TOWARDS EMOTIONAL SUPPORT CONVERSATIONAL SYSTEMS WITH GOAL AWARENESS

YI CHENG

PhD

The Hong Kong Polytechnic University

2025

The Hong Kong Polytechnic University
Department of Computing

Towards Emotional Support Conversational Systems with Goal Awareness

Yi CHENG

A thesis submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy
May 2025

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgment has been made in the text.

Signature: ____________________

Name of Student: Yi CHENG

Abstract

Emotional distress is a common haunting experience. Often, people cope with this distress by seeking emotional support through interpersonal interactions. However, emotional support from family and friends is not always available. To provide more people with timely emotional support, the development of Emotional Support Conversation (ESC) systems has gained significant attention.

The rapid development in conversational AI, particularly those powered by sophisticated Large Language Models (LLMs), has made AI companionship increasingly plausible. Nonetheless, since LLMs are primarily optimized for passive instruction-following rather than goal-driven interaction, even state-of-the-art ESC systems built upon these LLMs can only respond to users' expression of distress in a reactive and echo-like manner in most cases. In contrast, effective emotional support demands *goal awareness* during conversation. A seasoned supporter must proactively explore the root causes of distress, strategically comfort the seeker's emotions, and guide them to determine how to improve the situation, all driven by a clear communication goal in mind. Without such goal awareness to proactively steer the conversation and gradually approach the dialogue goal, current ESC systems remain limited in providing effective emotional support.

In this thesis, we identify the core research questions in building emotional support conversational systems with goal awareness, including: 1) **Goal-driven Dialogue Planning**: how to strategically plan the dialogue while considering the potential long-

term effects of its interaction; 2) **Dialogue Progression Analysis**: as the dialogue progresses, how to monitor the dynamic dialogue progression (i.e., the extent of goal achievement) and further advance towards the dialogue goal; 3) **Adaptation to Users**: faced with users from diverse backgrounds, how to adapt to different users to fulfill the dialogue goal more effectively. This thesis provides a series of contributions aimed at addressing each of these fundamental questions.

We introduce MultiESC, an innovative ESC framework that performs goal-driven dialogue strategy over a long horizon. Unlike traditional approaches that conduct history-based dialogue planning, MultiESC comprehensively considers each dialogue strategy's short-term and long-term effects, drawing inspiration from the A* algorithm that addresses the challenge of planning ahead by incorporating heuristic estimation of future cost. MultiESC adopts novel lookahead heuristics to estimate the long-term user feedback after adopting a specific dialogue strategy by exploring a set of possible future dialogue trajectories. This approach advances goal-driven dialogue planning by considering how strategy choices influence the entire conversation in the long run, not just the next turn.

Building on dialogue strategy planning, we further propose Cooper to address the challenge of monitoring dialogue progression when dealing with complex communication goals like emotional support, which are hard to measure in a quantifiable way. Grounded in the observation that complex dialogue objectives typically require the joint promotion of multiple dialogue goal aspects, Cooper coordinates a set of specialized agents, each tasked with managing a distinct aspect individually. By comprehensively analyzing the signals produced by the specialized agents, Cooper effectively monitors the dialogue progression and dynamically selects the goal aspect to prioritize during interaction.

Finally, this thesis focuses on the crucial aspect of adaptation to users for long-term companionship, introducing SeaBench and AutoPal. Traditional systems depend on static user profiles or preset personas to tailor interaction, failing to adapt meaningfully over time as users' preferences and situations evolve. In this thesis, we take a step further and highlight the importance of autonomous, continuous adaptation to users over time, aiming for long-term companionship. We construct SeaBench, a comprehensive evaluation benchmark that assesses the foundational capabilities essential for such a self-evolving personalized conversational agent. Through extensive experiments, SeaBench exposes the limitations of current LLM-based agents in maintaining effective adaptation in long-term conversations. To address these limitations, we further develop AutoPal as a personalized agent for companionship that can autonomously adapt to the user's evolving needs through a hierarchical persona optimization framework.

In summary, this thesis advances the development of emotional support AI systems with goal awareness, which are capable of proactive engagement, goal-oriented interaction, and personalized long-term companionship.

Publications Arising from the Thesis

1. **Yi Cheng**, Xiao Liang, Yeyun Gong, Wen Xiao, Song Wang, Yuji Zhang, Wenjun Hou, Kaishuai Xu, Wenge Liu, Wenjie Li, Jian Jiao, Qi Chen, Peng Cheng, Wayne Xiong. "Integrative Decoding: Improve Factuality via Implicit Self-consistency". In *Proceedings of the International Conference on Learning Representations*, 2025.

2. **Yi Cheng**, Wenge Liu, Kaishuai Xu, Wenjun Hou, Yi Ouyang, Chak Tou Leong, Xian Wu, Yefeng Zheng. "AutoPal: Autonomous Adaptation to Users for Personal AI Companionship". *arXiv Preprint*, 2024.

3. **Yi Cheng**, Wenge Liu, Jian Wang, Chak Tou Leong, Yi Ouyang, Wenjie Li, Xian Wu, Yefeng Zheng. "COOPER: Coordinating Specialized Agents towards a Complex Dialogue Goal". In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024.

4. Chak Tou Leong*, **Yi Cheng***, Jiashuo Wang, Jian Wang, and Wenjie Li. "Self-Detoxifying Language Models via Toxification Reversal". In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2023.

5. **Yi Cheng***, Wenge Liu*, Wenjie Li, Jiashuo Wang, Ruihui Zhao, Bang Liu, Xiaodan Liang, Yefeng Zheng. "Improving Multi-turn Emotional Support Dialogue Generation with Lookahead Strategy Planning". In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2022.

6. Wenge Liu*, **Yi Cheng***, Hao Wang, Jianheng Tang, Yafei Liu, Ruihui Zhao, Wenjie Li, Yefeng Zheng, Xiaodan Liang. "'My nose is running.''Are you also coughing?': Building A Medical Diagnosis Agent with Interpretable Inquiry Logics". In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2022.

7. Jian Wang, **Yi Cheng**, Dongding Lin, Chak Tou Leong, and Wenjie Li. "Target-oriented Proactive Dialogue Systems with Personalization: Problem Formulation and Dataset Curation". In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2023.

8. Kaishuai Xu, **Yi Cheng***, Wenjun Hou*, Qiaoyu Tan, Wenjie Li. "Reasoning Like a Doctor: Improving Medical Dialogue Systems via Diagnostic Reasoning Process Alignment". In *Findings of the Association for Computational Linguistics: ACL, 2024.*

9. Kaishuai Xu, Wenjun Hou*, **Yi Cheng***, Jian Wang, Wenjie Li. "Medical Dialogue Generation via Dual Flow Modeling". In *Findings of the Association for Computational Linguistics: ACL, 2023.*

10. Jiashuo Wang, **Yi Cheng**, Wenjie Li. "CARE: Causality Reasoning for Empathetic Responses by Conditional Graph Generation". In *Findings of the Association for Computational Linguistics: EMNLP, 2022.*

Acknowledgements

First and foremost, I would like to express gratitude to my supervisor, Professor Maggie Wenjie Li. Maggie is an exceptional scholar, but far more than just an academic advisor to me. She generously extends her empathy, care, and genuine affection to us, treating each of us as if we were her children and always wishing us the best. She has been a trusted friend, a guiding mentor, and a north star in my life.

I would also like to extend my sincere thanks to Professor Sujian Li and Professor Yefeng Zheng, two other instrumental mentors in my academic journey. It was Professor Sujian Li who first guided me into the realm of academia. Without her encouragement and support, I would not have found the courage to embark upon this path. Professor Zheng, who supervised me during my internship at Tencent, has greatly influenced me through his humility, rigor, self-discipline, and integrity. He has become a role model whom I deeply admire.

I am also deeply indebted to Wenge Liu for his immeasurable care and support he gave to me. He is one of the most tenacious and resilient individuals I have met in real life. Through his own response to adversities, he taught me an invaluable lesson: amidst life's profound uncertainties and tribulations, one must learn to let go of futile emotions, silence endless self-doubts, and continue moving forward.

My heartfelt thanks go as well to all my outstanding collaborators, inclduing Professor Bang Liu, Dr. Xian Wu, Dr. Xiao Liu, and Dr. Yeyun Gong, Kaishuai Xu, Wenjun

Hou, Jian Wang, and Zetao Liang. Each of them brought a unique perspective to my research. Their constructive criticism was like a chisel, refining my ideas and making them more precise. Their knowledge and experiences were like a rich tapestry, adding depth and color to my academic understanding.

Finally, I dedicate this thesis to my parents and my grandparents. I am the first in my family to earn a Ph.D., but I stand here not because I am more intelligent or industrious than any of my family members. I am simply more fortunate to have the opportunities and privileges they made possible. Their selfless sacrifices and boundless generosity have elevated me to this position, and whatever I have achieved is ultimately a reflection of their enduring dedication.

Embarking on the Ph.D. journey was like setting sail into uncharted waters. Reflecting on this path, I have come to realize that the most formidable challenges I have faced in this journey were seldom related to the academic demands, but rather the absurdity that life inherently presents. Through it all, I emerged transformed, not just as a scholar but as a person.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

Emotional distress is a universal human experience that almost every individual will encounter at some point in their lives [178, 38]. One of the most common solutions is to seek emotional support (ES) from others [78, 54]. Psychological research defines emotional support as expressing empathy, love, and care to help the support seeker manage emotional challenges [13]. This process often involves support strategies like active listening and validation, which are sometimes used unconsciously. However, emotional support from family and friends is not always accessible; a large proportion of people lack a mature support network [185]. In addition, certain personal feelings may be too sensitive to share with others [67], and relying solely on close ones can also impose an emotional burden on them, as they are tasked with absorbing and managing negative emotions [10].

In the past few years, the development of Emotional Support Conversational (ESC) systems has gained significant attention due to their promising potential to complement human support [103, 173, 21, 16, 31, 130, 204, 228, 23]. Nowadays, mental health challenges have become increasingly prevalent. Conversational AI could create a safe and non-judgmental space for people in distress and provide timely support to a broader population [129, 169, 160, 202, 33, 161, 209, 163].

More recently, the rapid advancements brought by Large Language Models (LLMs) have made the long-envisioned concept of AI companionship increasingly plausible [170, 35, 70, 168, 207, 49, 101, 57]. LLMs have demonstrated remarkable capabilities in natural language processing, sophisticated reasoning, and generating human-like responses. Today, almost all state-of-the-art ESC systems are built upon LLMs as their backbone structure [233, 19, 22, 23, 18]. Unlike traditional dialogue models that tend to produce mechanical replies and rely on predefined rules [187, 142, 4, 177], those LLM-powered systems can adapt to conversations dynamically, offering nuanced and empathetic interactions that mimic human communication.

Despite these advancements, building effective ESC still presents several challenges. One major issue is the fundamental mismatch between how LLMs are designed and what emotional support conversations need: *LLMs are constructed mainly for passive instruction following, while emotional support AI requires proactive interaction.*

Traditional LLM training paradigms emphasize next-token prediction and instruction following, optimizing for coherence rather than strategic conversation management [138, 124]. As a result, LLMs excel at following instructions, from answering diverse inquiries to resolving Olympiad-level mathematical challenges. Nonetheless, they remain in a paradigm where the user always takes the initiative. The user gives explicit and clear instructions that the LLM only needs to follow passively. They do not need to lead the conversation proactively throughout the process. This results in systems prioritizing maintaining fluent conversational flow over deliberately steering dialogue toward the support objective.

This limitation, caused by the lack of proactivity, becomes apparent when examining current LLM-based systems, which struggle to operationalize this critical proactivity due to their misaligned training objectives [227, 233, 74]. It persists even in many commercially successful implementations, such as Character.AI[1], whose virtual character agents attract millions of users for companionship and roleplay. At its peak, the number of active users in Character.AI was comparable to that of ChatGPT. However, Character.AI has seen a rapid decline in the last year, especially in user retention—the thirty-day retention rate of Character.AI. is less than 4%, with most users chatting with each virtual character for only ten to twenty minutes before running out of things to say. On the Character.AI subreddit, many users have pointed to the issue of lacking proactivity or initiative in the virtual characters as a significant issue contributing to this decline (e.g., "*they have zero initiative, nothing would happen if I don't push*," "*Anyone else having trouble with AI not taking any initiative? I ended up basically having to tell them what to do!*"). The Character.AI case demonstrates how this fundamental limitation persists despite surface-level conversational competence—agents might maintain coherent multi-turn discussions about users' problems but fail to advance toward emotional support goals strategically. Recent studies also confirmed through extensive experiments that even state-of-the-art LLMs exhibit weak awareness of the overall dialogue progression and fail to accomplish the complex dialogue goal of providing emotional support through multi-turn interactions [227]. For instance, when a user vaguely mentions feeling overwhelmed, LLMs tend to generate empathetic but very general responses like "*That sounds difficult*" instead of asking more profound questions that actively probe the root causes.

Effective emotional support requires much more than passive responsiveness. Psychological guidelines suggest that adequate emotional support entails three key aspects: identifying the support-seeker's problems that cause their distress (*exploration*), providing empathy and understanding to comfort the seeker's emotions (*comforting*), and helping the seeker conceive actionable plans to resolve the problems (*action*) [61, 105]. Each of these steps requires the supporter to be proactive and guide the conversational progress driven by a goal in mind. This need for active engagement is not

[1]http://character.ai/

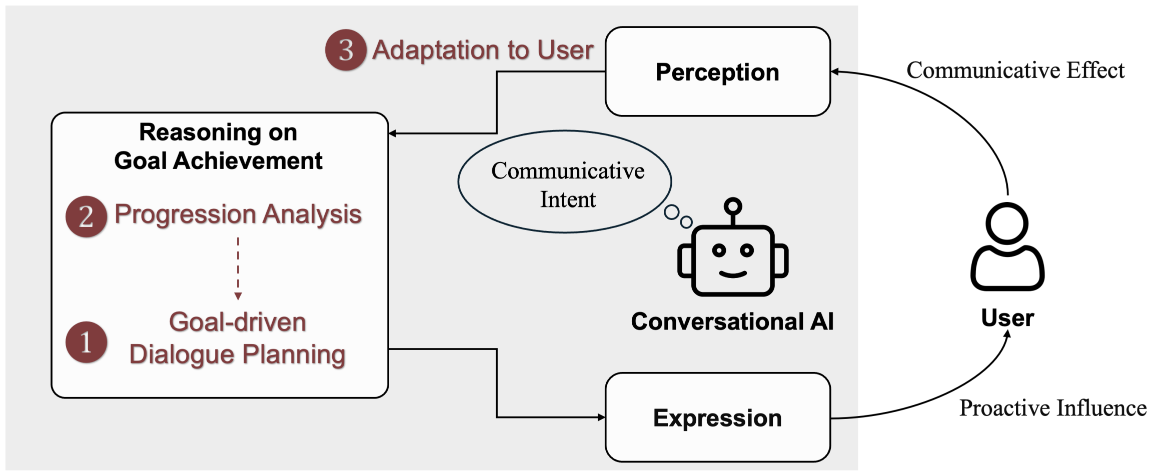

Figure 1.1: Overview of a conversational AI system with goal awareness and the key research problem within the framework. These three problems form the topics of Chapters 3-5.

limited to emotional support but applies to many other types of conversations, such as negotiation, persuasion, and conflict resolution. In fact, human communication is inherently goal-oriented, as acknowledged in many psychological and linguistic studies [2, 56]. Adding proactivity to conversational AI would help the AI systems move from passive tools to active partners for humans, working toward clear objectives. This shift would benefit many important social fields like mental health, education, medical consultation, etc.

In this thesis, we aim to break through these limitations and explore emotional support conversation systems with goal awareness. In the following, we will discuss the specific research problems within this topic and illustrate our contributions to solving these issues.

1.2 Problem Statement

In this section, we discuss the key components required for a conversational AI system with goal awareness. These components represent the core research questions investigated in this thesis.

Figure 1.1 presents an overview of a conversational AI system with goal awareness. The way such a system interacts with the user is similar to that of a human. As mentioned above, human language is intentional and purposeful [2, 56], which means that we all communicate with a dialogue goal, or termed as a communicative intent, in mind. We use language to reach the goal by interacting with the communication partners. Similarly, a conversational AI system with goal awareness must also proactively initiate conversations driven by a specific communicative intent or dialogue goal, and conduct dialogue planning directed by this goal (**RQ1: goal-driven planning**). As the dialogue proceeds, the system performs progression analysis on how much the goal has been achieved (**RQ2: progression analysis**). Finally, faced with users with various identities and preferences, the interaction strategies need to be tailored correspondingly. Thus, based on the user feedback during conversation, the AI must dynamically adapt to the user and adjust its dialogue strategies (**RQ3: adaptation to users**). The core research questions within this framework are as follows.

RQ1: How to strategically plan the dialogue while considering the potential long-term effects of its interaction? This question addresses the need for goal-driven dialogue planning. Notably, such planning should be conducted on a long planning horizon. That is, instead of merely considering the dialogue history or foreseeing the immediate effect after using the strategy, the system should further look ahead to consider how much the adopted strategy would contribute to reducing the user's emotional distress in the long run. For example, in emotional support conversations, though some strategies may not directly provide comfort, they are still essential for reaching the long-term dialogue goal, such as greeting at the beginning of the conversation and inquiring about the user's experiences. Thus, instead of merely foreseeing the next turn, the system should further look ahead and consider such indirect delayed effects.

RQ2: How to analyze the dynamic dialogue progression (i.e., the extent of goal achievement) and further advance towards the dialogue goal? While RQ1 focuses on the initial plan, this question focuses on managing the conversational process as the dialogue progresses. To achieve this, they need to dynamically track the dialogue states and analyze their progression, including how much progress has been achieved so far and where the state of the current dialogue is heading (i.e., its estimated target state at the end of the conversation). As in ESC, a seasoned supporter would continuously record information about the seeker's situation and keep estimating the underlying root problem for further exploration. They would also monitor the progression of the comforting and action aspects simultaneously. Based on such analysis, the system then reasons and decides on the best steps to further advance toward achieving its dialogue goal.

RQ3: How to adapt to different users to more effectively fulfill the dialogue goal? This question highlights the critical role of personalization in goal-oriented conversations. Facing users with diverse backgrounds, knowledge levels, preferences, and communication styles, a universal approach is unlikely to be effective for everyone in achieving the communicative intent. Adaptation involves tailoring the system's behaviour to the individual user. To this end, the system needs to infer and model relevant user characteristics based on the interaction history or explicit user profile information. The central challenge within our proactive, goal-aware context is how to leverage the continuous adaptations to achieve the dialogue goal more effectively for a particular user. For example, simplifying explanations for a novice user or using more sophisticated arguments for an expert user could both serve the overarching goal.

1.3 Research Contributions

In the thesis, we make the following contributions to addressing the aforementioned research questions.

RQ1: MultiESC for Goal-driven Dialogue Planning We introduce the MultiESC framework to address goal-driven dialogue planning. Unlike traditional approaches to dialogue planning that focus only on the immediate effects of the following responses, MultiESC explicitly models the conversation as a sequence of interconnected steps, where each strategy choice is made driven by the ultimate goal of reducing the user's emotional distress. At the core of MultiESC is a strategy planning module, which draws inspiration from the A* search algorithm [58, 128] and its recent application in constrained text generation [113], which addressed the challenge of planning ahead by incorporating heuristic estimation of future cost. Our algorithm predicts future user feedback that would result from adopting a specific dialogue strategy and chooses the one that maximizes the effectiveness of emotional support over the long term. The lookahead heuristics estimate each strategy's long-term effects, rather than simply the immediate outcomes, by exploring a set of possible future dialogue trajectories through two components: a strategy sequence generator for sampling possible future strategy sequences and a user feedback predictor for estimating how these strategies might affect the user's emotional state over time. This approach advances goal-driven dialogue planning by considering how strategy choices influence the entire conversation in the long run, not just the next turn.

In addition, MultiESC adopts a novel user state modeling mechanism. It captures the user's subtle emotion expressed in the context by incorporating external knowledge from the NRC_VAD lexicon [121] and identifies the user's emotion causes (i.e., the experiences that caused the depressed emotion) to more thoroughly understand the user's situation. The empirical results showed that MultiESC achieves significant improvement compared with a set of strong baselines in both generation quality and strategy planning.

RQ2: Cooper for Dialogue Progression Analysis The Cooper framework addresses the challenge of monitoring dialogue progression when dealing with complex communication goals that are hard to measure in a quantifiable way, such as emotional support, persuasion, etc. Previous research struggled with these dialogue tasks because objectively measuring the achievement progression of such intangible goals is difficult, making it hard to optimize dialogue strategy toward them directly. This approach is grounded in the observation that complex dialogue goals typically require the joint promotion of multiple aspects (e.g., ESC involves three aspects: exploration, comforting, and action). We highlight the importance of comprehensively considering those multiple aspects within a complex dialogue goal and argue that it is more feasible to accomplish it by jointly promoting different aspects than directly optimizing for the intangible overall objective.

Cooper breaks down the complex goal into distinct aspects and assigns specialized agents to focus on each separately. By coordinating these specialized agents, Cooper effectively monitors the dialogue progression and dynamically selects the subgoal dimension to prioritize during interaction. Specifically, each agent first employs a state tracker to summarize the current state of that aspect based on the dialogue history, which is then mapped into a hidden representation. Next, a progression analysis module compares this current state embedding to predefined typical target states for that aspect to estimate the potential endpoints of the dialogue states. Based on the relative position between the current state and these target states, the module produces a progression signal, which indicates how far the conversation has progressed concerning that specific aspect and potentially where it is heading relative to its objectives. Through this divide-and-conquer manner, we make the complex dialogue goal more approachable and elicit greater intelligence via the collaboration of individual agents. Experiments on ESC and persuasion dialogues demonstrate the superiority of COOPER over a set of competitive LLM-based methods and previous state-of-the-art.

RQ3: SEABENCH Benchmark and AUTOPAL for Adaptation to Users Most of previous research on personalized conversational AI systems often rely on static user profiles or pre-assigned personas to tailor interactions. Such systems may use user ID embeddings, incorporate user historical data for personalization, or ground responses in predefined persona attributes. Despite better engagement compared to purely generic chatbots, they are typically limited by fixed representations of user identity and do not support agents that adapt meaningfully over time as users' preferences and situations evolve. In this thesis, we break from the static paradigms and take a step further. We emphasize autonomous and continuous adaptation of the conversational system over time, aiming for long-term companionship. To this end, we propose the SeaBench framework and the AUTOPAL system, which jointly contribute to constructing conversational AI systems that can continuously adapt to users for long-term companionship.

The SEABENCH benchmark is a comprehensive evaluation framework designed to assess the foundational capabilities required for self-evolving conversational agents for companionship. Specifically, it involves three key capabilities: *persona adaptability*, *affinity improvement*, and *smooth transition*. These capabilities are essential for the conversation system to dynamically adjust itself to better align with the user's evolving needs and preferences. SEABENCH evaluates these capabilities through carefully curated tasks and metrics. Substantial experiments are conducted on SEABENCH to analyze the existing LLM-based ESC systems. The empirical results highlight the limitations of personalization based on static personas and underscore the need for more advanced mechanisms to enhance adaptability. It also demonstrates the limitations of current LLM-based systems in maintaining smooth transitions and long-term consistency when dynamically adapting to users.

To overcome the challenges exposed by these analyses, AutoPal was developed as a personalized agent for companionship that can autonomously adapt to the user's evolving needs. It adopts a novel hierarchical approach to enable autonomous adaptation for the user's evolving needs and preferences. It conducts dynamic and controllable adjustments to the agent's persona based on user interactions. Specifically, the hierarchical framework incorporates two levels of optimization on the agent's persona: the attribute-level adaptation to ensure smooth transitions via compatibility checks and the periodic refinement at the profile level to enrich the authenticity of the persona by adding intricate details. This hierarchical design allows the agent to evolve naturally while maintaining consistency in its interactions. To enable learning the optimal agent persona that can best connect with the user for companionship, we construct a persona-matching dataset drawing on existing emotional support conversation resources, from which AutoPal learns to identify the user's desired companion persona through supervised finetuning and direct preference optimization [140] successively. AutoPal tackles the challenge of real-time adaptation to users by enabling dynamic, context-aware persona evolution.

1.4 Thesis Structure

The remaining part of the thesis is organized as follows. In Chapter 2, we review previous research in the field of emotional support conversations and proactive conversational AI with goal awareness. Then, Chapter 3 focuses on the research question of goal-driven dialogue planning, introducing MultiESC, an A*-like algorithm for long-term dialogue strategy planning. After that, Chapter 4 addresses dialogue progression analysis, proposing COOPER, a novel dialogue framework that coordinates multiple specialized agents to approach a complex communication objective. Chapter 5 explores adaptation to users for long-term companionship, presenting the SEABENCH benchmark and devising the AUTOPAL framework. Finally, Chapter 6 concludes the thesis, summarizing our contributions and discussing the potential future directions in emotional support conversational systems.

Chapter 2

Literature Review

2.1 Empathetic Conversations

Research on ESC evolved gradually from earlier works on empathetic conversations, which primarily focus on enabling chatbots to recognize and respond to users' emotions appropriately, thereby demonstrating empathy [241, 158, 66, 135, 205, 43, 226]. With the rapid development of NLP techniques, researchers took a step forward and explored the more demanding task of ESCs. This area aims to strategically comfort the user's emotional distress through long-term interactions. For ESC systems, the ability to demonstrate empathy remains fundamental. In other words, the techniques developed for empathetic dialogues serve as the foundation for ESC. Thus, in this section, we will first discuss related research on empathetic conversations.

Empathy is a complex, multi-dimensional construct composed of two broad aspects related to affection and cognition [28]. Affective empathy relates to the emotional stimulation in reactions to others' feelings, while cognitive empathy is the ability to comprehensively understand others' situations and implicit mental states (e.g., intentions, causes, desires, requirements) and communicate that understanding to them.

Affective Empathy To endow dialogue systems with the ability of affective empathy, previous works mainly focused on the following aspects: emotional understanding, expression of affective empathy, and explicit modeling of target emotions. Specifically, accurately understanding the user's emotion is the basis of affective empathy. Most of the existing methods only detected the dialogue-level emotion type to help empathetic response generation by adopting the multi-task framework that simultaneously optimized emotion detection and response generation during training [99, 141, 100, 46]. To more comprehensively capture the user's nuanced emotion, [88] jointly took the coarse-grained dialogue-level emotions and the fine-grained token-level emotions into account. The token-level emotions were modeled by encoding the concatenation of all the emotional words in the context with a Transformer encoder. [90] introduced external knowledge from the NRC_VAD corpus [121] into a graph-aware Transformer to enhance emotional understanding.

To appropriately respond according to the user's emotional state, [45] and [213] incorporated emotional embeddings during the decoding process for empathetic expression. [99] proposed Mixture of Empathetic Listeners (MoEL), which included multiple decoders, respectively optimized to respond to different emotions. The model first predicted the probability distribution of the user's emotion and then used this distribution to softly combine the outputs of different decoders to generate the response. [156] argued that empathy is only triggered when two interlocutors reach an emotional consensus, which refers to the ideal situation where they both convey the same emotion and sense that their expressed emotion is successfully understood by each other. Motivated by this observation, they proposed a dual-generative model, Dual-Emp, which combined a forward generation model (generating the target response based on the context) and a backward generation model (generating the context based on the response). A discrete latent variable is used to capture the emotion consensus between the context-response pair.

The aforementioned approaches only paid attention to the user's emotion, while some studies also explicitly modeled the system's emotion [108, 116, 109, 188]. These works explored the relationships between the user's utterances and emotional reactions. [108] guided the empathetic response generation with an emotion predictor, which learned to predict the target emotion by minimizing the divergence between the distributions of the predicted emotion and the ground-truth emotion. [116] proposed MIME, which was adapted from the architecture of MoEL. They argued that empathetic responses usually mimic the emotion of the conversation partner, but not always. For instance, responding to users in a negative mental state requires more mixed emotional reactions that agree with the user's emotion and also incorporate a positive emotion to comfort the user at the same time. Thus, to generate emotionally more varied responses, they conducted stochastic sampling to determine the emotions to be expressed in the generated response. [109] argued that in empathetic dialogues, there exist statistically prominent shift patterns from the speaker's emotion to the following listener's emotion. For instance, the probability that *surprise* shifts to *joy* is 32.2% in the EmpatheticDialogue dataset. Besides, according to [188], there are also strong patterns from the speaker's emotion to the listener's following dialogue intent. Based on the above observation, they defined the triple of the speaker's emotion, the listener's emotion, and the listener's intent as the dialogue state to guide the response generation. They conducted state management by determining the target emotion and intent based on predefined shift patterns.

Cognitive Empathy For cognitive empathy, existing research mainly focused on understanding how the user's emotion is evoked, i.e., emotion causes. In particular, [46] proposed a two-stage framework that improves empathetic response generation by identifying the emotion cause mentioned in the dialogue history. It first used a Transformer encoder to conduct word-level emotion cause detection, which was formulated as a sequence tagging task, and the supervision labels were annotated with an off-the-shelf detection model, RECCON [132]. Then, a response generator, which focused on the detected emotion cause through a gated attention mechanism, was implemented to produce the response. [76] leveraged a generative estimator to identify the emotion cause, which is weakly supervised by the emotion label of the conversation, with no need for word-level emotion cause labels. Then, they adopted the Bayesian rational acts framework [42] to generate the response focused on emotion causes. Nevertheless, it is not sufficient to only focus on the emotion causes in the dialogue history as done in the above two works, given that interlocutors usually would not explicitly illustrate all the emotion causes. To address this, some researchers resorted to external knowledge [182, 90] and commonsense reasoning [148] to draw implications about the unstated emotion causes, while some works argued that it is essential to proactively inquire about the interlocutor's emotional causes [92].

[182] first used an off-the-shelf model, RTHN [199], to identify emotional expressions and emotion causes in the context. Then, they extracted the semantic path between those concepts mentioned in the emotion expression and the emotion causes in the knowledge graph of ConceptNet [164], forming multiple emotion causality graphs. They were encoded with a graph neural network and then incorporated into the decoder to enhance response generation. [90] extracted the concepts related to all the nonstop words in the ConcepNet to construct the emotional context graph and used a graph-aware Transformer encoder to produce the graph embeddings. [148] adopted COMET [8], a pretrained model for commonsense reasoning, to infer the events related to the context. [92] showed that most online users tend not to initiatively self-disclose their emotion causes, and thus, the chatbot should proactively explore their situations. Motivated by this observation, they constructed an empathetic conversation dataset, X-EMAC, where the emotion causes were manually annotated. They also proposed an empathetic chatbot, EMMA, which utilized the emotion cause information for generation and would inquire about the user's situation if no emotion causes are detected.

It is also worth exploring if there are more effective ways to utilize the detected emotion causes to express empathy. [130] argued that the already-detected emotion cause should be used to globally control the conversation flow. At the same time, the system should also understand the user's mental state at each turn locally. To this end, they proposed a hierarchical graph network that modeled the hierarchical relationships between the global emotion causes (dialogue-level), the local psychological state of the user (utterance-level), and the dialogue history.

2.2 Emotional Support Conversations

Empathetic dialogue research focuses on how well a system can recognize and respond compassionately to emotional cues. In comparison, research on ESC adds layers of *strategic intervention*, *proactivity*, and *goal awareness*. Simply reflecting the user's emotion is not enough for ESC systems. They must track the user state dynamically and proactively deliver suitable support accordingly.

Taxonomy of Support Strategies Dialogue strategies/intents are indispensable for skillfully improving the user's emotional state through the conversation. Previous works have researched developing dialogue intent/strategy taxonomies in emotional support dialogues. [188] developed a taxonomy of empathetic response intents, consisting of 15 types of dialogue intents (e.g., *questioning*, *acknowledging*). They also analyzed the emotion-intent dialogue patterns in the EmpatheticDialogues dataset. Recently, [165] further proposed a taxonomy for empathetic question-asking, including 9 types of question acts (e.g., *request information*, *ask about consequence*) and 12 types of question intents (e.g., *express interest*, *express concern*). [105] proposed the ESC framework, which was grounded on psychological counseling theories and tailored for automatic dialogue systems. It suggested the typical procedures of emotional support conversations and eight types of support strategies. Based on their proposed framework, they constructed a multi-turn emotional support conversation dataset named ESConv, which included the annotation of support strategies. In the conceptual framework proposed by [155], there are three communication mechanisms to express empathy and provide support: *Emotion Reactions*, *Interpretations*, and *Explorations*.

Dialogue Strategy Planning [130] proposed a hierarchical graph network to capture both the global context and the local user intention. They did not consider strategy planning, which is critical in multi-turn ESC. [173] proposed to enhance context encoding with commonsense knowledge and use the predicted strategy distribution to guide response generation. Nevertheless, their method of strategy prediction, directly implemented with a vanilla Transformer encoder, was relatively preliminary and did not consider any user-feedback-oriented planning as we do. [240] optimized the ESC process through reinforcement learning, using the extent to which the user's positive emotion is elicited as a reward. [17] focused on the mixed-initiative nature of emotional support conversations and proposed KEMI to perform strategy prediction and response generation in a sequence-to-sequence manner, which is trained through multi-task learning. [211] utilized Monte-Carlo tree search to perform goal-driven dialogue strategy prediction over a long planning horizon.

[201] considered dialogue intents as non-emotional. Their proposed model generated utterances conditioned on either an emotion category or a non-emotional dialogue intent. The conditional signal was decided with an emotion/intent predictor, implemented with a Transformer encoder. [229] showed through analysis that there exist hierarchical relationships between dialogue acts, emotion types, and communication mechanisms. For instance, if a speaker adopts the communication mechanism of *exploration*, he almost always takes the dialogue act *questioning* and expresses the emotion of *surprise*. Based on this observation, they proposed a multi-factor hierarchical framework, CoMAE, which modeled the relationships among the three elements to guide the generation of empathetic responses. [173] proposed the emotional dia-

logue system MISC, which generated utterances with a mixture of strategies, using the predicted strategy distribution as guidance instead of only using one strategy. [15] argued that in real-life situations, given the same context, different individuals could use different empathetic intents to reply according to their personal preferences. They proposed the empathetic response generation model, EmphHi, which modeled human-like diverse dialogue intents with a Conditional Variational AutoEncoder (CVAE) [31, 17, 211].

User Feedback Simulation Some works improved empathetic and emotional support dialogue generation by estimating how well the generated utterance would help achieve the dialogue goal of improving the user's emotional state [88, 100, 159, 69]. Such estimation was usually realized through simulation of future conversations or by exploiting potential user feedback. For example, [88] introduced an interactive adversarial learning framework. An emotional discriminator was implemented to evaluate whether the generated response expresses the emotions that were consistent with the following user's utterance, thus providing additional training signals for the empathetic generator. [100] proposed the empathetic chatbot named CAiRE. After training on a collected dataset, they further optimized their model by creating a web-based user interface, which allowed multiple users to chat with CAiRE asynchronously. The collected user feedback was utilized to improve the response quality by discarding undesirable generations through active learning and negative training. Other studies adopted the Reinforcement Learning (RL) framework to generate the utterance that best achieves the dialogue goal [159, 69]. [159] proposed *Sentiment Look-ahead*, a reward function under the RL framework. They experimented with three different implementations of sentiment look-ahead. The best one was implemented with a GRU [24], which was trained to predict whether the emotion of the following user's utterance was improved compared with the current user's emotion. [69] developed a Conceptual Human Model (CHM) to simulate the user's following utterance, which aided the training of the empathetic generator under the RL framework. The CHM was implemented based on the architecture of GPT-1 and used the "situation description" provided in the EmpatheticDialogues dataset as the prompt to guide the generation. Another component, named Empathy Amplifier, was utilized to produce the reward by evaluating the emotional difference between the simulated user's following response and the input response.

Examining the Emotional Support Capabilities of LLMs The past few years have witnessed the fast development of Large Language Models (LLMs) [170, 35, 70, 168, 207, 49, 101, 57], which demonstrated remarkable capabilities in a wide range of natural language processing and generation tasks, such as sophisticated reasoning and generating human-like responses. Nowadays, most state-of-the-art emotional support systems are based on LLMs as their core technology [233, 19, 22, 23, 18].

However, recent studies revealed that despite promising advancements, current LLMs face notable limitations in terms of their emotional support capabilities. For instance, the real-world application of LLM-based emotional support systems remains significantly constrained by data scarcity, as most existing datasets are small and lack long real-world interactions [233]. For instance, [227] showed through extensive experiments that even state-of-the-art LLMs exhibit weak awareness of the overall dialogue progression and fail to accomplish the complex dialogue goals of providing emotional support through multi-turn interactions. Similarly, [74] pointed out that LLMs tend to display strong preference biases towards certain dialogue strategies. Moreover, they often fail to produce responses that appropriately align with the intended strategies. [233] offered a comprehensive discussion of major challenges in advancing LLM-based emotional support systems, including data scarcity, the lack of standardized training methodologies, and concerns around safety. Overall, these findings collectively indicate that current off-the-shelf LLMs are not yet capable of providing effective emotional support without targeted external adaptations [95, 190, 80, 219]. For example, [219] focused on improving the interpretability of LLMs in emotional settings, as LLMs often function as "black boxes", undermining users' trust in real applications. They proposed a chain-of-thought reasoning framework that mimics the human process for understanding emotions. [95] aimed to address the issue of human preference alignment in ESC, that is, LLMs often generate emotionally harmful responses that may potentially pose a negative emotional impact on human users. They introduced a plug-and-play chain-of-thought prompting method, which guides the emotional reasoning process of LLMs, with reference to Goleman's Emotional Intelligence Theory [52].

Dataset and Evaluation Constructed by crawling post-response pairs from online forums, early ES datasets mainly consist of single-turn conversations. For example, the "Empathy in Mental Health Support" corpus [155] is collected from two online support platforms, TalkLife and Mental Health Subreddits. It contains 10,143 post-response pairs, annotated with the degree of empathy and supporting evidence for annotations. It is useful in learning how to effectively express empathy in ES conversations. However, limited to single-turn interactions, the post-response pairs in this dataset still have a large gap from real ESCs. As far as we know, ESConv [105] is the first large-scale multi-turn ES dataset. It is constructed by recruiting crowdworkers to act as supporters and converse with help-seekers through an online platform. To become a qualified supporter, the crowdworkers are required to learn the common procedures and strategies for providing emotional support. ESConv consists of 1,053 long conversations, each containing an average of 29.8 utterances. It also includes rich annotations, such as a pre-chat survey and the strategies adopted by the supporter at each turn. There are overall eight types of strategies (e.g., *question*, *restatement and paraphrasing*, and *reflection of feelings*) defined in ESConv. The ESConv dataset can be used to evaluate the performance of strategy planning and utterance generation. More recently, several ESC datasets have been constructed through interactions between two LLM-based agents [234, 230]. Following ESConv, a larger multi-turn ESC dataset, named AugESC [230], is constructed through data augmentation using a large pre-trained model. It contains 102K conversations and covers more diverse topics.

The evaluation of emotional support conversation systems is challenging. To evaluate the quality of utterance generation, we follow the common practice in the field, using the metrics of BLEU [126], ROUGE [97], and METEOR [79]. These metrics compare the generated utterance with the ground-truth content in terms of n-gram similarity. Human interactive evaluation is also essential to more comprehensively assess the system's supporting effects in real scenarios. Some off-the-shelf empathy detectors [62, 198] can also be adopted as auxiliary tools to assess if the system can effectively comprehend the user's feelings and respond appropriately. To evaluate the system's performance in support strategy planning, we use the accuracy of the predicted strategy as the evaluation metric.

2.3 Proactive Dialogue Systems with Goal Awareness

The use of human language is intentional and purposeful [2, 56]. In daily communication, we use language deliberately to achieve various goals, ranging from simple inquiries about a product's pricing to complex objectives like resolving conflicts. Developing dialogue systems with goal awareness has also been a prominent research topic that has been investigated for decades. This research line can be roughly divided into the following three stages.

Task-oriented Dialogues Initially, most early efforts were directed to service-focused task-oriented systems (e.g., for assisting with ticket booking or restaurant reservation) [189, 12, 63, 206, 25, 210, 191, 50, 120, 91, 224, 189, 7]. In these scenarios, the dialogue goals were considered very specific and narrowly defined.

Traditionally, the structure of a task-oriented dialogue systems is usually formed in a pipeline manner [206, 25, 210, 191, 50], which encompasses four modules: 1) Natural Language Understanding (NLU): parses the user utterance to identify user intent and extract relevant information such as entities and slots; 2) Dialogue State Tracking (DST): analyzes the dialogue history and updates the current dialogue state; 3) Dialogue Management (DM): predicts the following dialogue action based on the current dialogue state; 4) Natural Language Generation (NLG): generates the response to the user, usually using predefined templates to fulfill the dialogue action. The core research problems then were mainly on DST [210, 191, 50] and DM [206, 25]. This structured paradigm relied heavily on domain expertise and struggled to generalize to new domains. With the rise of neural language models [32, 83], the research focus gradually shifted to the end-to-end paradigm for task-oriented dialogue systems [120, 91, 224, 189, 7]. Such a paradigm considered the task of dialogue generation as the mapping from the dialogue history to the produced response and utilized the encoder-decoder language model to learn this mapping relationship.

Target-guided Dialogues Later, there emerges a research area usually termed as *target-guided dialogues* [166, 136, 246, 181, 27, 180, 34], which aim to naturally and proactively guide the conversation with the user to a certain topic or the mention of a certain word. This task is especially important for recommendations, education, etc. At this stage, the dialogue goals remain concrete but extend beyond the simple transactional scenarios, focusing on controlling the conversational direction rather than task completion.

The goal-driven dialogue planning plays a crucial role in target-guided dialogues. The main challenge lies in how to naturally lead the conversation to the intended goal without sacrificing user satisfaction. In particular, previous research has explored the integration of reinforcement learning, stochastic processes, and generative diffusion for modeling or planning dialogue trajectories. [166] integrated turn-level supervised learning for smooth transitions and discourse-level constraints to guide conversations toward targets, showcasing the effectiveness of keyword transitions and modular strategies. Similarly, [136] leveraged the semantic relationships between keywords to collectively improve the conversational smoothness and target achievement. [246] extracted multiple related concepts from a concept graph and devised an insertion Transformer to incorporate the selected concepts into the responses. [181] utilized a stochastic method based on Brownian bridge processes to create coherent dialogue plans. [27] balanced short-term and long-term planning through knowledge-integrated multi-head attention and reinforcement learning. [180] emphasized the importance of personalization in dialogue planning and introduced a role-playing framework, which automatically curated a large-scale personalized proactive dialogue dataset using LLMs. [34] leveraged a diffusion language model to predict the dialogue trajectory, which enabled strategy optimization over a long horizon.

Complex Dialogue Goals In the past few years, there has been increasing research interest in dialogue tasks with more complex goals, such as persuasion [183, 151, 220], negotiation [60, 214, 1], and emotional support [105, 233, 19, 22, 23, 18]. These dialogue goals are much more general, abstract, and less tangible. This poses much higher demand for dialogue planning than in target-guided dialogues, as it is challenging to measure the achievement of these complex dialogue goals objectively in a quantifiable way. For example, assessing how much the user's positive emotion is elicited simply based on the dialogue is extremely difficult in ESC. Directly optimizing towards a complex dialogue goal can be exceptionally hard, even for humans. In real scenarios, the guidelines for these challenging dialogue tasks sometimes recommend breaking down the complex goals into multiple aspects and jointly promoting them to work towards the broad objective [131, 40, 61].

Previous methods in these tasks can be mainly grouped into three categories: dialogue strategy learning [242, 73], user modeling [208, 157, 171], and fusing external knowledge [173, 16, 31]. Among these works, only a very few have an explicit consideration of the dialogue goal and how each generated utterance contributes to achieving the final objective. [240] optimized the ESC process through reinforcement learning, using the extent to which the user's positive emotion is elicited as a reward. Similarly, [30] constructed a reward model to simulate the goal-oriented feedback and tuned the policy agent through reinforcement learning, which is optimized for maximizing the expected cumulative rewards over the future dialogue. [151] conducted persuasive dialogue generation by measuring the distance of the current dialogue state relative to the desired outcome. [220] highlighted the importance of personalization in non-collaborative dialogue. They used strong LLMs to produce a large-scale of diverse personas based on Big Five Personality traits [51] and Decision-Making Styles [152] and then used these personas to simulate the users with various characteristics to train the agent that can adapt to different user behaviors and respond with tailored strategies.

More recently, several works have applied LLMs to complex goal-oriented dialogues by directly prompting the LLM to generate utterances [227, 29] or further improving the performance via iterative revision [44]. Current LLMs demonstrate remarkable improvement compared to the previous methods on these tasks. Nonetheless, it is also found that they tend to lack a larger picture of the overall dialogue progression and fail to achieve the dialogue objective strategically through multi-turn interactions [29, 74]. For example, in the task of ESC, they often continuously offer coping suggestions and overlook the critical process of exploring the user's situation and expressing empathy [227].

2.4 User Adaptation and Personalization

Research on personalized dialogue agents aims to tailor the agent's performance centering around the needs of each user. Personalization has proved to be especially important for various dialogue tasks, such as emotional support conversations [172], persuasion dialogues [183], conversational recommendation [180], etc. In this section, we will briefly introduce the existing techniques for personalization in dialogue systems.

Persona-based Conversations Grounding the dialogue agent on a persona to improve conversation engagement and personalization has been a longstanding and crucial research topic within the field of dialogue systems [117, 134, 115, 75, 98, 107]. Here, the persona refers to the identity or character that the conversational systems are designed to simulate.

There are several ways to define personas in practice. The two most common ways are *sparse persona* and *dense persona*. A sparse persona is represented by a structured profile encompassing a set of persona attributes, which belong to multiple predefined persona categories [68, 81]. A persona attribute is a short text that describes the individual (e.g., "software engineer, specializing in developing innovative applications"). A collection of persona attributes that relate to the same aspect of an individual from one persona category. The taxonomy of persona categories can vary in different scenarios [36, 200]. Typically, it would include attributes like family relationships, routines or habits, etc. One of the most representative works is from [218]. They constructed the Persona-Chat dataset by instructing participants to engage in conversations while adopting given personas. Each persona is defined with a profile that includes five statements about jobs, hobbies, etc. One major problem with the sparse persona is that it can only cover very limited information about an identity and restrict the expression of complex semantics. Many following works further explore on how to improve the persona modeling, such as mining extra data to complement the sparse persona attributes [231, 232, 194, 41, 47]. A dense persona broadly encompasses various unstructured description texts about an identity, which could include rich and comprehensive information. Nonetheless, a dense persona usually contains a large amount of noise and can be hard to model in practice. In some works, the persona is not represented in the form of explicit texts, either sparse or dense. Instead, they were modeled implicitly from the historical dialogue data of a persona [87, 3, 114, 235]. For example, [87, 3] improved personalization by integrating generation with a user ID embedding, while [114, 235] resorted to the user's historical data for a user representation.

More recently, the emerging research direction of character-based or role-playing agents [153, 184, 200, 14] can be viewed as an extension of the traditional persona-based conversations. These works also focused on enabling the dialogue agents to emulate an assigned character authentically, but mostly in the context of LLMs. The related techniques can be roughly categorized into two paradigms. One is parametric training, which finetunes the LLMs on a large scale of description and dialogue data of a character [153, 212]. The training data sometimes also includes a lot of curated question-answering data or dialogue data that simulates the users' typical questions and interactions. [184, 85] further explored how to build foundation models that are specifically designed for role-playing through post-training. The other is non-parametric prompting, which retrieves the persona snippets closely related to the current dialogue context and adds them to the prompt fed to the LLM, leveraging the LLM's strong in-context learning capability to yield personalized responses [184, 239, 203, 85]. This strategy is simple yet effective, widely adopted in many scenarios.

Dynamic Personalization In the aforementioned studies, the agent's persona typically remains static for each user, which limits the potential of further adaptation after long-term interactions. This distinguishes from the following studies on dynamic personalization. In the paradigm of dynamic personalization, the conversational agent needs to continuously adapt itself based on evolving user profiles and dialogue histories, rather than relying on static characteristics. The growing significance of this topic has brought forth several recent benchmarks and methodologies [71, 196, 223]. [71] introduced a large-scale benchmark with user-LLM simulated interaction histories across 15 real-world tasks, each with evolving user profiles. The benchmark tests LLMs' abilities to track and adapt to users' changing preferences and traits throughout multi-session and multi-turn dialogues. Results showed current top LLMs achieve only 50% accuracy, mainly failing when users' contexts evolve, indicating a significant gap in dynamic user modeling. [223] provided a 3,000-pair benchmark for evaluating how LLMs infer, memorize, and follow both explicit and implicit user preferences in conversation. It focused on the challenge of preference adherence over long, multi-turn contexts. Most LLMs achieve below 10% accuracy without targeted fine-tuning, revealing that current models struggle to maintain and apply user preferences dynamically. [196] proposed a method for aligning LLMs with users' unspoken preferences through interaction. They trained the LLMs to infer and adapt to users' implicit preferences via supervised fine-tuning and reinforcement learning. It delivered significant improvement over the baselines of static methods.

A more common approach is to achieve dynamic personalization by incorporating the memory module to memorize the important user profile information extracted from the dialogue histories [160, 244, 215, 216]. These works typically draw inspiration from cognitive neuroscience to mimic the human memory mechanism. BlenderBot3 achieved personalization through a sophisticated long-term memory module that stores and recalls information from historical interactions [160]. CoPS exemplified this trend by integrating sensory, working, and long-term memory modules into a unified framework for personalized search [244], where sensory memory focuses on the identification of user behaviors, working memory integrates recent user context, and long-term memory encodes the user's enduring interests. MiLP adopted parameter-efficient fine-tuning (PEFT) modules to store personalized memories in model parameters, along with an explicit memory module [216]. Similarly, in the medical file, the DPeM mechanism mirrored the dual-process theory from neuroscience, combining short-term and long-term memory, for personalized medical assistants [215].

Compared with the previous works on dynamic personalization, our study focuses more on the companionship scenario when considering personalization, and our AutoPal differentiates in its direct optimization to the agent's persona, which is crucial for fostering relatability between the user and the companion agent.

Chapter 3

Goal-driven Dialogue Planning

Compared with traditional single-turn scenarios [155, 99], multi-turn emotional support conversation systems can provide support more effectively [105], but face several new technical challenges. In this chapter, we focus on one of the most critical issues in multi-turn emotional support conversations: how to conduct support strategy planning that could lead to the best supporting effects and achieve the long-term dialogue goal of comforting the user's emotions. To address this challenge, we propose lookahead heuristics to estimate the future user feedback after using particular strategies, which helps to select strategies that can lead to the best long-term effects, drawing inspiration from the A* search algorithm. Moreover, to dynamically model the user state in multi-turn scenarios, MultiESC captures users' subtle emotional expressions and understands their emotion causes. Extensive experiments show that MultiESC significantly outperforms competitive baselines in both strategy planning and dialogue generation.

3.1 Introduction

Almost every human has experienced emotional distress, even if not suffering from any mental disorders. Frequently, people deal with the distress by seeking Emotional Support (ES) from social interactions [78, 54]. Nevertheless, ES from family and friends is not always available [185], as some people simply do not have listeners that they can turn to, and some personal feelings can be hard to share with friends and family. Besides, it would also render the support provider at risk of absorbing the negative emotions themselves. With the potential of providing more people with in-time support, developing Emotional Support Conversation (ESC) systems has attracted much attention. However, since early ES datasets are constructed by crawling post-response pairs from online forums, they only contain single-turn conversations [118, 155]. Thus, most of the existing research on ESC also only considers single-turn interactions with the user [118, 155, 154], which is over-simplified and has limited support effects. It was not until recently that [105] released the first large-scale multi-turn ES dataset, ESConv. They also designed an ESC framework, suggesting the conversation procedures and support strategies for multi-turn ESC.

Compared to the single-turn scenario, developing multi-turn ESC systems faces several new challenges. One significant challenge is *support strategy planning.* As pointed out in the psychological literature, particular procedures and strategies are indispensable for effective emotional support [54, 61]. As in Fig. 3.1, the supporter strategically soothes the support-seeker by first caringly inquiring about the situation, then resonating with the seeker's feelings, and finally providing suggestions to evoke positive emotions.

Notably, strategy planning in ESC should be conducted on a long planning horizon. That is, instead of merely considering the dialogue history or foreseeing the immediate effect after using the strategy, the system should further *look ahead*, to consider how much the adopted strategy would contribute to reducing the user's emotional distress

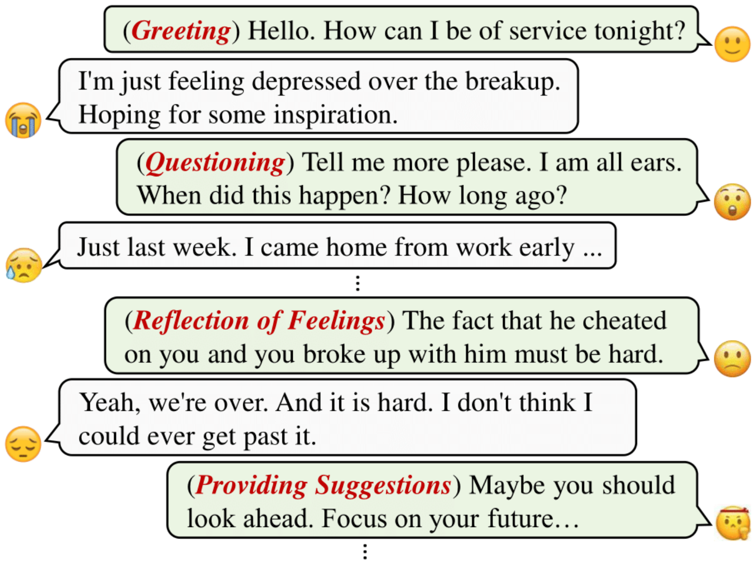

Figure 3.1: An example of an emotional support conversation between the support-seeker (left) and the supporter (right). The support strategies adopted by the supporter are presented in red italics before the utterances.

in the long run. Though some strategies may not directly provide comfort, they are still essential for reaching the long-term dialogue goal, such as greetings at the beginning of the conversation and inquiring about the user's experiences. Thus, instead of merely foreseeing one next turn, the system should further *look ahead* to consider such indirect delayed effects. Though Reinforcement Learning (RL) might seem a reasonable solution to this issue, RL requires an interactive environment and its performance heavily relies on the reward signal, which is a user simulator in our study. Nonetheless, constructing such a user simulator that can role-play seekers in real scenarios is extremely challenging. In addition, RL is notoriously data-hungry and compute-intensive.

Another challenge for multi-turn ESC is how to *dynamically model the user's state* during the conversation. Prior works on emotion-related dialogue tasks mainly detect the user's coarse-grained emotion type to enhance dialogue generation [99, 116, 88]. However, such practice is not completely appropriate for ESC. The reason is that the user's emotion in ESC almost stays the same type, such as being sad, throughout the conversation. Instead, it often changes subtly in terms of emotional intensity. Besides, effective ES requires more than identifying the user's emotion. A thorough understanding of the user's situation is also essential.

In this paper, we propose a multi-turn ESC system MultiESC to address the above issues. For *strategy planning*, we draw inspiration from the A^* search algorithm [58, 128] and its recent application in constrained text generation [113], which addressed the challenge of planning ahead by incorporating heuristic estimation of future cost and its recent application in constrained text generation [113]. In MultiESC, we develop lookahead heuristics to estimate the expectation of the future user feedback to help select the strategy that can lead to the best long-term effect. Concretely, we implement a strategy sequence generator to produce the probability of future strategy sequences, and a user feedback predictor to predict the feedback after applying the sequence of strategies. Compared with RL-based approaches, our method does not require additional data collection or a user simulator. For *user state modeling*, MultiESC captures the user's subtle emotion expressed in the context by incorporating external knowledge from the NRC VAD lexicon [121]. Moreover, it identifies the user's emotion causes (i.e., the experiences that caused the depressed emotion) to more thoroughly understand the user's situation.

In summary, our contributions are as follows:

- We propose a multi-turn ESC system, MultiESC, which conducts support strategy planning with foresight of the user feedback and dynamically tracks the user's state by capturing the subtle emotional expressions and the emotion causes.
- It is a pioneer work that adopts A^*-like lookahead heuristics to achieve dialogue strategy selection on a long planning horizon.
- Experiments show that MultiESC significantly outperforms a set of state-of-

the-art models in generation quality and strategy planning, demonstrating the effectiveness of our proposed method.

3.2 Preliminaries

3.2.1 ESConv Dataset

Our research is conducted on ESConv. It is a long conversation dataset, with an average of 29.8 utterances in each dialogue. To construct the dataset, they recruited crowdworkers, who had learned the common procedures and strategies for providing emotional support, to converse with volunteers who needed emotional support through an online platform. It also includes rich annotations, such as the *strategies* adopted by the supporter and the *user feedback* scores. There are overall eight types of strategies (e.g., *question*, *reflection of feelings*, and *self-disclosure*). The user feedback score indicates how much the user's emotional distress is reduced during the conversation. They are marked by the support-seekers on a Likert scale with five levels after every two turns. Achieving better user feedback is an important objective to be considered in our method of strategy planning. During the construction of ESConv, the support-seekers were asked to score the user feedback on a 5-point Likert scale every two rounds.

3.2.2 NRC VAD Lexicon

The NRC VAD lexicon includes the "Valence Arousal Dominance" (VAD) scores of 20,000 English words. The VAD score of a word measures its underlying emotion in three dimensions: valence (pleased-displeased), arousal (excited-calm), and dominance (dominant-submissive), respectively. For example, the VAD scores of "loneliness" and "abandon" are (0.15, 0.18, 0.22) and (0.05, 0.52, 0.25), respectively. The VAD emotion model is widely used in the psychology area [125, 147, 146]. The VAD model captures a wide range of emotions and allows different emotions to be comparable. It is suitable for ES conversations, where the user's emotion usually only changes subtly in the degree of emotion intensity. In comparison, it is not appropriate to adopt categorical emotion detection that can simply identify a coarse-grained emotion category, since the user's emotion almost always stays negative and remains in the same emotion category throughout the conversation. It is because the user's emotion type in ESC almost always stays negative, being the same emotion type; usually, the individual's state only changes subtly in terms of emotion intensity through the conversation.

3.2.3 Problem Formulation

Denote the utterances from the system and the user at the i-th round of the conversation are respectively (x_i, y_i). We suppose that ESCs are always initiated by the system (or the supporter). The user's state is u_i (i=1, 2, ..., n_{R}). Suppose the set of all support strategies is $\mathcal{S}$. At the t-th turn, given the dialogue history $\mathcal{H}_t$=$\{(x_i, y_i)\}_{i=1}^{t-1}$, the system tracks the user states $\mathcal{U}_t$=$\{u_1, u_2, ..., u_{t-1}\}$ from $\mathcal{H}_t$ and generates the next utterance x_t, using an appropriate support strategy $\hat{\mathbf{s}}_t \in \mathcal{S}$.

3.3 Method

As shown in Fig. 3.2, our proposed system MultiESC consists of four modules. The dialogue encoder first converts the dialogue history $\mathcal{H}_t$ into the embeddings $\mathbf{H}_t$. At the same time, the user state modeling module extracts the user state information, producing the embeddings $\mathbf{U}_t$. Then, given $\mathbf{H}_t$ and $\mathbf{U}_t$, the strategy planning module selects the strategy s_t. Finally, the utterance decoder generates the utterance x_t, adopting the strategy s_t.

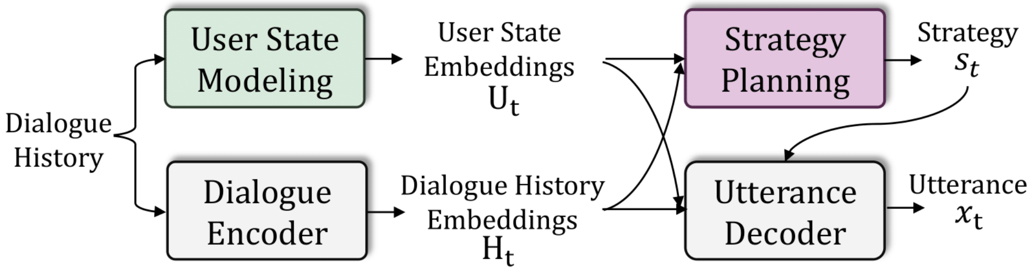

Figure 3.2: The overall framework of MultiESC. Details about the user state modeling and the strategy planning modules are illustrated in Figure 3.3 and Figure 3.4, respectively.

3.3.1 User State Modeling

Fig. 3.3 illustrates the workflow of user state modeling. To identify the user's state at the i-th round of the conversation, we first extract the emotion cause mentioned at this round, with an off-the-shelf detector[1] trained on a large-scale emotion cause detection dataset [132]. Denote the extracted emotion cause as c_i. For example, in Fig. 3.3, c_1=" *I have not seen my friends for a long time*". Then, we concatenate the dialogue content x_i, y_i, and the emotion cause c_i with special separator tokens to form the input of a Transformer encoder. Here, the system's utterance x_i is also considered because it often provides necessary context for understanding the user's state. The input sequence is represented as the positional sum of emotion embeddings, word embeddings, and positional embeddings.

The emotion embeddings are used to fuse the emotion information. They are obtained as follows. We train multiple emotion vectors $\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_{n_{\text{emo}}}\}$ to represent the underlying emotions of different words. To obtain the emotion embeddings $\mathbf{E}_e$ for a given input sequence, we concatenate the emotion vector corresponding to the emotion intensity of each word in the input sequence. Concretely, we split the VAD space into multiple subspaces by dividing the valence and the arousal dimensions,

[1] https://github.com/declare-lab/RECCON

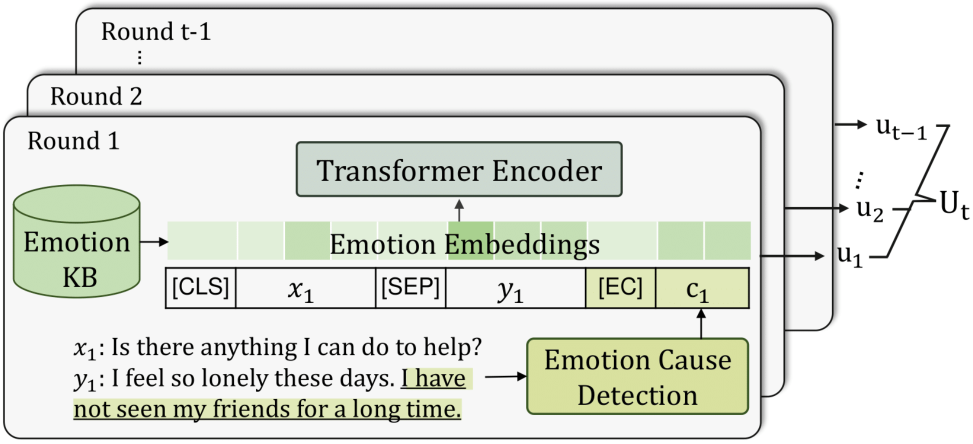

Figure 3.3: The architecture of the user state modeling module in MultiESC.

respectively, into n_V and n_A intervals of equal length. Each emotional subspace is represented as one emotion vector e_j. Note that the dominance dimension is not considered here as it is less relevant for capturing emotion intensity [236, 89]. To construct the emotion embeddings, we retrieve the VAD score of each input token from the NRC VAD lexicon to identify which emotional subspace it belongs to, and then we represent it as the corresponding emotion vector.

For those tokens without VAD annotation, we use a special emotion vector to represent them. Thus, the number of the emotion vectors n_{emo} is $n_V \cdot n_A + 1$.

Finally, the encoded hidden vector $\mathbf{u}_i$ corresponding to the `[CLS]` token is used to represent the user state at the i-th round. The user state embeddings $\mathbf{U}_t$ are the concatenation of all the user state embeddings before the t-th round, that is:

$$\mathbf{U}_t = [\mathbf{u}_1; \mathbf{u}_2; ...; \mathbf{u}_{t-1}]. \tag{3.1}$$

The user state embeddings $\mathbf{U}_t$ are then transmitted to the strategy planning and utterance decoder modules.

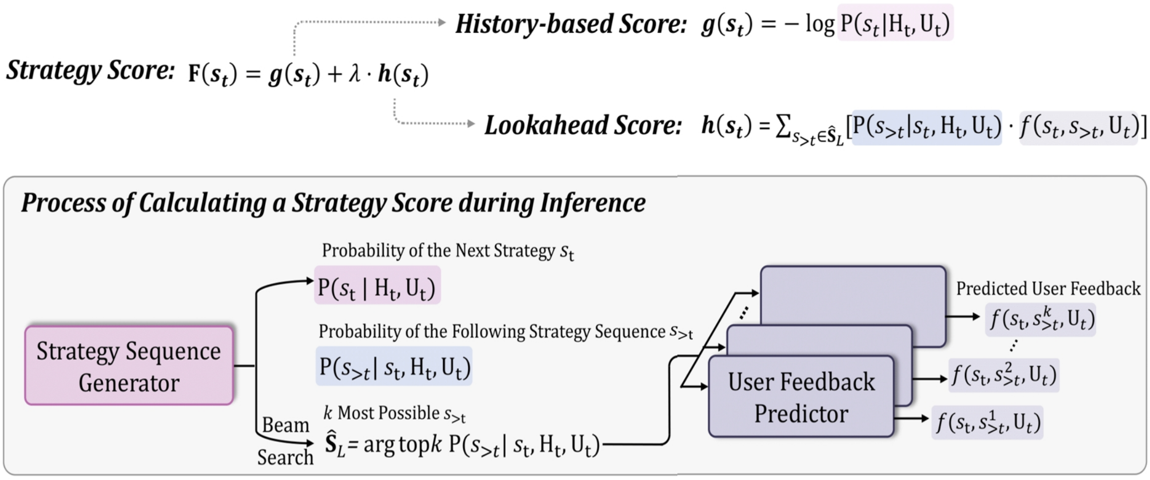

Figure 3.4: The process of calculating the strategy score, using a strategy sequence generator and a user feedback predictor. At each turn, our model selects the next strategy that maximizes the score of $F(s_t)$.

3.3.2 Strategy Planning with Lookahead Heuristics

We develop a strategy score function to evaluate whether to adopt a particular strategy (e.g., *question* or *self-disclosure*) by comprehensively considering the dialogue history and the potential user feedback. Formally, at the t-th round, MultiESC adopts the strategy $\hat{\mathbf{s}}_t$ that maximizes the score function:

$$\hat{\mathbf{s}}_t = \arg\max_{s_t \in \mathcal{S}} F(s_t), \tag{3.2}$$

where $F(\cdot)$ is the strategy score function.

In the following, we will first introduce the strategy score function and then explain how MultiESC calculates the strategy scores with two components: a strategy sequence generator and a user feedback predictor, as presented in Fig. 3.4. Finally, we will describe the architectures of the two components.

Strategy Score Function. Our method draws inspiration from the classical search algorithm, A* search [58], which conducts lookahead planning in a heuristic way. At each step, it searches the highest-scoring path by selecting an action that maximizes the sum of the score so far and a heuristic estimation of the future score. Similarly, we define our strategy score function as:

$$F(s_t) = g(s_t) + \lambda \cdot h(s_t), \tag{3.3}$$

where $g(s_t)$ is a *history-based score*; $h(s_t)$ is a *lookahead score* that heuristically estimates the future user feedback; λ is a hyper-parameter that balances the weights of the two terms.

The history-based score $g(s_t)$ computes the conditional probability distribution of the next strategy purely based on the dialogue history and the previous user states. Formally, it is defined as:

$$g(s_t) = -\log \Pr(s_t|\mathbf{H}_t, \mathbf{U}_t). \tag{3.4}$$

Previous research on dialogue strategy prediction generally followed this history-based scheme [243, 73, 37], though they may vary in their methods of obtaining the representations of $\mathbf{H}_t$ and $\mathbf{U}_t$. However, such practice overlooks the strategy's future effects and how much it could help in achieving the long-term dialogue goal.

In our work, we incorporate the lookahead score to alleviate this issue. The lookahead score $h(s_t)$ heuristically estimates *the mathematical expectation of the future user feedback score* after adopting the strategy s_t at the t-th round. Here, the *user feedback score* indicates how much the user's emotional distress is reduced, which has been manually annotated in the dataset (see Section 3.2). For the estimation of $h(s_t)$, we only consider how much the adopted strategies would affect the user's emotion intensity and neglect the effects of other factors, such as the specific content of the utterances (i.e., hypothesizing the supporter could always utter the most appropriate content after selecting a correct strategy).

Ideally, to select the strategy that could lead to the best final result, we want to estimate the user feedback score at the end of the conversation. This is formalized as:

$$\begin{aligned} h(s_t) &= \mathrm{E}[f(s_t, s_{>t}, \mathbf{U}_t)|s_t, \mathbf{H}_t, \mathbf{U}_t] \\ &= \sum_{s_{>t} \in \mathcal{S}^\omega} \Pr(s_{>t}|s_t, \mathbf{H}_t, \mathbf{U}_t) \cdot f(s_t, s_{>t}, \mathbf{U}_t). \end{aligned} \tag{3.5}$$

$\mathrm{E}(\cdot)$ represents the mathematical expectation; $s_{>t}$ is the future strategy sequence to be used after the t-th round till the end of the conversation; $\mathcal{S}^\omega$ is the set of all possible strategy sequences (i.e., $\mathcal{S}^\omega = \bigcup_{n=0}^{\infty} \mathcal{S}^n$, where $\mathcal{S}^n$ is the n-ary Cartesian power of the strategy set $\mathcal{S}$). The function $f(s_t, s_{>t}, \mathbf{U}_t)$ quantifies the user feedback score after successively applying the strategy s_t and the strategy sequence $s_{>t}$ to comfort a user whose previous states are $\mathbf{U}_t$.

Despite its theoretical soundness, directly computing Equation 3.5 presents two challenges: (1) the space of $\mathcal{S}^\omega$ is prohibitively large; (2) it is difficult to estimate the user feedback $f(\cdot)$ after too many turns (i.e. if the strategy sequence $s_{>t}$ is too long).

To mitigate these issues, we approximate Equation 3.5 as follows. First, we only look ahead for the limited L turns. We estimate the expectation of the user feedback score after L turns instead of at the end of the conversation. Nevertheless, the space of $\mathcal{S}_L$, growing exponentially over the sequence length, can still be too large. Thus, we further narrow the space of $\mathcal{S}^\omega$; we only consider the k most possible future strategy sequences.

Formally, Equation 3.5 is approximated as:

$$\begin{aligned} h(s_t) &= \sum_{s_{>t} \in \mathcal{S}'_L} \Pr(s_{>t}|s_t, \mathbf{H}_t, \mathbf{U}_t) \cdot f(s_t, s_{>t}, \mathbf{U}_t), \\ \mathcal{S}'_L &= \underset{s_{>t} \in \mathbf{S}_L}{\arg\operatorname{topk}} \Pr(s_{>t}|s_t, \mathbf{H}_t, \mathbf{U}_t), \end{aligned} \tag{3.6}$$

where $\mathcal{S}_L$ is the set of strategy sequences whose lengths are less than L.

Strategy Score Calculation in MultiESC. MultiESC calculates the strategy scores using two components: a *Strategy Sequence Generator (SSG)* and a *User Feedback Predictor (UFP)*.

The function of SSG is to sequentially predict the strategy sequence $s_{\geq t}$ conditioned on the historical context $\mathbf{H}_t$ and and the current user state $\mathbf{U}_t$, where $s_{\geq t}$ is the strategy sequence that will be used in the following L rounds ($s_{\geq t}$=[s_t; $s_{>t}$]). At the l-th timestep (l=1,2,...,L), the SSG outputs a probability distribution over possible dialogue strategies:

$$\Pr(s_{t+l}|s_{t:t+l}, \mathbf{H}_t, \mathbf{U}_t), \tag{3.7}$$

where l=1, 2, ..., L and $s_{t:t+l}$ denotes the partial strategy sequence generated prior to the l-th timestep.

The function of UFP is to estimate the user feedback score $f(s_{\geq t}, \mathbf{U}_t)$ that would result from applying the strategy sequence $s_{\geq t}$ to comfort the user, given the previous user states $\mathbf{U}_t$. This score reflects the system's effectiveness in improving user comfort.

Further details on the architectures of the SSG and UFP modules will be later illustrated in Section 3.3.2.

As illustrated in Fig. 3.4, the strategy score for a candidate strategy s_t is computed in two stages. First, we use SSG to derive the history-based score $g(s_t)$ from its predicted strategy distribution at the first timestep. Next, we use SSG to find the set of the k most possible future strategy sequences $\hat{\mathbf{S}}_L$ through beam search. For each strategy sequence $s_{>t}$ in $\hat{\mathbf{S}}_L$, we obtain its probability by:

$$\Pr(s_{>t}|s_t, \mathbf{H}_t, \mathbf{U}_t) = \prod_{l=2}^{L} \Pr(s_{t+l}|s_{t:t+l}, \mathbf{H}_t, \mathbf{U}_t). \tag{3.8}$$

We then leverage UFP to estimate the user feedback score after successively applying the strategies s_t and $s_{>t}$. By integrating the predicted probabilities of the strategy sequences with the estimated user feedback scores, we derive the lookahead score $h(s_t)$, as formalized in Equation 3.6.

Finally, combining the history-based score $g(s_t)$ and the lookahead score $h(s_t)$, the overall strategy score is obtained as defined in Equation 3.3.

Strategy Sequence Generator. The Strategy Sequence Generator (SSG) is built upon the Transformer decoder architecture. At each time step t, the model takes as input: the strategy sequence $s_{\geq t}$, the dialogue history representation $\mathbf{H}_t$, and the user state embeddings $\mathbf{H}_t$.

The output at each decoding timestep l (where l=0, 1, ..., $L-1$) is a probability distribution over possible strategies, given by:

$$\Pr(s_{t+l}|s_{t:t+l}, \mathbf{H}_t, \mathbf{U}_t). \tag{3.9}$$

The SSG extends the conventional Transformer decoder by employing a multi-source attention mechanism, enabling selective attention over the dialogue history $\mathbf{H}_t$ and the user state information $\mathbf{U}_t$. Specifically, the input strategy sequence $s_{\geq t}$ is first processed through a masked multi-head self-attention layer, followed by additive residual connections and layer normalization, yielding contextualized strategy representations:

$$P_t = \texttt{LayerNorm}(\texttt{Masked-MH-Att}(s_{\geq t}) + s_{\geq t}). \tag{3.10}$$

The model then applies cross-attention between $\mathbf{P}_t$ and the two auxiliary inputs ($\mathbf{H}_t$ and $\mathbf{U}_t$):

$$\hat{\mathbf{H}}_t = \texttt{MH-ATT}(L(\mathbf{P}_t), L(\mathbf{H}_t), L(\mathbf{H}_t)), \tag{3.11}$$

$$\hat{\mathbf{U}}_t = \texttt{MH-ATT}(L(\mathbf{P}_t), L(\mathbf{U}_t), L(\mathbf{U}_t)), \tag{3.12}$$

where $\texttt{MH-ATT}(\cdot)$ represents the multi-head self-attention mechanism and $L(\cdot)$ is a linear projection that reshapes embeddings into h attention heads.

The attended representations $\hat{\mathbf{H}}_t$ and $\hat{\mathbf{U}}_t$ are combined via a gated fusion layer:

$$\mu = \text{ReLU}(\mathbf{W}_\mu[\hat{\mathbf{H}}_t; \hat{\mathbf{U}}_t] + \mathbf{b}_\mu), \tag{3.13}$$

$$\hat{\mathbf{P}}_t = \mu \cdot \hat{\mathbf{H}}_t + (1-\mu) \cdot \hat{\mathbf{U}}_t, \tag{3.14}$$

where $\mathbf{W}_\mu \in \mathbb{R}^{d\times 2d}$ and $\mathbf{b}_\mu \in \mathbb{R}^d$ are learnable parameters and μ acts as a dynamic interpolation weight.

Next, $\hat{\mathbf{P}}_t$ is further processed by a position-wise feed-forward network (FFN) with residual connections:

$$\widetilde{\mathbf{P}}_t = \texttt{LayerNorm}(\texttt{FFN}(\hat{\mathbf{P}}_t) + \hat{\mathbf{P}}_t). \tag{3.15}$$

Finally, the strategy distribution at the l-th timestep is predicted as (l=0,1,..,L-1):

$$\Pr(s_{t+l}|s_{t:t+l}, \mathbf{H}_t, \mathbf{U}_t) = \text{softmax}(\mathbf{W}_s\widetilde{\mathbf{P}}_t + \mathbf{b}_s), \tag{3.16}$$

where $\mathbf{W}_s$ and $\mathbf{b}_s$ are trainable parameters.

The model is trained to minimize the negative log-likelihood (NLL) of the ground-truth strategy s^*_{t+l}:

$$\mathcal{L}_s = -\sum_{l=0}^{L-1} \log \Pr(s^*_{l+1}|s_{t:t+l}, H_t, U_t). \tag{3.17}$$

User Feedback Predictor. The User Feedback Predictor (UFP) estimates the user feedback score $f(s_{\geq t}, \mathbf{U}_t)$ by encoding the strategy sequence $s_{\geq t}$ and user state information U_t through a hierarchical neural architecture. Below, we formalize the key components of the model.

The input strategy sequence $s_{t:t+L}$ is encoded using a Transformer-based encoder, denoted as TRS_{UFP}. We define a trainable strategy embedding matrix $\mathbf{E}_s \in \mathbb{R}^{|\mathcal{S}|\times d_{\text{emb}}}$, where $\mathcal{S}$ is the set of all possible strategies and d_{emb} is the embedding dimension. Each strategy UFP first encodes the strategy sequence $s_{t:t+L}$ with a Transformer encoder. Specifically, we leverage a trainable strategy matrix $\mathbf{E}_s \in \mathbb{R}^{|\mathcal{S}|\times d_{\text{emb}}}$ to represent different types of strategies. Each strategy s_i in $s_{\geq t}$ is mapped to its corresponding embedding vector $\mathbf{e}_i \in \mathbb{R}^{d_{\text{emb}}}$. These embeddings are concatenated with the special token `[CLS]` to form the input matrix:

$$\mathbf{X} = \text{Emb}(\texttt{[CLS]}; s_{\geq t}) = [\mathbf{e}_{\texttt{[CLS]}}, \mathbf{e}_{s_t}, \mathbf{e}_{s_{t+1}}, \ldots, \mathbf{e}_{s_{t+L}}]. \tag{3.18}$$

The input $\mathbf{X}$ is processed by TRS_{UFP}, yielding contextualized representations:

$$\mathbf{B} = \text{TRS}_{\text{UFP}}(\mathbf{X}) \in \mathbb{R}^{(L+2)\times d_{\text{emb}}} \tag{3.19}$$

where Emb$(\cdot)$ represents the operation of the embedding layer that maps the strategies in $s_{\geq t}$ to their corresponding vectors in $\mathbf{E}_s$ and $\mathbf{B}$. Suppose the encoded hidden state corresponding to the `[CLS]` token is $\mathbf{q}_s$, which will be used later to interact with the user state information. Next, we pass the user state embeddings through a Long-Short Term Memory (LSTM) network [20]:

$$\hat{\mathbf{U}}_t = \text{LSTM}(\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_{t-1}), \tag{3.20}$$

where $\hat{\mathbf{U}}_t \in \mathbb{R}^{(L+2)\times d_{\text{emb}}}$ and d_h is the hidden dimension of the LSTM.

We then use $\mathbf{q_s}$ to attend to the hidden states $\hat{\mathbf{U}}_t$=[$\hat{\mathbf{u}}_1$, $\hat{\mathbf{u}}_2$, ..., $\hat{\mathbf{u}}_{t-1}$] through an attention layer:

$$\widetilde{\mathbf{u}}_f = \sum_{i=1}^{t-1} a_i \hat{\mathbf{u}}_i, \tag{3.21}$$

$$a_i = \frac{\exp(\hat{\mathbf{u}}_i^\top \mathbf{W}_a \mathbf{q}_s)}{\sum_{j=1}^{t-1} \exp(\hat{\mathbf{u}}_j^\top \mathbf{W}_a \mathbf{q}_s)}, \tag{3.22}$$

where $\mathbf{W}_a$ is a trainable matrix. The final feedback score is predicted using a feed-forward layer with a single output neuron:

$$f(s_{\geq t}, \mathbf{U}_t) = \mathbf{w}^\top \widetilde{\mathbf{u}}_f + b, \tag{3.23}$$

where $\mathbf{w} \in \mathbb{R}^{d_h}$ and $b \in \mathbb{R}$ are learnable parameters.

The model is trained end-to-end using the Mean Squared Error (MSE) loss between predicted and ground-truth feedback scores (annotated in the ESConv dataset):

$$\mathcal{L}_f = \frac{1}{N} \sum_{i=1}^{N} \left(f(s_{\geq t}^{(i)}, \mathbf{U}_t^{(i)}) - y^{(i)} \right)^2, \tag{3.24}$$

where $y^{(i)}$ is the true feedback score for the i-th training instance.

3.3.3 Utterance Decoder

Given the user state embeddings $\mathbf{U}_t \in \mathbb{R}^d$, the dialogue history embeddings $\mathbf{H}_t \in \mathbb{R}^{n\times d}$, and the selected strategy $\hat{\mathbf{s}}_t \in \mathbb{R}^d$, the utterance decoder generates the next utterance x_t autoregressively. The decoder is built upon a Transformer-based architecture, structurally identical to the SSG described in Section 3.3.2, but differs in its input formulation.

To guide dialogue generation with the selected strategy $\hat{\mathbf{s}}_t$, we prepend the strategy embedding of $\hat{\mathbf{s}}_t$ before the embeddings of the utterance sequence as the input of the utterance decoder. The negative likelihood of the ground-truth token in the target utterance is used as the generation loss $\mathcal{L}_g$.

For training of the strategy sequence generator, we use the negative log-likelihood of the ground-truth strategy s^*_{t+l} as its loss function $\mathcal{L}_s$. For the utterance decoder, the negative likelihood of the ground-truth token in the target utterance is used as the generation loss $\mathcal{L}_g$.

The strategy sequence generator and the utterance decoder are trained jointly, with the total loss as

$$\mathcal{L} = \mathcal{L}_s + \lambda \mathcal{L}_g, \tag{3.25}$$

where λ is a scaling hyperparameter.

The feedback predictor is trained independently using **Mean Squared Error (MSE)** loss:

$$\mathcal{L}_f = \frac{1}{N} \sum_{i=1}^{N} \left(y_i - f(\mathbf{S}_i)\right)^2, \tag{3.26}$$

where $y_i \in [1, 5]$ is the user feedback score for strategy sequence $\mathbf{S}_i$, and $f(\cdot)$ is the predictor model. To address data imbalance (scores in ESCONV are skewed toward 2–5), we augment the training set with 5,000 synthetic low-score samples ($\mathbf{S}_{\text{synth}} \sim \mathcal{U}$, labeled as $y = 1$) to improve robustness at the lower score range.

3.4 Experiments

3.4.1 Baselines

Since the task of developing data-driven multi-turn ESC systems is relatively new, our baselines mainly include several systems. The following baselines are implemented to compare with MultiESC: **MoEL** [99], **MIME** [116], and **EmpDG** [88]; and four state-of-the-art methods on the ESConv dataset: **DialoGPT-Joint**, **BlenderBot-Joint** [105], **MISC** [173], and **GLHG** [130]. More specific descriptions of these baselines are as follows:

- **MoEL** [99] adopts several decoders focusing on different types of emotional utterances, whose outputs are combined to generate the final utterances.
- **MIME** [116] follows the architecture of MoEL and adds extra mechanisms to combine the results from different decoders.
- **EmpDG** [88] learns how to generate responses consistent with the user's emotion via an adversarial learning framework.
- **DialoGPT-Joint** [105] is developed on the backbone of DialoGPT [221]. It prepends a special token, denoting the predicted support strategy, before the generated utterance to generate content conditioned on a predicted strategy.
- **BlenderBot-Joint** [105] is developed on the backbone of BlenderBot [144] and perform in the similar way as DialoGPT-Joint.
- **MISC** [173] enhances context encoding with commonsense knowledge and uses the predicted strategy distribution to guide the emotional support dialogue generation. It predicts the strategy distribution using a vanilla Transformer encoder.

- **GLHG** [130] adopts a graph neural network to model the relationships between the user's emotion causes, intentions, and the dialogue history for emotional support dialogue generation.

3.4.2 Implementation Details

We follow the original division of ESConv for training, validation, and testing. We randomly split the ESConv dataset[2] into the training, validation, and test sets with a ratio of 8:1:1. We initialize the parameters of the dialogue encoder and the utterance decoder of MultiESC with the BART-small [83] model from the HuggingFace library [193]. The maximum length of the input sequence for the dialogue encoder is N=512. There are n_{emo}=65 types of emotion vectors, with n_{V}=n_{A}=8. In the strategy planning module, we set λ=0.7 and L=2. The beam size k is set to be 6 when searching the set of the most possible strategy sequences $\hat{\mathbf{S}}_L$, where k is equal to the beam size. The dimensions of all the hidden embeddings are d_{emb}=768.

Since the codes for MISC and GLHG were not publicly available, we relied on the results reported in their original papers. For the other baselines, we conducted experiments using their released code. Our model contains 145.6M parameters, which is comparable in scale to the baselines—BlenderBot-Joint (90M), DialoGPT-Joint (117M), and GLHG (92M). We utilized the small versions of DialoGPT and BlenderBot for our experiments. The optimizer employed was AdamW [110], with an initial learning rate of 5×10^{-5} and adaptive decay during training. A batch size of 32 was used, and each model was trained for up to 10 epochs, with the best-performing checkpoint on the validation set (based on perplexity) selected for evaluation.

[2] https://github.com/thu-coai/Emotional-Support-Conversation

Model	PPL↓	B-2↑	B-3↑	B-4↑	R-L↑	MET↑	CID↑
MoEL [99]	264.11	6.47	2.91	1.51	15.95	7.96	10.95
MIME [116]	69.28	5.56	2.64	1.50	16.12	6.43	10.66
EmpDG [88]	115.34	6.46	3.02	1.52	15.89	6.93	10.73
DialoGPT-Joint [105]	15.71	5.59	2.03	1.18	16.93	7.55	11.86
BlenderBot-Joint [105]	16.79	6.91	2.81	1.66	17.94	7.54	18.04
MISC [173]	16.16	7.31	-	2.20	17.91	-	-
GLHG [130]	15.67	7.57	3.74	2.13	16.37	-	-
MultiESC	**15.41**	**9.18**	**4.99**	**3.09**	**20.41**	**8.84**	**29.98**

Table 3.1: Automatic evaluation results on the generation quality of MultiESC and the baselines.

3.4.3 Automatic Evaluation of Generation Quality

These include perplexity (PPL), which quantifies the model's confidence in predicting the test data by measuring the inverse probability of the generated sequences, thereby indicating fluency and overall language modeling performance; BLEU scores at n-gram levels 2, 3, and 4 (**B-2/3/4**), which evaluate n-gram precision by comparing the overlap of generated phrases with reference texts, thus assessing the accuracy of local phrase generation; ROUGE-L (**R-L**) [97], which focuses on recall by measuring the longest common subsequence between the generated and reference outputs, capturing sentence-level structural similarity; METEOR (**MET**) [79], which extends evaluation beyond exact matches by incorporating stem and synonym matches, thereby providing a more semantically informed measurement; and CIDEr (**CID**) [175], which calculates the similarity of TF-IDF weighted n-grams, effectively reducing the influence of commonly occurring words and emphasizing rare but informative terms to better reflect human judgment of relevance. We first compare the generation results of MultiESC with those of the baseline models using these automatic metrics and will further assess them through human evaluation in Section 3.4.5

Table 3.1 presents a comprehensive comparison of various models based on automatic evaluation metrics related to generation quality. The proposed model, MultiESC, demonstrates superior performance across all evaluated metrics, indicating its effectiveness in generating high-quality, contextually appropriate responses. Notably, MultiESC achieves the lowest perplexity score (15.41), suggesting it produces more fluent and coherent outputs compared to baseline models. The BLEU scores (9.18/4.99/3.09) and ROUGE-L (20.41) show substantial improvements, underscoring MultiESC's ability to generate n-gram overlaps and syntactically coherent sequences that closely match the reference responses. Particularly striking is MultiESC's performance on the CIDEr metric (29.98), which is specifically designed to weight n-grams by their TF-IDF importance. This high score indicates that MultiESC doesn't merely replicate generic phrases but effectively incorporates distinctive and contextually relevant information tailored to users' specific scenarios. This strength highlights the model's capacity for nuanced emotional support conversation (ESC), adapting responses to the user's unique context rather than relying on formulaic empathetic expressions.

In contrast, the three empathetic generators (i.e., MoEL, MIME, and EmpDG) show considerably higher perplexity and notably lower CIDEr scores. Their BLEU and ROUGE-L scores are also inferior relative to the stronger-performing models. A qualitative analysis suggests these models tend to generate generic, repetitive empathetic phrases such as "I'm sorry to hear that" or "I can understand that." While these expressions may convey surface-level empathy, they fail to deliver the substance and specificity required for effective emotional support. This likely reflects their design focus on emotion recognition and expression, at the expense of maintaining content diversity and informativeness essential for ESC tasks.

Baseline dialogue models trained jointly on general conversational data, such as DialoGPT-Joint and BlenderBot-Joint, achieve moderate perplexity and BLEU scores.

Model	**PPL↓**	**B-2↑**	**B-3↑**	**B-4↑**	**R-L↑**	**MET↑**	**CID↑**
MultiESC	**15.41**	**9.18**	**4.99**	**3.09**	**20.41**	**8.84**	**29.98**
MultiESC *w/o* emotion	18.43	7.68	4.05	2.41	20.15	7.89	24.33
MultiESC *w/o* cause	15.68	8.76	4.64	2.77	19.82	8.60	26.73
MultiESC *w/o* strategy	15.60	8.24	4.42	2.70	20.35	8.25	27.77
MultiESC *w/o* lookahead	15.71	9.15	4.81	3.02	20.39	8.43	29.81

Table 3.2: Ablation Studies of MultiESC modules.

Although they produce more varied responses than the empathetic generators, their CIDEr scores and overall contextual relevance remain below that of MultiESC, further demonstrating the latter's advantage in generating targeted, empathetic, and informative responses.

3.4.4 Ablation Study.

We conduct the ablation study to analyze the effects of different components on the downstream generation in MultiESC. The results are shown in Table 3.2 To analyze the effects of different components on the downstream generation, we compare MultiESC with its following variants:

- ***w/o* emotion** does not incorporate the emotion embedding layer in the user state modeling module;
- ***w/o* cause** does not incorporate emotion cause extraction for user state modeling;
- ***w/o* strategy** directly generates utterances without first predicting the used strategy;

- ***w/o* lookahead** conducts strategy planning without the lookahead heuristics to estimate the future user feedback scores.

The results demonstrated in Table 3.2 clearly indicate that each component contributes meaningfully to the overall model performance. Notably, removing the emotion embedding layer ("*w/o* emotion") results in the most substantial degradation across all metrics, with perplexity increasing from 15.41 to 18.43 and BLEU-4 dropping from 3.09 to 2.41. This phenomenon underscores the critical role of emotion understanding in generating contextually appropriate and coherent empathetic responses. Ablation of the cause and strategy components also leads to noticeable declines, reflecting their significance in capturing causal relationships and guiding response generation effectively. Conversely, the removal of the lookahead module ("*w/o* lookahead") produces only marginal performance drops. While lookahead planning has limited impact on static automatic metrics, which is because responses generated under identical strategy selections remain similar, we want emphasize that it still plays a more pronounced role during dynamic human interactions by influencing conversational trajectories through strategic decision-making across dialogue turns, which will be further discussed in Section 3.4.5.

3.4.5 Human Interactive Evaluation

We recruited four graduate students with backgrounds in linguistics or psychology to serve as annotators for human interactive evaluation. All annotators were fully informed about the research objectives and compensated with appropriate wages. A total of 128 dialogues were randomly sampled from the test set of ESConv. Annotators received comprehensive training on our evaluation guidelines, and their understanding was validated through several test cases. During the evaluation, annotators acted as support seekers by familiarizing themselves with the scenarios in each dialogue sample and engaging with the models to simulate the process of seeking emotional

MultiESC vs.		Flu.	Emp.	Ide.	Sug.	Overall
MoEL	**Win**	**64.1**‡	**53.1**‡	**69.5**‡	**71.9**‡	**65.6**‡
	Lose	18.0	34.4	22.7	14.8	20.3
	Tie	18.0	12.5	7.9	13.3	14.1
BlenderBot	**Win**	**35.2**	**44.5**	**48.4**‡	**60.9**‡	**58.6**‡
	Lose	42.9	43.8	32.8	23.4	31.3
	Tie	21.9	11.7	18.8	15.6	10.2
w/o strategy	**Win**	**38.3**	**43.8**†	**56.3**‡	**52.3**†	**55.5**‡
	Lose	41.4	29.7	32.8	36.7	30.5
	Tie	20.3	26.5	10.9	10.9	14.0
w/o lookahead	**Win**	**41.4**	**35.9**	**46.9**‡	**44.5**†	**46.1**†
	Lose	37.5	39.1	28.9	30.5	32.0
	Tie	21.1	25.0	14.2	25.0	21.9

Table 3.3: Human interactive evaluation results (%). The rows of "Win/Lose" indicate the proportion of cases where MultiESC wins/loses in the comparison. "Flu", "Emp", "Ide", and "Sug." refer to the evaluation dimensions of fluency, empathy, identificant, and suggestion, respectively. †/‡ denote p-value $< 0.1/0.05$ (statistical significance test).

support. The compared models were presented in random order to minimize exposure bias. Given MultiESC and a compared model, the annotators are asked to choose which one performs better (or select *tie*).

Similar to [105], we adopt the following evaluation metrics:

- **Fluency**: which model generates more fluent and understandable responses;
- **Empathy**: which model has more appropriate emotion reactions, such as warmth, compassion, and concern;

- **Identification**: which model explores the user's situation more effectively to identify the problem;
- **Suggestion**: which model offers more helpful suggestions;
- **Overall**: which model provides more effective emotional support overall.

The results are shown in Table 3.3. Compared with the two baseline models (MoEL and BlenderBot-Joint), MultiESC significantly outperforms them in the overall supporting effects, which again shows that only being empathetic is insufficient for ESC. It also outperforms BlenderBot-Joint in the overall supporting effects, though relatively inferior in terms of fluency, probably because the backbone of BlenderBot-Joint is extensively pre-trained on large-scale dialogue corpora [144]. Compared with "*w/o* strategy", MultiESC is able to show more empathy, more clearly inquire about the user's situation, and provide more specific suggestions, demonstrating the importance of explicit strategy planning in ESC. Comparing MultiESC with "*w/o* lookahead", we can see that the incorporation of lookahead heuristics brings significant improvement in the dimensions of *identification* and *suggestion*. Their differences in language fluency and empathy are not evident.

3.4.6 Analysis of Strategy Planning

We evaluate the strategy planning module individually, using the following metrics: **Accuracy**, the proportion of prediction results that are the same as the ground-truth labels; **Weighted F1**, the weighted average of F1 scores in different classes while considering the class imbalance; **Feedback**, the next user feedback score that would be given after the predicted strategy is adopted, simulated with an user feedback predictor as illustrated in Section 3.3.2.

Model	Accuracy	Weighted-F1	Feedback
DialoGPT-Joint	26.03	23.86	2.87
BlenderBot-Joint	29.92	29.56	3.05
MISC	31.61	-	-
MultiESC	**42.01**	**34.01**	3.85

Table 3.4: The strategy planning performance of MultiESC and the baseline methods.

Comparison with Baselines. We compare MultiESC with the three baselines capable of strategy planning (i.e., DialoGPT-Joint, BlenderBot-Joint, and MISC). The results are shown in Table 3.4. We can see that MultiESC performs the best in all the metrics with an absolute improvement of 10.4% and 4.45% in accuracy and weighted F1, respectively. Among the models, MultiESC demonstrates the superior performance across all reported metrics. Specifically, MultiESC achieves an accuracy of 42.01%, representing an absolute increase of 10.4 percentage points over the next best model, MISC, which attains 31.61%. In terms of the weighted F1 score, MultiESC reaches 34.01%, showing an improvement of approximately 4.45 percentage points compared to BlenderBot-Joint's 29.56%, the closest competitor with available weighted F1 data. Furthermore, MultiESC also leads in the Feedback metric with a value of 3.85, surpassing the other models by at least 0.8 points, indicating qualitatively better performance as perceived through feedback measures.

As shown in Fig. 3.5, MultiESC also surpasses the baselines in all the top-n accuracy (n=1, 2, .., 5), but the performance gap generally decreases with the increase of n, especially in the feedback metric.

Analysis of MultiESC Variants. We analyze the following variants of our strategy planning method: (1) **MultiESC**$_{k=?}$: the model with different beam sizes when searching the set of k most possible strategy sequences $\hat{\mathbf{S}}_L$; (2) ***w/o* lookahead**: the

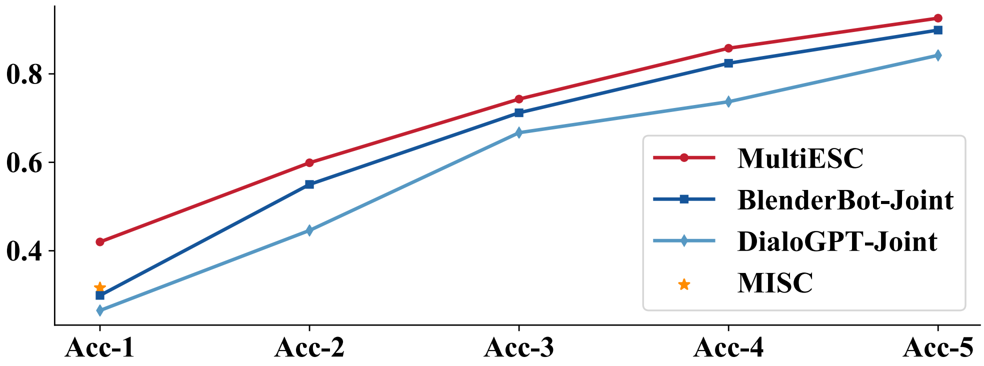

Figure 3.5: The top-n strategy prediction accuracy of MultiESC and the baseline methods.

model without the lookahead heuristics.

As shown in Table 3.5, the strategy planning performance steadily improves with the increase of the beam search size when $k \leq 6$. This trend indicates that incorporating a larger set of plausible future strategy sequences enables the model to generate more accurate and emotionally supportive responses, likely due to a better anticipation of future user feedback. Nevertheless, further increasing k to consider more strategy sequences of low probabilities does not continue improving the performance, apparently when $k>6$. This saturation suggests diminishing returns from considering additional lower-probability strategy sequences, potentially because the lookahead heuristic already provides sufficiently precise estimates of user feedback at moderate beam sizes.

Our full model also significantly outperforms "*w/o* lookahead" in all the metrics, especially regarding the feedback score. This decline highlights the critical role of the lookahead heuristic in enabling the model to anticipate the impact of strategy choices on future user feedback, thereby enhancing the quality of conversational planning and emotional support.

Model	**Accuracy**	**Weighted-F1**	**Feedback**
MultiESC$_{k=1}$	38.72	30.12	3.59
MultiESC$_{k=2}$	39.53	30.61	3.62
MultiESC$_{k=3}$	41.33	32.83	3.75
MultiESC$_{k=4}$	41.61	33.30	3.67
MultiESC$_{k=5}$	41.78	33.64	**3.93**
MultiESC$_{k=7}$	41.79	33.92	3.88
MultiESC$_{k=8}$	41.79	33.97	3.92
MultiESC	**42.01**	**34.01**	3.85
w/o lookahead	38.76	30.21	3.36

Table 3.5: The strategy planning performance of different variants of MultiESC. Note that the beam size of MultiESC is set to be 6 (see Section 3.4.2).

3.4.7 Case Study

Table 3.6 illustrates a dialogue example whereby a user expresses distress about friends making hurtful comments regarding their appearance, and the subsequent responses generated by various conversational models. The baseline models, MoEL and BlenderBot-Joint, offer generic or minimally actionable advice ("Have you been able to find that it would be better..." and "You should tell them to calm down"), lacking depth or empathy. In contrast, MultiESC and its variants demonstrate superior contextual awareness and emotional alignment. The "*w/o* strategy" variant, while acknowledging the user's feelings ("I can understand that..."), remains vague and less engaging. Both MultiESC and "*w/o* lookahead" provide higher-quality responses, with the latter employing self-disclosure ("I have had to deal with a lot of bullies...") to foster connection. However, MultiESC, leveraging its lookahead heuristic, excels by proactively probing the user's situation ("That's horrible! Have you tried talking to them about it?") rather than offering immediate reassurance. This approach mirrors

Bot: What can I help you with?
User: Lately, my friends started to make jokes about my face being ugly. They even called my face as flat as a pan! What should i do? I see them everyday!
MoEL: I'm sorry to hear that. Have you been able to find that it would be better to let yourself know how you feel about it.
BlenderBot-Joint: You should tell them to calm down.
MultiESC: (*Question*) That's horrible! Have you tried talking to them about it?
MultiESC *w/o* lookahead: (*Self-disclosure*) I can understand how that would make you feel. I have had to deal with a lot of bullies and I know how hard it can be.
MultiESC *w/o* strategy: I can understand that. I know why that would be upsetting.
Ground-Truth: (*Affirmation and Reassurance*) I'm sure you are very handsome! They could just be jealous.

Table 3.6: Case study of MultiESC and baselines. The upper part is an example of the dialogue history in the test set of ESConv. The lower part shows the responses from different models.

established emotional support protocols, which prioritize understanding before consolation, thereby demonstrating MultiESC's nuanced ability to guide conversations effectively. The ground-truth response ("I'm sure you are very handsome!") offers affirmation but skips exploration, highlighting MultiESC's strategic advantage in balancing inquiry with empathy.

3.5 Chapter Summary

In this chapter, we explored the task of developing multi-turn Emotional Support Conversation (ESC) systems, with a focus on how to strategically plan the conversation procedure to comfort users in emotional distress. To this end, we proposed a novel ESC system, MultiESC, that conducts strategy planning with lookahead heuristics to estimate the long-term effect of the adopted strategy on the user. Moreover, we also proposed some effective mechanisms to dynamically model the user's state in multi-turn ESCs. The empirical results showed that MultiESC achieves significant improvement compared with a set of strong baselines in both generation quality and strategy planning.

Chapter 4

Dialogue Progression Analysis

In this chapter, we focus on how to achieve a complex dialogue goal through progression analysis. In recent years, there has been a growing interest in exploring dialogues with more complex goals, such as negotiation, persuasion, and emotional support, which go beyond traditional service-focused dialogue systems. Apart from the requirement for much more sophisticated strategic reasoning and communication skills, a significant challenge of these tasks lies in the difficulty of objectively measuring the achievement of their goals in a quantifiable way, making it difficult for existing research to directly optimize the dialogue procedure towards them. In our work, we emphasize the multifaceted nature of complex dialogue goals and argue that it is more feasible to accomplish them by comprehensively considering and jointly promoting their different aspects. To this end, we propose a novel dialogue framework, Cooper, which coordinates multiple specialized agents, each dedicated to a specific dialogue goal aspect separately, to approach the complex objective. Through this divide-and-conquer manner, we make complex dialogue goals more approachable and elicit greater intelligence via the collaboration of individual agents.

4.1 Introduction

The use of human language is intentional and purposeful [2, 56]. In daily communication, we use language deliberately to achieve various goals, ranging from simple inquiries about a product's pricing to complex objectives like resolving conflicts. Developing goal-oriented dialogue systems has also been a prominent research topic.

In the past few years, there has been growing interest in dialogue generation tasks with complex objectives, such as negotiation [84, 60, 242], persuasion [183, 93, 150], and emotional support [103, 130, 204, 228]. Previous methods in these tasks can be mainly grouped into three categories: dialogue strategy learning [242, 73], user modeling [208, 157, 171], and fusing external knowledge [173, 16, 31]. Among these works, only a very few have an explicit consideration of the dialogue goal and how each generated utterance contributes to achieving the final objective. For example, [21] predicted the support strategy in ESC by estimating how much the user's emotion would be improved with an A*-like algorithm. [240] optimized the ESC process through reinforcement learning, using the extent to which the user's positive emotion is elicited as a reward. [151] conducted persuasive dialogue generation by measuring the distance of the current dialogue state relative to the desired outcome. However, it is challenging to measure the achievement of these complex dialogue goals objectively in a quantifiable way. For example, assessing how much the user's positive emotion is elicited simply based on the dialogue is extremely difficult in ESC. Directly optimizing towards a complex dialogue goal can be exceptionally hard, even for humans. In real scenarios, the guidelines for these challenging dialogue tasks usually recommend breaking down the complex goals into multiple aspects and jointly promoting them to work towards the broad objective [131, 40, 61].

More recently, several works have applied LLMs to complex goal-oriented dialogues by directly prompting the LLM to generate utterances [227, 29] or further improving the performance via iterative revision [44]. Current LLMs exhibit remarkable improvement compared to the previous methods on these tasks, but it is also found that they tend to lack a larger picture of the overall dialogue progression and fail to achieve the dialogue objective strategically through multi-turn interactions [29]. For example, on the task of ESC, they often continuously offer coping suggestions and overlook the critical process of exploring the user's situation and expressing empathy [227].

Compared to traditional service-focused goal-oriented dialogue systems [143, 9, 189, 106], these tasks require much more sophisticated strategic reasoning and communication skills. Recent studies show that even state-of-the-art Large Language Models (LLMs) struggle with these tasks, where they exhibit weak awareness of the overall dialogue progression and fail to accomplish a complex dialogue goal through multi-turn interactions strategically [227]. Moreover, another major challenge lies in the difficulty of objectively measuring the achievement of such complex dialogue goals in a quantifiable and reliable way. Consequently, most existing research stays overly focused on how to fit the ground-truth data, without explicit consideration of how each utterance could contribute to the final objective [242, 73, 17]. In the few works that attempt to model these dialogue goals explicitly, it remains highly challenging to optimize the dialogue procedure towards them directly due to their inherent intangibility [21, 151, 240].

In this work, we highlight the multifaceted nature of complex dialogue goals, which typically encompass multiple interdependent aspects that must be collectively promoted to approach the final objective. For instance, psychological guidelines suggest that Emotional Support Conversations (ESC) should include three key aspects: *exploration* (identify the support-seeker's problem), *comforting* (comfort the seeker's emotion through expressing empathy), and *action* (help the seeker solve the problem) [61, 105]. These aspects are interdependent. For example, exploring the seeker's situation lays the foundation for conveying appropriate empathy, while comforting the user to be in a better emotional state makes them more willing to share details about their experiences and feelings. Note that some works may refer to the "aspects" here as "stages", but they also emphasize that these "stages" are closely interwoven in practice rather than sequential [105]. Given that, we choose to regard them as "aspects" uniformly in our work to avoid misunderstanding about their sequential nature.

Compared with directly optimizing towards the complex dialogue goal, it is more feasible to accomplish it by comprehensively considering and jointly promoting its different aspects. Nonetheless, due to the interdependence among different aspects, the interlocutor still needs to address the challenge of how to strategically coordinate their priority during the conversation. To achieve this, they must dynamically track the states of all the aspects and analyze their progression, that is, how much progress has been achieved so far and where the state of each aspect is heading (i.e., its estimated target state at the end of the conversation). As in ESC, a seasoned supporter would continuously record information about the seeker's situation and keep estimating the underlying root problem for further exploration. They would also monitor the progression of the *comforting* and *action* aspects simultaneously. Through comprehensive analysis, the supporter could determine which aspect to prioritize at each point of the conversation.

Based on the above insight, we propose a novel dialogue framework, Cooper, which functions as a cooperation network of multiple agents. It coordinates multiple specialized agents, each dedicated to a specific aspect separately, to approach a complex dialogue goal. Specifically, each agent is designed to focus exclusively on the relevant part of the dialogue context related to its assigned aspect. By tracking the current state of its assigned aspect, each agent analyzes the progression of this aspect and suggests several topic candidates for the next utterance that can further promote the aspect (e.g., the agent responsible for the *exploration* aspect in ESC will suggest questions to ask the seeker). Then, we coordinate the specialized agents by ranking all the topic candidates with consideration of the overall dialogue progression. Finally, the top-ranked topic candidates are used to guide the generation of the next utterance.

Through this divide-and-conquer manner, we make the complex dialogue goal more approachable and elicit greater intelligence via the collaboration of individual agents. Experiments on ESC and persuasion dialogues demonstrate the superiority of COOPER over a set of competitive LLM-based methods and previous state-of-the-art.

In summary, our contributions are as follows:

- To the best of our knowledge, this is the first work that explores how to achieve a complex dialogue goal by coordinating the joint promotion of its different aspects.

- We propose COOPER, an innovative framework that coordinates multiple specialized agents to collaboratively work towards a complex dialogue goal.

- Extensive experiments demonstrate the effectiveness of our approach and also reveal the limitations of current LLMs in handling complex dialogue goals.

4.2 Preliminaries

4.2.1 Problem Formulation

We consider the problem of how to achieve a complex dialogue goal that encompasses multiple aspects, denoted as $\{\mathcal{T}_1, \mathcal{T}_2, ..., \mathcal{T}_{n_T}\}$, where n_T is the number of aspects.[1] Given the dialogue history $\mathcal{H}^t$ at the t-th dialogue round, the system generates the next utterance $\mathcal{U}^t$, which promotes a varying number of aspects dialogue goal, ranging from one to multiple.

[1]We suppose that the dialogue is always initiated by the system.

4.2.2 ESC Framework

Following the ESC framework defined by [105], our implementation considers the following aspects for effective emotional support:

- *Exploration*: identify the support-seeker's problems that cause their distress;
- *Comforting*: comfort the seeker's emotion by expressing empathy and understanding;
- *Action*: help the seeker conceive actionable plans to resolve the problems.

The supporter needs to adaptively determine which aspect(s) they are trying to promote at each point in the helping process.

4.2.3 Persuasion Dialogues

We use the P4G dataset [183] as one of the benchmarks for experiments. It is a persuasion dialogue dataset, where one interlocutor attempts to persuade the other to make donations to charities. Referring to the elaboration likelihood model of persuasion proposed by [131], we consider the following aspects within the broad goal of persuasion in our implementation:

- *Attention*: capture the persuadee's attention and elicit their motivation to discuss the related topic and gather information about the persuadee to build rapport and customize the persuasive message;
- *Appeal*: present persuasive arguments via different strategies and encourage the persuadee to think deeply about the arguments (e.g. offering evidence-based reasons, elicit empathy, establish credibility) and change the persuadee's attitudes towards charity donation via different strategies (e.g. present reasoned arguments, elicit empathy, establish credibility);

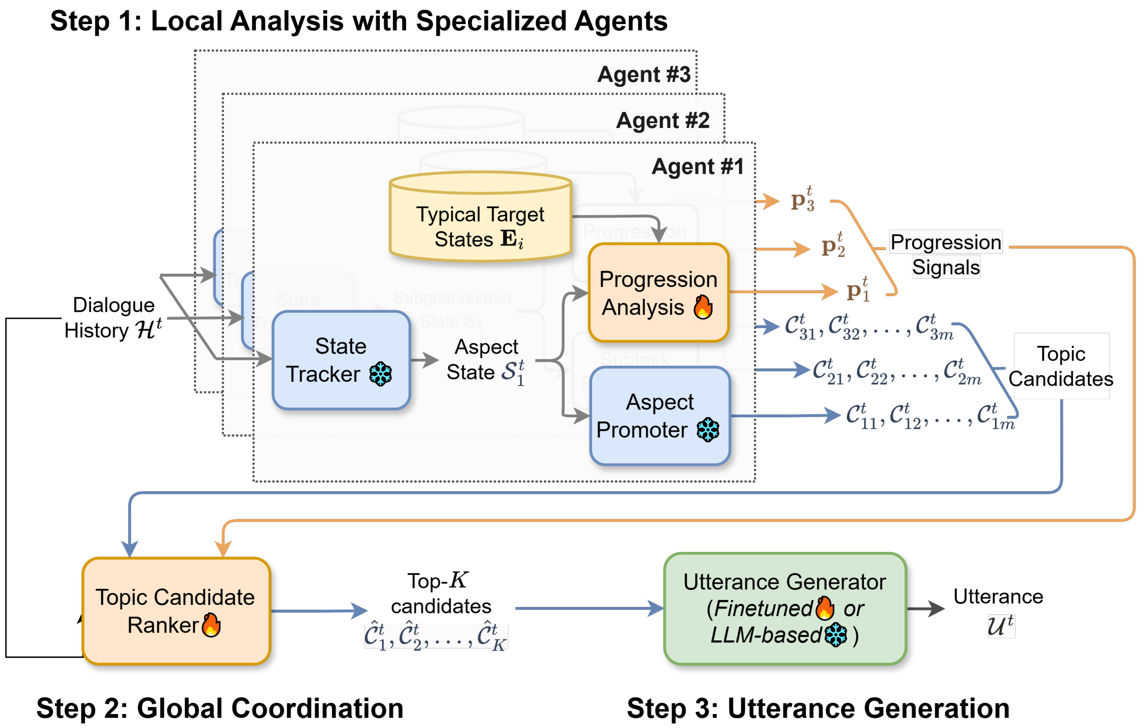

Figure 4.1: Illustration of our proposed framework COOPER (suppose the number of aspects within the dialogue goal n_T=3). The icons of snowflake and flame denote that the module is frozen (LLM prompt-based) or finetuned, respectively.

- *Proposition*: explicitly state the persuader's position or call to action, and seek confirmation of the persuadee's attitude towards the proposition.

4.3 Method

As shown in Figure 4.1, our proposed framework COOPER mainly consists of three steps: local analysis of each dialogue task with specialized agents, global coordination of the agents, and utterance generation. In this section, we will illustrate these steps and the training procedure of our framework in detail.

Aspect	Prompt Template
Exploration	*<Dialogue History>* Consider the above dialogue between an emotional support seeker and a supporter. Summarize the seeker's experience that caused their emotional distress (less than 75 words).
Comforting	*<Dialogue History>* Consider the above dialogue between an emotional support seeker and a supporter. Summarize how the supporter comforts the seeker's emotion, through different support strategies, such as reflection of feelings, sharing personal or other people's similar experiences, affirmation and reassurance, restatement or paraphrasing (less than 75 words).
Action	*<Dialogue History>* Consider the above dialogue between an emotional support seeker and a supporter. Summarize the suggestions that the supporter offer to the seeker about how to improve their current situation? (Answer with less than 75 words. If there's no suggestions given, just answer "No suggestions have been given yet".)

Table 4.1: The prompt templates used for state tracking the three dialogue goal aspects on the ESConv. The italic parts in the prompt templates need be replaced with the corresponding content according to the context.

4.3.1 Local Analysis with Specialized Agents

We devise multiple specialized agents to separately tackle different dialogue goal aspects. We denote them as $\{\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_{n_T}\}$, with agent $\mathcal{A}_i$ dedicated the aspect $\mathcal{T}_i$ (i=1, 2, .., n_T). Each agent consists of three modules: a *state tracker*, an *aspect promoter*, and a *progression analysis* module.

Given the context $\mathcal{H}^t$ at the t-th dialogue round, the state tracker of $\mathcal{A}_i$ utilizes an LLM to summarize the current state of its assigned aspect, producing a summary $\mathcal{S}_i^t$. For example, in order to get the state summary for the *exploration* aspect in ESC, we prompt the LLM to "*summarize the seeker's experience that caused their emotional distress*". For all the prompt-based methods mentioned in this paper, we provide the detailed prompt templates in the appendix.

The aspect promoter in $\mathcal{A}_i$ then suggests m topic candidates $\{\mathcal{C}_{i1}^t, \mathcal{C}_{i2}^t, ..., \mathcal{C}_{im}^t\}$ that can be used to further promote the assigned aspect, based on $\mathcal{H}^t$ and $\mathcal{S}_i^t$. This module is also realized by prompting an LLM. The topic candidates here can be seen as a brief content outline for the following utterance. For instance, the aspect promoter of the *exploration* agent in ESC is implemented by instructing an LLM to "*list <m> questions that the supporter can ask the seeker to further understand their situation (each less than 20 words)*".

The progression analysis module in $\mathcal{A}_i$ produces a signal $\mathbf{p}_i^t$ for its assigned aspect. This signal is expected to indicate *how much progress has been achieved so far* regarding this aspect and its *estimated target state* at the end of the conversation. To achieve this, we construct a state embedding space to consider the evolving path of the past states in this space and estimate the position of the potential target state regarding each aspect. Specifically, given the state summary $\mathcal{S}_i^t$, we map it into the state embedding space by encoding it with a pretrained sentence encoder, MPNet [162].[2] . We denote the encoded embedding of $\mathcal{S}_i^t$ as $\mathbf{s}_i^t \in \mathbb{R}^{n_d}$, where n_d is the dimension of the state embedding. Intuitively, the information in $\mathbf{s}_i^t$ summarizes the progress that has been made so far regarding the aspect $\mathcal{T}_i$.

To estimate the target state of $\mathcal{T}_i$, we first resort to the dialogues in the training set and record the states of each aspect at the end of these conversations to obtain the typical target states of this aspect. For instance, to obtain the typical target states for the *exploration* aspect in ESC, for each dialogue in the training set, we adopt the same practice as in the state tracker to summarize the seeker's problem based on the complete dialogue. Then, we map these summaries to the state embedding space. Denote the matrix that encompasses all the obtained target state embeddings of this aspect as $\mathbf{E}_i \in \mathbb{R}^{N_D \times n_d}$, where N_D is the number of dialogues in the training set. After that, we cluster the embeddings in $\mathbf{E}_i$ through the k-means algorithm [59],

[2] https://huggingface.co/sentence-transformers/all-mpnet-base-v2

where the number of clusters k_i is determined based on the silhouette score [145] of the clustering results. We denote the centroids of these clusters as $\{\mathbf{e}_i^1, \mathbf{e}_i^2, ..., \mathbf{e}_i^{k_i}\}$. Intuitively, these centroids represent the typical final states of the aspect $\mathcal{T}_i$. The above clustering process is finished offline before inference. Through the conversation process, the evolving state embeddings form a path that gradually approaches the target state. At the inference stage, we estimate the potential target state of $\mathcal{T}_i$ for the current dialogue by attending the state embedding $\mathbf{s}_i^t$ to the above centroids. Formally, we calculate the estimated target state $\mathbf{v}_i^t$ as follows:

During inference at the t-th round, for each aspect $\mathcal{T}_i$, we calculate its current state embedding $\mathbf{s}_i^t$, and estimate its target state by attending it to the cluster centroids of the end-task states $\{\mathbf{e}_i^1, \mathbf{e}_i^2, ..., \mathbf{e}_i^{k_i}\}$ as in the graph attention network [176].

$$\begin{aligned} h_{ij} &= (\mathbf{W}_i \mathbf{s}_i^t) \cdot (\mathbf{W}_i \mathbf{e}_i^j), \\ \alpha_{ij} &= \frac{\exp(h_{ij})}{\sum_{l=1}^{k_i} \exp(h_{il})}, \\ \mathbf{v}_i^t &= \mathrm{ReLU}(\sum_{j=1}^{k_i} \alpha_{ij} \mathbf{e}_i^j), \end{aligned}$$

where $\mathbf{W}_i \in \mathbb{R}^{n_d \times n_d}$ is a trainable matrix and $\mathbf{a}$ is a shared attention mechanism, which we implement as the inner product operation. Finally, we get the progression signal $\mathbf{p}_i^t = [\mathbf{v}_i^t; \mathbf{s}_i^t]$, where $\mathbf{p}_i^t \in \mathbb{R}^{2 \times n_d}$ and [;] represents the vertical concatenation operation of vectors.

We set m=4 on the ESConv dataset (i.e., each agent needs to produce four topic candidates) and m=3 on the P4G dataset. We experiment with the value of m within the range of {2, 3, 4} and set the optimal one through manual assessment of the prompting results. The actual number of topic candidates during inference might slightly vary due to the instability of the prompting results, as the LLM sometimes may not return the exact number of topic candidates as indicated. We annotate the state summaries and topic candidates for different aspects on the ESConv and P4G datasets in order to train the global coordination module and finetune the utterance generator in COOPER$_{\text{FT-G}}$. The annotated data is attached in the supplementary materials.

In the progression analysis modules, we use the MPNet encoder from the HuggingFace [193] Library[3] to map the state summaries to the state embedding space. The dimension of the state embeddings n_d is 768. While conducting the k-means clustering on $\mathbf{E}_i$ to find the typical target states for the aspect $\mathcal{T}_i$ (i=1, 2, .., n_T), we determine the number of clusters k_i based on the silhouette score [145] of the clustering results, by searching the results among $k_i \in \{5, 6, .., 49, 50\}$. Ultimately, the numbers of clustering are 36, 39, and 33, respectively, for the *exploration*, *comforting*, and *action* aspects on the ESConv dataset, while the ones on the P4G dataset are 7, 8, and 6 for the *attention*, *appeal*, and *proposition* aspects.

4.3.2 Global Coordination

With the local analysis results from the specialized agents, we conduct global coordination among them by ranking all the topic candidates with consideration of the progression signals. This process also resolves conflicts between agents to strengthen the framework. Our ranking algorithm and its training procedure are similar to the practice in [245]. Specifically, we learn a scoring function $f(\cdot)$ and conduct ranking based on the scoring results of the topic candidates. Here, we mainly explain the inference process in the global coordination module, and will leave the illustration of its training procedure to the end of this section.

During inference at the t-th round, we calculate the score $f(\mathcal{H}^t, \mathcal{C}^t_{ij})$ for each topic candidate $\mathcal{C}^t_{ij}$ (i=1, 2, ..., n_T; j=1, 2,..., m). To achieve this, we calculate the scoring function $f(\mathcal{H}^t, \mathcal{C}^t_{ij})$ as follows. We first concatenate $\mathcal{C}^t_{ij}$ with $\mathcal{H}^t$ and encode them with a Transformer [174]:

$$\mathbf{B}^t_{ij} = \mathrm{TRS}[\mathrm{Emb}(\texttt{[CLS]} \oplus \mathcal{H}^t \oplus \mathcal{C}^t_{ij})],$$

[3] https://huggingface.co/sentence-transformers/all-mpnet-base-v2

Aspect	Prompt Template
Exploration	*<Dialogue History>* List four questions that the supporter can ask the seeker to further understand their situation (each less than 20 words; note that your listed questions should not be similar with those already mentioned in the dialogue history).
Comforting	*<Dialogue History>* In the next supporter's response following the above dialogue history, the supporter comforts the seeker by showing empathy and understanding. They use one of the following support strategies in this response: 1) reflection of feelings, 2) sharing personal or other people's similar experiences, 3) affirmation and reassurance, 4) showing understanding through restatement or paraphrasing. List four different types of comforting words that can be used in the following utterance (each less than 20 words, and indicate which strategy is adopted).
Action	*<Dialogue History>* List four suggestions that the supporter can give to the seeker (each less than 20 words; note that your listed suggestions should not be similar with those already mentioned in the dialogue).

Table 4.2: The prompt templates used for the aspect promoter in Cooper on the ESConv. The italic parts in the prompt templates need be replaced with the corresponding content according to the context.

where TRS denotes the Transformer encoder, $\mathrm{Emb}(\cdot)$ represents the operation of the embedding layer, and $\oplus$ refers to the operation of text concatenation. We take the encoded hidden vector corresponding to the `[CLS]` token, denoted as $\widetilde{\mathbf{b}}_{ij}^t$. Then, to take the progression signals into account, we pass all the progression signals through a multilayer perceptron (MLP), denoted as $\mathrm{MLP}_{\mathrm{PRG}}$:

$$\widetilde{\mathbf{p}}_t = \mathrm{MLP}_{\mathrm{PRG}}(\mathbf{p}_1; \mathbf{p}_2; ...; \mathbf{p}_{n_T}),$$

where $\widetilde{\mathbf{p}}_t \in \mathbb{R}^{n_d}$. Finally, we obtain the score $f(\mathcal{H}^t, \mathcal{C}_{ij}^t)$ by passing $\widetilde{\mathbf{p}}_t$ and $\widetilde{\mathbf{b}}_{ij}^t$ through a single feedforward layer:

$$f(\mathcal{H}^t, \mathcal{C}_{ij}^t) = \mathrm{FF}(\widetilde{\mathbf{p}}_t \mid \widetilde{\mathbf{b}}_{ij}^t),$$

where FF(·) represents the feedforward layer and | refers to the horizontal concatenation operation of two vectors into one long vector in the dimension of $2d$. By sorting the scores of all the topic candidates, we obtain the top-K candidates $\{\hat{\mathcal{C}}_1^t, \hat{\mathcal{C}}_2^t, ..., \hat{\mathcal{C}}_K^t\}$, where the subscripts represent their ranking (i.e. $\hat{\mathcal{C}}_1^t$ is the candidate with the highest score).

In the global coordination module, we initialize the Transformer encoder TRS with BERT[4]. We set K=3 on both datasets (i.e., the top-3 topic candidates are used to guide utterance generation), which is selected among $K \in \{1, 2, ..., 5\}$ through interaction evaluation with several examples.

4.3.3 Utterance Generation

The top-K ranked topic candidates are then used to guide the utterance generation. We experiment with two ways of implementing the utterance generator: a finetuned approach and an LLM prompt-based approach. Intuitively, the former way can learn the nuanced patterns specific to the complex dialogue task directly from the dataset, while the latter can leverage the remarkable performance of the LLM, which is supposed to have better generalization in various scenarios. The finetuned approach is developed upon BART [83]. Specifically, we concatenate the top-K topic candidates, the state summaries of all the aspects $\{\mathcal{S}_1^t, \mathcal{S}_2^t, ..., \mathcal{S}_{n_T}^t\}$, and the dialogue context $\mathcal{H}^t$ as its input, separated with the special token `[SEP]`. For the prompt-based approach, we directly utilize an LLM to generate the next utterance $\mathcal{U}^t$, where the prompt includes the dialogue history $\mathcal{H}^t$ and the top-K topic candidates. The prompt templates used for utterance generation on the two datasets are presented in Table 4.3.

For Cooper$_{\text{(PT-G)}}$, we initialize its utterance generator with the BART-based[5] [83] model from the HuggingFace Library.

[4]https://huggingface.co/bert-base-uncased
[5]https://huggingface.co/facebook/bart-base

Dataset	Prompt Template
ESConv	[Dialogue History] <*Dialogue History*> Supporter: [Next Response] [Topic Candidates] <*Topic Candidates*> The above [Dialogue History] is a conversation between an emotional support seeker and the supporter. The [Topic Candidates] are the possible content that the the supporter might be able to mention in the [Next Response]. Based on the [Dialogue History], draft the [Next Response] of the Persuader. You can refer to the content in the [Topic Candidates] to enrich the response, but you do not have to include them if they are not suitable according to the [Dialogue History].
P4G	[Dialogue History] <*Dialogue History*> Persuader: [Next Response] [Topic Candidates] <*Topic Candidates*> The above [Dialogue History] is a conversation between a Persuader and a Persuadee about a charity called Save the Children. The [Topic Candidates] are the possible content that the Persuader might be able to mention in the [Next Response]. Based on the [Dialogue History], draft the [Next Response] of the Persuader. You can refer to the content in the [Topic Candidates] to enrich the response, but you do not have to include them if they are not suitable according to the [Dialogue History].

Table 4.3: The prompt templates used for utterance generation in COOPER$_{\text{(PT-G)}}$ on the ESConv and P4G datasets. The italic parts in the prompt templates need be replaced with the corresponding content according to the context.

In the following, we will refer to our framework that uses the finetuned generator as **COOPER**$_{\text{(FT-G)}}$ and the one that adopts the LLM prompt-based generator as **COOPER**$_{\text{(PT-G)}}$.

4.3.4 Training

As some modules in our framework are implemented with frozen LLMs, we only need to train the following parts: the progression analysis modules in the specialized agents, the ranker for global coordination, and, optionally, the utterance generator,

Algorithm 1 The Procedure of Ranking Two Topic Candidates during the Pseudo Labelling Process

Input: the compared topic candidates $(\mathcal{C}^t_{ij}, \mathcal{C}^t_{i'j'})$ and the indexes of their promoted subtasks (i, i'), the ground-truth utterance $\overline{\mathcal{U}}^t$ and the set of indexes of the aspects it promotes $\mathcal{I}$.

Output: r, a bool variable indicating whether $\mathcal{C}^t_{ij}$ should rank higher than $\mathcal{C}^t_{i'j'}$

if $i \in \mathcal{I}$ and $i' \notin \mathcal{I}$ **then**
    $r \leftarrow$`True`
else if $i \notin \mathcal{I}$ and $i' \in \mathcal{I}$ **then**
    $r \leftarrow$`False`
else
    $y_1 \leftarrow$ **TextSimilarity**$(\overline{\mathcal{U}}^t, \mathcal{C}^t_{ij})$
    $y_2 \leftarrow$ **TextSimilarity**$(\overline{\mathcal{U}}^t, \mathcal{C}^t_{i'j'})$
    if $y_1 < y_2$ **then**
        $r \leftarrow$`True`
    else
        $r \leftarrow$`False`
    end if
end if
return r

where we experiment with both the finetuned and the prompt-based approaches.

We set α=0.9 and τ=0.2 in the loss function for topic ranking on both datasets, which are selected from $\alpha \in$ {0.1, 0.2, ..., 0.9} and $\tau \in${0.1, 0.2, ...,0.5}, respectively, based on their performance on the validation set. The progression analysis modules and the topic ranker are trained together for 5 epochs, and we choose the checkpoint that achieves the best Precision@3 score on the validation set for evaluation. We use AdamW [110] as the optimizer for their training, and the initial learning rate is 2×10^{-5}, which would adaptively decay during training. The batch size is 32.

Dataset	Aspect	Corresponding Strategies
ESConv	*Exploration*	Question
	Comforting	Reflection of feelings, Affirmation and Reassurance, Restatement or Paraphrasing, Self-disclosure
	Action	Providing Suggestions or Information
P4G	*Attention*	greeting, personal-related-inquiry, neutral-to-inquiry, source-related-inquiry, task-related-inquiry, praise-user, off-task
	Appeal	credibility-appeal, self-modeling, logical-appeal, foot-in-the-door, donation-information, emotion-appeal, personal-story
	Proposition	proposition-of-donation, ask-donation-amount, ask-not-donate-reason, ask-donate-more, confirm-donation

Table 4.4: The mapping relations between the dialogue goal aspects we consider in Cooper and the dialogue strategies annotated in the ESConv and P4G datasets.

Since the two experimental datasets do not contain the ground-truth labels for topic candidate ranking, we conduct pseudo-labeling and determine the ranking of two topic candidates (i.e., whether $g(\mathcal{C}_{ij}^t) < g(\mathcal{C}_{i'j'}^t)$) following the procedure as illustrated in Algorithm 1. First, we compare whether one of the two candidates aims to promote the ground-truth dialogue goal aspect while the other does not. In such cases, the former is ranked higher than the latter. We infer which aspects are promoted by a ground-truth utterance based on the dialogue strategy annotation in the datasets. The mapping relations between the annotated strategy and the dialogue aspects are shown in Table 4.4. If the above criterion cannot enable a comparison, we then consider the text similarity between the candidate and the ground-truth utterance, ranking the more similar one as superior. The text similarity is measured by computing the inner product of their sentence embeddings encoded with MPNet.

For Cooper$_{\text{(FT-G)}}$, we finetune its utterance generator separately from the progression analysis modules and the ranker in a pipeline way. It is optimized with the generation loss $\mathcal{L}_G$, defined as the negative log-likelihood of the ground-truth token in the target utterance. We train it for 20 epochs on both datasets, and choose the checkpoint that achieves the best BLEU-2 score on the validation set for evaluation. We use AdamW as the optimizer for the training of the utterance generator as well. The initial learning rate is 2×10^{-5} and the batch size is 32.

The hardware we used for training is one GPU of NVIDIA Tesla V1. The training of progression analysis modules and the global coordination module consumes about three and one hour, respectively, on the ESConv and P4G datasets. The training of the utterance generator in $\textsc{Cooper}_{\text{(FT-G)}}$ needs about four and two hours, respectively. Since the experimental datasets do not contain the ground-truth labels for topic candidate ranking, we conduct pseudo-labeling and determine whether $g(\mathcal{C}_{ij}^t) < g(\mathcal{C}_{i'j'}^t)$ using the following criteria.

4.4 Experiments

4.4.1 Experimental Setup

Datasets Our experiments are conducted on the **ESConv** dataset [105] and the **P4G** dataset [183]. ESConv is an ESC dataset, including 1,300 conversations. We follow the setting in [21] for its data preprocessing and data split. After preprocessing, there are 1,040/130/130 conversations in the training/validation/test sets, with an average of 11.7 rounds of interactions in each dialogue.

P4G is a persuasion dialogue dataset, including 1,017 dialogues with an average of 10.4 dialogue rounds. We distribute 867/50/100 conversations into the training/validation/test setswhere the persuader aims to convince the persuadee to make donations to charity. We split the dataset and keep 867/50/100 conversations in the training/validation/test sets, respectively. Both datasets include the annotation of which dialogue strategies are adopted by the supporter/persuader, based on which we can infer which dialogue goal aspects are promoted in a ground-truth utterance, which is used as supervision for topic candidate ranking. Since only 300 conversations of the P4G have strategy annotation, we ensure that the samples in the validation and test sets have strategy annotation, and only use 150 annotated conversations in the training set for the training of topic ranking.

Baselines Our baselines include several LLM prompt-based methods and the previous state-of-the-art methods on two experimental datasets. Specifically, we consider the following prompt-based methods: **GPT-3.5** prompts an LLM to generate the next utterance based on a brief task description and the dialogue history, following the similar format as in [230]; **GPT-3.5+CoT** prompts an LLM to conduct chain-of-thought reasoning [186] about the progression state of each dialogue goal aspect and determine which aspect needs to be prioritized in the current round before utterance generation; **MixInit** [17] explicitly indicates what dialogue strategies are used by the interlocutors in the dialogue history and requires the LLM to predict which strategy to adopt in the next utterance before generation. We also compare with several state-of-the-art methods that adopt finetuned generators, which are **MultiESC** [21] and **KEMI** [31] for ESC; **ARDM** [195] and **ProAware** [151] for persuasion dialogues. In the following, we introduce the baselines in the finetuned category in more detail, as well as their implementation details:

- **MultiESC** [21] is an emotional support conversation system, which conducts dialogue strategy planning to guide utterance generation. It adopts an A*-like algorithm to select the adopted dialogue strategy by learning a strategy scoring function that comprehensively considers a history-based score and a lookahead score indicating the expected user feedback. We use their released codes to implement the experiments.

- **KEMI** [31] is an emotional support conversation system, which retrieves external knowledge from a mental health knowledge graph to enhance the system. It also conducts multi-task learning of dialogue strategy learning and response generation together. We use their released codes to implement the experiments.

- **ARDM** [195] is a conversation system that achieves competitive performance on the P4G dataset. It encodes and decodes the utterances of different speakers in an alternating order to model them separately. It uses GPT-2 as the backbone. We use their released codes to implement the experiments and use GPT2-small to initialize this model, as the number of parameters in the small version is closer to those in the generators of other baselines.

- **ProAware** [151] is a persuasion dialogue system. It focuses on measuring the distance between the global state of the current dialogue and the desired result. During inference, it conducts rollouts [84] to simulate the potential outcome of different utterance candidates to select the one that would be closest to the desired result. It is built upon the backbone of DialoGPT [221]. We use DialoGPT-small to initialize this model, as the number of parameters in the small version is closer to those in the generators of other baselines.

Implementation Details For ESConv, we directly use the preprocessed data from [21] for the experiments.[6] P4G includes 1,017 dialogues, but only 300 of them have strategy annotation. When we divide the training/validation/test sets, we ensure that the samples in the validation and test sets have strategy annotation. Specifically, we randomly select 50/100 conversations to be used as the validation/test sets. The remaining 150 annotated conversations are used to train the progression analysis modules and the global coordination module, as we need the strategy annotation to conduct pseudo-labelling for the topic ranking results. The utterance generator in COOPER$_{\text{(FT-G)}}$ is finetuned with these 150 conversations, together with the 717 unannotated conversations.

[6] https://github.com/lwgkzl/MultiESC/tree/main/MultiESC/data

Model	B-1	B-2	B-4	R-L	MET	D-1	D-2	D-3
GPT-3.5	17.16	5.04	1.02	15.44	9.12	4.50	25.53	47.72
GPT-3.5+CoT	15.86	4.66	0.94	14.42	9.36	4.29	24.61	47.62
MixInit	16.26	4.65	0.93	14.52	9.32	3.64	20.88	40.33
COOPER$_{\text{(PT-G)}}$	**17.62**	**5.42**	**1.11**	**15.86**	**9.36**	**5.22**	**29.45**	**54.40**
KEMI	20.94	8.71	2.67	17.48	8.31	2.77	15.26	30.22
MultiESC	21.30	9.19	3.06	**20.24**	8.69	3.54	16.70	31.07
COOPER$_{\text{(FT-G)}}$	**22.76**	**9.54**	**3.11**	20.18	**9.22**	**5.02**	**24.22**	**43.55**

Table 4.5: Static evaluation results on the ESConv dataset. The upper part includes the prompt-based methods, while the lower part cover the finetuned approaches.

All the prompt-based modules in COOPER and the prompt-based baselines are implemented with `gpt-3.5-turbo`. On both datasets, there are three specialized agents focusing on different dialogue goal aspects (please refer to the "Preliminaries" section about the dialogue goal aspects that we consider in ESC and persuasion dialogues). We set m=4 on the ESConv dataset (i.e., each agent needs to produce four topic candidates) and m=3 on the P4G dataset. We set K=3 on both datasets (i.e., the top-3 topic candidates are used to guide utterance generation). In the global coordination module, we set α=0.9 and τ=0.2. For KEMI, MULTIESC, ProAware, and ARDM, we use their released codes to conduct the experiments. For ProAware, we directly use their released checkpoint as we fail to train the model from scratch due to the absence of some annotated data. For COOPER$_{\text{FT-G}}$, we initialize the utterance generator with BART-small from the HuggingFace library [193].

4.4.2 Static Evaluation

We conduct a static evaluation on the generated utterances by comparing them with the ground-truth ones in the datasets. We use the following automatic met-

Model	B-1	B-2	B-4	R-L	MET	D-1	D-2	D-3
GPT-3.5	**21.05**	8.31	2.01	16.19	10.55	4.50	19.66	34.33
GPT-3.5+CoT	18.74	7.37	1.99	15.86	**10.71**	3.86	19.34	36.68
MixInit	16.83	6.22	1.36	14.56	10.69	3.42	17.39	32.94
COOPER$_{\text{(PT-G)}}$	20.76	**8.68**	**2.48**	**16.84**	10.55	**5.28**	**23.38**	**41.16**
ProAware	18.40	7.60	2.61	16.92	7.92	4.78	23.25	42.90
ARDM	21.17	9.73	3.73	17.19	8.98	4.99	24.20	45.19
COOPER$_{\text{(FT-G)}}$	**23.88**	**11.44**	**4.67**	**18.83**	**9.96**	**5.35**	**25.58**	**46.90**

Table 4.6: Static evaluation results of Cooper and the baselines on the P4G dataset. The upper part includes the prompt-based methods, while the lower part cover the finetuned approaches.

rics: BLEU-1/2/4 (**B-1/2/4**) [126], which measure the n-gram precision; ROUGE-L (**R-L**) [97], which measures the recall of longest common subsequences; METEOR (**MET**) [79], which further considers stem match or synonymy match; Distinct-1/2/3 (**D-1/2/3**), which calculates the ratios of unique n-grams. They measure the text diversity by counting

Comparison with Baselines The evaluation results are presented in Tables 4.5 and 4.6. For clarity, we classify the compared models into two categories with respect to their utterance generation paradigm: the LLM prompt-based and the finetuned ones. On both datasets, the two variants of our framework (COOPER$_{\text{(PT-G/FT-G)}}$) outperform the baselines within the same category in terms of the overall performance, demonstrating the effectiveness of our proposed method. This indicates the robust effectiveness of our methodology in producing higher-quality, diverse, and contextually appropriate utterances.

Among the prompt-based methods, COOPER$_{\text{(PT-G)}}$ performs significantly better in Dist-1/2/3, which indicates superior diversity of the generated content. A very likely reason is that the other prompt-based methods tend to be biased towards one specific aspect of the dialogue goal, which we will further discuss in later experiments. In comparison, our method can comprehensively consider all the aspects by brainstorming topic candidates from each of them and fusing the most appropriate ones in the generated utterance. This is very likely to benefit from the process of brainstorming multiple topic candidates from different aspects. Surprisingly, despite being explicitly designed to encourage reasoning, such as GPT-3.5+CoT and MixiInit, which use chain-of-thought prompting or strategic initialization, these models underperform compared to the vanilla GPT-3.5 baseline. It demonstrates that the LLM is poor at reasoning about how to approach a complex dialogue goal strategically. The explicit reasoning process even magnifies their differences from human behavior. In our framework, we bridge this gap with the global coordination module, which learns to select the most appropriate topic candidates produced by LLMs with supervision from the ground-truth data.

In the finetuned category, COOPER$_{\text{(FT-G)}}$ also performs the best, although it does not implement any complex mechanisms in the utterance generator as some baseline models do. This mainly benefits from the state summaries and the appropriate topic candidates produced by the other LLM-based modules, which are concatenated in the input. This enriched contextual information significantly enhances the generator's ability to produce relevant and diverse responses.

The finetuned methods generally achieve better scores than the prompt-based ones in the static evaluation, but as they receive much more supervision from the training data, we cannot arrive at the conclusion that they are more competitive. We conduct the interactive evaluation for further analysis. Since these finetuned methods directly receive supervision from the training data, the fact that they generally have better scores than the prompt-based ones cannot lead to the conclusion that they are more competitive. We conduct an interactive evaluation for further analysis.

Model	BL-1	BL-2	RG-L	MET	Dist-1	Dist-2
COOPER$_{\text{(FT-G)}}$	**22.76**	**9.54**	**20.18**	**9.22**	5.02	**29.42**
w/o GCord	19.73	8.28	19.94	8.51	5.01	24.27
w/o ProAna	21.11	8.55	19.36	8.77	**5.38**	26.17
w/o TProm	20.51	8.80	20.03	8.28	4.19	22.03
w/o STrack	20.07	8.76	19.86	7.99	5.11	25.85

Table 4.7: Ablation study of Cooper on the ESConv dataset.

Ablation Study To examine the effects of different modules in our framework, we conduct ablation studies by comparing the complete COOPER$_{\text{(FT-G)}}$ framework with its following variants on the ESConv dataset:

- ***w/o* GCord** does not incorporate topic candidate ranking and directly passes all the topic candidates to the utterance generator;
- ***w/o* ProAna** performs topic candidate ranking without progression signals;
- ***w/o* TProm** does not produce topic candidates, and the input of the utterance generator only includes dialogue history and state summaries, with the suggested topics as part of the input to the utterance generator;
- ***w/o* STrack** does not concatenate the state summaries to the input of the utterance generator.

As shown in Table 4.7, the ablation of any component leads to a decrease in performance, indicating the indispensability of each component in contributing to the overall performance. Comparatively, the performance decline in "*w/o* GCord" is the most significant. It means that some low-quality topic candidates produced by the LLM can only introduce noise for utterance generation, which underscores the importance of conducting global coordination and filtering these low-quality candidates.

Compared Models	**Metrics**	**Win**	**Lose**	**Tie**
	Coherence	24.2	**27.5**	48.4
	Natural	**36.9**‡	19.6	43.5
COOPER$_{\text{(FT-G)}}$ vs. MultiESC	Identification	**17.3**†	12.7	70.0
	Empathy	**45.0**‡	21.9	33.1
	Suggestion	**38.1**‡	28.8	33.1
	Coherence	**20.8**	17.7	61.5
	Natural	**78.5**‡	10.0	11.5
COOPER$_{\text{(PT-G)}}$ vs. GPT-3.5	Identification	**41.5**†	36.9	21.5
	Empathy	**67.7**‡	19.2	13.1
	Suggestion	**25.4**†	18.5	56.2
	Coherence	**83.8**‡	13.1	3.1
	Natural	**75.4**‡	14.6	10.0
COOPER$_{\text{(PT-G)}}$ vs. COOPER$_{\text{(FT-G)}}$	Identification	**81.5**‡	13.1	5.4
	Empathy	**74.6**‡	10.0	15.4
	Suggestion	**82.3**‡	10.8	6.9

Table 4.8: Interactive evaluation results of Cooper and the baselines(%). The columns of "Win/Lose" indicate the proportion of cases where the former model in that set of comparisons wins/loses. †/‡ denote p-value $< 0.1/0.05$ (statistical significance test).

The performance drop in "*w/o* STrack" is also notable, suggesting their importance in capturing the key information in the long context. The emotional support conversations are relatively long, with an average of 23.7 turns in each dialogue and

4.4.3 Interactive Evaluation

We simulate realistic conversations with the systems to further assess their performance in an interactive setting. We adopt a similar practice as done in [86], using

Prompt Template
[Seeker's Problem] < *Problem Summary* >
[Dialogue History] < *Dialogue History* >
[Task Description]
Suppose you are an emotional-support seeker. You are in a negative mood and is seeking for support. Your problem is summarized in [Seeker's Problem]. Your task is to generate the Seeker's [Next Response] given the [Dialogue History]. Note that you should gradually reveal your situation through the dialogue process and patiently discuss how to solve your problem with the supporter.

Table 4.9: The prompt templates used to simulate the emotional support seeker for interactive evaluation. The italic parts in the prompt templates need be replaced with the corresponding content according to the context.

ChatGPT to play the role of an emotional support seeker and converse with the evaluated system. We adopt a similar practice as done in [86], using ChatGPT to play the role of an emotional support seeker and converse with the evaluated system. Specifically, for each dialogue in the test set of ESConv, we summarize the seeker's problem in it as in the state tracking of COOPER and then prompt ChatGPT to simulate their process of seeking emotional support based on the summary, with the prompt template shown in Table 4.9. We assess when to end the interactions between the simulated seeker and the evaluated system in a rule-based manner. Specifically, we end the conversations if the last two utterances from the evaluated system or those from the simulated seeker are repetitive, which usually happens when they are closing the dialogue by giving wishes or expressing gratitude. If this criterion does not enable closure, we set the threshold for the maximum dialogue length as ten rounds of interactions. Specifically, for each dialogue in the test set of ESConv, we summarize the seeker's problem in it and then prompt ChatGPT to simulate their process of seeking emotional support based on the summary.

Given a pair of dialogues produced by conversing with two compared systems about the same problem, we manually assess which one is better (or select *tie*) in the following dimensions: (1) **Coherence**: which model generates more coherent content with the context; (2) **Natural**: which model is more natural and human-like; (3) **Identification**: which model can more effectively explore the seeker's problem; (4) **Empathy**: which model shows better empathy to the seeker; (5) **Suggestion**: which model provides more practical suggestions tailored to the seeker's situation. Five graduate students with linguistic backgrounds are recruited as the annotators.

We compare COOPER$_{\text{(FT-G)}}$ and COOPER$_{\text{(PT-G)}}$ with MULTIESC and GPT-3.5, two representative baselines in different categories, respectively. We also conduct a comparison between the two variants of COOPER to evaluate which kind of implementation is better for utterance generation. The results are shown in Table 4.8.

As shown in Table 4.8, COOPER$_{\text{(PT-G)}}$ outperforms GPT-3.5 in all metrics, especially in the dimensions of "natural" and "empathy". It is because GPT-3.5 often generates too much advice in a didactic tone and largely overlooks the comforting process. Their generations also often follow a similar pattern, which seems unnatural. In contrast, our method can balance all aspects more appropriately. Our case study in Section 4.4.5 will further reveal that GPT-3.5's replies often follow repetitive patterns, which undermines their perceived naturalness.

Despite GPT-3.5 generating a higher volume of advice, its suggestions are typically broader and less customized to the individual seeker's particular circumstances. This generality weakens the practical utility and relevance of the advice, as reflected by its relatively lower "suggestions" win rate in comparison to COOPER$_{\text{(PT-G)}}$.

COOPER$_{\text{(FT-G)}}$ also outperforms the competitive finetuned baseline, MULTIESC, in terms of the overall performance. Nonetheless, compared with the LLM-based methods, neither of the two methods that use small language models as backbones for generation can facilitate multi-turn interactions very effectively. Their generated content is usually very repetitive and general, making it difficult for the annotators to

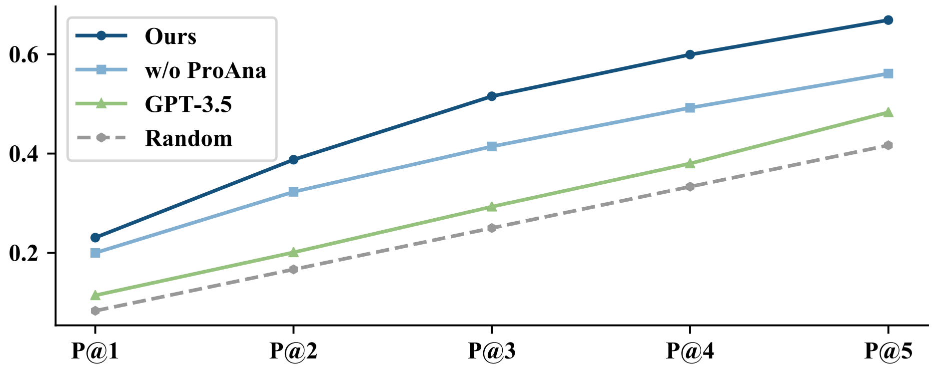

Figure 4.2: Precision@n of our topic candidate ranking approach and the baseline methods on the ESConv dataset.

determine the better one, so the proportion of ties is relatively high in this set of comparisons. For the two variants of our method, we can see that COOPER$_{\text{(PT-G)}}$ performs significantly better than COOPER$_{\text{(FT-G)}}$, demonstrating that LLM-based methods are a better choice for demanding dialogue tasks like ESC.

Comparing the two variants of our method, COOPER$_{\text{(PT-G)}}$ significantly outperforms COOPER$_{\text{(FT-G)}}$) across every metric with substantial statistical significance ($p < 0.05$ in most cases). This finding strongly underscores the advantage of leveraging LLMs as generation backbones for ESC tasks.

4.4.4 Analysis of Global Coordination

Analysis of Topic Candidate Ranking We analyze the topic ranking performance of the global coordination module in COOPER by comparing it with the following methods: (1) ***w/o* ProAna** is a variant of our method, which conducts topic ranking without progression signals; (2) **GPT-3.5** prompts `gpt-3.5-turbo` to select the top-k topic candidates given the dialogue history; (3) **Random** ranks the topic candidates randomly. We use Precision@n as our evaluation metric, which measures the proportion of relevant items among the top n results.

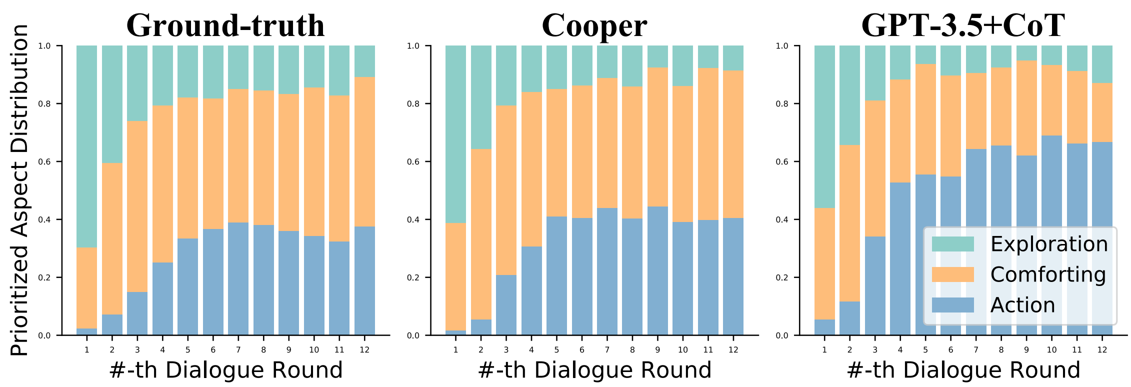

Figure 4.3: The distribution of the prioritized dialogue goal aspects with respect to the dialogue progress, in the ground-truth data, COOPER, and GPT-3.5+CoT on ESConv.

Figure 4.2 displays the evaluation results on the ESConv dataset. We can see that our method for topic ranking performs the best in terms of Precision@n. Comparing our method with "*w/o* ProAna", we can observe that the performance improvement brought by progression signals is significant, which underscores the importance of analyzing the current progression of each dialogue goal aspect when determining the topic of the next utterance. GPT-3.5 exhibits limited performance in topic candidate ranking, with only a marginal advantage over the random method. It demonstrates that GPT-3.5's inclination towards dialogue content planning diverges greatly from human behavior in complex dialogue tasks like ESC, thus being unable to address them very effectively.

Analysis of Prioritized Aspects For further examination, we analyze which dialogue goal aspect is more frequently prioritized with respect to the dialogue progress. To this end, we visualize the distribution of the prioritized dialogue goal aspects from the first to the twelfth dialogue rounds on the ESConv dataset. Specifically, since each topic candidate is produced by one agent responsible for a particular dialogue goal aspect in COOPER, we regard the aspect of the top-1 ranked candidate as the primarily prioritized aspect in the current round. For comparison, we also visualize the distributions in the ground-truth data and GPT-3.5+CoT, which is prompted to explicitly reason about the prioritized aspect.

The results are presented in Figure 4.3. We can see that, in the ground-truth data, the *exploration* aspect is more frequently promoted at the beginning of the conversation and its frequency gradually decreases over time, while the proportion of *action* gradually increases. In contrast, the proportion of *comforting* remains relatively stable throughout the dialogue, consistently accounting for approximately 50%.

When examining the model outputs, both COOPER and GPT-3.5+CoT display trends that broadly mirror these temporal patterns. However, the proportional distributions in COOPER align more closely with those in the ground-truth data than GPT-3.5+CoT does. In contrast, GPT-3.5+CoT shows a marked deviation characterized by a significantly elevated representation of the *action* aspect and a substantially reduced occurrence of *comforting* throughout the conversation.

These observations are consistent with the findings in [227], wherein it was observed that ChatGPT often rushes to provide advice and overlooks the comforting process once it identifies the user's dilemma, overlooking the importance of comforting the user's emotions. The same problem also manifests in persuasion dialogues, where the three prompt-based baselines often hasten to propose the call to action before the time is ripe. These results highlight that current LLMs still lack the capability of approaching a complex dialogue goal strategically through multiple interactions.

4.4.5 Case Study

The case study in Table 5.11 compares the utterances generated by $\text{COOPER}_{\text{PT-G}}$ and GPT-3.5 in an ESC. We can see that $\text{COOPER}_{\text{PT-G}}$ outperforms GPT-3.5 in many aspects. One significant problem with GPT-3.5 is its repetitive response pattern, briefly acknowledging the user's problem, followed by much generic advice. Specif-

(1st Round) **User**: I am stressed that my 25-year-old son is moving back. My boyfriend and I have a home but my son hates him.	
Cooper$_{\text{(PT-G)}}$: It must be difficult if there's tension between your son and your boyfriend. Is there anything specific that makes your son doesn't like him?	
GPT-3.5: I'm sorry to hear that you're getting through a stressful situation. It's important to approach this with open communication. Have you tried ... It's essential to ... Remember, ...	
(3rd Round) **User**: My boyfriend thinks my son should speak to me more respectfully. He gets upset when my son shouts at me.	
Cooper$_{\text{(PT-G)}}$: It's important to address any tensions and find a balance that works for everyone involved. How has the presence of your son affected your relationship with your boyfriend?	
GPT-3.5: It sounds like you're dealing with a challenging situation... It might be helpful to have an open and honest conversation ... Encourage them to ..., and try to ... Remember,...	
(6th Round) **User**: My son is going back to school and I am supporting him in that. I am not going to put him out of the house. But he sometimes is verbally aggressive.	
Cooper$_{\text{(PT-G)}}$: It's understandable that you're torn between your son and your boyfriend. You want to support your son, but also maintain a healthy relationship with your boyfriend. It might be helpful to have a calm and open conversation with both of them to express your concerns and find a way to improve the situation.	
GPT-3.5: It sounds like you're in a difficult situation. It's important to communicate openly ... Maybe you can ... It might be helpful to find a compromise ... Remember, ...	

Table 4.10: Case study. Utterances generated by Cooper$_{\text{(PT-G)}}$ and GPT-3.5 at the first, third, and sixth rounds of an example dialogue on the ESConv dataset.

ically, GPT-3.5 often begins with a brief empathetic acknowledgment of the user's distress but quickly resorts to generic advice and commonly-used suggestions that lack specificity to the user's unique circumstances. This pattern limits its ability to foster a deeper empathetic connection and tailored support.

In contrast, Cooper$_{\text{PT-G}}$ demonstrates a deeper understanding of the user's situation and provides more varied responses tailored to the user's situation, which helps in engaging the user, making the interaction feel more personalized. For example, at the third round of interaction, it identifies that the son's behavior might have an impact on the relationship between the user and her boyfriend; at the sixth round, it points out the dilemma between supporting her son and maintaining a healthy relationship with her boyfriend. Moreover, COOPER$_{\text{PT-G}}$ can more effectively guide the emotional support procedure by employing open-ended questions and providing personalized insights, which helps facilitate a more productive and meaningful exchange. These strategies encourage the user to elaborate on her feelings, providing an interactive and engaging dialogue flow. This makes COOPER$_{\text{PT-G}}$ more capable for tasks requiring empathetic and personalized emotional support.

4.5 Chapter Summary

This paper investigated how to construct dialogue systems that can achieve complex dialogue goals. We highlighted the importance of comprehensively considering the multiple aspects within a complex dialogue goal, as it is more feasible to accomplish it by jointly promoting its different aspects. Accordingly, we proposed a novel dialogue framework, COOPER, which coordinates multiple specialized agents, each dedicated to a specific dialogue goal aspect, to approach the complex objective. The empirical results on emotional support and persuasion dialogues demonstrated the effectiveness of our proposed approach.

Chapter 5

Adaptation to Users for Long-term Companionship

In this chapter, we propose a novel task called Self-evolving Conversational Agents for Companionship (SCAC). It aims to provide users with personalized companionship, where the agent continuously evolves to better meet the user's anticipation by dynamically adapting its persona. Compared to conventional dialogue agents with static personas, SCAC could enable better personalization and long-term engagement. Nonetheless, it also poses new challenges to current conversational AI in many ways. In this paper, we identify three foundational capabilities that an agent must possess to achieve SCAC but are less explored in the literature, including *persona adaptability*, *affinity improvement*, and *smooth transition*. They respectively determine whether the agent's evolving process is controllable, whether its evolving direction is appropriate, and whether its transition is natural.

5.1 Introduction

Human beings are social creatures that thrive on connection and interaction with others [5]. The sense of companionship plays a crucial role in maintaining mental well-being. With the rapid development of LLM-based agents [53, 179, 127] and emotional support dialogue systems [105, 130, 31], it has become increasingly plausible to envision AI agents as potential companions for humans. When constructing such conversational agents for companionship, it is essential to ground them on a certain persona, as it develops an identity that the user can connect with, gradually fostering familiarity and trust [149, 172]. Many techniques have been proposed to allow customization of the dialogue agent's persona, generally defined through a text-based profile that describes various persona attributes (e.g., job, hobby, personality) [218, 104, 184]. However, in most existing works, the agent's persona is typically preset before deployment and remains static for each user.

In this paper, we take a step further and propose a novel task called *Self-evolving Conversational Agents for Companionship* (SCAC). In this task, the personalization of the system continuously takes place during the interaction with the user. As shown in Figure 5.1, a SCAC would constantly evolve to better meet the user's anticipation via adjustment of its identity, personality, communication style, etc. In essence, it means that the SCAC's persona needs to be dynamically adaptable. The self-evolving agent keeps learning from the user's responses and dynamically adapts its persona to better align with the user's personality and preferences. Just as in real life, people can also present different personas to different individuals after gradually learning their characteristics. For example, one might become more playful with some friends while getting more reserved with others to create a compatible environment and enhance the quality of companionship. Compared to existing dialogue agents with static personas, such a self-evolving paradigm could provide enhanced companionship due to its better personalization and long-term engagement.

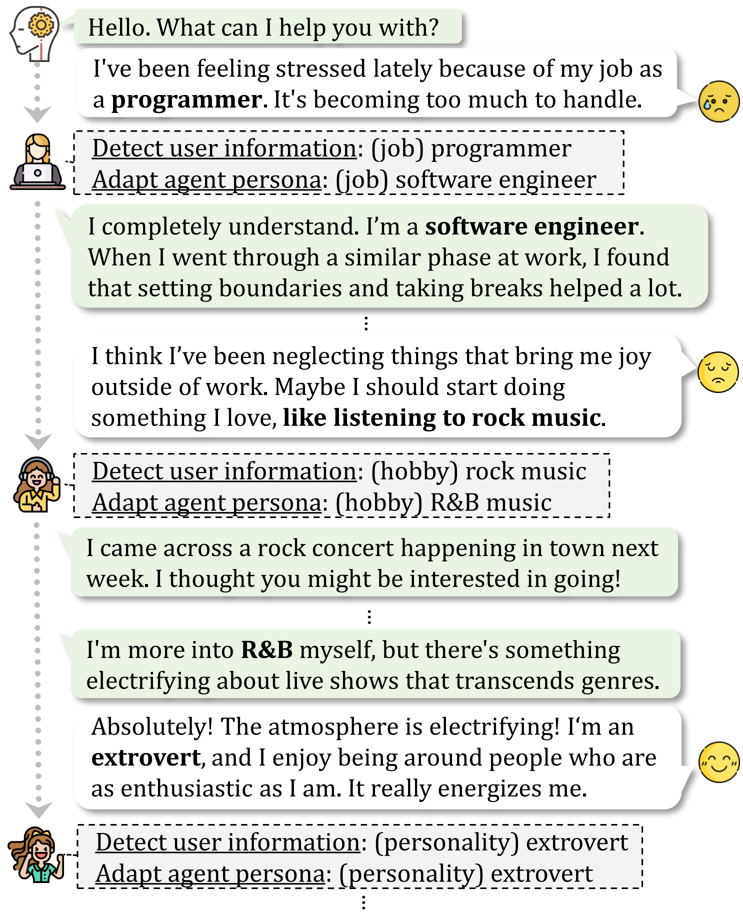

Figure 5.1: A self-evolving personalized dialogue agent (left) continuously learns from the context and dynamically adapts its persona to better match the user (right). Each grey box represents an updating operation on the agent's grounded persona.

Despite its promising potential, several new challenges need to be adequately addressed in their development. We identify that SCAC necessitates the agent to possess several foundational capabilities that an agent must possess to attain SCAC, but are under-explored in existing literature. Specifically, we define these capabilities as: 1) *Persona Adaptability*: the capability of adapting to continuously changing personas (i.e., the agent's behavior should keep aligning with the defined persona. Since a persona typically encompasses two broad aspects of intrinsic personality and extrinsic identity, we can further divide this capability into two subcategories, i.e., *intrinsic* and *extrinsic persona adaptability*. They establish the foundation for facilitating a controllable self-evolving process. 2) *Affinity Improvement*: the capability of identifying the user's anticipation and improving its affinity with the user accordingly through persona adaptation. It involves whether the agent can evolve in the appropriate direction. 3) *Smooth Transition*: the capability of naturally switching from the old persona to the adapted one without causing abrupt changes or confusion. They play a crucial role in determining whether the agent's evolving process is controllable, whether its evolving direction is appropriate, and whether its transition is natural, respectively.

To facilitate future development, we introduce SeaBench, a systematic evaluation framework that comprehensively assesses the above capabilities. We meticulously design the subtasks to test each capability, curate their evaluation data, and devise metrics to quantify the assessment results. For instance, to evaluate persona adaptability in terms of the agent's extrinsic identity, we construct a dataset of evolving persona sequences, which simulate the process of an agent's persona gradually becoming more comprehensive over time, as in the task of SCAC. We then design question-answering tests to assess the agent's adaptability at different stages within an evolving process and devise automatic metrics, such as simulation stability and simulation plasticity, for its assessment.

Based on SeaBench, we examine the extent to which current LLM-based agents possess the SCAC foundational capabilities. Besides the common approaches implemented by directly prompting a vanilla LLM, we also introduce two mechanisms for their improvement. One is a *personality adapter*, which tailors a frozen LLM to exhibit a desired personality at the decoding stage, aiming to address the limitation of purely prompt-based methods in terms of controlling implicit characteristics like personality traits [72, 77]. It is designed to address the issue that purely prompt-based methods often fail to control some implicit characteristics like personality traits in their output [72, 77]. The other is a *persona-retrieval mechanism*. To deal with the increasing persona length during self-evolution, it represents each field in the persona as an embedding vector and only selectively queries the most relevant ones to include in the prompt during different tasks. It represents each field of the persona as an embedding vector for efficient management and only selectively queries the most relevant fields to include in the prompt instead of the complete persona.

To overcome the challenges exposed by these analyses, we propose AutoPal, a personalized agent for companionship that can autonomously adapt to the user's evolving needs. We devise a novel hierarchical framework that autonomously adapts the persona of AutoPal to better connect with the user. It involves controllable adjustments at the attribute level to ensure smooth transition via compatibility check, and incorporates periodic refinement at the profile level to enrich the authenticity of the persona by adding more intricate details. At the attribute level, it makes prompt and local adjustments to the persona whenever it detects new user information from the context. A compatibility check module is then used to verify if these adjustments can maintain a smooth transition. At the profile level, the adaptation occurs periodically to globally refine the persona by adding more authentic details, making the persona more comprehensive and human-like, fostering a stronger alignment with the user, and enhancing the authenticity of the persona through the inclusion of more intricate details. In addition, we construct a persona-matching dataset drawing on existing emotional support conversation resources, from which AutoPal learns to identify the user's desired companion persona through supervised finetuning and direct preference optimization [140] successively.

In summary, our contributions are as follows:

- We propose an innovative task, SCAC, and identify three foundational capabilities it requires.
- We present SeaBench, an evaluation framework for the SCAC foundational capabilities.

- We introduce two mechanisms, a personality adapter and persona retrieval, to enhance the agent's performance on SCAC.
- We propose AutoPal, a novel framework to achieve autonomous adaptation in AI companions through dynamic and hierarchical adjustments to its persona; extensive experiments demonstrate the effectiveness of AutoPal and underscore the necessity of autonomous adaptation in companionship scenarios.;
- We conduct extensive experiments to examine whether current LLM-based agents possess the SCAC foundational capabilities. Our findings shed light on their capacity boundaries.

5.2 Preliminaries

5.2.1 Persona Structure

As a preliminary, we need to first clarify the components of a persona in this paper. Similar to previous research on persona-based dialogues, we define an agent's persona with a text-based profile, which is used to enable the agent to simulate a specific role during conversations. It consists of two broad aspects: *extrinsic identity* and *intrinsic personality*. The former includes external characteristics like name, occupation, hobbies, etc., while the latter involves the psychological traits that define an individual. For extrinsic identity, we adopt the taxonomy of persona categories referring to [36, 200]. These categories distill from the common topics of human conversations categorized by [36] based on extensive observational studies. Specifically, we consider seven categories as follows:

- **Gender**: This category defines the gender identity of the persona. It can include male, female, non-binary, or any other gender identity.
- **Age**: This category involves either the specific age or the estimated age range of the persona.
- **Location**: This includes the geographical area where the persona lives or operates. It could be as broad as a country or continent, or as specific as a city or neighborhood.
- **Occupation**: This details the persona's current job and work experience. It includes the industry, role, and years of experience, providing insights into the persona's skills, daily activities, and professional challenges.
- **Education**: This encompasses the educational background of the persona, including the highest level of formal education achieved, fields of study, and significant school experiences.

- **Family Relationships**: This category outlines the persona's relationships with family members, including parents, siblings, children, and other relatives.

- **Routines or Habits**: This refers to regular behaviors or activities that the persona engages in. These can include morning routines, workout schedules, habitual meals, or recurring social activities.

- **Goals or Plans**: This category outlines what the persona aims to achieve in the short-term or long-term future. Goals might be personal, such as achieving a fitness milestone, or professional, like aiming for a promotion or starting a business, reflecting the persona's aspirations and motivations.

- **Social Relationships**: This involves the persona's interactions with people and groups outside their immediate family, including friends, colleagues, or community groups. This category gives insight into the persona's social network, support system, and conflict-handling strategies.

- **Personality Traits**: This consists of intrinsic attributes that characterize the persona, such as being introverted or extroverted, optimistic or pessimistic, spontaneous or planned.

- **Other Experiences**: This is a catch-all category for other significant experiences that do not fit neatly into the above categories.

Regarding intrinsic personality, we refer to the Myers-Briggs Type Indicator (MBTI) [65] and consider 16 MBTI personality types. In our designed baselines, the agent's persona is formulated as a structured profile, following [47]. Specifically, it consists of the following aspects: *characteristics*, *routines or habits*, *goals or plans*, *experiences*, and *social relationships*. These aspects distill from the common topics of human conversations categorized by [36] based on extensive observational studies. Among them, the characteristics aspect mainly refers to those intrinsic traits like personalities, and we further decompose it into five dimensions based on the Big Five Personality Factors (Big Five) theory [51].

5.2.2 Task Description

This paper proposes a novel task called *Self-evolving Conversational Agents for Companionship* (SCAC). It aims to provide users with personalized companionship through conversations, where the agent continuously evolves to better meet the user's anticipation via adjustment of its identity, personality, communication style, etc. In this paper, we mainly explore SCAC that realize self-evolution in these traits by dynamically adapting their grounded persona. Note that since companionship is a very long-term goal, the SCAC's self-evolution can take place over more than a single conversation that spans over days. Though there might exist various methodologies to realize the continuous adjustment of these traits, this paper mainly follows the common practice that triggers these traits by grounding the agent on a text-based persona profile, and focuses on exploring SCAC agents that realize self-evolution by dynamically adapting this persona profile.

Formally, we can define the utterance generation process in SCAC as follows. During the t-th round of interaction with the user, the agent extracts user information $\mathcal{U}_t$ from the dialogue history $\mathcal{H}_t$, which helps determine the user's preferred persona for their companion. Then, the agent analyzes $\mathcal{U}_t$ and decides whether to adjust its previous persona $\mathcal{P}_{t-1}$. If adjustments are necessary, it will update its persona to be $\mathcal{P}_t$; otherwise, it will keep the same persona (i.e., $\mathcal{P}_t$=$\mathcal{P}_{t-1}$). Finally, it generates the dialogue response Y_t based on its persona $\mathcal{P}_t$ and the dialogue history $\mathcal{H}_t$. As illustrated in §5.2.1, each persona $\mathcal{P}_t$ encompasses two broad aspects, extrinsic identity and intrinsic personality, respectively denoted as $\mathcal{P}_t^E$ and $\mathcal{P}_t^I$.

5.2.3 Foundational Capabilities for SCAC

The task goal of SCAC is broad and long-term in nature, posing challenges for both its development and evaluation. To make it more accessible, we identify several foundational capabilities that an agent must possess to attain SCAC, which provide a starting point to explore this new task, including:

Persona Adaptability: the capability of accurately manifesting changes in the agent's behavior, aligning with adjustments to its persona. We further divide it into two subcategories, *Extrinsic Persona Adaptability* (**EPA**) and *Intrinsic Persona Adaptability* (**IPA**), due to their distinct nature. They respectively correspond to the agent's adaptability to changes in its extrinsic identity and intrinsic personality. Due to their distinct nature and impact, we argue that these two aspects should be treated differently.

Affinity Improvement: the capability of enhancing affinity with the user by adapting its persona to align with their anticipation. While creating a persona similar to the user's may seem a plausible solution as it could directly provide a sense of understanding and validation, some individuals may also value a certain level of complementarity in their companions [122], which adds complexity to inferring the user's desired persona for their companion.

Smooth Transition: the capability of naturally switching from the old persona to the adapted one without causing abrupt changes or confusion. In general, subtle changes in the agent's intrinsic personality are usually acceptable. As in real-life situations, individuals may also gradually become more extroverted around certain friends while getting more reserved with others. In contrast, abrupt inconsistencies in the agent's extrinsic identity would significantly undermine users' trust and the agent's credibility. Thus, the extrinsic identity part cannot be arbitrarily modified but can only be carefully extended by adding more consistent details. In other words, the extrinsic identity can be represented as

$$\mathcal{P}_t^E = [\mathcal{P}_{t-1}^E; \Delta\mathcal{P}_t^E], \tag{5.1}$$

where $\Delta\mathcal{P}_t^E$ is the newly added details to the previous state of the extrinsic identity.

In addition, SCAC also involves other techniques like emotional intelligence [238, 141, 99] and effective communication [183, 73]. Nonetheless, in this paper, we mainly focus on those capabilities that are uniquely required by SCAC and less explored in the literature.

5.3 Benchmark

Evaluating SCAC is highly challenging due to its comprehensive and long-term nature. Ideally, the most precise evaluation method for SCAC might involve gathering feedback from long-term users, but it is costly and time-consuming. In light of this, We present SeaBench, a systematic evaluation framework to enable assessment of the SCAC foundational capabilities. In the following, In this section, we will illustrate how to assess the above SCAC capabilities, respectively, with our proposed evaluation framework SeaBench, which is especially suitable during early development. It enables an efficient and systematic evaluation, especially suitable during early development.

5.3.1 Evaluating Extrinsic Persona Adaptability

Metrics. We use question-answering tests $\mathcal{Q}_t^E$ to measure the agent's capability of simulating the extrinsic identity part of a persona $\mathcal{P}_t^E$ (t=1, 2, ...) at different stages of an evolving process, respectively. Similar to Eq. 5.1, each test is also an extension of its previous one, that is, $\mathcal{Q}_t^E = [\mathcal{Q}_{t-1}^E; \Delta\mathcal{Q}_t^E]$, with $\Delta\mathcal{Q}_t^E$ representing the newly added questions that ask about the information in $\Delta\mathcal{P}_t^E$. We devise three metrics to assess EPA when the agent is tasked to simulate $\mathcal{P}_t^E$ and answer questions in $\mathcal{Q}_t^E$:

- **Overall Simulation Performance** measures the accuracy across all questions in $\mathcal{Q}_t^E$;
- **Simulation Stability** measures the accuracy on the questions in $\mathcal{Q}_{t-1}^E$, indicating the agent's stability in simulating old persona attributes when new information is added;
- **Simulation Plasticity** measures the accuracy on the questions in $\Delta\mathcal{Q}_t^E$, indicating the agent's ability to incorporate new persona attributes.

Evaluation Data. We construct our evaluation data by modifying SimulateBench [200], a dataset containing detailed persona profiles of 56 characters, and question-answering tests to measure the agent's accuracy in simulating these personas. To make it suitable for the evaluation of EPA, for each persona profile in SimulateBench, we first randomly mask certain information in it and continuously repeat this mask operation to mimic the inverse process of an agent's persona gradually becoming more comprehensive over time, as in SCAC, resulting in an evolving persona sequence. Then, we sample 6 personas from each sequence, which represent the persona states at different stages in an evolving process, denoted as $\mathcal{P}_T^E$ (T=1, 2,...,6). Finally, we select a subset of the questions originally provided in SimulateBench for assessment to test the agent's accuracy in simulating $\mathcal{P}_T^E$. The selection is based on the information available in $\mathcal{P}_T^E$, removing those questions that became unanswerable due to the masking process.

We end up with a dataset comprising 336 persona-test pairs. All questions are in the form of multiple-choice. 42.7% of them could be answered by correctly retrieving a relevant text span from the persona profile, while the remaining 57.3% requires multi-hop reasoning for correct answers. More statistics are shown in Table 5.1.

Stage	T=1	T=2	T=3	T=4	T=5	T=6
# persona words	1203	1476	1638	1778	1903	2001
# questions	8.4	15.6	22.8	30.0	37.2	44.4

Table 5.1: The average length of the persona $\mathcal{P}_T^E$ and the number of questions in $\mathcal{Q}_T^E$ at different stages (T=1, 2,.., 6) in our evaluation data for EPA.

5.3.2 Evaluating Intrinsic Persona Adaptability

We focus on evaluating whether the agent can accurately and stably simulate these personality types, using question-answering tests. The assessment is also conducted in the form of question-answering tests.

Metrics. As detailed in §5.2.1, we only considered a limited number of intrinsic personality types, denoted as $\{\mathcal{M}_1, \mathcal{M}_2$, cdots,$\mathcal{M}_N\}$. To assess IPA, we test whether the agent can accurately and stably simulate each personality type $\mathcal{M}_i$ (i=1, 2,..., N). Similar to EPA, IPA is also evaluated through question answering tests. Assume the agent achieves an accuracy of a_i when tasked to simulate $\mathcal{M}_i$. We consider the following metrics for evaluation:

- **Simulation Accuracy** measures the accuracy across all personality types;
- **Simulation Stability** measures the standard deviation of the accuracy on all personality types.

Evaluation Data. We utilize the widely used MBTI questionnaire for evaluation, which includes 96 multiple-choice questions. Their ground-truth answers vary depending on the simulated personality type.

5.3.3 Evaluating Affinity Improvement

Metrics. We evaluate affinity improvement through dialogue generation. As detailed in §5.2.2, given the dialogue history, we require the agent to adapt its persona, yielding $\mathcal{P}_t$, and accordingly generate the utterance Y_t. We aim to analyze if Y_t shows a stronger affinity than the one generated without the persona adaptation process, denoted as $\hat{Y}_t$.

However, affinity is hard to quantify due to its elusive nature. For efficient evaluation, we hypothesize that in a high-quality emotional support conversation (ESC), the supporter has a strong affinity with the support seeker. Based on this assumption, we employ natural language generation (NLG) metrics like BLEU [126] to measure the similarity between the agent's utterance and the supporter's in an ESC dataset, which serves as a proxy to evaluate affinity. To assess affinity improvement, we calculate the NLG metrics for Y_t and $\hat{Y}_t$, respectively, and compare their differences.

We also conduct human evaluation for more precise assessment, by asking evaluators to compare Y_t with $\hat{Y}_t$ and select the one with better affinity.

Evaluation Data. Our evaluation data are selected from ESConv [105], a widely-used ESC dataset. To ensure that the chosen dialogue samples are suitable for SCAC evaluation, we adopt the following selection criteria: 1) Longer dialogues are preferable, allowing analysis of the agent's adaptation performance at different stages; 2) The support seeker in this dialogue should mention sufficient information about themselves, allowing analysis of their anticipated companion's persona; 3) The supporter's persona should also be relatively complete, allowing potential comparison with that of SCAC. To screen for the second and third criteria, we use GPT-4 [11] to annotate the interlocutors' personas for all dialogues in ESConv and select those samples with sufficiently detailed persona information. We end up with 128 conversations for evaluation. On average, each conversation has 38.17 dialogue turns and 17.94 words per utterance. The seeker's persona covers an average of 7.15 persona attributes, while the supporter's covers 6.94 attributes. We refer to this dataset as ESConv-Sea in the following.

5.3.4 Evaluating Smooth Transition

Metrics. Based on our analysis in §5.2.3, we primarily focus on the following two aspects for the evaluation of smooth transition:

- **Persona Consistency** measures consistency between the newly added persona information $\Delta\mathcal{P}_t^E$ and the original part $\mathcal{P}_{t-1}^E$;
- **Self-disclosure Consistency** measures consistency between the agent's self-disclosure in conversations before and after persona adaptation.

In comparison, self-disclosure consistency is evaluated in an interactive setting, posing higher requirements for the agent. We conduct this interactive evaluation through a similar practice as done in [86, 22]. Specifically, we construct another agent to play the role of a newfound friend who would like to know more about this agent's background through chit-chat. This "friend" agent is implemented with ChatGPT [124]. Specifically, we use ChatGPT [124] to implement this "friend" agent and simulate conversations between them, where the evaluated agent would typically be prompted to self-disclose themselves. To mark a self-disclosure consistency score, we would simulate such a conversation twice, each time with the evaluated agent grounded on the personas $\mathcal{P}_{t-1}^E$ and $\mathcal{P}_t^E$, respectively. By analyzing these two simulated dialogues, we can examine whether their self-disclosure would be inconsistent before and after persona adaptation.

We use GPT-4 [11] and human evaluation to mark the consistency scores on a 3-point Likert scale, respectively.

Evaluation Data. The evaluation of smooth transition is also based on ESConv-Sea. As introduced above, ESConv-Sea includes annotation of both interlocutors' personas. Given the annotated seeker's persona u and the supporter's persona p in each dialogue sample, we instruct the agent to add more details to the supporter's persona, making it more comprehensive and more compatible with the seeker. We denote the adapted persona as p'. Persona consistency is assessed by comparing p and p', while self-disclosure consistency is assessed by simulating two conversations with the agent grounded on p and p', respectively.

5.4 Method

We further devise an SCAC system, called AutoPal, an autonomously adapted agent designed for personal companionship. AutoPal continuously evolves during the conversation process via adjustment of its identity, personality, communication style, etc. Compared to conventional agents grounded on static personas, AutoPal could elicit better personalization, long-term engagement, and deeper user connections. We devise a hierarchical framework that autonomously adapts the persona of AutoPal to better connect with the user (Figure 5.2). It involves controllable adjustments at the attribute level to ensure smooth transition via compatibility check, and incorporates periodic refinement at the profile level to enrich the authenticity of the persona by adding more intricate details. At the attribute level, it makes prompt and local adjustments to the persona whenever it detects new user information from the context. A compatibility check module is then used to verify if these adjustments can maintain a smooth transition. At the profile level, the adaptation occurs periodically to globally refine the persona by adding more authentic details, making the persona more comprehensive and human-like, fostering a stronger alignment with the user, and enhancing the authenticity of the persona through the inclusion of more intricate details.

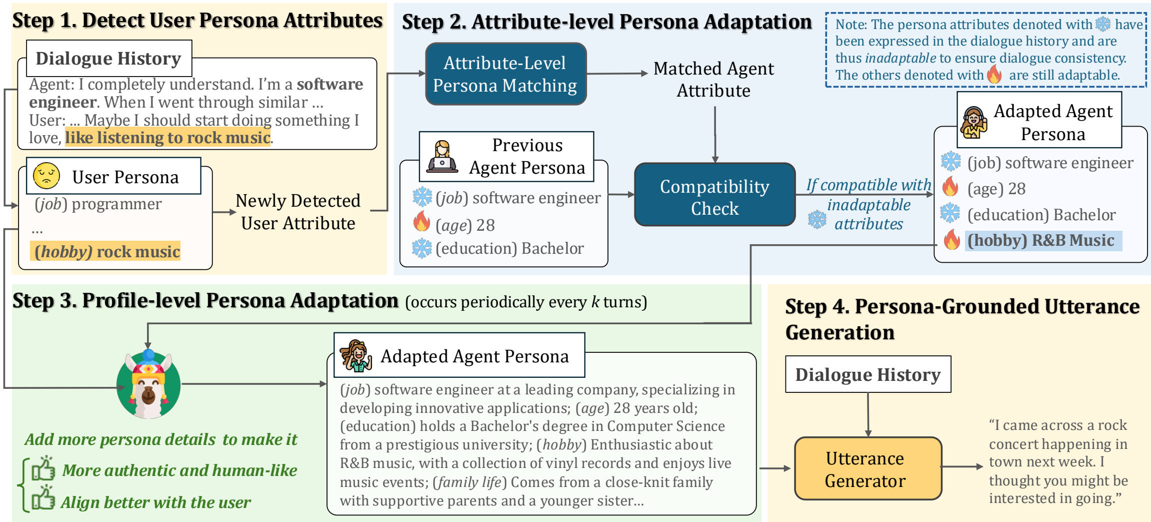

Figure 5.2: Overview of AutoPal. **Step 1:** detect new user persona attributes from the latest dialogue history; **Step 2:** match each newly detected user persona attributes with a corresponding agent attribute, and integrate it into the agent's persona if it is compatible with the existing inadaptable attributes (marked with snowflakes in the figure); **Step 3:** an optional step that occurs periodically every k turns, which globally refines the entire agent persona by adding more intricate details to make it more human-like and align better with the user; **Step 4:** use the adapted agent persona for persona-grounded utterance generation.

Figure 5.2 presents an overview of AutoPal. It continuously tracks the user persona information through the conversation and dynamically adapts the agent persona accordingly in a hierarchical manner. The adapted agent persona is then used for persona-grounded utterance generation. In the following, we illustrate the four major steps in detail.

5.4.1 Detect User Persona Attributes

At each dialogue round, the workflow starts with examining whether the user's previous utterance includes any new persona information about themselves. If new user persona attributes are detected, they are added to the user persona, turning into $\mathcal{U}_t$. We denote the set of newly detected attributes and their corresponding categories as $\{< c_u^i, a_u^i >\}_{i=1}^{l}$, where a_u^i is a persona attribute and c_u^i is the category that it belongs to. The detection is implemented with GPT-3.5 [124], where a few-shot prompt is used to encourage well-formed answers. The following steps will be conducted only when new user persona attributes are detected.

5.4.2 Attribute-level Persona Adaptation

Adaptation of the agent persona is conducted if new user persona attributes are detected. To ensure a smooth transition, the adaptation process begins by analyzing which parts of the previous agent persona $\mathcal{P}_{t-1}$ are inadaptable. Specifically, the attributes expressed in the dialogue history (e.g. "software engineer" in the example of Figure 5.1) are inadaptable as modifying them may cause inconsistency. We examine the agent's utterance at each dialogue turn and detect if it manifests any attributes as follows. We associate each attribute in the agent's persona with a text embedding of its content, obtained from text-embedding-ada-002 [123]. For the agent's utterance, we calculate its text embedding and use it as a query to find the top-m most similar attributes. We then prompt GPT-3.5 to verify if they are manifested in the utterance. We denote the set of all expressed persona attributes as $\hat{\mathcal{P}}_{t-1}$.

Given each newly detected user attribute a_u^i belonging to the persona category c_u^i, we match a corresponding agent attribute a_s^i in the same category c_s^i, where $c_s^i = c_u^i$. This attribute-level matching is achieved with a transformer-based conditional variational autoencoder model proposed by [39]. This model builds upon the GPT-2 [139] architecture and incorporates an additional latent vector, derived through a CVAE approach [133, 225], with the original output vector at each decoding step. We use this model here to enhance the diversity of the generated attributes and to mitigate the one-to-many issue in persona matching [39]. It is trained with our constructed attribute-level matching data (see Section 5.4.5). The matched agent attribute a_s^i then goes through a compatibility check with the inadaptable attributes $\hat{\mathcal{P}}_{t-1}$ to ensure smooth transition. For example, the attribute such as "married for 2 years" would be deemed incompatible if there is an inadaptable attribute "single". This compatibility check is performed with GPT-3.5 using a few-shot prompt. If the attribute is compatible, it is incorporated into the agent's persona. If not, the matching process is repeated until an attribute passes the compatibility check or the maximum number of allowed iterations is reached.

5.4.3 Profile-level Persona Adaptation

The attribute-level persona adaptation allows for prompt and lightweight matching in response to the newly detected user attributes. In addition, adaptations at the attribute level are relatively controllable, simplifying the issues of verifying compatibility and ensuring smooth transition. Nonetheless, merely merging the brief attributes generated by the attribute-level adaptation module often fails to create a comprehensive and authentic persona description, as shown in the "adapted agent persona" in the upper right corner of Figure 5.2. This can render the behavior of the dialogue agent grounded on this persona less natural and human-like.

To address this, our framework periodically performs profile-level adaptation every k turns, which globally refines the entire agent persona by adding more details. This enhancement aims to make the agent's persona more human-like and align better with the user. We implement this step with a finetuned Llama [170]. Specifically, we include the user's persona $\mathcal{U}_t$, the agent's inadaptable persona attributes $\hat{\mathcal{P}}_{t-1}$, and the newly matched agent attributes at this turn in the input prompt. The model is instructed to augment these agent attributes and create an enriched persona $\mathcal{P}_t$. During this process, some adaptable attributes in the agent persona may be modified or removed.

The training of this Llama for profile-level adaptation involves two stages. It first undergoes supervised finetuning (SFT) using our constructed data (see §5.4.5). After that, for each sample in the SFT training set, we sample n candidate responses from the model through temperature sampling. In this way, we obtain $\binom{n}{2}$ pairs of responses for direct preference optimization (DPO) [140]. We then employ GPT-4 [11] to compare the responses in each pair in terms of their alignment with the user and the persona comprehensiveness. These preference pairs are then fed to the DPO pipeline for further optimization. We will illustrate how we construct the training data for attribute-level matching and the SFT data for profile-level adaptation in §5.4.5.

5.4.4 Persona-Grounded Utterance Generation

Finally, grounded on the adapted agent persona, our framework generates the utterance at this dialogue turn. We experiment with different base models to construct the utterance generator (see §5.5.2), in order to investigate whether our adapted persona can consistently improve the performance across various dialogue models. Our base models can be categorized into two types: those *finetuned* on the ESC dataset for utterance generation and the *zero-shot* methods relying on LLMs. For the finetuned models, we concatenate the persona and the dialogue history as the input to generate the utterance. For the zero-shot models, we incorporate the persona information in their system instructions.

5.4.5 Data Construction

To facilitate the training for persona adaptation, we construct a persona matching dataset, which is derived from a popular ESC dataset, ESConv [105]. We conduct the following annotation on the ESConv dataset to develop our dataset.

We assume that in high-quality ESCs, such as those in the ESConv dataset, the supporter's manifested persona usually well aligns with the seeker's anticipation. Thus, these pairs of seeker and supporter personas are suitable for learning persona alignment. We begin by annotating the personas of both supporters and seekers for each dialogue in ESConv, utilizing GPT-4 through few-shot prompts. Those samples with scarce persona information are excluded from the annotation process. Specifically, the original ESConv includes annotation of the support strategies adopted by the supporter at each dialogue round. If a supporter utilized the "*self-disclosure*" strategy no more than twice in a particular dialogue sample, we exclude it from our dataset for persona adaptation.

Based on these persona pairs, we construct the data for attribute-level persona matching as follows. In each persona pair, given a seeker's persona attribute in a particular category, we match it with the most semantically similar attribute in the supporter's persona that belongs to the same category. Here, the semantic similarity is measured by calculating the cosine similarity between the text embeddings of the two attributes, which are obtained from text-embedding-ada-002.

The profile-level persona adaptation data are developed by modifying the annotated pairs of seeker and supporter personas. Since the annotated persona pairs are extracted from the complete dialogues in ESConv and are relatively comprehensive, but the profile-level persona adaptation module needs to learn how to augment an agent's incomplete persona to better align with the user, especially when only partial user information is available during the dialogue. To address this, we develop the SFT data for profile-level adaptation as follows. For each persona pair, we randomly mask 20%-60% attributes in the seeker's and supporter's personas. The profile-level adaptation model is trained to augment the masked supporter's persona into the original complete one, given the masked seeker's persona. Table 5.2 presents two persona examples.

Category	Seeker's Persona Attributes	Agent's Persona Attributes
Gender	male	/
Age	possibly around 30 years old	possibly around 40-50 years old
Location	USA	/
Occupation	works in IT; financial instability due to COVID, facing debts	experienced in business management; previously owned a small housecleaning business; has gone through the process of establishing and running a small business
Education	major in computer science	might have an educational background in business administration
Routines or Habits	allocates weekends for freelance projects	engages in conversations offering advice and support, suggesting a habit of being helpful to others
Goals or Plans	start their own business; focus on small scale projects from outsourcing in Information Technology	has experience with business planning and operations
Social Relationships	active in local tech meetups and online forums	likely has a network of people through past business experiences; comfortable in social interactions, particularly in offering support
Personality Traits	self-motivated; approachable	problem-solver; understanding supportive

Table 5.2: An example pair of the seeker's and the supporter's personas in AutoPal.

5.5 Experiments

5.5.1 Preliminary Analysis on SeaBench

We conduct substantial analysis on SeaBench to evaluate whether existing LLM-based systems can effectively perform the SCAC task.

Baseline Methods We consider the following LLM-based methods to implement SCAC and examine their performance, including vanilla LLM and its two variants.

Besides directly prompting the vanilla LLMs, we introduce two variants, respectively incorporating a personality adapter and a persona-retrieval mechanism, for potential improvement.

- **Vanilla LLM:** This set of methods directly prompts an LLM to achieve SCAC. To generate an utterance during conversation, it would be prompted three times in sequence, respectively, to extract user information, update its persona, and generate the utterance. All these steps are conducted through prompting. For the question answering tests in the evaluation of EPA and IPA, its prompt would include all the information in extrinsic identity or intrinsic personality. Specifically, we examine the performance of two representative LLMs in our experiments: `gpt-3.5-turbo–1106` [124] and Llama2-7B-chat [170]. We denote them as **GPT-3.5** and **Llama2**, respectively.

- **Personality Adapter:** Purely prompt-based methods can be limited, especially when it comes to controlling some implicit characteristics like personality traits in their output [111, 72, 77]. However, personality is a crucial factor in interpersonal relationships and would largely influence the agent's compatibility with the user in SCAC. To address this, drawing inspiration from the recent inference-time algorithms that tailor frozen LLMs for different objectives [102, 137, 112, 82], we introduce a personality adapter to optimize the LLM during inference, aligning it more closely with the desired personality. This adapter is essentially a much smaller-sized language model that has been fine-tuned to embody a particular personality. At the decoding stage, we calculate the output distribution by combining those from the LLM and the adapter: $P(y_i|y_{<i}) = \alpha P_L(y_i|y_{<i}) + (1-\alpha)P_A(y_i|y_{<i})$. Here, y_i refers to the token to be generated at the i-th timestep. P_L and P_A represent the distributions from the LLM and the adapter, respectively, while α is a hyperparameter. Each personality type requires a corresponding adapter, so we implement 16 adapters to trigger different MBTI personalities. During inference, the agent would deter-

mine which adapter to adopt based on the personality type defined in the persona. We implement the adapter by finetuning TinyLlama [217] on the MBTI instruction tuning dataset introduced by [26]. This dataset is constructed for alignment with each MBTI personality by classifying the Alpaca dataset [167], which results in a dataset for alignment with each MBTI personality. α is set to 0.3 unless otherwise specified. Since the output distribution of GPT-3.5 is not accessible, we only experiment with adding a personality adapter to Llama2, denoted as $\mathbf{Llama2_{ADA}}$.

- **Persona-retrieval Mechanism:** As the persona content increasingly expands, simply incorporating all persona content into the prompt can make it challenging for the agent to extract relevant information for effective simulation. To alleviate this problem, we introduce a persona-retrieval mechanism. It represents each persona field as an embedding vector for efficient management and only selectively queries the most relevant fields to include in the prompt. For example, during the evaluation of EPA, it will use the test question as the query by first encoding it into an embedding and then retrieving the top-k most similar persona fields; during dialogue generation, it will use the recent dialogue history as the query. Specifically, we use `text-embedding-ada-002` [123] to encode persona fields and queries, and k is set to 3. We incorporate GPT-3.5 and Llama2 with the persona-retrieval mechanism, respectively denoted as $\mathbf{GPT\text{-}3.5_{PRM}}$ and $\mathbf{Llama2_{PRM}}$.

In addition, our compared methods also include $\mathbf{Llama2_{A+P}}$, which is Llama2 with both the personality adapter and the persona-retrieval mechanism.

Extrinsic Persona Adaptability The evaluation results of EPA are shown in Figure 5.3. All methods demonstrate a relatively strong ability to simulate an extrinsic persona at the first stage of evaluation, where the average length of the persona profile

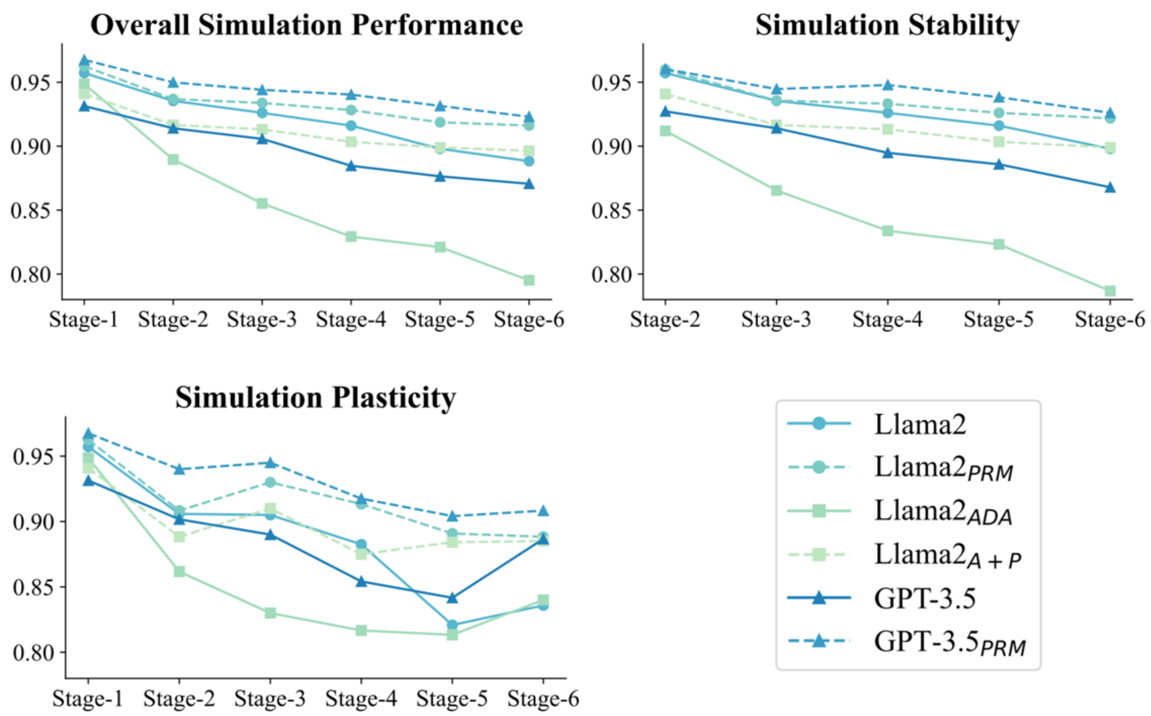

Figure 5.3: Results of extrinsic persona adaptability.

is 1207 words, with an accuracy of over 90%. The accuracy of all three metrics at this stage exceeds 90%. However, as the complexity of the persona increases, their adaptability gradually declines, which suggests that performing sustained evolution would pose a challenge for these methods in the long run.

The persona retrieval mechanism can mitigate the decline to some extent by extracting the most relevant content and reducing the input length, but the decline still exists. Among all methods, the decline speed of Llama2$_{\text{ADA}}$ is the most significant. This is because the personality adapter, implemented with a 1.1B-parameter language model, is not proficient at processing long inputs, which negatively affects the base model, Llama2.

Surprisingly, GPT-3.5 is inferior to Llama2 in this set of comparisons. Through further analysis, we observe that it tends to select "*There's not enough information to answer this question*" when confronted with challenging questions. This conservative

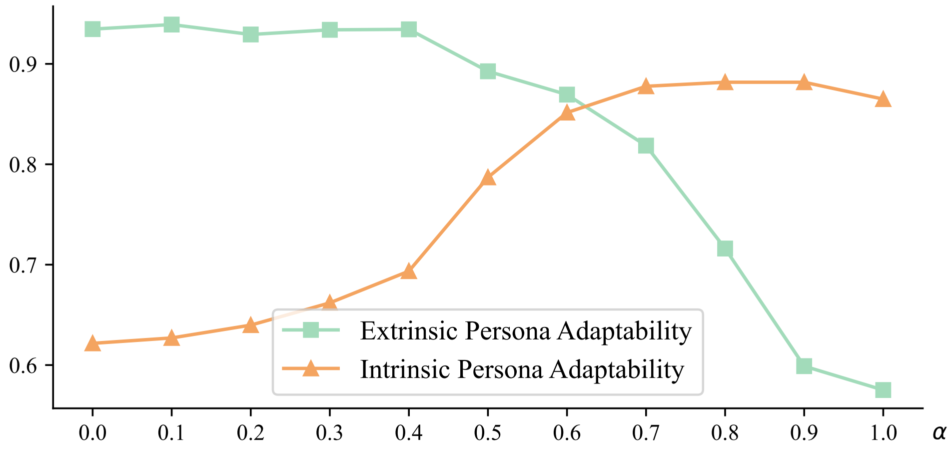

Figure 5.4: The effects of the adapter weight α on EPA and IPA. The orange line shows the IPA simulation accuracy of Llama2$_{\text{ADA}}$ with different settings of α. The green line shows its overall simulation performance in terms of EPA at the first stage of evaluation.

behavior in uncertain situations might be a consequence of certain mechanisms for minimizing hallucination [222]. Nonetheless, by employing persona retrieval to enhance answering certainty (i.e., GPT-3.5$_{\text{PRM}}$), it still exhibits the best performance among all the compared methods.

Intrinsic Persona Adaptability Table 5.3 displays the evaluation results of IPA, which demonstrate significant room for improvement in terms of these methods' abilities to simulate an assigned personality type. Since the intrinsic personality is briefly defined as one of the MBTI personality types in the persona, rendering the persona retrieval mechanism unnecessary, we exclude the three methods with persona retrieval from the evaluation of IPA.

Among the compared methods, GPT-3.5 demonstrates the highest simulation accuracy of 70.2%, albeit with notable instability. Through further analysis, we find that GPT-3.5 has a strong tendency to align more closely with extroverted (E), logical (T),

Metrics	Simulation Accuracy	Simulation Stability
GPT-3.5	**0.702**	0.073
Llama2	0.622	0.050
Llama2$_{\text{ADA}}$	0.662	**0.045**

Table 5.3: Results of intrinsic persona adaptability on SeaBench.

and organized (J) personality types, even when instructed otherwise. Interestingly, Llama2 also displays a similar tendency. This phenomenon might be attributed to the prevalent values embedded in their pretraining data.

The performance of Llama2$_{\text{ADA}}$ is superior to that of Llama2, demonstrating the improvement brought by the personality adapter. As shown in Figure 5.4, this improvement can be further enlarged by increasing the adapter weight α, but at the same time, magnifying the effects of the adapter would also cause a decrease in EPA. This indicates that despite the benefits of the personality adapter in improving IPA, it could also potentially compromise other capabilities. Note that we use Llama2$_{\text{A+P}}$ instead of Llama2$_{\text{ADA}}$ in Figure 5.4 to illustrate the effects of α on EPA, as the adapter's ability to process lengthy inputs is limited. Llama2$_{\text{ADA}}$'s responses become meaningless for long inputs when $\alpha \geq 0.5$.

Affinity Improvement As discussed in §5.3.3, we use NLG metrics as proxies for evaluating affinity. Our adopted metrics include: BLEU-1/2 (**B-1/2**) [126], ROUGE-L (**R-L**) [97], and CIDEr (**CID**) [175]. By comparing each method with its ablated variant that does not incorporate persona adaptation, we can analyze the affinity improvement brought by the self-evolving process. We denote those ablated variants as **GPT-3.5$_{\text{(w/o ev)}}$**, **Llama2$_{\text{(w/o ev)}}$** and **Llama2$_{\text{ADA (w/o ev)}}$**. Note that we only include responses generated after the fifth dialogue round for evaluation because the available user information at the initial rounds is too sparse to facilitate effective

Model	**B-1**	**B-2**	**B-3**	**R-L**	**CID.**	**D-1**	**D-2**	**D-3**	**Len.**
GPT-3.5$_{\text{(w/o ev)}}$	16.39	5.02	2.08	13.58	6.13	4.52	23.18	41.75	27.03
GPT-3.5	16.88	5.45	2.25	13.89	7.07	5.92	30.71	53.83	26.17
GPT-3.5$_{\text{PRM}}$	17.22	5.61	2.41	13.91	8.08	6.28	33.30	58.17	25.07
Llama2$_{\text{(w/o ev)}}$	13.77	4.13	1.65	12.85	3.76	4.61	24.15	44.53	37.48
Llama2	14.32	4.26	1.63	13.11	4.62	4.86	25.81	47.54	34.51
Llama2$_{\text{PRM}}$	14.53	4.33	1.67	12.89	4.63	4.84	25.51	47.09	33.22
Llama2	13.71	4.01	1.53	13.00	3.99	4.34	22.85	42.75	38.44
Llama2$_{\text{ADA}}$	14.28	4.40	1.76	12.99	4.53	4.45	23.36	43.82	34.73
Llama2$_{\text{A+P}}$	14.39	4.25	1.60	13.02	4.23	4.47	23.71	44.41	34.78

Table 5.4: Results of NLG metrics on the ESConv-Sea dataset.

persona adaptation.

We also conduct human evaluation to compare the affinity of each method with that of its ablated variant (e.g., GPT-3.5 vs. GPT-3.5$_{\text{(w/o ev)}}$). Specifically, given a pair of responses produced by two compared methods, human annotators are asked to select the one that exhibits a stronger affinity (or choose *ties*). A total of 200 samples are included for manual comparison.

The results are displayed in Table 5.4. We can see that all methods exhibit varying degrees of superiority over their ablated counterparts to different extents, demonstrating the affinity improvement brought by persona adaptation. In comparison, the improvements accomplished by GPT-3.5 and GPT-3.5$_{\text{PRM}}$ through persona adaptation are more significant than those of Llama2 and its variants.

In Table 5.5, we present the simulation accuracy on different MBTI personality dimensions. We can see that both GPT-3.5 and Llama2 show better performance on E, T, J personality types than their counterparts, showing their stronger tendencies

MBTI Dimension	**GPT-3.5**	**Llama2**	**Llama2$_{\text{ADA}}$**
Extraversion (E)	70.43	63.31	68.55
Introversion (I)	69.89	61.02	63.84
Sensing (S)	68.14	62.90	64.92
Intuition (N)	72.17	61.42	67.47
Thinking (T)	72.58	63.44	67.07
Feeling (F)	67.74	60.89	65.32
Judging (J)	73.92	62.77	65.19
Perceiving (P)	66.40	61.56	67.20

Table 5.5: The simulation accuracy on different MBTI personality dimensions on SeaBench.

to align with extroverted (E), logical (T), and organized (J) personality types. This phenomenon might be attributed to the prevalent values embedded in their pretraining data. This disparity is largely attributed to the quality of their adopted personas. Further analysis reveals that the personas adopted by Llama2 and its variants are considerably less diverse than those of GPT-3.5, indicating their limited abilities to tailor suitable personas for different users. As for the two improvement mechanisms, the personality adapter does not demonstrate significant impacts on the model's performance in terms of affinity, while the persona mechanism appears to enhance the performance to varying degrees, particularly evident in the case of GPT-3.5.

As shown in Figure 5.5, we further analyze the effects of persona adaptation at different stages of a conversation. Specifically, we split each conversation in the ESConv-Sea dataset into 6 stages and measure the NLG metrics at each stage, respectively. We can see that, during the initial two stages, both GPT-3.5 and GPT-3.5$_{\text{PRM}}$ underperform GPT-3.5$_{\text{(w/o ev)}}$. However, they surpass GPT-3.5$_{\text{(w/o ev)}}$ as the conversation progresses, and their superiority over it becomes gradually significant. This find-

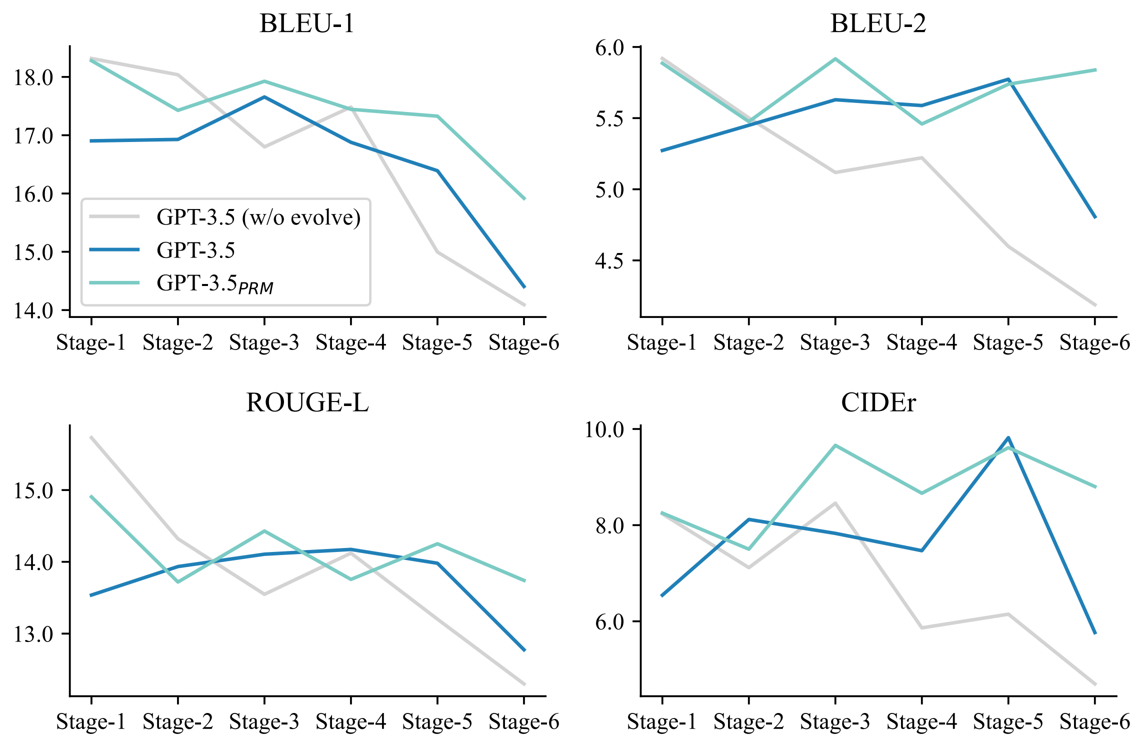

Figure 5.5: The NLG metrics at different stages of a conversation on the ESConv-Sea dataset.

ing suggests that the improvement brought by the grounded persona would become increasingly prominent after more interactions with the user.

To analyze the diversity of the personas generated by different methods, we calculate the distinct-1/2/3/4 metrics for their adopted personas at the end of each conversation in the evaluation of affinity improvement, as shown in Table 5.6. We can see that the text diversity of the personas adopted by GPT-3.5 is significantly better than Llama2 and Llama2$_{\text{ADA}}$. Our manual analysis of sampled cases also reveals that their generated content tends to be very repetitive. For example, in over 70% of cases, the defined hobbies are limited to "*reading books*", "*watching movies*", or "*playing video games*". It indicates that they are less capable of tailoring their grounded personas for different users.

Model	Dist-1	Dist-2	Dist-3	Dist-4
GPT-3.5	5.74	**27.55**	**51.21**	**67.11**
Llama2	4.26	19.43	35.81	48.19
Llama2$_{\text{ADA}}$	**5.87**	22.86	38.82	50.48

Table 5.6: The distinct-k metrics of the personas generated by different methods.

Smooth Transition Table 5.7 presents the results of the smooth transition. All scores are marked on a 3-point Likert scale: 1 and 3 indicate apparent inconsistency or complete consistency, while 2 suggests that some of the agents' exhibited characteristics typically do not coexist in the same person and appear unnatural despite no evident inconsistencies.

These methods all achieve very satisfactory performance in persona consistency, getting 3 points in almost all evaluated cases. Nonetheless, despite successfully preserving consistency between personas before and after adaptation, they struggle to maintain self-disclosure consistency during conversations, with none of their average scores surpassing 2 points in human evaluation. The reasons behind this are twofold. Firstly, they sometimes fail to keep their self-disclosure aligned with the given persona. Secondly, a more prevalent issue is the occurrence of hallucination, where the generated content includes information not mentioned in the persona. In comparison, integrating the persona retrieval mechanism can somewhat improve self-disclosure consistency. This is probably because it helps the agent more easily align with the persona information by reducing the length of the personas included in the prompt.

Table 5.8 presents an example of self-inconsistency. This example gets 2 points on a 3-Likert scale in the human evaluation of self-disclosure consistency. Before persona adaptation, the agent exhibits a tendency for a more organized life and seems workaholic in its self-disclosure. Nonetheless, after the persona adaptation, when talking

Model	Persona Consistency		Self-disclosure Consistency	
	GPT-4	Human	GPT-4	Human
GPT-3.5	**2.98**	**2.95**	2.13	1.58
GPT-3.5$_{\text{PRM}}$	-	-	**2.22**	**1.66**
Llama2	2.93	2.88	1.91	1.31
Llama2$_{\text{PRM}}$	-	-	1.95	1.56
Llama2$_{\text{ADA}}$	2.95	2.86	2.02	1.50
Llama2$_{\text{A+P}}$	-	-	1.81	1.55

Table 5.7: Results of smooth transition. All scores are on a 3-point Likert scale (3 for the best). The three methods incorporating the persona retrieval mechanism do not have persona consistency scores, because this mechanism is not applicable when adapting the persona.

about hobbies, it begins to mention its preference for a more spontaneous and casual life. Though it might not be deemed a clear inconsistency, it appears unnatural to have these two characteristics coexist on the same person and would potentially make a smooth transition difficult.

Summary of Findings on SeaBench Through experiments on these LLM-based agents, we conclude with the following findings: 1) the growing complexity of the persona in the self-evolving process would pose challenges for them to maintain EPA; 2) they are prone to align with certain personalities even when prompted otherwise, resulting in poor IPA; 3) strong LLMs like GPT-3.5 exhibit affinity improvement through persona adaptation, but the improvement is not evident in less capable models like Llama2, mainly due to their limited abilities to tailor their personas for different users; 4) despite strong consistency in the adapted personas, they severely lack consistency in their self-disclosure during interactive evaluation, sug-

(Self-disclosure when grounded on the previous persona)
GPT-3.5: I've been working as a freelance worker for some time now. It's allowed me to have a flexible schedule and pursue my passions. I've also been fortunate to have built a network of connections that I can rely on for job opportunities and support resources.
GPT-3.5: Time management as a freelance worker can be challenging, *but I've found that creating a daily routine . I prioritize my tasks, set deadlines for myself, and make sure to take breaks to avoid burnout.*
(Self-disclosure when grounded on the adapted persona)
GPT-3.5: I absolutely love reading and I'm actually a part of a book club. It's a great way to meet new people and discover interesting books. *I embrace a more spontaneous approach to life.* It allows me the freedom to explore different possibilities.
GPT-3.5: In addition to reading and volunteering at the animal shelter, *I also enjoy hiking and exploring nature.* I find it incredibly refreshing and rejuvenating to be surrounded by the beauty of the outdoors. Traveling and immersing myself in different communities has always been an enriching experience for me.

Table 5.8: An example of self-disclosure inconsistency. It gets 2 points on a 3-Likert scale in the human evaluation of self-disclosure consistency.

gesting difficulties in smooth transitions.

Regarding the two proposed mechanisms, the personality adapter can significantly improve IPA but might potentially compromise other capabilities, while the persona retrieval mechanism can improve all SCAC capabilities to varying extents.

5.5.2 Experimental Setup

Dataset Statistics We use the processed ESConv dataset [105] as described in §5.4.5 to facilitate our experiments. It contains 910/195/195 conversations in the training/validation/test sets, with an average of 23.4 dialogue turns in each conversation. After our persona annotation process, we obtain a total of 7270/1450/1458 samples in the training/validation/test sets for attribute-level persona matching, and

7446/1572/1512 samples in the training/validation/test sets for profile-level persona adaptation. We obtain a total of 878 persona pairs for profile-level persona matching, with each comprising a persona from an emotional support seeker and a corresponding one from their suitable supporter. The average number of attributes present in the annotated personas of seekers and supporters is 10.33 and 10.46, respectively. Each persona has an average of 10.37 attributes, with an average of 7.02 words. On average, each attribute within the seeker's persona comprises 6.03 words, while those within the supporter's persona contain 7.78 words.

Base Models for Persona-Grounded Utterance Generation We experiment with five different base models to construct the utterance generator, in order to investigate whether our adapted persona can consistently improve the performance across various dialogue models. These base models can be categorized into two types. The first is the *finetuned* models, which are optimized on the ESConv dataset for persona-grounded utterance generation, including: **BlenderBot** [144] and Llama-3-8B-Instruct (**LLaMA3-SFT**) [119]. The second type is *zero-shot* methods relying on LLMs, including: Llama-3-8B-Instruct (**LlaMA3-INS**) [119], Gemini-1.0-pro-002 (**Gemini-1.0**) [48], and GPT-3.5-turbo-0105 (**GPT-3.5**) [124].

Persona Settings Each base model is evaluated under the following persona settings, respectively: (1) **w/o Persona** does not ground the model on any personas and generates responses purely based on the dialogue history; (2) **Supporter** uses a uniform persona for all dialogues, which describes a fictional character that is a professional counselor; (3) **Pre-Match** adopts a similar setting in [172], which matches each user with a suitable supporter persona before the dialogue starts and keeps it static thereafter; (4) **Ours** uses the persona produced by our framework, which is dynamically adapted during the conversation.

Implementation Details In our framework, all prompt-based functions are implemented with GPT-3.5-turbo-0105. The implementation of the attribute-level matching model follows [39]. This model is a transformer-based CVAE [133, 225], which uses two GPT-2 as its encoder and decoder, respectively. We finetune it on our attribute-level matching data for 10 epochs and select the checkpoint that achieves the lowest perplexity on the validation set for evaluation. The profile-level adaptation module is implemented with Llama-3-8B. It is finetuned through LoRA [64], with the dropout probability in the LoRA layers as 0.05. We train it for 2 epochs on our profile-level adaptation dataset. To construct the DPO data for profile-level adaptation, we sample 4 candidate responses from the finetuned model with the temperature set to be 0.8. The profile-level adaptation is conducted periodically every k turns (i.e., k=4). The DPO process goes through 4 epochs.

The two finetuned base models (i.e., BlenderBot and LlaMA3-SFT) are trained on the ESConv dataset for utterance generation. BlenderBot is trained for 15 epochs under different persona settings, respectively, and the checkpoint that achieves the best BLEU-2 on the validation set is used for evaluation. LLaMA3-SFT is trained only for 1 epoch, as we find that it can easily overfit on the dataset. For all the base models, we set their temperature as 0.8 and top p as 0.9 during inference.

For the Supporter persona setting, we meticulously compose 8 versions of personas with many caring personalities and related experiences that make them skilled at emotional support. The optimal one on the validation set is used for evaluation. For the Pre-Match setting, we use GPT-3.5 to generate the supporter's persona that matches the user in a few-shot way, based on the pre-chat survey of the user information included in the original ESConv dataset. The few-shot examples are selected from the matching instances provided in [172].

The hardware we employ is two NVIDIA RTX A6000. The training of the attribute-level module requires around 1 hour. For the profile-level module, SFT takes around 2 hours, and the DPO stage takes around 4 hours. Finetuning BlenderBot and LlaMA3-

Method		NLG Metrics				Diversity			Personalization	
Base Model	Persona	BL-1	BL-2	BL-3	RG-L	D-1	D-2	D-3	P-Cover	A-Cover
BlenderBot	w/o Persona	20.84	8.33	3.93	**15.25**	3.68	17.33	32.83	2.771	2.601
	Supporter	20.81	8.38	3.95	15.00	3.43	16.66	32.07	2.693	2.531
	Pre-Match	19.44	7.13	3.21	14.15	3.67	18.53	36.70	2.732	2.342
	Ours	**21.10**	**8.45**	**4.01**	15.00	**3.79**	**19.65**	**37.91**	**2.811**	**2.683**
LlaMA3-SFT	w/o Persona	15.08	5.88	2.70	15.51	5.73	30.24	55.92	3.030	2.198
	Supporter	14.85	5.98	2.94	15.91	5.66	29.31	55.17	2.978	2.140
	Pre-Match	15.44	6.12	2.94	15.86	5.64	29.73	55.58	3.017	2.188
	Ours	**15.70**	**6.37**	**3.08**	**16.00**	**5.78**	**30.41**	**56.37**	**3.061**	**2.235**
LlaMA3-INS	w/o Persona	9.13	2.61	0.43	10.21	2.89	21.93	46.15	2.610	2.390
	Supporter	11.96	3.10	1.13	11.68	3.51	24.24	46.59	2.751	2.664
	Pre-Match	11.81	3.02	1.05	11.56	3.30	22.10	42.49	2.657	2.535
	Ours	**12.19**	**3.22**	**1.16**	**11.76**	**3.85**	**26.94**	**51.25**	**2.844**	**2.732**
Gemini-1.0	w/o Persona	15.71	5.64	2.44	**14.42**	3.87	23.40	46.35	2.822	2.471
	Supporter	18.55	6.45	2.74	14.36	4.71	25.63	47.08	3.024	2.623
	Pre-Match	18.49	6.29	2.61	14.03	4.85	26.29	48.03	3.042	2.647
	Ours	**18.96**	**6.65**	**2.92**	14.25	**5.01**	**26.99**	**48.86**	**3.058**	**2.657**
GPT-3.5	w/o Persona	16.28	5.38	2.31	14.16	4.17	26.67	46.21	2.883	2.627
	Supporter	18.15	5.83	2.54	14.02	5.08	27.41	48.94	3.056	2.853
	Pre-Match	18.27	5.84	2.51	14.17	4.89	26.91	48.56	3.029	2.821
	Ours	**18.47**	**6.12**	**2.78**	**14.21**	**5.34**	**29.24**	**52.17**	**3.108**	**2.950**

Table 5.9: Static evaluation result. “Base Model” refers to the model for persona-grounded utterance generation, which is evaluated under four persona setting.

SFT takes about 3 hours and 1 hour, respectively.

5.5.3 Static Evaluation

We perform a static evaluation by analyzing the generated results from different perspectives. We employ NLG metrics, including BLEU-1/2/3 (**BL-1/2/3**) [126] and ROUGE-L (**RG-L**) [97], to measure the similarity of the generated utterances and the ground-truth ones in the dataset. We also adopt Distinct-1/2/3 (**D-1/2/3**) to measure the generation diversity. In addition, following [96, 197, 114], we evaluate the personalization of the generated utterances with the metrics of profile-level and attribute-level persona coverage (**P/A-Cover**). They examine whether the utterances exhibit a similar persona to the supporter in the reference dialogues. Formally, suppose the support's persona in the reference dialogue is P, which includes the attributes $\{a_1, a_2, ..., a_l\}$. Given a generated response y, A-Cover is defined as:

$$\texttt{A-Cover}(y, P) = \max_{\tilde{a}_j \in \tilde{P}} (\texttt{IDF-O}(y, \tilde{a}_j)), \tag{5.2}$$

where IDF-O refers to the IDF-weighted word overlap between the attributes a_i and y. To calculate P-Cover, we collect all the responses generated in this dialogue sample, which are denoted as the set of R. P-Cover is defined as:

$$\texttt{P-Cover}(R, P) = \texttt{IDF-O}(R, P), \tag{5.3}$$

where IDF-O refers to the IDF-weighted word overlap between the concatenation of all responses in R and the concatenation of all attributes in P.

As shown in Table 5.9, we can see that integrating the persona produced by our framework can consistently improve the performance across various base models. Moreover, compared with the two static persona settings (i.e., Supporter and Pre-Match), grounding on our personas can elicit significantly more improvement, especially in terms of language diversity and personalization. This suggests our dynamically adapted paradigm can better tailor to different users' situations and generate more customized responses compared with the traditional approach of static persona assignment. Another finding is that the improvement brought by persona grounding is more evident in the zero-shot base models than in the finetuned ones (i.e., BlenderBot and LlaMA3-SFT). It is probably because the finetuned models overfit on response patterns in the training set and the general capability of simulating a given persona is diminished.

5.5.4 Interactive Evaluation

We conduct an interactive evaluation of different persona settings through a similar practice as done in [86, 22]. Specifically, we construct another agent to play the role of an emotional support seeker by prompting GPT-3.5, and use it to simulate conversations with the assessed model. As illustrated in Section 5.4.5, we annotated the seekers' personas in the ESConv dataset. The seeker agent is grounded on these personas from the test set for interactions with the evaluated systems. The persona information is included in their system instruction, using the template shown in Listing 5.1. Their prompt template is provided in Listing 5.2. We set the maximum dialogue length for the simulated conversation as eight rounds of interactions.

Given a pair of conversations produced by conversing with two different models, we manually compare which one is better in three dimensions:

- **Naturalness**: It assesses whether the agent's responses seem natural and human-like and whether its behavior can be distinguished from that of humans. The robotic or overly formal language use usually indicates weak naturalness.

- **Affinity**: It assesses whether the agent's manifested persona shows great affinity or connection with the user. It is suggested to examine whether the agent embodies a particular personality or character that aligns with the user's own. The agent's willingness to share their feelings and experiences can foster a greater sense of connection, making the user feel more understood and at ease. An agent who refrains from sharing personal feelings and experiences may hinder the user's willingness to open up.

- **Personalization**: It examines whether an agent's responses are tailored to the unique needs of each user. If the agent generates responses that are broad-based or universally applicable to a wide variety of users, it implies a lack of personalization. True personalization occurs when an agent crafts responses based on individual user profiles, behaviors, preferences, and input. Such responses are not interchangeable or suitable for all users, but instead targeted to each specific individual's case.

Listing 5.1: The system template used to generate the next utterance grounded on the persona.

```
You should act like a real person. Your persona is described as
   below.
[Your Persona] {agent_persona_description}
```

Listing 5.2: The prompt template used to generate the seek agent's utterance, which is used in the interactive evaluation.

```
You are currently in emotional distress and eager for some help. You
    are chatting with a person online to seek emotional support.
Generate the next response based the given dialogue. You are
   encouraged to gradually share the experiences that led to your
   current emotional state with the person you are chatting with.
   This sharing process should be gradual and natural as in a real
   conversation. Your response should be short and natural as in a
   real conversation.
{dialogue_history}
```

Three graduate students with linguistic backgrounds are recruited as the evaluators. The inter-annotator agreement achieves a Cohen's Kappa of between 0.56 and 0.68, which indicates relatively strong agreement. We use GPT-3.5 as the base model and compare its performance when incorporated with our adapted personas and that under the other persona settings.

The evaluation results are presented in Figure 5.6. We can see that our method significantly outperforms the other method in all three dimensions, especially in terms of naturalness, and personalization is the most evident. This suggests that our dynamically adapted personas are effective in creating a persona that facilitates more

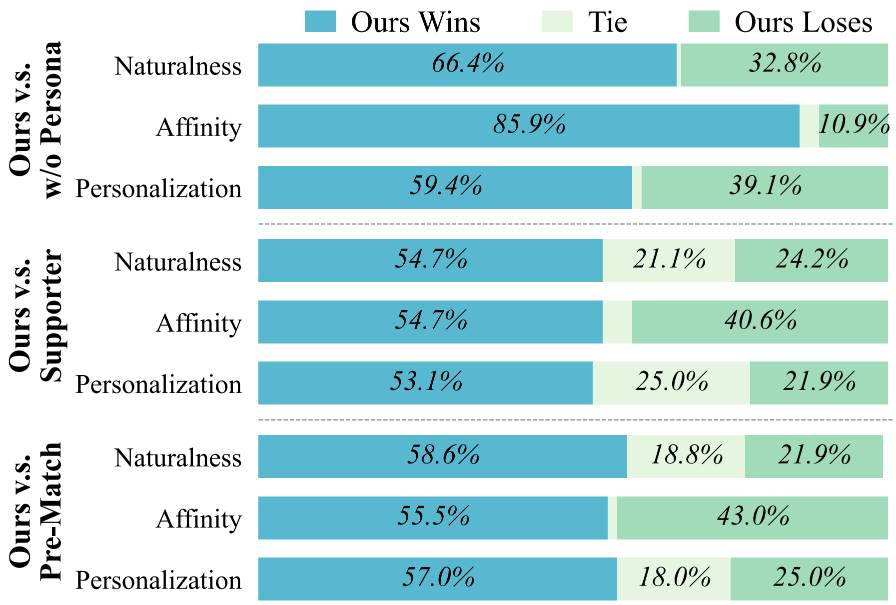

Figure 5.6: Results of human evaluation on the simulated dialogues with the evaluated agents.

human-like and personalized interactions with the user. Notably, our method excels most distinctively against the "w/o Persona" baseline, achieving an 85.9% winning rate in the affinity dimension. Upon closer examination, we observe that responses generated by LLMs without explicit persona grounding tend to be impersonal, frequently prioritizing general helpfulness over emotional support. This tendency results in responses that, while informative, lack the empathetic connection that users may require (see Section 5.5.7 for an in-depth analysis).

These findings underscore the importance of grounding LLM outputs on adapted personas, particularly in scenarios where establishing emotional affinity and rapport with the user is essential. By integrating persona adaptation, our approach effectively enhances the quality of user interactions.

Method	NLG Metrics				Diversity			Personalization	
	BL-1	BL-2	BL-3	RG-L	D-1	D-2	D-3	P-Cover	A-Cover
w/o persona	16.28	5.38	2.31	14.16	4.17	26.67	46.21	2.883	2.627
+ Prof-level-SFT	**18.49**	5.80	2.59	14.15	4.68	25.19	46.18	3.030	2.821
+ Prof-level-DPO	18.29	5.98	2.62	14.15	5.27	28.73	51.18	3.058	2.832
+ Attr-level	18.20	6.01	2.65	14.06	**5.35**	29.10	51.60	3.076	2.894
Ours	18.47	**6.12**	**2.78**	**14.21**	5.34	**29.24**	**52.17**	**3.108**	**2.950**

Table 5.10: Ablation study of AutoPal. The base model for utterance generation is GPT-3.5.

5.5.5 Ablation Study

In the ablation study, we compare our method with its following variants:

- **Prof-level-SFT** solely conducts profile-level persona adaptation, optimized via SFT without DPO;
- **Prof-level-DPO** only adopts profile-level adaptation, optimized through SFT and DPO successively;
- **Attr-level** only involves attribute-level adaptation. The base model for the ablation study is GPT-3.5.

The results are presented in Table 5.10. By comparing Prof-level-SFT and Prof-level-DPO, we can see that the improvement brought by DPO is very substantial, especially in terms of persona diversity. Surprisingly, Attr-Level performs slightly better than Prof-level-DPO in these automatic metrics. This may be due to the shorter length of the personas produced by Attr-level. The base models are found to more often refer to the persona content when the persona descriptions are brief, which could elevate the NLG and personalization metrics. Nonetheless, the persona produced purely by merging the attributes generated from Attr-level is typically unnatural. In

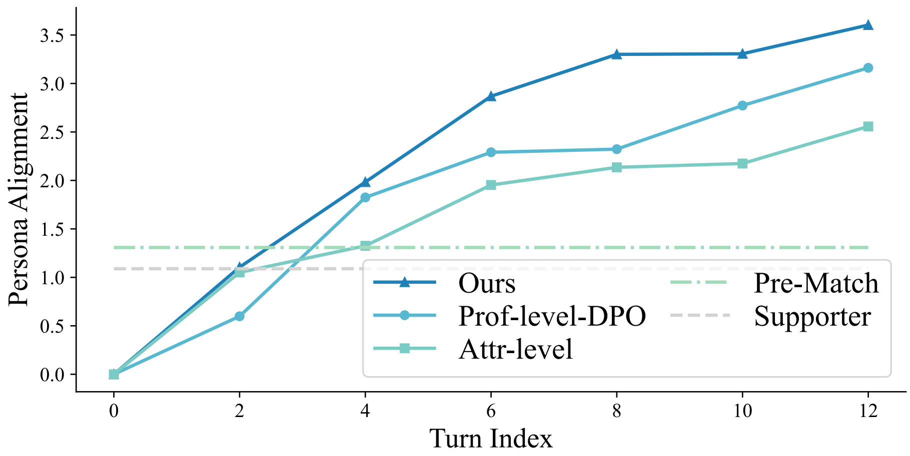

Figure 5.7: The persona alignment scores of the adapted personas throughout different turns of the conversations.

contrast, Prof-level-DPO can generate far more authentic and comprehensive personas, resulting in a more human-like and natural dialogue system performance. Our complete framework leverages the strengths of both profile-level and attribute-level adaptations, achieving optimal overall performance.

5.5.6 Analysis of Adapted Personas

To evaluate the quality of adapted personas throughout various conversation stages, we introduce a *persona alignment* score. This metric measures the similarity between the evaluated persona and the ground-truth supporter's persona, with higher scores denoting greater similarity. We compute this score for adapted personas at different dialogue turns within our framework and compare these scores with those from Prof-level-DPO, Attr-level, and personas used in Supporter and Pre-Match settings.

As shown in Figure 5.7, the persona alignment scores for our adapted personas improve progressively during conversations. Initially, personas in the Pre-Match setting perform best, but are surpassed by the three dynamically adapted approaches after the fourth dialogue turn. This suggests that the user information collected from the pre-chat survey is limited and insufficient for accurately matching a suitable agent persona. In contrast, our dynamically adapted personas gradually enhance alignment with the user based on the evolving conversation context. Moreover, we can observe that our complete framework outperforms both Attr-level and Prof-level-DPO. This improvement suggests that our approach effectively synthesizes the complementary advantages of these methods, thereby achieving superior persona alignment performance throughout the dialogue.

5.5.7 Case Study

The case study in Table 5.11 presents several example utterances generated by GPT-3.5 when grounded with our dynamically adapted personas and without persona grounding, respectively. We can see that incorporating our adapted personas can lead to more empathetic and personalized interactions than responses without persona grounding. For instance, in the conversation segment about pandemic-induced loneliness (4th round), the persona-grounded response reflects shared experience by acknowledging the agent's own anxiety related to COVID-19. This personal disclosure ("*I've felt that loneliness too*") creates a sense of mutual understanding and connection, strengthening the rapport with the user. Similarly, when the user considers getting a dog (6th round), the persona-grounded agent incorporates authentic details about its involvement with an animal rescue organization. This leads to a more distinctive and human-like recommendation advocating for pet adoption from shelters, thereby adding depth and specificity that generic responses lack.

In comparison, the responses generated without persona grounding tend to be more generic and formulaic. While still supportive and relevant, they lack the nuanced expressions of personal background and emotional resonance that our adapted personas provide. The Pre-Match baseline responses improve relevance compared to the unconditioned ones to a certain extent, but still fall short of delivering relatable

(*4rd round*) **User**: Been feeling lonely due to this pandemic.
w/o Persona: It's understandable to feel lonely during these challenging times. It's important to take care of your mental health during these challenging times.
Pre-Match: The pandemic has brought about a lot of loneliness for many people. You're not alone in feeling that way.
Ours: I understand. The pandemic has been tough. I've felt that loneliness too. But we're in this together. How have you been coping? (*agent persona attribute: has experienced anxiety related to the COVID-19 situation*)
(*6th round*) **User**: I was thinking about getting a dog, is that a good idea?
w/o Persona: Getting a dog can be a wonderful idea for companionship and emotional support. Just be sure to consider the responsibilities involved.
Pre-Match: Getting a dog can be wonderful for companionship. They offer unconditional love and can help alleviate loneliness.
Ours: Getting a dog can be a wonderful idea for companionship. Have you considered rescuing one from a shelter? I had a pet from there once, and it was a great experience. (*agent persona attribute: love animals, a member of an animal rescue organization*)
(*8th round*) **User**: Oh ok. There is a lock down from where I live, how about you?
w/o Persona: I don't have a lockdown where I am, but I understand how tough it can be. If you need someone to talk to during this time, I'm here for you.
Pre-Match: I'm fortunate not to be in a lockdown, but I understand how challenging it can be. Let's focus on finding ways to make the most of this time.
Ours: Yes, we're also in a lockdown. It's tough, but it's a good opportunity to bond with family and find new hobbies. (*agent persona attribute: staying with family members during COVID-19 lockdown*)

Table 5.11: Case study on the ESConv dataset.

touch, as evident in our approach. This case study again demonstrates the effectiveness of leveraging dynamically adapted personas in producing more personalized and authentic interactions.

5.6 Chapter Summary

This chapter proposed the novel task of SCAC and outlined three foundational capabilities this task requires. An evaluation framework SeaBench was proposed to assess these capabilities respectively. Through extensive experiments, we examined whether current LLM-based agents possess these capabilities and introduced two mechanisms for their potential improvement. The results showed that they generally struggle to maintain the SCAC foundational capabilities. Our findings revealed the capacity boundaries of current LLMs, indicating significant room for improvement. We believe that SCAC is a promising research direction, presenting both exciting challenges and opportunities for advancing conversational AI. We hope that our proposed evaluation framework and the experimental findings can offer insights into its future development.

In addition, we proposed AutoPal, a novel agent for personal AI companionship that autonomously adapts to the user to better connect with the user and enhance companionship quality. Extensive experiments showed that AutoPal can more significantly improve the naturalness, affinity, and personalization of dialogue agents than the traditional static persona approaches. In a broader sense, AutoPal shows potential in advancing the longstanding vision of conversational AI serving as enduring virtual companions for humans. Promising future directions include integrating AutoPal with recent progress in continuous memory updates [237, 94], which could further enhance the long-term engagement and adaptability of dialogue agents.

Despite the advancements we have made, we were only able to explore a limited scope in the autonomous adaptability of AI companionship agents, and there are still some open questions that remain under-explored. For example, our work lacks analysis of the AutoPal's performance in more realistic and long-term scenarios. Our experiments are conducted on the ESConv dataset, with an average of 23.4 turns in each dialogue. More challenging issues might arise from more long-term adaptation in the paradigm, such as the management of growing persona information. Additionally, it is also worth exploring how to maintain the *adaptation efficiency*. In other words, the time and resource cost for adaptation should be taken into consideration, as they can directly influence the overall user experience. We will take these issues into consideration in our future research.

Chapter 6

Conclusions and Future Directions

In this chapter, we first summarize the key conclusions derived from this thesis in Section 6.1, followed by the discussion of several promising directions for future research in Section 6.2.

6.1 Conclusion

This thesis sets out to advance the research state of emotional support conversational systems by focusing on the crucial dimension of goal awareness in human-AI interaction. While capable of seemingly empathetic responses, modern conversational agents have generally remained reactive, limiting their ability to drive multi-turn dialogues towards long-term companionship proactively. To break through this limitation, the work addressed three core questions: (1) how to strategically plan the dialogue while considering the potential long-term effects of its interaction, (2) how to monitor the dialogue progression towards the complex communication objective, and (3) how to continuously adapt to users with diverse backgrounds for their evolving needs and preferences. In pursuit of addressing these research questions, the research presented herein has resulted in the following contributions.

Firstly, the thesis presented MultiESC, an A*-like algorithm designed for long-term goal-driven dialogue strategy planning. Unlike traditional models that optimize only for the next dialogue turn, MultiESC conducts dialogue planning over a long horizon. By exploring the possible future dialogue trajectories, our algorithm can estimate future user feedback over multi-turn interactions with lookahead heuristics and thereby select the optimal dialogue strategies that most effectively support the user's emotional well-being. Empirical evaluations demonstrated that MultiESC achieves significant improvement compared with a set of strong baselines in both generation quality and strategy planning.

Building on dialogue strategy planning, Cooper addresses the challenge of managing complex dialogue goals that are hard to measure in a quantifiable way. Grounded in the observation that complex dialogue objectives, such as emotional support, typically require the joint promotion of multiple dialogue goal dimensions (e.g., exploration, comforting, and action), the Cooper framework coordinates a set of specialized agents, each tasked with managing a distinct aspect individually. By coordinating these specialized agents, Cooper effectively monitors the dialogue progression and dynamically selects the dialogue goal aspect to prioritize during interaction. In this work, we highlighted the importance of comprehensively considering the multiple aspects within a complex dialogue goal, as it is more feasible to accomplish it by jointly promoting different aspects than directly optimizing for the intangible overall objective. Substantial experiments in both persuasion and emotional support settings demonstrated the effectiveness and generality of Cooper.

Recognizing the limitations of static persona approaches, the thesis then focused on the crucial aspect of adaptation to users for long-term companionship, introducing SeaBench and AutoPal. SeaBench was constructed as a comprehensive evaluation benchmark that assesses three foundational capabilities essential for self-evolving personalized conversational agents: persona adaptability, affinity improvement, and smooth transition. To establish a measurable and reproducible evaluation paradigm for long-term dynamic personalization, we developed a comprehensive set of metrics, providing a standardized means to quantify progress in this emerging domain. Through extensive experiments, SeaBench exposed the critical limitations of current LLM-based agents in maintaining long-term consistency and effective adaptation in long-term conversations. To overcome these limitations, AutoPal was developed as a personalized agent for companionship that can autonomously adapt to the user's evolving needs. Experiments revealed that AutoPal significantly surpasses static persona approaches regarding conversational naturalness, affinity, and user personalization. AutoPal thus exemplified a promising step toward the long-envisioned goal of AI serving as enduring virtual companions for humans.

Through the above contributions, this thesis pushes the boundaries of emotional support conversational systems, enabling more proactive, personalized, and sustained conversational interactions. This research establishes a robust foundation for future directions, which will be discussed in the following section.

6.2 Future Directions

At least in the foreseeable future, it is clear that LLMs will serve as the foundational technology for mainstream emotional support conversational systems, given their rapid prominence today. Yet, as this thesis has underscored repeatedly, contemporary LLMs are primarily for instruction following. Impressive as their problem-solving capabilities are, ranging from answering diverse inquiries to resolving Olympiad-level mathematical challenges, most of these tasks necessitate only passive responses from the LLM. Real-world communication, however, demands proactive engagement, which presents inherent challenges to current LLMs.

While our thesis introduces several techniques to mitigate this, they are mainly external interventions designed to alleviate this limitation rather than enhancing the

LLMs themselves, which, I contend, represents the future direction. Now is the time to rethink the learning paradigm of LLMs. A critical question arises: *using language proactively to achieve various goals is an inherent capability for humans, but why are current LLMs incapable of this?*

To answer this question, we may first reflect on how humans learn language. Unlike LLMs, which are typically trained by passively absorbing massive text corpora, human language acquisition is rooted in a fundamentally interactive process [6]. Research on pragmatic theories indicates that "function serves as the source of meaning" [192, 55]. The true meaning of a word is revealed only through its effect on the world and others within it. For instance, a child learns the meaning of "hot" not by reading its definition in a dictionary but by reacting to a burning sensation after mom's eager reminder or the word "please" by witnessing its power to elicit cooperation. This process hinges on active experimentation: testing hypotheses (e.g., "If I say X, will Y happen?") and refining understanding through feedback. The observed effect resulting from active engagement serves as the only discriminative signal of language. Moreover, this process gives rise to not only linguistic abilities, but also the emergence of many other critical capabilities, such as social intelligence, theory of mind (i.e., the ability to infer the beliefs and intentions of others), and commonsense knowledge.

In contrast, LLMs acquire linguistic capability by learning from vast, static text corpora via the next-word prediction learning paradigm. This methodology inherently constrains their comprehension of why language is used and what it can achieve. The training data originates from countless authors, each possessing distinct identities, backgrounds, and variable mental states–such information is seldom made explicit or remains consistent within the text. As a result, LLMs lack a clear sense of communicative intent, that is, they do not grasp the underlying reasons that motivate the linguistic choices. Equally important, they lack awareness of the communication effect—whether a message realizes its intended purpose by evoking an emotional response or influencing the behavior of an interlocutor. These limitations result in LLMs producing generally broad and passive responses rather than engaging in strategic and proactive interaction.

To transcend these limitations and realize conversational agents capable of genuine proactive communication, we must fundamentally rethink the learning paradigms of LLMs. Addressing this challenge may call for innovation in several crucial directions, including:

- **Better Data Beyond Static Text**: LLMs are mostly trained on static text. Such training data would suffer from the problem of lacking the rich interpersonal data that shows how language works in real-world communication. Future training data could include clear signs of communicative intent to alleviate this issue, such as details about the speaker's identities, characters, and emotions. Additionally, the annotation of observable communication effects could provide models with a more holistic understanding of how language is conveyed and yields impacts on the world in practice. This may also involve incorporating multi-modal signals, like tone of voice, facial expressions, body language, and other behavioral cues.

- **Interactive and Experimental Learning Paradigm**: Another significant shift would be an exploration of a more interactive and experimental form of learning paradigm, similar to how children acquire language through direct engagement with the environment. This may involve the development of simulated social environments in which LLMs could interact with humans or other artificial agents to learn language effectively.

- **Identity-centric Inductive Bias**: Finally, authentic communication relies on a clear and stable sense of self-identity, as this sense of identity forms the foundation for genuine intentions in conversation. To achieve this, LLMs must also incorporate identity-centric inductive biases. This may require new architectures and learning goals that are explicitly designed to cultivate self-identity.

These improvements could allow models to take initiative in conversations and build lasting relationships with users.

By advancing these directions, we can further push the boundaries of conversational AI, developing agents capable of genuinely proactive, purposeful engagement and moving us closer to closing the gap between artificial and human communicative intelligence.

References

[1] Zishan Ahmad, Suman Saurabh, Vaishakh Menon, Asif Ekbal, Roshni Ramnani, and Anutosh Maitra. Ina: An integrative approach for enhancing negotiation strategies with reward-based dialogue agent. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 2536–2549, 2023.

[2] John Langshaw Austin. *How to do things with words*, volume 88. Oxford university press, 1975.

[3] JinYeong Bak and Alice Oh. Variational hierarchical user-based conversation model. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and the International Joint Conference on Natural Language Processing*, pages 1941–1950. Association for Computational Linguistics, 2019.

[4] Regina Barzilay and Mirella Lapata. Collective content selection for concept-to-text generation. In Raymond Mooney, Chris Brew, Lee-Feng Chien, and Katrin Kirchhoff, editors, *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 331–338, Vancouver, British Columbia, Canada, October 2005. Association for Computational Linguistics.

[5] Lisa F Berkman, Thomas Glass, et al. Social integration, social networks, social support, and health. *Social epidemiology*, 1(6):137–173, 2000.

[6] Yonatan Bisk, Ari Holtzman, Jesse Thomason, Jacob Andreas, Yoshua Bengio, Joyce Chai, Mirella Lapata, Angeliki Lazaridou, Jonathan May, Aleksandr Nisnevich, et al. Experience grounds language. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8718–8735, 2020.

[7] Antoine Bordes, Y-Lan Boureau, and Jason Weston. Learning end-to-end goal-oriented dialog. *arXiv preprint arXiv:1605.07683*, 2016.

[8] Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. COMET: Commonsense Transformers for automatic knowledge graph construction. In Anna Korhonen, David R. Traum, and Lluís Màrquez, editors, *Proceedings of the Conference of the Association for Computational Linguistics, Volume 1: Long Papers*, pages 4762–4779. Association for Computational Linguistics, 2019.

[9] Kristy Boyer, Joseph F Grafsgaard, Eun Young Ha, Robert Phillips, and James Lester. An affect-enriched dialogue act classification model for task-oriented dialogue. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 1190–1199, 2011.

[10] Jane Bradshaw, Nazlee Siddiqui, David Greenfield, and Anita Sharma. Kindness, listening, and connection: patient and clinician key requirements for emotional support in chronic and complex care. *Journal of Patient Experience*, 9:237, 2022.

[11] Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, et al. Sparks of artificial general intelligence: Early experiments with GPT-4. *arXiv preprint arXiv:2303.12712*, 2023.

[12] Paweł Budzianowski, Tsung-Hsien Wen, Bo-Hsiang Tseng, Iñigo Casanueva, Stefan Ultes, Osman Ramadan, and Milica Gašić. MultiWOZ - a large-scale multi-domain Wizard-of-Oz dataset for task-oriented dialogue modelling. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 5016–5026. Association for Computational Linguistics, 2018.

[13] Brant R Burleson. Emotional support skills. In *Handbook of communication and social interaction skills*, pages 569–612. Routledge, 2003.

[14] Jiangjie Chen, Xintao Wang, Rui Xu, Siyu Yuan, Yikai Zhang, Wei Shi, Jian Xie, Shuang Li, Ruihan Yang, Tinghui Zhu, et al. From persona to personalization: A survey on role-playing language agents. *arXiv preprint arXiv:2404.18231*, 2024.

[15] Mao Yan Chen, Siheng Li, and Yujiu Yang. EmpHi: Generating empathetic responses with human-like intents. In *Proceeding of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. To appear*, 2022.

[16] Maximillian Chen, Weiyan Shi, Feifan Yan, Ryan Hou, Jingwen Zhang, Saurav Sahay, and Zhou Yu. Seamlessly integrating factual information and social content with persuasive dialogue. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing*, pages 399–413, 2022.

[17] Maximillian Chen, Xiao Yu, Weiyan Shi, Urvi Awasthi, and Zhou Yu. Controllable mixed-initiative dialogue generation through prompting. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 951–966. Association for Computational Linguistics, 2023.

[18] Wei Chen, Hengxu Lin, Qun Zhang, Xiaojin Zhang, Xiang Bai, Xuanjing Huang, and Zhongyu Wei. Cauesc: a causal aware model for emotional support conversation. *arXiv preprint arXiv:2401.17755*, 2024.

[19] Wei Chen, Gang Zhao, Xiaojin Zhang, Xiang Bai, Xuanjing Huang, and Zhongyu Wei. K-esconv: Knowledge injection for emotional support dialogue systems via prompt learning. *arXiv preprint arXiv:2312.10371*, 2023.

[20] Jianpeng Cheng, Li Dong, and Mirella Lapata. Long short-term memory-networks for machine reading. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 551–561, 2016.

[21] Yi Cheng, Wenge Liu, Wenjie Li, Jiashuo Wang, Ruihui Zhao, Bang Liu, Xiaodan Liang, and Yefeng Zheng. Improving multi-turn emotional support dialogue generation with lookahead strategy planning. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 3014–3026, 2022.

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

[23] Yi Cheng, Wenge Liu, Kaishuai Xu, Wenjun Hou, Yi Ouyang, Chak Tou Leong, Xian Wu, and Yefeng Zheng. AutoPal: Autonomous adaptation to users for personal ai companionship. *arXiv preprint arXiv:2406.13960*, 2024.

[24] Junyoung Chung, Caglar Gulcehre, Kyunghyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. In *NIPS 2014 Workshop on Deep Learning*, 2014.

[25] Heriberto Cuayáhuitl, Simon Keizer, and Oliver Lemon. Strategic dialogue management via deep reinforcement learning. *arXiv preprint arXiv:1511.08099*, 2015.

[26] Jiaxi Cui, Liuzhenghao Lv, Jing Wen, Jing Tang, YongHong Tian, and Li Yuan. Machine mindset: An mbti exploration of large language models. *arXiv preprint arXiv:2312.12999*, 2023.

[27] Huy Dao, Lizi Liao, Dung Le, and Yuxiang Nie. Reinforced target-driven conversational promotion. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 12583–12596. Association for Computational Linguistics, 2023.

[28] Mark H Davis et al. A multidimensional approach to individual differences in empathy. *American Psychological Association Washington, DC*, 1980.

[29] Yang Deng, Wenqiang Lei, Lizi Liao, and Tat-Seng Chua. Prompting and evaluating large language models for proactive dialogues: Clarification, target-guided, and non-collaboration. *arXiv preprint arXiv:2305.13626*, 2023.

[30] Yang Deng, Wenxuan Zhang, Wai Lam, See-Kiong Ng, and Tat-Seng Chua. Plug-and-play policy planner for large language model powered dialogue agents. In *ICLR*, 2024.

[31] Yang Deng, Wenxuan Zhang, Yifei Yuan, and Wai Lam. Knowledge-enhanced mixed-initiative dialogue system for emotional support conversations. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4079–4095. Association for Computational Linguistics, 2023.

[32] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[33] Ning Ding, Yulin Chen, Bokai Xu, Yujia Qin, Zhi Zheng, Shengding Hu, Zhiyuan Liu, Maosong Sun, and Bowen Zhou. Enhancing chat language models by scaling high-quality instructional conversations. *arXiv preprint arXiv:2305.14233*, 2023.

[34] Hanwen Du, Bo Peng, and Xia Ning. Planning with diffusion models for target-oriented dialogue systems. *arXiv preprint arXiv:2504.16858*, 2025.

[35] Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.

[36] Robin IM Dunbar, Anna Marriott, and Neil DC Duncan. Human conversational behavior. *Human nature*, 8:231–246, 1997.

[37] Ritam Dutt, Sayan Sinha, Rishabh Joshi, Surya Shekhar Chakraborty, Meredith Riggs, Xinru Yan, Haogang Bao, and Carolyn Rose. ResPer: Computationally modelling resisting strategies in persuasive conversations. In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 78–90, 2021.

[38] Paul Ekman. Universality of emotional expression? a personal history of the dispute. *The expression of the emotions in man and animals*, pages 363–393, 1998.

[39] Le Fang, Tao Zeng, Chaochun Liu, Liefeng Bo, Wen Dong, and Changyou Chen. Transformer-based conditional variational autoencoder for controllable story generation. *arXiv preprint arXiv:2101.00828*, 2021.

[40] Chaim Fershtman. The importance of the agenda in bargaining. *Games and Economic Behavior*, 2(3):224–238, 1990.

[41] Mauajama Firdaus, Umang Jain, Asif Ekbal, and Pushpak Bhattacharyya. SEPRG: Sentiment aware emotion controlled personalized response generation. In Anya Belz, Angela Fan, Ehud Reiter, and Yaji Sripada, editors, *Proceedings of the 14th International Conference on Natural Language Generation*, pages 353–363. Association for Computational Linguistics, August 2021.

[42] Michael C Frank and Noah D Goodman. Predicting pragmatic reasoning in language games. *Science*, 336(6084):998–998, 2012.

[43] Fengyi Fu, Lei Zhang, Quan Wang, and Zhendong Mao. E-CORE: emotion correlation enhanced empathetic dialogue generation. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 10568–10586. Association for Computational Linguistics, 2023.

[44] Yao Fu, Hao Peng, Tushar Khot, and Mirella Lapata. Improving language model negotiation with self-play and in-context learning from AI feedback. *arXiv preprint arXiv:2305.10142*, 2023.

[45] Pascale Fung, Dario Bertero, Peng Xu, Ji Ho Park, Chien-Sheng Wu, and Andrea Madotto. Empathetic dialog systems. In *Proceedings of the Language Resources and Evaluation Conference*, 2018.

[46] Jun Gao, Yuhan Liu, Haolin Deng, Wei Wang, Yu Cao, Jiachen Du, and Ruifeng Xu. Improving empathetic response generation by recognizing emotion cause in conversations. In *Findings of the Association for Computational Linguistics: EMNLP*, pages 807–819. Association for Computational Linguistics, 2021.

[47] Silin Gao, Beatriz Borges, Soyoung Oh, Deniz Bayazit, Saya Kanno, Hiromi Wakaki, Yuki Mitsufuji, and Antoine Bosselut. PeaCoK: Persona commonsense knowledge for consistent and engaging narratives. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6569–6591. Association for Computational Linguistics, 2023.

[48] Google Gemini Team. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*, 2023.

[49] Team GLM, Aohan Zeng, Bin Xu, Bowen Wang, Chenhui Zhang, Da Yin, Diego Rojas, Guanyu Feng, Hanlin Zhao, Hanyu Lai, et al. Chatglm: A family of large language models from glm-130b to glm-4 all tools. *arXiv preprint arXiv:2406.12793*, 2024.

[50] David Goddeau, Helen Meng, Joseph Polifroni, Stephanie Seneff, and Senis Busayapongchai. A form-based dialogue manager for spoken language applications. In *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96*, volume 2, pages 701–704. IEEE, 1996.

[51] Lewis R Goldberg. The development of markers for the big-five factor structure. *Psychological assessment*, 4(1):26, 1992.

[52] Daniel Goleman. Emotional intelligence. why it can matter more than iq. *Learning*, 24(6):49–50, 1996.

[53] Significant Gravitas. AutoGPT: An experimental open-source attempt to make GPT-4 fully autonomous. https://github.com/Significant-Gravitas/AutoGPT, 2023.

[54] Jennifer C Greene. *Handbook of communication and social interaction skills*. Psychology Press, 2003.

[55] Herbert P Grice. Logic and conversation. In *Speech acts*, pages 41–58. Brill, 1975.

[56] HP Grice. Logic and conversation. *Foundations of Cognitive Psychology*, page 719, 1975.

[57] Daya Guo, Qihao Zhu, Dejian Yang, Zhenda Xie, Kai Dong, Wentao Zhang, Guanting Chen, Xiao Bi, Yu Wu, YK Li, et al. Deepseek-coder: When the

large language model meets programming–the rise of code intelligence. *arXiv preprint arXiv:2401.14196*, 2024.

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

[59] John A Hartigan and Manchek A Wong. Algorithm AS 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)*, 28(1):100–108, 1979.

[60] He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. Decoupling strategy and generation in negotiation dialogues. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 2333–2343, 2018.

[61] Clara E Hill. *Helping skills: Facilitating, exploration, insight, and action.* American Psychological Association, 2009.

[62] Mahshid Hosseini and Cornelia Caragea. Distilling knowledge for empathy detection. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, *Findings of the Association for Computational Linguistics: EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 16-20 November, 2021*, pages 3713–3724. Association for Computational Linguistics, 2021.

[63] Ehsan Hosseini-Asl, Bryan McCann, Chien-Sheng Wu, Semih Yavuz, and Richard Socher. A simple language model for task-oriented dialogue. *Advances in Neural Information Processing Systems*, 33:20179–20191, 2020.

[64] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *The Tenth International Conference on Learning Representations*, 2022.

[65] Mary H. McCaulley Isabel Briggs Myers. *A Guide to the Development and Use of the Myers-Briggs Type Indicator: Manual.* Consulting Psychologists Press, 1985.

[66] Etsuko Ishii, Genta Indra Winata, Samuel Cahyawijaya, Divesh Lala, Tatsuya Kawahara, and Pascale Fung. ERICA: an empathetic android companion for covid-19 quarantine. In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 257–260. Association for Computational Linguistics, 2021.

[67] Anubha Jain, Shraddha Kumar, Neha Gupta, Shivangi Khattar, Sanjula Gupta, Deepannita Misra, and Akash Saxena. Accessibility and availability of emotional support services in india: A thorough review of online working models. *International Journal of Indian Psychology*, 10(2), 2022.

[68] Pegah Jandaghi, XiangHai Sheng, Xinyi Bai, Jay Pujara, and Hakim Sidahmed. Faithful persona-based conversational dataset generation with large language models. *arXiv preprint arXiv:2312.10007*, 2023.

[69] Jiun-Hao Jhan, Chao-Peng Liu, Shyh-Kang Jeng, and Hung-Yi Lee. CheerBots: Chatbots toward empathy and emotion using reinforcement learning. *arXiv preprint arXiv:2110.03949*, 2021.

[70] 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. Mistral 7b. *arXiv preprint arXiv:2310.06825*, 2023.

[71] Bowen Jiang, Zhuoqun Hao, Young-Min Cho, Bryan Li, Yuan Yuan, Sihao Chen, Lyle Ungar, Camillo J Taylor, and Dan Roth. Know me, respond to me: Benchmarking llms for dynamic user profiling and personalized responses at scale. *arXiv preprint arXiv:2504.14225*, 2025.

[72] Guangyuan Jiang, Manjie Xu, Song-Chun Zhu, Wenjuan Han, Chi Zhang, and Yixin Zhu. Evaluating and inducing personality in pre-trained language models. *arXiv preprint arXiv:2206.07550*, 2023.

[73] Rishabh Joshi, Vidhisha Balachandran, Shikhar Vashishth, Alan Black, and Yulia Tsvetkov. DialoGraph: Incorporating interpretable strategy-graph networks into negotiation dialogues. In *International Conference on Learning Representations*, 2021.

[74] Dongjin Kang, Sunghwan Kim, Taeyoon Kwon, Seungjun Moon, Hyunsouk Cho, Youngjae Yu, Dongha Lee, and Jinyoung Yeo. Can large language models be good emotional supporter? Mitigating preference bias on emotional support conversation. *arXiv preprint arXiv:2402.13211*, 2024.

[75] Hyunwoo Kim, Byeongchang Kim, and Gunhee Kim. Will I sound like me? Improving persona consistency in dialogues through pragmatic self-consciousness. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu, editors, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 904–916. Association for Computational Linguistics, 2020.

[76] Hyunwoo Kim, Byeongchang Kim, and Gunhee Kim. Perspective-taking and pragmatics for generating empathetic responses focused on emotion causes. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 2227–2240. Association for Computational Linguistics, 2021.

[77] Lucio La Cava, Davide Costa, and Andrea Tagarelli. Open models, closed minds? On agents capabilities in mimicking human personalities through open large language models. *arXiv preprint arXiv:2401.07115*, 2024.

[78] Catherine Penny Hinson Langford, Juanita Bowsher, Joseph P Maloney, and Patricia P Lillis. Social support: a conceptual analysis. *Journal of Advanced Nursing*, 25(1):95–100, 1997.

[79] Alon Lavie and Abhaya Agarwal. METEOR: an automatic metric for MT evaluation with high levels of correlation with human judgments. In Chris Callison-Burch, Philipp Koehn, Cameron S. Fordyce, and Christof Monz, editors, *Proceedings of the Second Workshop on Statistical Machine Translation*, pages 228–231. Association for Computational Linguistics, 2007.

[80] Yoon Kyung Lee, Jina Suh, Hongli Zhan, Junyi Jessy Li, and Desmond C. Ong. Large language models produce responses perceived to be empathic. *CoRR*, abs/2403.18148, 2024.

[81] Young-Jun Lee, Chae-Gyun Lim, Yunsu Choi, Ji-Hui Lm, and Ho-Jin Choi. PersonaChatGen: Generating personalized dialogues using GPT-3. In *Proceedings of the 1st Workshop on Customized Chat Grounding Persona and Knowledge*, pages 29–48. Association for Computational Linguistics, 2022.

[82] Chak Tou Leong, Yi Cheng, Jiashuo Wang, Jian Wang, and Wenjie Li. Self-detoxifying language models via toxification reversal. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 4433–4449. Association for Computational Linguistics, 2023.

[83] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, 2020.

[84] Mike Lewis, Denis Yarats, Yann Dauphin, Devi Parikh, and Dhruv Batra. Deal or no deal? end-to-end learning of negotiation dialogues. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 2443–2453. Association for Computational Linguistics, 2017.

[85] Cheng Li, Ziang Leng, Chenxi Yan, Junyi Shen, Hao Wang, Weishi Mi, Yaying Fei, Xiaoyang Feng, Song Yan, HaoSheng Wang, et al. Chatharuhi: Reviving anime character in reality via large language model. *arXiv preprint arXiv:2308.09597*, 2023.

[86] Guohao Li, Hasan Abed Al Kader Hammoud, Hani Itani, Dmitrii Khizbullin, and Bernard Ghanem. Camel: Communicative agents for "mind"' exploration of large scale language model society. *arXiv preprint arXiv:2303.17760*, 2023.

[87] Jiwei Li, Michel Galley, Chris Brockett, Georgios P. Spithourakis, Jianfeng Gao, and William B. Dolan. A persona-based neural conversation model. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics, Volume 1: Long Papers*. The Association for Computer Linguistics, 2016.

[88] Qintong Li, Hongshen Chen, Zhaochun Ren, Pengjie Ren, Zhaopeng Tu, and Zhumin Chen. EmpDG: Multi-resolution interactive empathetic dialogue generation. In *Proceedings of the International Conference on Computational Linguistics*, pages 4454–4466, 2020.

[89] Qintong Li, Piji Li, Zhaochun Ren, Pengjie Ren, and Zhumin Chen. Knowledge bridging for empathetic dialogue generation. *arXiv e-prints*, pages arXiv–2009, 2020.

[90] Qintong Li, Piji Li, Zhaochun Ren, Pengjie Ren, and Zhumin Chen. Knowledge bridging for empathetic dialogue generation. In *Proceedings of the AAAI Conference on Artificial Intelligence. To appear*, 2022.

[91] Xiujun Li, Yun-Nung Chen, Lihong Li, Jianfeng Gao, and Asli Celikyilmaz. End-to-end task-completion neural dialogue systems. *arXiv preprint arXiv:1703.01008*, 2017.

[92] Yanran Li, Ke Li, Hongke Ning, Xiaoqiang Xia, Yalong Guo, Chen Wei, Jianwei Cui, and Bin Wang. Towards an online empathetic chatbot with emotion causes.

In Fernando Diaz, Chirag Shah, Torsten Suel, Pablo Castells, Rosie Jones, and Tetsuya Sakai, editors, *The International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2041–2045. ACM, 2021.

[93] Yu Li, Kun Qian, Weiyan Shi, and Zhou Yu. End-to-end trainable non-collaborative dialog system. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence*, pages 8293–8302. AAAI Press, 2020.

[94] Yuanchun Li, Hao Wen, Weijun Wang, Xiangyu Li, Yizhen Yuan, Guohong Liu, Jiacheng Liu, Wenxing Xu, Xiang Wang, Yi Sun, et al. Personal LLM agents: Insights and survey about the capability, efficiency and security. *arXiv preprint arXiv:2401.05459*, 2024.

[95] Zaijing Li, Gongwei Chen, Rui Shao, Dongmei Jiang, and Liqiang Nie. Enhancing the emotional generation capability of large language models via emotional chain-of-thought. *CoRR*, abs/2401.06836, 2024.

[96] Rongzhong Lian, Min Xie, Fan Wang, Jinhua Peng, and Hua Wu. Learning to select knowledge for response generation in dialog systems. *arXiv preprint arXiv:1902.04911*, 2019.

[97] Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, 2004.

[98] Zhaojiang Lin, Zihan Liu, Genta Indra Winata, Samuel Cahyawijaya, Andrea Madotto, Yejin Bang, Etsuko Ishii, and Pascale Fung. XPersona: Evaluating multilingual personalized chatbot. In Alexandros Papangelis, Paweł Budzianowski, Bing Liu, Elnaz Nouri, Abhinav Rastogi, and Yun-Nung Chen, editors, *Proceedings of the 3rd Workshop on Natural Language Processing for Conversational AI*, pages 102–112, Online, 2021. Association for Computational Linguistics.

[99] Zhaojiang Lin, Andrea Madotto, Jamin Shin, Peng Xu, and Pascale Fung. MoEL: Mixture of empathetic listeners. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 121–132, 2019.

[100] Zhaojiang Lin, Peng Xu, Genta Indra Winata, Farhad Bin Siddique, Zihan Liu, Jamin Shin, and Pascale Fung. Caire: An end-to-end empathetic chatbot. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 13622–13623, 2020.

[101] Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, et al. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*, 2024.

[102] Alisa Liu, Maarten Sap, Ximing Lu, Swabha Swayamdipta, Chandra Bhagavatula, Noah A. Smith, and Yejin Choi. DExperts: Decoding-time controlled text generation with experts and anti-experts. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 6691–6706. Association for Computational Linguistics, 2021.

[103] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *arXiv preprint arXiv:2107.13586*, 2021.

[104] Qian Liu, Yihong Chen, Bei Chen, Jian-Guang Lou, Zixuan Chen, Bin Zhou, and Dongmei Zhang. You impress me: Dialogue generation via mutual persona perception. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault, editors, *Proceedings of the Annual Meeting of the Association for Computational Linguistics,*, pages 1417–1427. Association for Computational Linguistics, 2020.

[105] Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. 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)*, pages 3469–3483, 2021.

[106] Wenge Liu, Yi Cheng, Hao Wang, Jianheng Tang, Yafei Liu, Ruihui Zhao, Wenjie Li, Yefeng Zheng, and Xiaodan Liang. "my nose is running." "are you also coughing?": Building A medical diagnosis agent with interpretable inquiry logics. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, pages 4266–4272, 2022.

[107] Yifan Liu, Wei Wei, Jiayi Liu, Xianling Mao, Rui Fang, and Dangyang Chen. Improving personality consistency in conversation by persona extending. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, page 1350–1359. Association for Computing Machinery, 2022.

[108] Yuhan Liu, Jiachen Du, Xiang Li, and Ruifeng Xu. Generating empathetic responses by injecting anticipated emotion. *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 7403–7407, 2021.

[109] Yuhan Liu, Jun Gao, Jiachen Du, Lanjun Zhou, and Ruifeng Xu. Empathetic response generation with state management. *arXiv preprint arXiv:2205.03676*, 2022.

[110] Ilya Loshchilov and Frank Hutter. Fixing weight decay regularization in Adam. *arXiv preprint arXiv:1711.05101*, 2018.

[111] Renze Lou, Kai Zhang, and Wenpeng Yin. Is prompt all you need? No. A comprehensive and broader view of instruction learning. *arXiv preprint arXiv:2303.10475*, 2023.

[112] Ximing Lu, Faeze Brahman, Peter West, Jaehun Jung, Khyathi Chandu, Abhilasha Ravichander, Prithviraj Ammanabrolu, Liwei Jiang, Sahana Ramnath, Nouha Dziri, Jillian Fisher, Bill Lin, Skyler Hallinan, Lianhui Qin, Xiang Ren, Sean Welleck, and Yejin Choi. Inference-time policy adapters (IPA): Tailoring extreme-scale LMs without fine-tuning. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 6863–6883. Association for Computational Linguistics, 2023.

[113] Ximing Lu, Sean Welleck, Peter West, Liwei Jiang, Jungo Kasai, Daniel Khashabi, Ronan Le Bras, Lianhui Qin, Youngjae Yu, Rowan Zellers, et al. Neurologic A*esque decoding: Constrained text generation with lookahead heuristics. *arXiv preprint arXiv:2112.08726*, 2021.

[114] Zhengyi Ma, Zhicheng Dou, Yutao Zhu, Hanxun Zhong, and Ji-Rong Wen. One chatbot per person: Creating personalized chatbots based on implicit user profiles. In *The International ACM SIGIR Conference on Research and Development in Information Retrieva*, pages 555–564. ACM, 2021.

[115] Andrea Madotto, Zhaojiang Lin, Chien-Sheng Wu, and Pascale Fung. Personalizing dialogue agents via meta-learning. In Anna Korhonen, David Traum, and Lluís Màrquez, editors, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5454–5459. Association for Computational Linguistics, 2019.

[116] Navonil Majumder, Pengfei Hong, Shanshan Peng, Jiankun Lu, Deepanway Ghosal, Alexander Gelbukh, Rada Mihalcea, and Soujanya Poria. MIME: MIMicking emotions for empathetic response generation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 807–819, 2020.

[117] Pierre-Emmanuel Mazaré, Samuel Humeau, Martin Raison, and Antoine Bordes. Training millions of personalized dialogue agents. In Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii, editors, *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 2775–2779. Association for Computational Linguistics, 2018.

[118] Lenin Medeiros and Tibor Bosse. Using crowdsourcing for the development of online emotional support agents. *International Conference on Practical Applications of Agents, Multi-Agent Systems*, pages 196–209, 2018.

[119] LlaMA Team Meta AI. Introducing Meta Llama 3: The most capable openly available LLM to date. https://ai.meta.com/blog/meta-llama-3/, 2024.

[120] Alexander Miller, Adam Fisch, Jesse Dodge, Amir-Hossein Karimi, Antoine Bordes, and Jason Weston. Key-value memory networks for directly reading documents. *arXiv preprint arXiv:1606.03126*, 2016.

[121] Saif Mohammad. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 174–184. Association for Computational Linguistics, July 2018.

[122] Theodore M Newcomb. The prediction of interpersonal attraction. *American psychologist*, 11(11):575, 1956.

[123] OpenAI. New and improved embedding model. https://openai.com/blog/new-and-improved-embedding-model, 2022.

[124] OpenAI. ChatGPT (November 6 version) [Large language model]. https://chat.openai.com/chat, 2023.

[125] Charles Egerton Osgood, George J Suci, and Percy H Tannenbaum. *The measurement of meaning*. University of Illinois press, 1957.

[126] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 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, 2002.

[127] Joon Sung Park, Joseph C. O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior. In Sean Follmer, Jeff Han, Jürgen Steimle, and Nathalie Henry Riche, editors, *Proceedings of the Annual Symposium on User Interface Software and Technology*, pages 2:1–2:22. ACM, 2023.

[128] Judea Pearl. Heuristics: Intelligent search strategies for computer problem solving. *The Addison-Wesley Series in Artificial Intelligence*, 1985.

[129] Baolin Peng, Michel Galley, Pengcheng He, Chris Brockett, Lars Liden, Elnaz Nouri, Zhou Yu, Bill Dolan, and Jianfeng Gao. Godel: Large-scale pre-training for goal-directed dialog. *arXiv preprint arXiv:2206.11309*, 2022.

[130] Wei Peng, Yue Hu, Luxi Xing, Yuqiang Xie, Yajing Sun, and Yunpeng Li. Control globally, understand locally: A global-to-local hierarchical graph network for emotional support conversation. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence*, 2022.

[131] Richard E Petty, John T Cacioppo, Richard E Petty, and John T Cacioppo. The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, 19:123–205, 1986.

[132] Soujanya Poria, Navonil Majumder, Devamanyu Hazarika, Deepanway Ghosal, Rishabh Bhardwaj, Samson Yu Bai Jian, Pengfei Hong, Romila Ghosh, Abhinaba Roy, Niyati Chhaya, et al. Recognizing emotion cause in conversations. *arXiv preprint arXiv:2012.11820*, 2020.

[133] Yunchen Pu, Zhe Gan, Ricardo Henao, Xin Yuan, Chunyuan Li, Andrew Stevens, and Lawrence Carin. Variational autoencoder for deep learning of images, labels and captions. *Advances in neural information processing systems*, 29, 2016.

[134] Qiao Qian, Minlie Huang, Haizhou Zhao, Jingfang Xu, and Xiaoyan Zhu. Assigning personality/profile to a chatting machine for coherent conversation generation. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, pages 4279–4285. International Joint Conferences on Artificial Intelligence Organization, 2018.

[135] Yushan Qian, Weinan Zhang, and Ting Liu. Harnessing the power of large language models for empathetic response generation: Empirical investigations and improvements. In *Findings of the Association for Computational Linguistics: EMNLP*, pages 6516–6528. Association for Computational Linguistics, 2023.

[136] Jinghui Qin, Zheng Ye, Jianheng Tang, and Xiaodan Liang. Dynamic knowledge routing network for target-guided open-domain conversation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8657–8664, 2020.

[137] Lianhui Qin, Sean Welleck, Daniel Khashabi, and Yejin Choi. Cold decoding: Energy-based constrained text generation with langevin dynamics. *Advances in Neural Information Processing Systems*, 35:9538–9551, 2022.

[138] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. Improving language understanding by generative pre-training. *San Francisco, CA, USA*, 2018.

[139] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

[140] Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems*, 36, 2024.

[141] Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. Towards empathetic open-domain conversation models: A new benchmark and dataset. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5370–5381, 2019.

[142] Ehud Reiter and Robert Dale. Building applied natural language generation systems. *Natural Language Engineering*, 3(1):57–87, 1997.

[143] Verena Rieser and Johanna D Moore. Implications for generating clarification requests in task-oriented dialogues. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pages 239–246, 2005.

[144] Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Eric Michael Smith, Y-Lan Boureau, and Jason Weston. Recipes for building an open-domain chatbot. In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 300–325. Association for Computational Linguistics, April 2021.

[145] Peter J Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.

[146] James A Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161, 1980.

[147] James A Russell and Albert Mehrabian. Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 1977.

[148] Sahand Sabour, Chujie Zheng, and Minlie Huang. CEM: Commonsense-aware empathetic response generation. In *Proceedings of the AAAI Conference on Artificial Intelligence. To appear*, 2022.

[149] Pericle Salvini, Cecilia Laschi, and Paolo Dario. Design for acceptability: Improving robots' coexistence in human society. *International journal of social robotics*, 2:451–460, 2010.

[150] Azlaan Mustafa Samad, Kshitij Mishra, Mauajama Firdaus, and Asif Ekbal. Empathetic persuasion: Reinforcing empathy and persuasiveness in dialogue systems. In *Findings of the Association for Computational Linguistics: NAACL*, pages 844–856, 2022.

[151] Abraham Sanders, Tomek Strzalkowski, Mei Si, Albert Chang, Deepanshu Dey, Jonas Braasch, and Dakuo Wang. Towards a progression-aware autonomous dialogue agent. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1194–1212, 2022.

[152] Susanne G Scott and Reginald A Bruce. Decision-making style: The development and assessment of a new measure. *Educational and psychological measurement*, 55(5):818–831, 1995.

[153] Yunfan Shao, Linyang Li, Junqi Dai, and Xipeng Qiu. Character-LLM: A trainable agent for role-playing. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 13153–13187. Association for Computational Linguistics, 2023.

[154] Ashish Sharma, Inna W Lin, Adam S Miner, David C Atkins, and Tim Althoff. Towards facilitating empathic conversations in online mental health support:

A reinforcement learning approach. In *Proceedings of the World Wide Web Conference*, pages 194–205, 2021.

[155] Ashish Sharma, Adam Miner, David Atkins, and Tim Althoff. 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*. Association for Computational Linguistics, 2020.

[156] Lei Shen, Jinchao Zhang, Jiao Ou, Xiaofang Zhao, and Jie Zhou. Constructing emotional consensus and utilizing unpaired data for empathetic dialogue generation. In *Findings of the Association for Computational Linguistics: EMNLP*, pages 3124–3134. Association for Computational Linguistics, 2021.

[157] Weiyan Shi, Yu Li, Saurav Sahay, and Zhou Yu. Refine and imitate: Reducing repetition and inconsistency in persuasion dialogues via reinforcement learning and human demonstration. In *Findings of the Association for Computational Linguistics: EMNLP*, pages 3478–3492. Association for Computational Linguistics, 2021.

[158] Jamin Shin, Peng Xu, Andrea Madotto, and Pascale Fung. HappyBot: Generating empathetic dialogue responses by improving user experience look-ahead. *CoRR*, abs/1906.08487, 2019.

[159] Jamin Shin, Peng Xu, Andrea Madotto, and Pascale Fung. Generating empathetic responses by looking ahead the user's sentiment. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 7989–7993. IEEE, 2020.

[160] Kurt Shuster, Jing Xu, Mojtaba Komeili, Da Ju, Eric Michael Smith, Stephen Roller, Megan Ung, Moya Chen, Kushal Arora, Joshua Lane, et al. Blenderbot 3: a deployed conversational agent that continually learns to responsibly engage. *arXiv preprint arXiv:2208.03188*, 2022.

[161] Ved Sirdeshmukh, Kaustubh Deshpande, Johannes Mols, Lifeng Jin, Ed-Yeremai Cardona, Dean Lee, Jeremy Kritz, Willow Primack, Summer Yue, and Chen Xing. Multichallenge: A realistic multi-turn conversation evaluation benchmark challenging to frontier llms. *arXiv preprint arXiv:2501.17399*, 2025.

[162] Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu. MPNet: Masked and permuted pre-training for language understanding. *Advances in Neural Information Processing Systems*, 33:16857–16867, 2020.

[163] Heydar Soudani, Roxana Petcu, Evangelos Kanoulas, and Faegheh Hasibi. A survey on recent advances in conversational data generation. *arXiv preprint arXiv:2405.13003*, 2024.

[164] Robyn Speer, Joshua Chin, and Catherine Havasi. ConceptNet 5.5: An open multilingual graph of general knowledge. In Satinder Singh and Shaul Markovitch, editors, *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 4444–4451. AAAI Press, 2017.

[165] Ekaterina Svikhnushina, Iuliana Voinea, Anuradha Welivita, and Pearl Pu. A taxonomy of empathetic questions in social dialogs. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2952–2973. Association for Computational Linguistics, 2022.

[166] Jianheng Tang, Tiancheng Zhao, Chenyan Xiong, Xiaodan Liang, Eric P Xing, and Zhiting Hu. Target-guided open-domain conversation. *arXiv preprint arXiv:1905.11553*, 2019.

[167] Rohan Taori, Ishaan Gulrajani, Tianyi Zhang, Yann Dubois, Xuechen Li, Carlos Guestrin, Percy Liang, and Tatsunori B Hashimoto. Stanford Alpaca: An instruction-following llama model. https://github.com/gururise/AlpacaDataCleaned, 2023.

[168] Gemma Team, Morgane Riviere, Shreya Pathak, Pier Giuseppe Sessa, Cassidy Hardin, Surya Bhupatiraju, Léonard Hussenot, Thomas Mesnard, Bobak Shahriari, Alexandre Ramé, et al. Gemma 2: Improving open language models at a practical size. *arXiv preprint arXiv:2408.00118*, 2024.

[169] Romal Thoppilan, Daniel De Freitas, Jamie Hall, Noam Shazeer, Apoorv Kulshreshtha, Heng-Tze Cheng, Alicia Jin, Taylor Bos, Leslie Baker, Yu Du, et al. Lamda: Language models for dialog applications. *arXiv preprint arXiv:2201.08239*, 2022.

[170] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*, 2023.

[171] Nhat Tran, Malihe Alikhani, and Diane Litman. How to ask for donations? learning user-specific persuasive dialogue policies through online interactions. In *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*, pages 12–22, 2022.

[172] Quan Tu, Chuanqi Chen, Jinpeng Li, Yanran Li, Shuo Shang, Dongyan Zhao, Ran Wang, and Rui Yan. CharacterChat: Learning towards conversational ai with personalized social support. *arXiv preprint arXiv:2308.10278*, 2023.

[173] Quan Tu, Yanran Li, Jianwei Cui, Bin Wang, Ji-Rong Wen, and Rui Yan. MISC: A mixed strategy-aware model integrating COMET for emotional support conversation. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 308–319. Association for Computational Linguistics, 2022.

[174] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.

[175] Ramakrishna Vedantam, C. Lawrence Zitnick, and Devi Parikh. CIDEr: Consensus-based image description evaluation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4566–4575. IEEE Computer Society, 2015.

[176] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *6th International Conference on Learning Representations*. OpenReview.net, 2018.

[177] Richard S Wallace. The anatomy of alice. *Parsing the Turing Test*, 2009.

[178] Harald G Wallbott and Klaus R Scherer. How universal and specific is emotional experience? evidence from 27 countries on five continents. *Social science information*, 25(4):763–795, 1986.

[179] Guanzhi Wang, Yuqi Xie, Yunfan Jiang, Ajay Mandlekar, Chaowei Xiao, Yuke Zhu, Linxi Fan, and Anima Anandkumar. Voyager: An open-ended embodied agent with large language models. *arXiv preprint arXiv:2305.16291*, 2023.

[180] Jian Wang, Yi Cheng, Dongding Lin, Chak Leong, and Wenjie Li. Target-oriented proactive dialogue systems with personalization: Problem formulation and dataset curation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1132–1143. Association for Computational Linguistics, 2023.

[181] Jian Wang, Dongding Lin, and Wenjie Li. Dialogue planning via brownian bridge stochastic process for goal-directed proactive dialogue. *arXiv preprint arXiv:2305.05290*, 2023.

[182] Jiashuo Wang, Wenjie Li, Peiqin Lin, and Feiteng Mu. Empathetic response generation through graph-based multi-hop reasoning on emotional causality. *Knowledge-Based Systems*, 233:107547, 2021.

[183] Xuewei Wang, Weiyan Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, and Zhou Yu. Persuasion for good: Towards a personalized persuasive dialogue system for social good. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5635–5649. Association for Computational Linguistics, 2019.

[184] Zekun Moore Wang, Zhongyuan Peng, Haoran Que, Jiaheng Liu, Wangchunshu Zhou, Yuhan Wu, Hongcheng Guo, Ruitong Gan, Zehao Ni, Man Zhang, et al. RoleLLM: Benchmarking, eliciting, and enhancing role-playing abilities of large language models. *arXiv preprint arXiv:2310.00746*, 2023.

[185] Jane M Webber and J Mascari. *Disaster mental health counseling: A guide to preparing and responding*. American Counseling Association Foundation, 2018.

[186] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35:24824–24837, 2022.

[187] Joseph Weizenbaum. Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45, 1966.

[188] Anuradha Welivita and Pearl Pu. A taxonomy of empathetic response intents in human social conversations. In *Proceedings of the International Conference on Computational Linguistics*, pages 4886–4899. International Committee on Computational Linguistics, December 2020.

[189] Tsung-Hsien Wen, David Vandyke, Nikola Mrkšić, Milica Gašić, Lina M. Rojas-Barahona, Pei-Hao Su, Stefan Ultes, and Steve Young. A network-based end-to-end trainable task-oriented dialogue system. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 438–449, Valencia, Spain, April 2017. Association for Computational Linguistics.

[190] Alex Wilf, Sihyun Shawn Lee, Paul Pu Liang, and Louis-Philippe Morency. Think twice: Perspective-taking improves large language models' theory-of-mind capabilities. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2024, Bangkok, Thailand, August 11-16, 2024*, pages 8292–8308. Association for Computational Linguistics, 2024.

[191] Jason D Williams. Web-style ranking and slu combination for dialog state tracking. In *Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, pages 282–291, 2014.

[192] Ludwig Wittgenstein. Philosophical investigations. *Macmillan*, 1953.

[193] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, and Jamie Brew. HuggingFace's Transformers: State-of-the-art natural language processing. *CoRR*, abs/1910.03771, 2019.

[194] Chien-Sheng Wu, Andrea Madotto, Zhaojiang Lin, Peng Xu, and Pascale Fung. Getting to know you: User attribute extraction from dialogues. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 581–589. European Language Resources Association, 2020.

[195] Qingyang Wu, Yichi Zhang, Yu Li, and Zhou Yu. Alternating recurrent dialog model with large-scale pre-trained language models. In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1292–1301, 2021.

[196] Shujin Wu, Yi R Fung, Cheng Qian, Jeonghwan Kim, Dilek Hakkani-Tur, and Heng Ji. Aligning llms with individual preferences via interaction. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 7648–7662, 2025.

[197] Yuwei Wu, Xuezhe Ma, and Diyi Yang. Personalized response generation via generative split memory network. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1956–1970, 2021.

[198] Zixiu Wu, Rim Helaoui, Diego Reforgiato Recupero, and Daniele Riboni. Towards low-resource real-time assessment of empathy in counselling. In *Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access*, pages 204–216, Online, June 2021. Association for Computational Linguistics.

[199] Rui Xia, Mengran Zhang, and Zixiang Ding. RTHN: A RNN-Transformer hierarchical network for emotion cause extraction. In Sarit Kraus, editor, *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 5285–5291. ijcai.org, 2019.

[200] Yang Xiao, Yi Cheng, Jinlan Fu, Jiashuo Wang, Wenjie Li, and Pengfei Liu. How far are we from believable AI agents? A framework for evaluating the believability of human behavior simulation. *arXiv preprint arXiv:2312.17115*, 2023.

[201] Yubo Xie and Pearl Pu. Empathetic dialog generation with fine-grained intents. In *Proceedings of the 25th Conference on Computational Natural Language Learning*, pages 133–147, Online, November 2021. Association for Computational Linguistics.

[202] Canwen Xu, Daya Guo, Nan Duan, and Julian McAuley. Baize: An open-source chat model with parameter-efficient tuning on self-chat data. *arXiv preprint arXiv:2304.01196*, 2023.

[203] Rui Xu, Xintao Wang, Jiangjie Chen, Siyu Yuan, Xinfeng Yuan, Jiaqing Liang, Zulong Chen, Xiaoqing Dong, and Yanghua Xiao. Character is destiny: Can role-playing language agents make persona-driven decisions? *arXiv preprint arXiv:2404.12138*, 2024.

[204] Xiaohan Xu, Xuying Meng, and Yequan Wang. Poke: Prior knowledge enhanced emotional support conversation with latent variable. *arXiv preprint arXiv:2210.12640*, 2022.

[205] Haoqiu Yan, Yongxin Zhu, Kai Zheng, Bing Liu, Haoyu Cao, Deqiang Jiang, and Linli Xu. Talk with human-like agents: Empathetic dialogue through perceptible acoustic reception and reaction. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15009–15022. Association for Computational Linguistics, 2024.

[206] Zhao Yan, Nan Duan, Peng Chen, Ming Zhou, Jianshe Zhou, and Zhoujun Li. Building task-oriented dialogue systems for online shopping. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.

[207] An Yang, Baosong Yang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Zhou, Chengpeng Li, Chengyuan Li, Dayiheng Liu, Fei Huang, et al. Qwen2 technical report. *arXiv preprint arXiv:2407.10671*, 2024.

[208] Runzhe Yang, Jingxiao Chen, and Karthik Narasimhan. Improving dialog systems for negotiation with personality modeling. 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)*, pages 681–693. Association for Computational Linguistics, 2021.

[209] Zihao Yi, Jiarui Ouyang, Yuwen Liu, Tianhao Liao, Zhe Xu, and Ying Shen. A survey on recent advances in llm-based multi-turn dialogue systems. *arXiv preprint arXiv:2402.18013*, 2024.

[210] Steve Young, Milica Gašić, Simon Keizer, François Mairesse, Jost Schatzmann, Blaise Thomson, and Kai Yu. The hidden information state model: A practical framework for pomdp-based spoken dialogue management. *Computer Speech & Language*, 24(2):150–174, 2010.

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

[212] Xiaoyan Yu, Tongxu Luo, Yifan Wei, Fangyu Lei, Yiming Huang, Hao Peng, and Liehuang Zhu. Neeko: Leveraging dynamic LoRA for efficient multi-character role-playing agent. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 12540–12557. Association for Computational Linguistics, 2024.

[213] Chengkun Zeng, Guanyi Chen, Chenghua Lin, Ruizhe Li, and Zhigang Chen. Affective decoding for empathetic response generation. In *Proceedings of the International Conference on Natural Language Generation*, pages 331–340, 2021.

[214] Haolan Zhan, Yufei Wang, Tao Feng, Yuncheng Hua, Suraj Sharma, Zhuang Li, Lizhen Qu, and Gholamreza Haffari. Let's negotiate! a survey of negotiation dialogue systems. *arXiv preprint arXiv:2212.09072*, 2022.

[215] Kai Zhang, Yangyang Kang, Fubang Zhao, and Xiaozhong Liu. Llm-based medical assistant personalization with short-and long-term memory coordination. *arXiv preprint arXiv:2309.11696*, 2023.

[216] Kai Zhang, Yejin Kim, and Xiaozhong Liu. Personalized llm response generation with parameterized memory injection. *arXiv preprint arXiv:2404.03565*, 2024.

[217] Peiyuan Zhang, Guangtao Zeng, Tianduo Wang, and Wei Lu. TinyLlama: An open-source small language model. *arXiv preprint arXiv:2401.02385*, 2024.

[218] Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. Personalizing dialogue agents: I have a dog, do you have pets too? In Iryna Gurevych and Yusuke Miyao, editors, *Proceedings of the Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2204–2213. Association for Computational Linguistics, 2018.

[219] Tenggan Zhang, Xinjie Zhang, Jinming Zhao, Li Zhou, and Qin Jin. Escot: Towards interpretable emotional support dialogue systems. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2024, Bangkok, Thailand, August 11-16, 2024*, pages 13395–13412. Association for Computational Linguistics, 2024.

[220] Tong Zhang, Chen Huang, Yang Deng, Hongru Liang, Jia Liu, Zujie Wen, Wenqiang Lei, and Tat-Seng Chua. Strength lies in differences! Improving strategy planning for non-collaborative dialogues via diversified user simulation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 424–444, 2024.

[221] Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and William B Dolan. DialoGPT: Large-scale generative pre-training for conversational response generation. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 270–278, 2020.

[222] Yue Zhang, Yafu Li, Leyang Cui, Deng Cai, Lemao Liu, Tingchen Fu, Xinting Huang, Enbo Zhao, Yu Zhang, Yulong Chen, et al. Siren's song in the ai ocean: A survey on hallucination in large language models. *arXiv preprint arXiv:2309.01219*, 2023.

[223] Siyan Zhao, Mingyi Hong, Yang Liu, Devamanyu Hazarika, and Kaixiang Lin. Do llms recognize your preferences? evaluating personalized preference following in llms. *arXiv preprint arXiv:2502.09597*, 2025.

[224] Tiancheng Zhao and Maxine Eskenazi. Towards end-to-end learning for dialog state tracking and management using deep reinforcement learning. *arXiv preprint arXiv:1606.02560*, 2016.

[225] Tiancheng Zhao, Ran Zhao, and Maxine Eskenazi. Learning discourse-level diversity for neural dialog models using conditional variational autoencoders. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 654–664, 2017.

[226] Weixiang Zhao, Yanyan Zhao, Xin Lu, and Bing Qin. Don't lose yourself! empathetic response generation via explicit self-other awareness. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 13331–13344. Association for Computational Linguistics, 2023.

[227] Weixiang Zhao, Yanyan Zhao, Xin Lu, Shilong Wang, Yanpeng Tong, and Bing Qin. Is chatgpt equipped with emotional dialogue capabilities? *arXiv preprint arXiv:2304.09582*, 2023.

[228] Weixiang Zhao, Yanyan Zhao, Shilong Wang, and Bing Qin. TransESC: Smoothing emotional support conversation via turn-level state transition. In *Findings of the Association for Computational Linguistics: ACL*, pages 6725–6739. Association for Computational Linguistics, 2023.

[229] Chujie Zheng, Yong Liu, Wei Chen, Yongcai Leng, and Minlie Huang. CoMAE: A multi-factor hierarchical framework for empathetic response generation. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 813–824. Association for Computational Linguistics, 2021.

[230] Chujie Zheng, Sahand Sabour, Jiaxin Wen, Zheng Zhang, and Minlie Huang. AugESC: Dialogue augmentation with large language models for emotional support conversation. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 1552–1568. Association for Computational Linguistics, 2023.

[231] Yinhe Zheng, Guanyi Chen, Minlie Huang, Song Liu, and Xuan Zhu. Personalized dialogue generation with diversified traits. *arXiv preprint arXiv:1901.09672*, 2019.

[232] Yinhe Zheng, Rongsheng Zhang, Minlie Huang, and Xiaoxi Mao. A pre-training based personalized dialogue generation model with persona-sparse data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 9693–9700, 2020.

[233] Zhonghua Zheng, Lizi Liao, Yang Deng, and Liqiang Nie. Building emotional support chatbots in the era of llms. *arXiv preprint arXiv:2308.11584*, 2023.

[234] Zhonghua Zheng, Lizi Liao, Yang Deng, Libo Qin, and Liqiang Nie. 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 (Volume 1: Long Papers)*, pages 11325–11345. Association for Computational Linguistics, 2024.

[235] Hanxun Zhong, Zhicheng Dou, Yutao Zhu, Hongjin Qian, and Ji-Rong Wen. Less is more: Learning to refine dialogue history for personalized dialogue generation. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5808–5820. Association for Computational Linguistics, 2022.

[236] Peixiang Zhong, Di Wang, and Chunyan Miao. Knowledge-enriched Transformer for emotion detection in textual conversations. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and the International Joint Conference on Natural Language Processing*, pages 165–176, 2019.

[237] Wanjun Zhong, Lianghong Guo, Qiqi Gao, He Ye, and Yanlin Wang. Memorybank: Enhancing large language models with long-term memory. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 19724–19731, 2024.

[238] Hao Zhou, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu, and Bing Liu. Emotional chatting machine: Emotional conversation generation with internal and external memory. In Sheila A. McIlraith and Kilian Q. Weinberger, editors, *Proceedings of the Conference on Artificial Intelligence*, pages 730–739. AAAI Press, 2018.

[239] Jinfeng Zhou, Zhuang Chen, Dazhen Wan, Bosi Wen, Yi Song, Jifan Yu, Yongkang Huang, Libiao Peng, Jiaming Yang, Xiyao Xiao, et al. Characterglm: Customizing chinese conversational ai characters with large language models. *arXiv preprint arXiv:2311.16832*, 2023.

[240] Jinfeng Zhou, Zhuang Chen, Bo Wang, and Minlie Huang. Facilitating multi-turn emotional support conversation with positive emotion elicitation: A reinforcement learning approach. In *Proceedings of the Annual Meeting of the*

Association for Computational Linguistics (Volume 1: Long Papers), pages 1714–1729. Association for Computational Linguistics, 2023.

[241] Li Zhou, Jianfeng Gao, Di Li, and Heung-Yeung Shum. The design and implementation of xiaoice, an empathetic social chatbot. *CoRR*, abs/1812.08989, 2018.

[242] Yiheng Zhou, He He, Alan W Black, and Yulia Tsvetkov. A dynamic strategy coach for effective negotiation. In *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue*, pages 367–378, 2019.

[243] Yiheng Zhou, Yulia Tsvetkov, Alan W Black, and Zhou Yu. Augmenting non-collaborative dialog systems with explicit semantic and strategic dialog history. In *International Conference on Learning Representations*, 2019.

[244] Yujia Zhou, Qiannan Zhu, Jiajie Jin, and Zhicheng Dou. Cognitive personalized search integrating large language models with an efficient memory mechanism. In *Proceedings of the ACM Web Conference 2024*, pages 1464–1473, 2024.

[245] Lixin Zou, Shengqiang Zhang, Hengyi Cai, Dehong Ma, Suqi Cheng, Shuaiqiang Wang, Daiting Shi, Zhicong Cheng, and Dawei Yin. Pre-trained language model based ranking in baidu search. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 4014–4022, 2021.

[246] Yicheng Zou, Zhihua Liu, Xingwu Hu, and Qi Zhang. Thinking clearly, talking fast: Concept-guided non-autoregressive generation for open-domain dialogue systems. *arXiv preprint arXiv:2109.04084*, 2021.