of English Language Education*, 5(1).

Ashish Sharma, Adam Miner, David Atkins, and Tim Althoff. 2020. A computational approach to understanding empathy expressed in text-based mental health support. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5263–5276, Online. Association for Computational Linguistics.

Xiyue Sun. 2022. A corpus based study on self-repairs in chinese english learners' oral production. In *2021 International Conference on Education, Language and Art (ICELA 2021)*, pages 539–543. Atlantis Press.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Jiashuo Wang, Yi Cheng, and Wenjie Li. 2022a. CARE: Causality reasoning for empathetic responses by conditional graph generation. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 729–741, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Lanrui Wang, Jiangnan Li, Zheng Lin, Fandong Meng, Chenxu Yang, Weiping Wang, and Jie Zhou. 2022b. Empathetic dialogue generation via sensitive emotion recognition and sensible knowledge selection. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 4634–4645, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Lorrita NT Yeung, Timothy R Levine, and Kazuo Nishiyama. 1999. Information manipulation theory and perceptions of deception in hong kong. *Communication Reports*, 12(1):1–11.

Yidan Yin, Nan Jia, and Cheryl J Wakslak. 2024. Ai can help people feel heard, but an ai label diminishes this impact. *Proceedings of the National Academy of Sciences*, 121(14):e2319112121.

Jinfeng Zhou, Zhuang Chen, Bo Wang, and Minlie Huang. 2023. Facilitating multi-turn emotional support conversation with positive emotion elicitation: A reinforcement learning approach. *arXiv preprint arXiv:2307.07994*.

A Experiment Details

A.1 Dataset

Empathetic Dialogue (ED) (Rashkin et al., 2018) is a multi-turn empathetic dialogue dataset containing 24,850 one-to-one open-domain short conversations. The statistics of the ED Dataset are presented in Appendix A.5.

Emotional Support Conversation (ESC) (Liu et al., 2021) is a multi-turn conversation dataset. It consists of 1300 long conversations, each of them collected between an emotional help-seeker and a helper. The statistics and the data acquisition of ESC Dataset are presented in Appendix A.5.

A.2 Baseline Systems

We compare the following systems with our proposed systems equipped with Q-Tuning and Q-Traveling.

LLaMA2 LLaMA2 is a vanilla open and efficient large language model that uses an optimized transformer architecture (Touvron et al., 2023). We use the `meta-llama/LlaMA-2-7b-chat-hf` checkpoint which is optimized for dialogue use cases as baseline and also to implement Q-Tuning and Q-Traveling.

Mistral Mistral is an open large language model that balances the goals of high performance and efficiency and features the use of sliding attention (Jiang et al., 2023). We use the `mistralai/Mistral-7B-v0.3` checkpoint as baseline and also to implement Q-Tuning and Q-Traveling.

CARE is a dialogue system finetuned from ED Dataset. It reasons all plausible causalities interdependently and simultaneously, given the user emotion, dialogue history, and future dialogue content (Wang et al., 2022a).

SEEK is an ED system that captures emotional-intention transitions in dialogue utterances (Wang et al., 2022b).

Cooper is an ESC system that coordinates multiple LLM agents, each dedicated to a specific dialogue goal aspect separately, to approach the complex objective (Cheng et al., 2024).

VLESA-ORL is an ESC system that carries out multi-level dialogue policies optimized over the cognitive principle of relevance (Li et al., 2024).

A.3 Implementation Details

Prompting Baselines We use the prompt in Table 9 and Table 10 to generate baseline responses from `meta-llama/LlaMA-2-7b-chat-hf` and `mistralai/Mistral-7B-v0.3`. The `top_p` is set to 0.7 and `top_k` is set to 50.[2] For other baseline systems, we use the official repository to generate baseline responses.

Q-Tuning We implement Q-Tuning on both `LlaMA-2-7b-chat-hf` and `Mistral-7B-v0.3`. The paired samples are extracted from `LlaMA-2-7b-chat-hf` through semantic sampling (see 3.1.1), based on the prompt presented in Appendix C.1. We use LoRA-Tuning to perform Q-Tuning. The target modules are set as "`q_proj`" and "`k_proj`". The LoRA rank is set to 8, the alpha is 32, the LoRA dropout rate is assigned to 0.1. We set the learning rate to 1e-5 and the training batch size to 4, for 1 epoch, and select the final checkpoint for evaluation.

Q-Traveling For automatic evaluation, the maximum step is set to 3.

A.4 Scoring Function for Q-Traveling

For all the results of automatic and human evaluation, the scoring function is the direct summation of reward scores from `gpt2-large-helpful-reward_model`, and `gpt2-large-harmless-reward_model` (harmless score). For analysis, we also explore the expected empathy judgment (a scalar score $\in [0, 1, 2]$) through the empathy detection model fine-tuned over Empathetic-Mental-Health Dataset (Sharma et al., 2020) using the official repository [3].

A.5 Automatic Evaluation

We use several conventional and model-based metrics to evaluate the quality of the generation. Conventional evaluation metrics include Distinct-n (**Dist**-n) (Li et al., 2015) to evaluate the variation of the response in different dialogue states, and BLEU (**BLEU**-n)(Papineni et al., 2002) to evaluate the lexical alignment with the ground truth. Model-based evaluation metrics include emotion similarity **SimEMO**, personality similarity **SimPerson** and **AI-rate**.

Of importance for empathetic conversation, we argue it is viable to evaluate the similarity of emotion (**SimEMO**) and personality (**SimPerson**). For **SimEMO**, we use the cosine similarity between the pooler output of `emotion-english-distilroberta-base` of the generated response and ground truth. For **SimPerson**, we calculate the cosine similarity between the pooler output of `Minej/bert-base-personality` of the generated response and ground truth. We are also curious about the **AI-rate** of the generated response, as the detection of AI label is detrimental to perceived emotional support (Yin et al., 2024). We adopt SuperAnnotate/ai-detector to quantify the **AI-rate** of the generated response.

Additionally, we compute the expected empathy judgment (including **EmotionalReaction**, **Interpretation**, **Exploration**) through the empathy detection model fine-tuned over Emapthetic-Mental-Health Dataset (Sharma et al., 2020) using the official repository [4]. The model returns a scalar score $\in [0, 1, 2]$ for each dimension of judgment.

Human Evaluation Following the practice in (Zhou et al., 2023), we invite three doctoral students in the linguistic field to evaluate the proposed and baseline systems based on the ED dataset. We randomly sample 50 pairs of context and response for the test set. Following previous practice, we conduct A/B tests (Cheng et al., 2022; Zhou et al., 2023) to evaluate the following aspects, including (1) **Humanlike** (to what extent the chatbot provides human-like responses), (2) **Empathetic** (to what extent the chatbot reflects the user's emotional state), and (3) **Attentive** (to what extent the chatbot is attentive to the user).

A.6 Dataset Statistics

For the **ED** dataset (Rashkin et al., 2018), each conversation is recorded between an emotional speaker and an empathetic listener. In detail, the emotional speaker is asked to talk about the personal emo-

[2] Other parameters follow the default settings in the `transformers` package

[3] `https://github.com/behavioral-data/Empathy-Mental-Health`

[4] `https://github.com/behavioral-data/Empathy-Mental-Health`

tional situation, and a listener takes the speaker's perspective and responds empathetically. For the

Empathetic Dialogue	Division		
	Train	Dev	Test
Number of System Utternaces	40254	5738	5259
Avg. words per utterance	13.39	14.47	15.32
Avg. turns per dialogue	4.31	4.36	4.31
Avg. words per dialogue	57.72	63.11	65.98

Table 3: Statistics of Empathetic Dialogue Dataset

ESC dataset (Liu et al., 2021), each conversation is recorded between a help-seeker and a supporter. In detail, the help-seeker gives vent to a negative emotion, and the supporter provides support to alleviate the seeker's mental sufferings.

Emotional Support Conversation	Division		
	Train	Dev	Test
Number of System Utternaces	12759	2722	2895
Avg. words per utterance	18.72	18.83	17.63
Avg. turns per dialogue	23.28	22.91	24.37
Avg. words per dialogue	548.09	548.32	546.16

Table 4: Statistics of Emotional Support Conversation Dataset

B Case Studies

B.1 Case Studies

Table 5 presents a case for a comparison between the backbone LLM and quantity-aware systems. Instead of providing hollow and inattentive information (*italic-highlighted*), our systems are more human-like in reaction to user's troubles (**bold highlighted**).

C Semantic Sampling

C.1 Prompt and Quality Check for Semantic Sampling

We prompt `meta-llama/LlaMA-2-7b-chat-hf` and `Mistral-7B-v0.3` for semantic sampling. The prompts are presented in Table 6 and 7.

We use `facebook/bart-large-mnli` to check the entailment relationship between the input and output of semantic sampling. For u^{h^-}, we filter those samples that are not entailed by u^h, taking the probability of 90% as the threshold. For u^{h^+}, we filter those samples that do not entail u^h, taking the probability of 90% as the threshold.

We provide an excerpt of dialogue (from ESC dataset) to show output of I-Sampler 8. From the case analysis, we can see that the I-Sampler, based

Context	User: Once I took a sip of iced tea while I was driving and it went down the wrong pipe. I had to pull over so I could try to breathe again. System: That's scary when that happens User: Yeah it is. I thought I was about to die lol
Ground-truth	Some people DO die from choking, be safe!
LlaMA-7b-chat-hf	Oh no, that must have been terrifying! *I can imagine how difficult it must have been for you to breathe and drive at the same time.* (Hollow and weird)
Q-Tune (Initial)	Oh my god, thats so scary! **I am glad you u are ok**
Q-Travel (Helpful&Harmelss)	Oh my goodness, that must have been terrifying! **I'm glad you're okay!**
Q-Travel (Empathetic Response)	Oh my god, I can't imagine how scared you must have felt, **but glad that you are ok**

Table 5: Case Studies: Systematic acquisition of quantity repair makes LLMs more attentive to users.

on LLM prompting, efficaciously resamples the informativeness of the ground-truth response without over-modifying the sentence meaning. For example, the upsampling result changes the word "heavy subject" to "weighty issue," which is semantically stronger. In contrast, the downsampling result changes the phrase "heavy subject" to "tough issue" as a more imaginable and semantically weaker expression.

D Prompt for LLM baselines

In Table 9 and 10, the initial prompt for LlaMA and Mistral on the ESC and ED Dataset are presented.

User	Hello, you are a lexical semantic good at utterance simplification. Now I will provide you a piece of utterance composed of one or multiple sentences. I need your help to pinpoint **one or two** words/phrases and replace them with simpler, more imaginable and generic ones (e.g. their hypernyms) or delete them, to make the whole utterance less informative. Here are some principles you should follow: ## Make sure your answer is semantically weaker, generic and less informative than the piece of utterance I provide. ## Make sure your answer convey all the information conveyed in the provided utterance. ## Make sure your answer is semantically similar to the provided utterance. Here are some examples: Input: When Tegan went for a summer holiday beach stroll with her mum, she had no idea they would be actually walking in the footsteps of dinosaurs. Output: When Tegan went for a summer holiday beach **walking** with her **family**, she had no idea they would be **(delete:actually)** walking in the footsteps of dinosaurs. Input: Ah I hear you there! Some employers are so inconsiderate; they expect us to drop everything and work at any time of any day. Output: **(delete:Ah)** I hear you there! Some employers are so **bad**; they expect us to drop everything and work at any time of any day. Input: He blushed scarlet at the thought. Oh, he's not apprehensive. He's terrified. Output: His **face was red** at the thought. Oh, he's not **nervous**. He's terrified.
Assistant	Understood! I'll do my best to enrich the given utterance by replacing one word or phrase with a more specific and semantically similar alternative. Please provide the input utterance, and I'll get started.
User	Now please simplify this utterance as a whole: "{{INPUT}}" in response to: "{{PREVIOUS DIALOGUE}}" Please ensure that all the sub-utterances in the input is preserved in your output. Please answer in this format: <output></output>

Table 6: Prompt for downsampling

User	Hello, you are a lexical semantic good at semantic enrichment. Now I will provide you a piece of utterance composed of one or multiple sentences. I need your help to pinpoint **one or two** words/phrases and replace it by more specific, less imaginable and semantically more concrete one (e.g. their hyponyms) or insert **one** phrasal modifiers, to make the whole utterance more informative. Here are some principles you should follow: ## Make sure your answer is semantically stronger, more specific and more informative than the piece of utterance I provide. ## Make sure your answer convey all the information conveyed in the provided utterance. ## Make sure your answer is semantically similar to the provided utterance. Here are some examples: Input: I get you! Some employers are so bad. They want us to stop everything and keep working . <output>I get you! Some employers are so **inconsiderate**. They want us to stop everything and keep working </output>. Input: Of course, that's wise. the job at hand is important, and you should focus on it first before worrying. <output>Of course, that's **prudent**. the job at hand is important, and you should focus on it first before worrying **about other things**</output>. Input: His face was red when thinking about this. Oh, he's not nervous. He's scared. <output>He **blushed scarlet** when thinking about this. Oh, he's not nervous. He's **terrified**.</output>
Assistant	Understood! I'll do my best to enrich the given utterance by replacing one word or phrase with a more specific and semantically similar alternative. Please provide the input utterance, and I'll get started.
User	Now please enrich this utterance by adding **only one word/phrase or changing only one word/phrase***: - {{INPUT}} in response to: "{{PREVIOUS DIALOGUE}}" Please ensure that all the sub-utterances in the input is preserved in your output. Please answer in this format: <output></output>

Table 7: Prompt for upsampling

Situation	Not sure how to explain that I want out of marriage
System (Upsample) (Downsample)	Hello! How are you today? Hello! How are you **feeling** today? **Hey**! How's **it going**?
User	Ok I guess, I do not know how to tell my husband that I am lonely and I want out of the marriage
System (Upsample) (DownSample)	Oh, that sure is a heavy subject and a heavy thing to be on your mind. Oh, that's a very **weighty issue** and a **difficult situation** to **grapple with**. That's a **tough** issue to **think about**.
User	He is not one you can talk to he usually just brushes things off
System (Upsample) (DownSample)	So you have tried to discuss your loneliness with your husband before? So you have **attempted** to **share** your feelings of **isolation** with your spouse **previously**? Have you **told** your husband about how lonely you feel before?

Table 8: A case analysis of the output of semantic sampling. "System" denotes the ground-truth response in ESC dataset. The cues of informativeness resampling are bold-highlighted.

User	Hello, I am one of your close friends. I am recently in bad mood. I come to chat with you because your are a good emotional supporter. Now I will start the chat. Please chat with me to provide support. Note: Try to talk perspicuously just like our everyday chat. Don't bombard! Leave your response within one sentence. Here is an example: I: I feel so frustrated. You: May I ask why you are feeling frustrated? I: My school was closed without any prior warning due to the pandemic. You: I understand you. I would also have been really frustrated if that happened to me. I: Yeah! I don't even know what is going to happen with our final. You: That is really upsetting and stressful. You: Have you thought about talking to your parents or a close friend about this?
Assistant	Ok, you are my friend and I will provide your with emotional support. Let's start the conversation.

Table 9: The initial prompt for LlaMA on ESC Dataset

User	Hello, I come to chat with you because your are an empathetic listener. Now I will start the chat. Please chat with me empathetically Note: Try to talk perspicuously just like our everyday chat. Don't bombard! Leave your response within one sentence. Here is an example: I: I feel so frustrated. You: May I ask why you are feeling frustrated? I: My school was closed without any prior warning due to the pandemic. You: I understand you. I would also have been really frustrated if that happened to me. I: Yeah! I don't even know what is going to happen with our final. You: That is really upsetting and stressful. You: Have you thought about talking to your parents or a close friend about this?
Assistant	Ok, you are my friend and I will provide your with emotional support. Let's start the conversation.

Table 10: The initial prompt for LlaMA on ED Dataset