REVIEW ARTICLE

The future of cognitive strategy-enhanced persuasive dialogue agents: new perspectives and trends

Mengqi CHEN[1], Bin GUO (✉)[1], Hao WANG[1], Haoyu LI[1], Qian ZHAO[1], Jingqi LIU[1], Yasan DING[1], Yan PAN[2], Zhiwen YU[1]

1 School of Computer Science, Northwestern Polytechnical University, Xi'an 710129, China
2 Science and Technology on Information Systems Engineering Laboratory, National University of Defense Technology, Changsha 410073, China

Abstract Persuasion, as one of the crucial abilities in human communication, has garnered extensive attention from researchers within the field of intelligent dialogue systems. Developing dialogue agents that can persuade others to accept certain standpoints is essential to achieving truly intelligent and anthropomorphic dialogue systems. Benefiting from the substantial progress of Large Language Models (LLMs), dialogue agents have acquired an exceptional capability in context understanding and response generation. However, as a typical and complicated cognitive psychological system, persuasive dialogue agents also require knowledge from the domain of cognitive psychology to attain a level of human-like persuasion. Consequently, the cognitive strategy-enhanced persuasive dialogue agent (defined as **CogAgent**), which incorporates cognitive strategies to achieve persuasive targets through conversation, has become a predominant research paradigm. To depict the research trends of CogAgent, in this paper, we first present several fundamental cognitive psychology theories and give the formalized definition of three typical cognitive strategies, including the persuasion strategy, the topic path planning strategy, and the argument structure prediction strategy. Then we propose a new system architecture by incorporating the formalized definition to lay the foundation of CogAgent. Representative works are detailed and investigated according to the combined cognitive strategy, followed by the summary of authoritative benchmarks and evaluation metrics. Finally, we summarize our insights on open issues and future directions of CogAgent for upcoming researchers.

Keywords persuasive dialogue, cognitive strategy, cognitive psychology, persuasion strategy

1 Introduction

Building intelligent human-machine dialogue agents that can conduct natural and engaging conversations with humans is the long-standing goal of artificial intelligence (AI) [1,2]. Moreover, the persuasive ability of dialogue agents has garnered extensive attention from researchers. Persuasion is one of the crucial abilities in human communication. The Elaboration Likelihood Model (ELM) theory [3] suggests that people tend to engage with persuasive messages when communicating with others. It is a prevalent phenomenon for individuals to hold diverse perspectives on a given topic and endeavor to influence others in altering their viewpoints, attitudes, or behaviors through conversational interactions [4,5]. There are massive persuasive scenarios in the real world [6,7], such as persuasion for social good [8], winning debates [9], healing negative emotions [10], and recommending items to users [11]. A persuasive conversion includes two distinct parties, corresponding to persuader and persuadee, respectively [12]. The goal of the persuader is to change the persuadee's viewpoint on a specific topic by combining cognitive strategies, the personality of the persuadee, and other context features [8,13]. The development of intelligent persuasive dialogue agents that can persuade users to accept certain standpoints is emerging as a promising research field [14,15].

Modern dialogue agents have arrived at the era characterized by large language models (LLMs) [16,17]. Driven by an immense scale of parameters and an abundance of training data, dialogue agents (e.g., ChatGPT, LLaMA [18]) have acquired an exceptional capability of context understanding and response generation [19,20], reaching a satisfactory level of fluency, logic, and personalization when conversing with humans [21,22]. However, the persuasive process aims to alter individuals' cognitive perspectives on specific events, individuals, or situations through ongoing communication, rather than mere fluency in language expression. Research on persuasion mechanisms suggests that achieving persuasion requires understanding the cognitive states of the persuadees and employing appropriate cognitive strategies to effectively change their viewpoints in a readily acceptable manner [23–26]. Therefore, designing persuasive dialogue agents necessitates drawing from cognitive

psychology and incorporating cognitive strategies to logically organize the content, logic, and language style of persuasive responses [13,27,28]. There have been significant advancements in persuasive dialogue systems, which primarily enhance persuasiveness from three perspectives: integrating persuasion strategies [8,29,30], planning topic paths [31–33], and extracting argument structures [34–36]. In this paper, we argue that persuasion is a cognitive psychology activity and that the persuasion strategy, the topic path planning strategy, and the argument structure prediction strategy can all be categorized as cognitive strategies. We define cognitive strategy-enhanced persuasive dialogue agent as **CogAgent**.

CogAgent aims to integrate various cognitive strategies to ensure that dialogue contents can effectively influence the persuadee in terms of their perceptions, opinions, or attitudes [8,37]. Figure 1 depicts a persuasive dialogue example. The dialogue agent persuades the user to reduce anxiety from job crises using various persuasion strategies. The social and communicative dynamics behind persuasive dialogue contexts are complex. Effective and successful persuasive dialogue does not mechanically convey target viewpoints to persuadees but rather empathetically addresses persuadees through social and emotional communications [38]. Thus, persuasive dialogues are not strictly task-oriented but are carried around tasks with additional cognitive strategies to build trust and empathy with persuadees, leading to a smooth persuasive process.

As an emerging research area, an in-depth survey of the existing academic efforts is necessary. Duerr et al. [39] broadly review the works that use natural language generation technologies to automatically detect and generate persuasive texts. Zhan et al. [40] concentrate on the negotiation dialogue system, a typical type of persuasive dialogue system, and summarizes corresponding benchmarks, evaluations, and methodologies. Deng et al. [41] provide an overview of the prominent problems and advanced designs in proactive dialogue systems, which treats persuasive dialogue as the subset of the proactive dialogue. Compared with these surveys, we provide a comprehensive review of concepts, challenges, methodologies, and applications in the field of cognitive strategy-enhanced persuasive dialogue. We formalize the definition of cognitive strategies extended from cognitive psychology theory. Based on the formalized concept model and generic system architecture, we summarize representative research in CogAgent from a systematic perspective. Furthermore, benchmarks, evaluation metrics, and thoughts on promising research trends are analyzed to promote the research progress. In conclusion, our contributions are summarized as follows.

- Drawing from cognitive psychology theories, we formalize the definition of cognitive strategies, and present the concept model and generic system architecture of CogAgent, to provide an overall picture for the summary of research works.
- We make a profound investigation of the development in CogAgent by presenting the core contributions of each work. Besides, we also comprehensively summarize available datasets and evaluation metrics.
- We further discuss some open issues and promising research trends in CogAgent, including model adaptivity/generality of CogAgent, multi-party CogAgent, etc., to promote the development of the research community.

The rest of the paper is organized as follows. In Section 2, we first summarize the typical cognitive psychology theories and present the definition of cognitive strategies. Then we formalize the concept model of CogAgent and design a generic system architecture, followed by typical application scenarios of CogAgent. In Section 3, we first introduce the challenges faced by CogAgent and then summarize the key techniques to achieve CogAgent based on the used cognitive strategies. In Section 4, we summarize the available datasets and evaluation metrics, followed by open issues and promising research trends in Section 5.

2 Formalized concept model and system architecture for CogAgent

In this section, we first summarise the typical cognitive psychology theories involved in human conversations, as the theoretical foundation for the design of CogAgent. Then we formalize the concept model for CogAgent and present the generic system architecture to visualize the overall picture in CogAgent.

2.1 The cognitive psychology theory

As a typical cognitive-psychological activity, the persuasion process requires the support of cognitive psychology theories to effectively model the mental changes that people experience during conversations, thus promoting the design of CogAgent. This section summarises typical cognitive psychology theories to inspire subsequent CogAgent researchers.
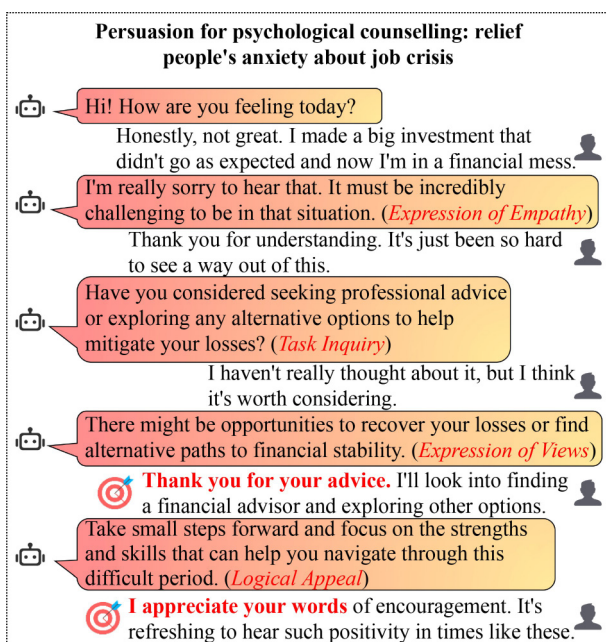

Fig. 1 The persuasive dialogue example, where the dialogue agent persuades the user to relieve for job crisis using various persuasion strategies

2.1.1 Pre-suasion

The concept of Pre-suasion [42], proposed by the renowned authority on persuasion, Robert Cialdini, is a prominent theory in the persuasion field. Pre-suasion means that the success rate of persuasion can be significantly enhanced by attracting the attention of the persuadee through appropriate choices of words and actions before communication or requests are conducted. Pre-suasion emphasizes that the timing of persuasion is as important as persuasive content. When we intend to persuade others to accept our points, we need to consider others' perspectives and organize our conversational arguments at the appropriate time to effectively complete the persuasion process.

2.1.2 Principle of consistency

The principle of consistency suggests that people usually try to maintain consistency based on what they have expressed and the commitments they have made in the past [43]. By planning topic paths, one can think about and define one's opinions and arguments in advance to maintain consistency and increase persuasiveness when communicating with others. The principle of consistency plays an important role in persuading others. Through consistency of statements, the persuadees will recognize that the points raised are consistent with their beliefs or opinions and will effectively increase the effectiveness of persuasion.

2.1.3 Theory of mind

The theory of mind (ToM) [44] suggests that effective questions and answers in communications are based on a shared world of experiences and referents between interlocutors. To communicate effectively, people model both the mental states of their listeners and the effects of their behavior on the world, and then react to and predict the behavior of others. This ability to understand and infer human intentions is defined as a ToM. One way to imitate ToM is to observe others' perspectives in various situations and to derive a set of rules that affect their perspectives and emotions. When the same or highly similar scenarios reoccur, we can make reasonable behavioral or emotional predictions accordingly. Many researchers explicitly model ToM as a concrete cognitive process to ensure that dialogue agents can access potential human psychological states and cognitive processes [45–47].

2.1.4 Rhetoric

Aristotle, one of the earliest masters of the art of persuasion, proposes three basic elements of persuasion: ethos (credibility), pathos (emotions), and logos (logic) in his work, The Philosophy of Rhetoric [48]. These principles serve as a guide to effective persuasive communication. By establishing credibility, appealing to emotions, and applying logical reasoning, one can effectively persuade others to accept his propositions. Aristotle's insights in Rhetoric remain highly influential not only in the field of persuasive dialogue but also in shaping our understanding of aesthetics and related concepts.

Credibility represents the identification of persuaders, including their identity and moral character, which influences the persuasiveness of the speaker. Aristotle in his Rhetoric explains in detail the three elements that affect credibility, namely wisdom, virtue, and goodwill. Wisdom includes elements such as breadth of knowledge, expertise, and authority. Virtue includes elements such as fairness, honesty, and dignity. By demonstrating wisdom, virtue, and goodwill, persuaders can enhance their persuasiveness and foster the trust and reliability of persuadees. Combining essential elements of credibility can greatly enhance the effectiveness of CogAgent. Emotion refers to the expression of sentiments during the persuasion process, thus lowering people's psychological defenses in accepting persuasive content. Aristotle stated that we cannot persuade others through rationality, but can achieve it with emotion. Emotional expressions play an important role in changing the cognitive decisions of others. The use of emotionally charged content and expressions can be more effective in eliciting agreement and empathy from the persuadee. Logic refers to the use of inherent factual logic, causality, or other rational factors in expressions to gain persons' trust and persuade them to change their perceptions. By presenting coherent logical arguments, supported by factual data and authoritative sources, persuaders can establish credibility, gain persuadees' perceptions, and change their opinions.

The cognitive psychology theories, which can be used to model the dynamics of human cognitive psychological status, provide a solid foundation for CogAgent. Under the guidance of cognitive psychology, we can comprehensively investigate and model explicit cognitive factors and strategies that can change users' cognitive psychological states, such as logical expressions and emotional appeals. These cognitive strategies can facilitate CogAgent to understand the psychological state of the persuadee and enhance the persuasiveness of responses from multiple perspectives to achieve more efficient persuasion processes.

2.2 Cognitive strategy

In-depth research on the persuasion process has demonstrated that to achieve effective persuasion, it is necessary to understand the persuadees' cognitive states and combine cognitive strategies to organize persuasive content reasonably [23–26]. Therefore, designing persuasive dialogue agents necessitates drawing from cognitive psychology and incorporating cognitive strategies to logically organize the content, logic, and language style of persuasive responses [13,27,28]. Evolved from cognitive psychological theories, we categorize cognitive strategies into three aspects, **persuasion strategy**, **topic path planning strategy**, and **argument structure prediction strategy**, detailed as follows.

2.2.1 Persuasion strategy

The persuasion strategy aims to influence or change the perceptions, opinions, attitudes, or behaviors of persuadees from a psychological standpoint, through the use of linguistic techniques of expression, such as logical appeal, foot-in-the-door, and self-disclosure [4,8,49]. Based on existing research, we construct a comprehensive and effective set of persuasion strategies that can achieve persuasive goals, inspired by the theory of mind, the rhetoric, and other psychology theories.

We formalize the definitions and examples of expressions of persuasion strategies, as shown in Tables 1 and 2.

Numerous studies [8,29,30,57] have demonstrated that persuasion strategies can effectively enhance the persuasiveness of the dialogue content. How to reasonably select the appropriate strategies according to the dialogue context and the perusadee's psychological state to generate a persuasive dialogue response is crucial to achieving high-quality CogAgent.

2.2.2 Topic path planning strategy

The topic path planning strategy aims to plan the topic transition sequence during the persuasive dialogue process, to ensure the dialogue coherence and the progress of the dialogue towards the persuasive target. The persuasive dialogue agent should smoothly navigate between topics to reduce irrelevant associations of the persuadee and the difficulty of the persuasion process [58,59]. The topic path planning strategy is widely employed in target-guided persuasive dialogue systems [60,61]. Starting from the topic of interest to the persuadee, the persuasive dialogue agent needs to gradually and smoothly transfer the conversation topic to the persuasive target to improve the persuadee's psychological acceptability and ensure the persuasive effect. How to plan the reasonable topic path and generate an in-depth multi-turn persuasive conversation according to the corresponding topics is to be explored.

2.2.3 Argument structure prediction strategy

Argument structure prediction strategy is designed to predict

Table 1 Part of definitions and examples of persuasion strategies

Strategy	Definition	Example
Present of facts [8]	Using factual evidence (e.g., official news reports, statistics) and a credible reasoning process to persuade others	In recent months, the demand for residential properties has become extremely high. The price of residential property has risen almost twenty percent.
Challenges and inquiries [8]	Expressing disbelief or opposition to the other side's viewpoints and providing strong rebuttal evidence to enhance persuasiveness	Really? I don't agree. This Star Wars episode was incredible!
Emotional resonance [8]		Advanced special effects are the main reason for the success of previous episodes, so audiences have high expectations for this one, and I don't think they will be disappointed.
Eliciting anger [8]	Eliciting specific emotions to influence others' attitudes	If that's the case, there's not much point in further discussion. We might as well call the whole deal off.
Eliciting guilt [8]		Come on, you can at least try a little, besides your cigarette.
Self-modeling [8]	Expressing one's intention to act and opting to serve as a role model for the persuadee to emulate	That still leaves a gap of 20 dollars to be covered. Let's meet each other halfway once more, then the gap will be closed and our business completed.
Building trust [15]	Building rapport and psychological trust through a harmonious conversation	I'm glad we've agreed on price. We'll go on to the other terms and conditions at our next meeting.
Courtesy tips [50]	Expressing gratitude, approval, praise, etc. to lower the other party's psychological defenses	I know exactly what you mean. Hearing that song gives me a nostalgic feeling.
Compromise [51]	Expressing concessions on time to avoid being too intense in the guidance process and causing the other party to end the conversation	I think it unwise for either of us to insist on his price. How about meeting each other halfway so that business can be concluded?
Attachment of views [50]	Expressing kindness and concern through active listening and to some extent seconding the other person's point of view	Better late than never.

Table 2 Part of definitions and examples of persuasion strategies

Strategy	Definition	Example
Problem decomposition [52]	Decomposing the ultimate persuasion goal into sub-issues and stepping through the persuasion process	Let me get down some information about your apartment first. what is your property's address?
Social identity [53]	Gaining psychological support from the other person by emphasizing group and identity belonging	I know. I have been a subscriber for the past two years.
Herd mentality [54]	Presenting a viewpoint that is recognized or accepted by the majority of people and persuading the other side to accept it	There was always a good round of applause every time she sang.
Expression of disgust [55]		Oh, my god! I look so old. I look as if I were 40. I think it's time for some plastic surgeries.
Expression of empathy [38]	Expressing a particular point of view or emotion to emphasize the persuasive content	I know, dear. I am too. But we've just been too busy to look for a house.
Expression of views [15]		That means the apartment has furniture in it.
Logical appeal [8]	Enhancing the credibility of persuasive content through the logical and reasoning process	It certainly is. But to tell you the truth, the room is so large that I can share it with someone else, and that will decrease the total amount of the rent.
Task Inquiry [8]	Asking questions related to persuasive goals	That might be going overboard a bit. How about just that scarf with a bracelet?
Personal story [8]	Using narrative examples to illustrate the positive outcomes of your actions to inspire others to follow suit	Yes, I'm sure I've done a lot of house painting in my life. If I got even a tiny drop of paint on her furniture, she would get furious. So I learned to be very picky.
Refutation of objections [56]	Directly refuting the other side's point of view	Not necessary. If we use a realtor to find a house, it will be more expensive.
Greeting [8]	Greeting at the beginning of a dialogue	Hi there! How are you doing today?

persuasive and authoritative argument surrounding the discussed topic, thereby enhancing the credibility of persuasive dialogue contents and convincing the persuadee of the plausibility of the proposed claims [62–64]. Persuasive dialogue agents need to be equipped with a large-scale library of arguments and counter-arguments. By predicting reasonable argument structures based on specific persuasive topics, dialogue agents can incorporate coherent argumentation skills, such as citing authorities and providing convincing arguments and evidence, to effectively enhance the plausibility of dialogue contents and the credibility of the persuasion process. The argument structure prediction strategy has been extensively explored in the field of debate dialogue, where debaters often consider argument structures to express viewpoints with clarity, logical coherence, and compelling evidence [9,65,66]. With the argumentative structure, the whole persuasive process can be progressed incrementally, and the overall organization, logical coherence, and credibility of the persuasive process can be significantly increased. How to mine the supporting argument structures based on the dialogue context and reasonably integrate the argument structures into dialogue contents to enhance the credibility of persuasive dialogue is to be investigated.

2.3   Formalized concept model for CogAgent

Based on the definitions of cognitive strategies, we define the dialogue system that is incorporated with cognitive strategies to accomplish persuasive tasks through smooth and accessible conversations as Cognitive Strategy-enhanced Persuasive Dialogue (**CogAgent**). We introduce the formalized concept model of CogAgent as follows.

Typically, given the dialogue context sequence $H = \{(Q_1, A_1), ..., (Q_{S-1}, A_{S-1})\}$ with $S-1$ turns, where Q_i and A_i are the dialogue query and response at the i-th dialogue turn, and the current dialogue query $Q_S = (q_1, ..., q_m)$ with m words, the objective of general dialogue system is to generate the dialogue response $A_S = (a_1, ..., a_n)$ with n words. The modern dialogue systems usually follow the encoder-decoder architecture [2,67] or decoder-only architecture [18,68]. For the encoder-decoder architecture, the encoder aims to transform input text sequence into vector representations using LSTM [69], Transformer [70] or other advanced neural models, as shown in Eq. (1).

$$\mathbf{Q_S}, \mathbf{H} = \mathbf{Encoder}(Q_S, H). \tag{1}$$

Based on the semantic vectors of dialogue context and input query, the decoder generates the dialogue response word by word in an auto-regressive manner, as shown in Eq. (2), where a_t is the t-th words in the response.

$$P(A|\mathbf{Q_S}, \mathbf{H}) = \prod_{t=1}^{n} p(a_t|\mathbf{Q_S}, \mathbf{H}, a_{<t}). \tag{2}$$

For the decoder-only architecture, all input text sequences will concatenated into a uniform sequence with special tokens, and then the decoder also generates the response in a word-by-word manner. The general dialogue system can generate smooth and fluent responses based on the dialogue context. To generate persuasive dialogue content, it is essential to combine

three kinds of cognitive strategies.

Based on the definitions of three cognitive strategies, we give the formalized definition of CogAgent. Given the dialogue context and the current query, CogAgent needs to first predict the persuasion strategy Per, conversation topic Top, and the argument content Arg based on the current dialogue content, as follows.

$$Per, Top, Arg = Str\_Pre(\mathbf{Q_S}, \mathbf{H}), \tag{3}$$

where Str_Pre refers to the cognitive strategy predictor. Then the dialogue decoder generates the dialogue response word by word conditioned on additional cognitive strategies, as shown in Eq. (4).

$$P(A|\mathbf{Q_S}, \mathbf{H}, Per, Top, Arg) =$$
$$\prod_{t=1}^{n} p(a_t|\mathbf{Q_S}, \mathbf{H}, Per, Top, Arg, a_{<t}). \tag{4}$$

2.4   Generic system architecture

After the concept model of CogAgent, we present the generic system architecture of CogAgent, as shown in Fig. 2. The overall process of CogAgent starts from the semantic understanding of dialogue context and the persuasive target, powered by LLMs (e.g., ChatGPT, LLaMa, Claude, ChatGLM). The input text will be encoded into semantic embeddings for subsequent processes. The *Cognitive Strategy Mining* part is responsible for mining cognitive strategies, including persuasion strategies, topic paths over knowledge graph, and argument structure of topics. The *Cognitive Strategy Prediction for Dialogue Modelling* part predicts appropriate cognitive strategies based on dialogue context and enhances the linguistic expression, logical structure, and other persuasive aspects of responses.

The persuasion strategy mining process first mines various kinds of persuasion strategies through crowd strategy emergence based on cognitive psychology theories. According to the dialogue context, the persuasion strategies to be used in subsequent rounds of dialogue will be predicted. The topic graph construction process constructs topic graphs or topic paths and then plans the wandering paths of topics for persuasion according to the dialogue context and persuasion strategies. The argument mining process first constructs a complete argument structure from credible data sources and then predicts the arguments needed for persuasion based on the above cognitive strategies. Finally, the cognitive strategies-enhanced dialogue context will be fed into LLMs to generate persuasive dialogue responses, for numerous applications, such as psychological counseling, bargaining, and persuasion for social good.

2.5   Application scenarios

Persuasive dialogue system has widespread applications in daily life. It is an ongoing effort of the academic/industry researchers to conduct persuasive dialogue with users to achieve persuasive targets, summarized as follows.

Persuasion for social good Persuasion for social good is a typical persuasive dialogue scenario where people are
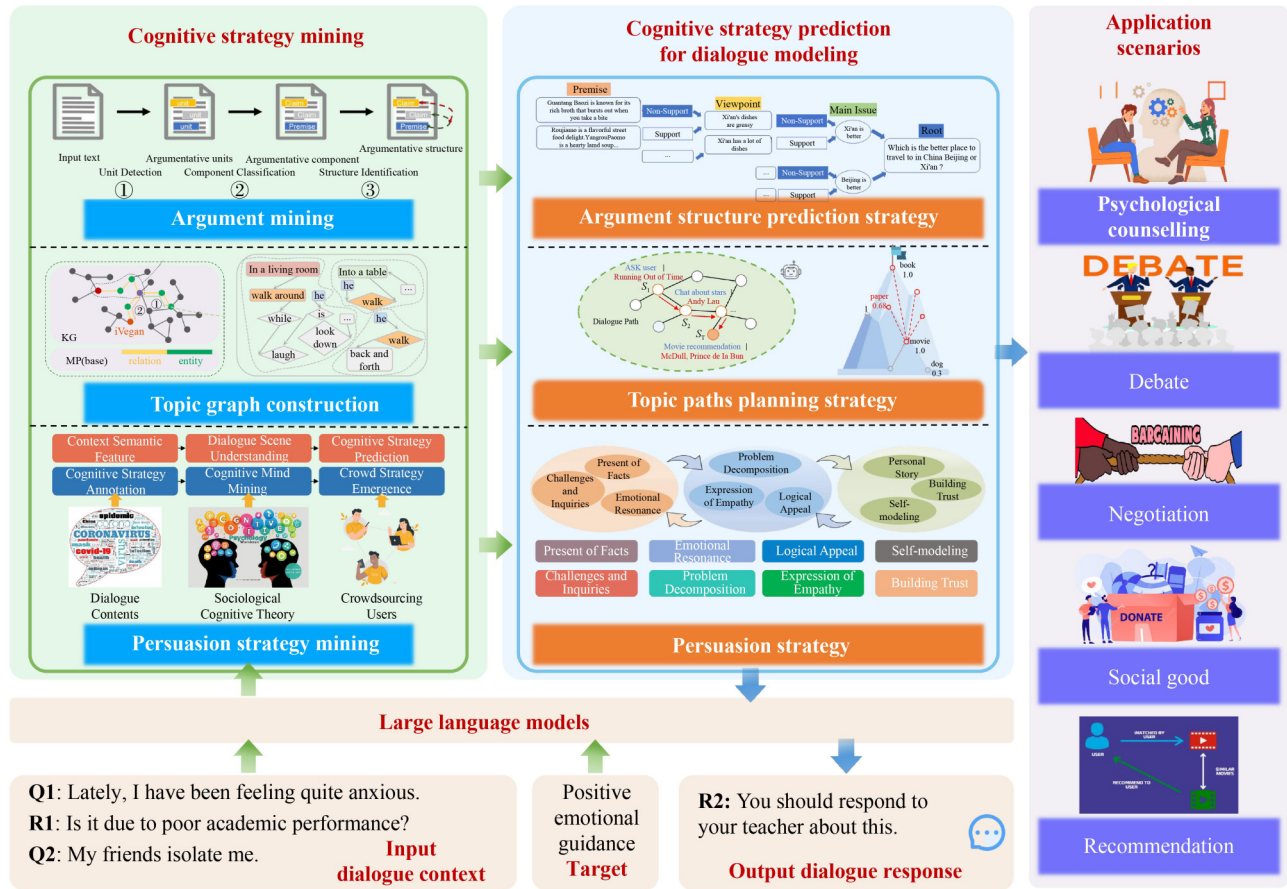
Fig. 2   The generic system architecture of CogAgent

persuaded to donate money or goods to charities for social good purposes, such as children's aid and natural disaster relief. Many researchers have explored combining persuasion strategies to promote users' donation behavior. For example, Wang et al. [8] provide an insightful analysis of what persuasion strategies are effective for what types of personal characteristics of users. Mishra et al. [71] propose a Reinforcement Learning (RL) based persuasive dialogue system with an efficient reward function consisting of five different sub rewards, Persuasion, Emotion, Politeness-Strategy Consistency, Dialogue-Coherence, and Non-repetitiveness. Chen et al. [38] produce a modular persuasive dialogue system that seamlessly integrates factual information and persuasive content into generated dialogue response using the conditional language model.

Persuasion for psychological counseling The frequent occurrence of mental diseases, such as depression, makes mental health gradually receive extensive attention from society [72–74]. Psychological counseling aims at reducing people's emotional distress and helping them understand and work through the challenges that they face. Relieving the psychological pressure of the persuaded through conversation holds profound significance for the persuasive dialogue system. Extensive studies have explored the possibility of using persuasive dialogue systems to provide psychological counseling. For example, Liu et al. [49] collect the Emotion Support Conversation dataset (ESConv) with well-designed

persuasion strategy annotation to train dialogue system to provide emotional support through dialogue interactions. Zhou et al. [75] build a commonsense cognition graph and an emotional concept graph based on commonsense knowledge from COMET [76] and concept knowledge from ConceptNet [77]. The two kinds of knowledge are aligned to generate dialogue responses for emotional support.

Persuasion for negotiation Negotiation is a common real-life persuasion scenario in which two parties negotiate through ongoing conversations to persuade the other party to accept the terms or demands they make to maximize their interests. Negotiation is a necessary means of facilitating agreements among people and improving the efficiency of society. There have been several studies using persuasive dialogue systems to achieve negotiation. For instance, Joshi et al. propose DIALOGRAPH [15], a negotiation dialogue system that explicitly incorporates dependencies between sequences of strategies into graph neural networks. Nortio et al. [78] embark on an exploration of persuasive techniques in international negotiations, emphasizing the significance of persuasion strategies during the negotiation process.

Persuasion for debate Debate is a professional persuasive scenario in which debaters persuade the opponent and the audience to accept their viewpoints by planning their arguments wisely and arguing their points from multiple perspectives. Many researchers have explored the automatic generation of persuasive arguments from online discussions or

debate competitions [65,79,80]. Slonim et al. [9] introduce Project Debater, an autonomous debating system that can engage in a competitive debate with humans.

Persuasion for recommendation Engaging in dialogue-based recommendations for movies, products, and other such aspects proves to be a highly practical application of a persuasive dialogue system. To achieve successful recommendations, it is crucial to employ a persuasion strategy to facilitate rapid user comprehension and acceptance of the recommendations. For example, Gupta et al. [81] propose to decompose the recommendation response generation process into first generating explicit commonsense paths between the source and persuasive target followed by generating responses conditioned on the generated paths. In addition, some researchers adopt LLMs to enhance the persuasiveness of recommender systems [82–84]. For example, Hua et al. [82] enhance the personalized recommendation capability of recommender systems by sending persuasive linguistic commands for various recommendation tasks such as rating prediction, sequential recommendation, direct recommendation, and generating explanations to LLMs to generate more accurate and relevant recommendations.

3   Research challenges and key techniques

Due to the complexity of modeling the psychological changes in the persuasive conversation, many critical challenges in CogAgent need to be addressed. In this section, we first detail these challenges faced by CogAgent, and then conduct a comprehensive investigation of representative works of CogAgent according to the adopted cognitive strategies, i.e., the persuasion strategy, the topic path planning strategy, and the argument structure prediction strategy.

3.1   Research challenges in CogAgent

Exhaustive mining of cognitive strategies Psychology defines human cognition as the process by which a person encounters, perceives, and understands things [85,86]. The formation and evolution of human cognition is an extremely complex process involving knowledge, personality, emotion, and many other aspects. Effective persuasive dialogue changes people's feelings and perceptions about things through persuasion strategies that convince people to change their opinions and behaviors [8,87]. Therefore, it is a great challenge to build a complete set of cognitive strategies from the perspective of cognitive psychology by mining cognitive strategies that can effectively change the way human beings perceive and understand things. Several researchers have defined some persuasion strategies based on cognitive psychology theories (e.g., *logical appeal* and *emotion appeal* from [8], *self-disclosure* from [49]). However, most of these strategies are task-specific and not exhaustive enough to cope with generalized persuasion scenarios. How to construct well-defined cognitive strategies from multiple perspectives needs to be explored in depth.

Modeling and selecting of cognitive strategies In persuasive dialogues, people usually dynamically choose different persuasion strategies depending on different persuasive goals and the evolving conversational contexts. persuasion strategies contain complex semantic patterns, rather than mere names or descriptions [88,89]. How to model the implicit associations between strategy definitions and linguistic expressions, and precisely select cognitive strategies according to the dialog context to facilitate the smooth flow of the persuasive dialog process is a serious challenge. Some researches have explored how to select appropriate cognitive strategies based on the dialogue context [15,30]. The appropriate selection of cognitive strategies is a critical step for CogAgent to simulate humans in persuasive conversations and is essential for achieving high-quality persuasive conversations.

Integrating cognitive strategies into models As defined at the cognitive psychology level, cognitive strategies are more abstract semantic concepts. Data-driven neural network models (DNNs), even LLMs, remain superficial in the understanding of cognitive strategies. How to facilitate DNNs to learn the profound semantics of cognitive strategies, to rationally integrate cognitive strategies into the generation of persuasive dialogues, and to improve the persuasiveness of CogAgent, is quite challenging. Graph-based, [15], reinforcement learning-based [90] and other advanced methods are investigated to integrate cognitive strategies into persuasive dialog generation. It is promising to integrate cognitive strategies with the outstanding language comprehension ability of LLMs.

Absence of evaluation metrics To improve the quality of persuasive dialogue, the performance of CogAgent needs to be evaluated accurately and comprehensively. However, existing evaluation metrics for dialog systems (e.g., BLEU [91], METEOR [92], ROUGE-L [93]) are usually evaluated at the level of word similarity or semantic similarity between generated responses and ground truth, without taking into account the effectiveness of persuasion strategies, the rationality of persuasive path planning, and the richness of argument structure. It is a challenge to develop comprehensive and reasonable evaluation metrics to accurately evaluate the quality of CogAgent, incorporating the characteristics of persuasive dialog systems.

3.2   Persuasion strategy-based CogAgent

Incorporating persuasion strategies to enhance the persuasiveness of dialog responses is an important research direction in CogAgent. By using specific persuasion strategies, CogAgent can express the persuasive content in a way that is more acceptable to the persuadees, thus accomplishing the persuasive goals more smoothly. As abstract psychological concepts, how to select appropriate persuasion strategies according to the dialogue context and guide the generation of responses is an important research question. In this section, we investigate the employment of persuasion strategies in CogAgent, summarized in Table 3.

To enhance CogAgent's capacity to combine persuasion strategies, current research primarily focuses on two aspects. The first involves accurately selecting persuasion strategies. As delineated in Tables 1 and 2, various categories of

Table 3   Representative works of persuasion strategy-based CogAgent

Solution	Work	Description	Scenario
Strategy classifying based on dialogue context	Wang et al. [8]	Proposing a classifier to predict persuasion strategies in dialogue using context and sentence features.	Social good
	He et al. [57]	Decoupling strategy selection and response generation in CogAgent for predicting strategy and generating responses based on dialogue history.	Negotiation
Persuasion strategy planning	Cheng et al. [30]	Proposing lookahead heuristics to estimate future user feedback after using the specific strategy.	Psychological counselling
	Yu et al. [94]	Using Monte Carlo Tree Search for persuasion strategy planning without model training.	Social good
Graph-based strategy incorporation	Joshi et al. [29]	Using GNNs to model strategies, dialogue acts, and dependencies in graph structures for response generation.	Negotiation
	Zhou et al. [95]	Modeling both dialogue context semantic and persuasion strategy history with finite state transducers.	Negotiation and Social good
Knowledge-enhanced strategy modeling	Jia et al. [96]	Introducing a knowledge-enriched encoder and memory-enhanced strategy module for dynamic emotion and semantic pattern modeling.	Psychological counselling
	Chen et al. [38]	Designing RAP for dynamic factual and persuasive responses based on knowledge and individual persuasion strategies.	Social good
Novel integration mechanism	Mishra et al. [71]	Creating an RL reward function to enhance consistency in politeness strategy, persuasiveness, and emotion acknowledgment in persuasive dialogue.	Social good
	Tu et al. [97]	Proposing a novel model MISC, which firstly infers the user's fine-grained emotional status, and then responds skillfully using a mixture of strategies.	Psychological counselling

persuasion strategies exist to achieve persuasive goals in diverse ways. Determining the appropriate strategy for specific dialogue contexts is a key research focus. Developing strategy classifiers based on dialogue context [8,57] and planning strategies over a long horizon [30,94] are two main research directions, discussed in Section 3.2.1 and Section 3.2.2, respectively. Classifier-based methods can map semantic representations of the dialogue context to a specific strategy category. However, they often overlook the long-term impact of strategy selection, making it challenging to ensure the optimal choice of strategy during a long-term conversation. Conversely, strategy planning-based methods excel in selecting strategies over the long term, thereby facilitating the smooth achievement of persuasive goals.

Following the selection of an appropriate strategy, the next step for CogAgent is to integrate the corresponding strategies into the model to generate persuasive dialogue responses, thereby achieving persuasion goals. Given the abstract nature of strategies (typically expressed in a few brief words), researchers are interested in how to model the complex semantic patterns and features (expression manner, language habits) implicit in the strategy definitions, and then generate responses aligned with the strategies. Graph-based strategy incorporation methods [29,95,98], knowledge-enhanced strategy modelling algorithms [38,96], and some emerging integration mechanisms [71,99] have been explored and are summarized in Section 3.2.3, Section 3.2.4, and Section 3.2.5, respectively. The graph-based strategy incorporation methods model the dependency relationships between strategies, enabling a more comprehensive understanding of the incorporation process of strategies in long-term conversations. On the other hand, knowledge-enhanced incorporation methods can significantly enhance the specific strategy patterns in generated responses. In summary, developing persuasion strategy-based CogAgent requires both the accurate selection of persuasion strategies and the comprehensive modeling of these strategies to select the appropriate strategies at different conversation stages and produce persuasive expressions within the constraints of

persuasion strategies. We summarize typical works on the above two aspects below.

3.2.1 Strategy classification based on dialogue context
A straightforward approach to fusing persuasion strategies in persuasive conversations is to predict a strategy label (e.g., Present of Facts) based on the dialogue context and feed the strategy into the decoder with the dialogue context to generate the dialogue response.

For example, Wang et al. [8] propose a persuasion strategy classifier to predict 10 persuasion strategies based on the dialogue context information and sentence-level features. The authors also analyze the impact that different people's backgrounds on strategy prediction, laying the groundwork for research on personalized persuasive dialogue agents. He et al. [57] decouple strategy selection and response generation in CogAgent. The dialogue manager predicts a persuasion strategy based on the persuasion strategies in dialogue history by a sequence to sequence model and the response generator produces a response conditioned on the strategy and dialogue history.

3.2.2 Persuasion strategy planning
Persuasive dialogue is usually a process that lasts multiple turns, supported by successive strategies [100,101]. Consequently, strategy planning within a long planning horizon in CogAgent is quite important, rather than predicting a specific strategy based on the dialogue history. Several studies focus on long-term planning of persuasion strategies, making CogAgent more efficient in reaching persuasion goals.

For instance, Cheng et al. [30] firstly adopt an A* search algorithm for persuasion strategy planning. When predicting the appropriate strategy in each dialogue turn, look-ahead heuristics are proposed to estimate future user feedback after using the specific strategy, thus considering the long-term effect of persuasion strategies. The proposed lookahead method requires abundant annotated data, affecting the application to broader persuasive dialogue scenarios. To overcome this bottleneck, Yu et al. [94] prompts LLMs to perform persuasion strategy planning by simulating future

dialogue interactions using the Monte Carlo Tree Search (MCTS) algorithm. This method requires no model training and can therefore be adapted to any persuasion scenario.

3.2.3   Graph-based strategy incorporation

Graph Neural Networks (GNNs) [102–104] can combine the benefits of interpretability and expressivity, benefiting from encoding graph-structured data through message propagation. Due to the human brain's reasoning process to capture semantic associations, graph-based methods have been widely used in various tasks [105,106]. Numerous researchers have embarked on exploring the potential of graph-based methods for incorporating persuasion strategies in CogAgent. For example, Joshi et al. [29] introduce DIALOGRAPH, a persuasive dialogue system that incorporates persuasion strategies using GNNs in the negotiation dialogue scenario. The model architecture of DIALOGRAPH is shown in Fig. 3, which consists of three main components: (1) hierarchical dialogue encoder, responsible for encoding each utterance in the dialogue context into vector representations; (2) structure encoder tasked with modeling the graph representations of persuasion strategies and dialogue acts using GNNs; (3) utterance decoder for generating output responses. The sequences of persuasion strategies (e.g., *Negotiate side offers*, *Communicate Politely*) and dialogue acts (e.g., *Inquiry*, *Insist*) embodied in the multi-turn dialogue context are modeled into graph structures for better capturing the dependence and influence of strategies and acts on each other and predicting the next adopted strategy and act for response generation. Zhou et al. [95] propose to model both dialogue context semantic and persuasion strategy history finite state transducers (FSTs). To model the persuasion factors affecting the persuasive content of dialogues, Liu et al. [98] present persuasion-factor graph convolutional layers to encode and learn representations of the persuasion-aware interaction data.

3.2.4   Knowledge-enhanced strategy modeling

As concepts in cognitive psychology, persuasion strategies encompass complex semantic information and various intricate linguistic features [101,107]. To comprehensively represent the complex semantics embedded within persuasion strategies, researches investigate combining external knowledge to model and mimic the the intricate patterns in strategies.

For example, Jia et al. [96] propose a knowledge-enriched dialogue context encoder to model the dynamic emotion state and a memory-enhanced strategy modeling module to model the semantic patterns of persuasion strategies. The same-strategy responses are stored in the memory bank to provide more specific guidance for the strategy- constrained response generation. Chen et al. [38] design the Response-Agenda Pushing Framework (RAP) to dynamically produce factual responses based on knowledge facts and persuasive responses conditioned on individual persuasion strategies.

3.2.5   Novel integration mechanisms

In addition to the above studies to model and integrate persuasion strategies, researchers propose some novel integration mechanisms to improve the performance of CogAgent, summarized as follows.

Combined with RL, Yang et al. [99] propose two variants of ToM-based persuasive dialog agent, where the explicit version that outputs the opponent type as an intermediate prediction, and an implicit version that models the opponent type as a latent variable. Both models are optimized using reinforcement learning. Similarly, Mishra et al. [71] design an efficient reward function in RL to improve the politeness-
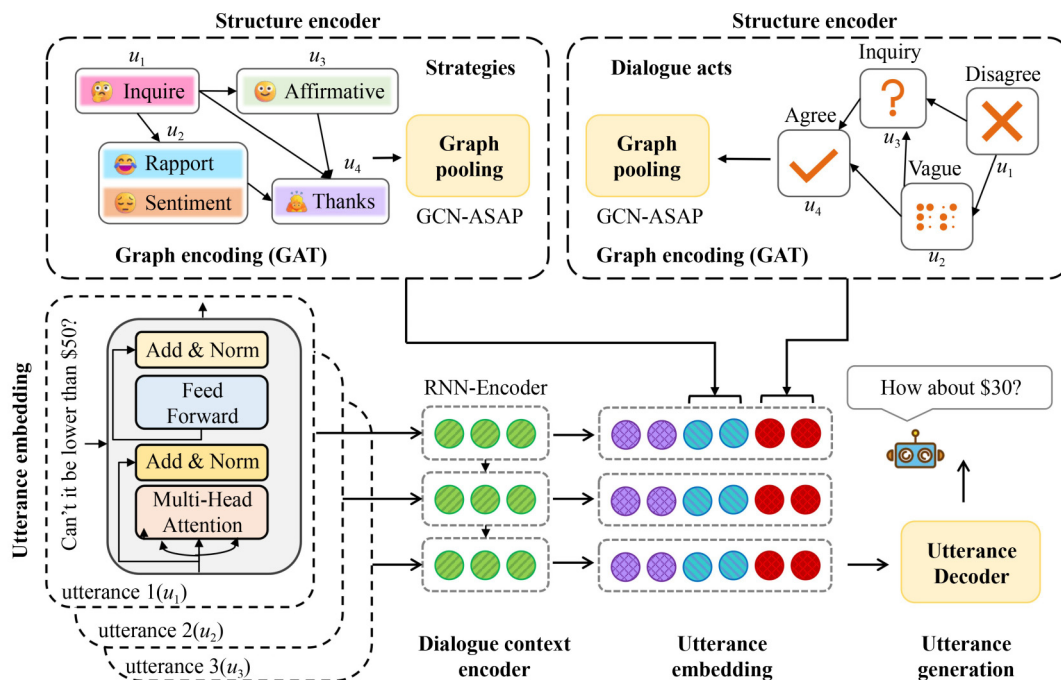

Fig. 3 The model architecture of DIALOGRAPH [29], which models the sequences of persuasion strategies and dialogue acts into graph structures for better modeling of strategy dependence and enriching dialogue context representations

strategy consistency, persuasiveness, and emotional acknowledgement in persuasive dialogue.

To increase the expressed empathy and learn the gradual transition in the long response, Tu [97] introduce a MIxed Srategy-aware model (MISC) integrating COMET, a pre-trained generative commonsense reasoning model, for emotional persuasive dialogue. The COMET knowledge tuples are adopted to enhance the fine-grained emotional understanding of users. Then MISC formulates persuasion strategy as a probability distribution over a strategy codebook to use a mixture of strategies for persuasive response generation.

To investigate the potential of LLMs in persuasive conversations, Zheng et al. [108] first construct a large-scale persuasive dialogue dataset in the emotional support domain, leveraging the generative capabilities of LLMs. Then several advanced tuning techniques (fine-tuning, adapter-tuning, LoRA-tuning) are employed to to showcase the superiority of LLMs in persuasive dialogue generation.

3.3 Topic path planning strategy-based CogAgent

In persuasive dialogues, generating engaging responses through effective topic path planning is critical to achieving persuasive targets. Topic path planning strategy is a navigation tool that enhances the coherence of the persuasion process by continuously leading users to discuss different points and topics until reaching persuasive targets. This section deepens into the intricate details of the topic path planning strategy, summarized in Table 4. As a planning problem, reinforcement learning represents a typical solution for topic path planning. Reinforcement learning excels in gradually selecting optimal actions for agents based on specific goals to complete tasks. However, due to the complexity and subjectivity of persuasive dialogue processes, it is challenging for dialogue agents to determine rewards under different conversation states. This challenge can lead to a risk of getting stuck in local optimality or being unable to find the optimal path for topic planning. Typical works [109,110,115] that incorporate reinforcement learning to achieve topic path planning are summarized in Section 3.3.1. Topic path planning also necessitates guidance from knowledge graphs. By utilizing a knowledge graph, we can discover implicit semantic relationships between different

topics, thereby guiding the selection of topic paths. However, due to the generally large scale of knowledge graphs, accurately locating knowledge in a specific graph to achieve topic path planning is difficult. Some graph-based topic planning works [32,33,112] are presented in Section 3.3.2. Furthermore, some novel topic path planning mechanisms are also discussed in Section 3.3.3.

3.3.1 Reinforcement learning-based planning

In the context of topic paths planning strategy, Reinforcement Learning (RL) serves as a dynamic framework for guiding persuasive dialogue systems in a goal-oriented manner. The core of RL is to learn the optimal sequence of actions according to the reward function and is therefore ideally suited for planning coherent topic paths in CogAgent.

For example, to achieve coherent topic path planning, Xu et al. [109] introduce a three-layer Knowledge aware hierarchical RL-based model (KnowHRL). The upper layer of KnowHRL plans a high-level topic sequence to track user interests toward persuasive targets. The lower layers are responsible for generating multi-turn persuasive dialogue responses. Similarly, Liu et al. [110] propose Goal-oriented Chatbots (GoChat), a frame for end-to-end training chatbots to maximize the long-term reward of topic transitions. The model architecture of GoChat is depicted in Fig. 4, which utilizes hierarchical reinforcement learning (HRL) to concurrently learn a high-level policy for guiding the conversation and a low-level policy for generating responses based on the guidance from the high-level policy. Specifically, at each dialogue turn, the high-level policy (the manager) observes the previous conversation as its state, selects a subtopic planning goal as its action, and awaits a reward from the environment indicating whether the final goal is achieved. The low-level policy (the worker) observes the state along with the sub-goal from its manager, generates the corresponding response to fulfill the sub-goal and receives a reward from the manager. The GoChat framework is trained with Advantage Actor-Critic (A2C) [116], an algorithm widely employed in modern reinforcement learning. To plan topic paths from a global perspective, Yang et al. [115] introduce the global planning method integrated with a commonsense knowledge graph (KG). The key advancement is introducing a global RL

Table 4 Representative works of topic path planning strategy-based CogAgent

| Solution	Work	Description	Scenario
Reinforcement learning based planning	Xu et al. [109]	Presenting KnowHRL, a three-layer Knowledge-aware hierarchical RL-based model for coherent topic path planning and multi-turn persuasive dialogue responses.	Recommendation
	Liu et al. [110]	Hierarchical RL for conversation topic path planning, using high-level strategies and low-level responses.	Social good
	Lei et al. [111]	Introducing four persuasion-related factors in the reward function to achieve persuasive goals efficiently.	Recommendation
Graph-based planning	Zhong et al. [32]	Using commonsense knowledge graphs and GNNs to enhance semantic relations between topic keywords, improving keyword-augmented response retrieval.	Recommendation
	Zou et al. [112]	Employing a concept graph for topic planning, utilizing an Insertion Transformer for persuasive response generation based on multi-concept paths.	Social good
	Wang et al. [113]	Introducing a Transformer-based network for target-driven topic path planning with knowledge-target mutual attention and set-search decoding.	Recommendation
Novel planning mechanism	Tang et al. [61]	Combining various planning algorithms for robust and smooth topic path planning, incorporating a sampling strategy, flow generator, and global planner.	Recommendation
	Ren et al. [114]	Modelling a user's preference in a finer granularity by identifying the most relevant relations from a structured knowledge graph.	Recommendation
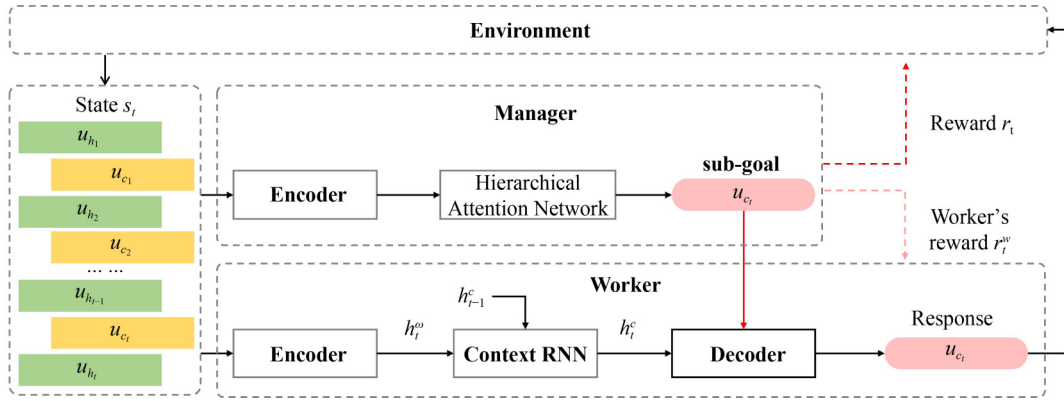
Fig. 4 The overall framework of GoChat [110], which utilizes hierarchical reinforcement learning (HRL) to concurrently plan high-level and low-level policy for efficiently achieving topic path planning

framework that utilizes topic path planning on KG to guide the local response generation model toward persuasive targets, resulting in more coherent conversations. To achieve persuasive goals more effectively, Lei et al. [111] consider four factors (dialogue turn, goal completion difficulty, user satisfaction estimation, and cooperative degree) in the reward function. The targets of achieving persuasive targets quickly and maintaining the engagingness of users.

3.3.2   Graph-based planning

Knowledge is essential to the cognitive reasoning processes of human beings. We humans usually perform common reason during persuasive conversation to enhance the logic and persuasiveness of dialog contents [117,118]. Therefore, relying on commonsense knowledge graphs for topic path planning can produce more persuasive target-related topic paths for CogAgent, thus reaching persuasive targets more efficiently.

Initially, the semantic knowledge relations among topic keywords are captured to perform next-turn topic prediction during conversation [31,119]. Then the predicted topic keywords are used to retrieve appropriate candidate responses for persuasive targets. Furthermore, Zhong et al. [32]

introduce commonsense knowledge graphs and GNNs to model the semantic relations between topic keywords and enhance the keyword-augmented response retrieval,

To plan topic paths more reasonably, Zou et al. [112] introduce a concept graph based on the dialogue data, where the vertices represent concepts and edges are concept transitions between utterances. The multi-concept planning module obtains the topic sequence containing multiple concepts and an Insertion Transformer generates a persuasive response according to the planned topic paths. Wang et al. [33] present a target-driven planning network (TPNet) designed to facilitate transitions between different conversation topics in dialogue systems. Illustrated in Fig. 5, TPNet employs various encoders to learn representations of different input types, such as the Graph Attention Transformer for input domain knowledge graphs and the Memory Network for input user profiles. The Target-driven Conversation Planner within TPNet treats the topic path planning process as a sequence generation task, generating a dialogue path comprising dialog topics and the dialogue actions to present these topics. The planner is constructed based on the transformer architecture, incorporating a knowledge-target mutual attention mechanism. The planned topic path is
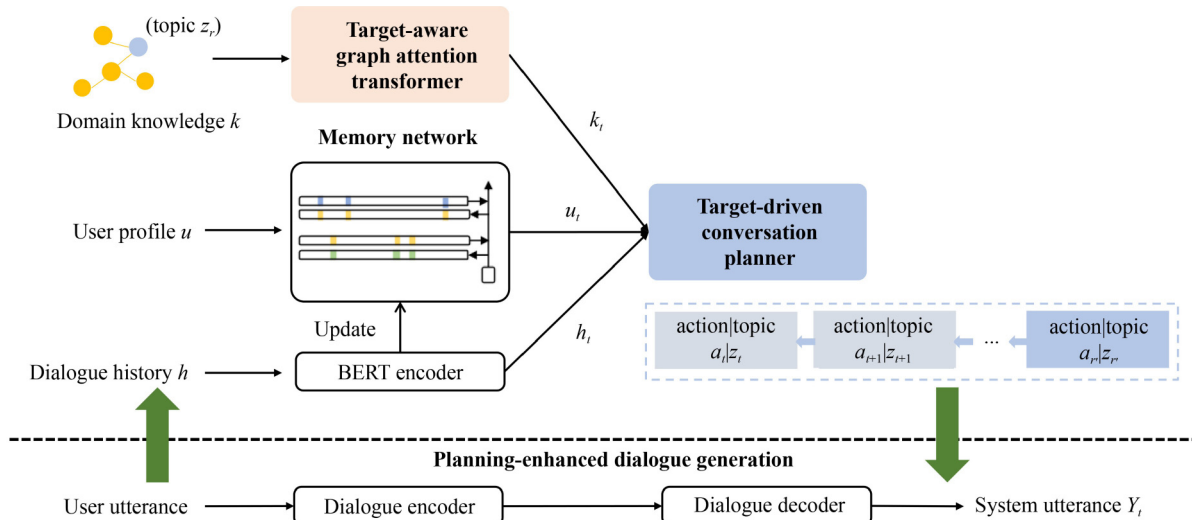

Fig. 5 The overview framework of TPNet [33], which models the topic path planning process as a sequence generation task empowered by the designed target-driven conversation planner

subsequently utilized to extract knowledge facts and generate final responses.

3.3.3 Novel planning mechanism

In addition to the above research to plan topic paths in CogAgent, there are some novel planning mechanisms to be explored, summarized as follows.

Combining the strengths of multiple topic planning algorithms, Tang et al. [61] propose an EAGLE model for topic path planning. Comprising a topic path sampling strategy, topic flow generator, and global planner, EAGLE achieves robustness to unseen target topics and smooth transitions. The model demonstrates enhanced global planning ability through its integrated approach, addressing limitations in existing topic-planning conversation models.

To ensure the smooth and coherent progression toward persuasive goals across different turns, Wang et al. [113] introduce a consistency-driven dialogue planning approach that utilizes stochastic processes to model the temporal evolution of the conversation path dynamically. Firstly, a latent space is defined, and Brownian bridge processes are employed to capture the continuity of goal-oriented behavior, allowing for more flexible integration of user feedback into dialogue planning, and explicitly generating conversation paths. Ultimately, these paths are employed as natural language prompts to guide the generation of persuasive dialogue.

3.4 Argument structure prediction strategy-based CogAgent

CogAgent entails an ongoing conversation between a dialogue agent and a user at the cognitive level, where the dialogue agent proactively steers the conversation. As the conversation progresses, the contents presented by the dialogue agent to support its opinions undergo dynamic transformations. Consequently, the reasonable selection and application of arguments and evidence play a pivotal role in the persuasiveness of the dialogue. The utilization of arguments and evidence is imperative in the process of persuasion. Firstly, employing arguments and evidence allows for the gradual decomposition and progressive reasoning of persuasive targets, thereby facilitating a logical and sequential flow of the conversation that enhances the users' acceptance of viewpoints [120]. Secondly, the provision of factual support elevates the credibility of persuasive discourse, thereby augmenting the persuasiveness of the conversation. In this section, we provide an investigation of the crucial techniques for argument structure prediction-strategy-based CogAgent, as summarized in Table 5.

To achieve the argument structure-enhanced CogAgent, existing research primarily focuses on two aspects. The first is argument mining. Arguments are often implicit in vast amounts of interactive content, such as online social conversations, which need to be extracted from existing content to provide further argumentative structure and factual knowledge to support the generation of persuasive dialogue. Existing argument mining works include various techniques, such as multi-task learning technology [35] and argument boundary prediction technology [36], detailed in Section 3.4.1. After mining a rich set of arguments, the next step is to predict the corresponding argument structure from a massive argument database based on the dialogue context [65,79,124]. This enhances the rationality and logic of persuasive dialogue content. In summary, first mining rich arguments and then predicting argument structures relevant to the dialogue context, and thus integrating them into the persuasive dialogue generation process, is the approach to achieving the argument-enhanced CogAgent.

3.4.1 Argument mining

To integrate the argument structure into CogAgent, it is first necessary to perform argument mining according to conversation topics. Researchers embark on mining argumentative text from dialogues for CogAgent.

Debate involves the explicit use of argumentative content for dialogue expression, making it an important source of argument mining. For example, Khatib et al. [121] utilize online debate portals to acquire both controversial and non-controversial text snippets related to several contentious topics. These snippets are organized in a semi-structured format. Eventually, by employing a diverse set of vocabulary, grammar, and metric feature types, the arguments are classified and structurally modeled. Hua et al. [122] propose a framework for generating arguments to opposing viewpoints. The retrieval module of this framework comprises Query

Table 5 Representative works of argument structure prediction strategy-based cogAgent

Solution	Work	Description	Scenario
Argument mining	Khatib et al. [121]	Classifying and structurally modeling arguments from online debate portals based on diverse vocabulary, grammar, and metric features.	Debate
	Hua et al. [122]	Proposing an argument generation framework with retrieval modules and a sentence-level LSTM for generating viewpoints.	Negotiation
	Srivastava et al. [36]	Using attention-based link prediction and Transformer encoder to model hierarchical causal relationships and discover associations in online argument structures.	Negotiation
	Niculae et al. [123]	Introducing factor graph model for argument mining, concurrently learning fundamental unit types classification and argument relationship prediction.	Negotiation
Argument structure prediction	Rach et al. [65]	Proposing argument search technique using supervised learning-based relation classification to retrieve arguments for debate dialogue system	Debate
	Sakai et al. [79]	Introducing an approach to consider the human agreement and disagreement, resulting in a persuasive argument with a hierarchical argumentation structure.	Debate
	Prakken et al. [34]	Enhancing argument modeling with a five-layer graph, serving as a knowledge base for a chatbot to identify user focal points and select rebuttal points.	Debate
	Li et al. [124]	Using factor graphs to extract online debate features, incorporating them into an LSTM model to predict persuasive arguments.	Debate

Formulation, Keyphrase Extraction, and Passage Ranking and Filtering. Subsequently, a sentence-level LSTM is trained to generate a sequence of sentences.

In online discussion platforms, people also use argumentative texts to enhance their expressions. For instance, Tran et al. [35] and others [125–127] employ multi-task learning to unearth arguments and evidence at both the micro and macro levels, enhancing persuasive power in online discussions. Srivastava et al. [36] employs an attention-based link prediction embedding model to model the hierarchical causal relationships within common argument structures in online discussions. They then utilize Transformer encoder layers to discover the associations and boundaries between arguments. Furthermore, they employ AMPERSAND [64] and SMOTE [128] to address data imbalance issues, thereby improving model accuracy. Furthermore, Niculae et al. [123] introduce a factor graph model for argument mining, wherein the model concurrently learns the classification of fundamental unit types and prediction of argument relationships. Furthermore, the parameter structures of structured SVM and RNN can enforce structural constraints (e.g., transitivity), while also representing dependencies between adjacent relationships and propositions.

3.4.2 Argument structure prediction

Dialogue systems of persuasive tasks commonly rely on structured knowledge concerning arguments and their relationships. Numerous researchers have demonstrated that predicting argument structures and integrating them into CogAgent can significantly enhance topic consistency, content coherence, and persuasiveness of persuasive dialogue contents [9,79,80].

For example, Rach et al. [65] propose an argument search technique for a debate dialogue system, which utilizes supervised learning-based relation classification to retrieve arguments mapped to a generic tree structure for the dialogue model. Sakai et al. [79] introduce an approach to consider human agreement and disagreement, resulting in a persuasive argument with a hierarchical argumentation structure. The dialogue agent selects the next action based on the user's agreement or disagreement and sends the chosen action to the response generation module to generate logically consistent and persuasive dialogue.

For more intensive argument modeling, Prakken et al. [34] equip dialogue agents with a five-layer argument graph, consisting of 1288 nodes, with an average of three counterarguments per node. This graph serves as the knowledge base for the proposed chatbot, allowing it to dynamically identify and annotate the user's focal points on the parameters, enabling the selection of appropriate rebuttal points. Li et al. [124] employ a factor graph model to depict the influence of discourse structure on the persuasiveness of arguments extracted from online debates, illustrated in Fig. 6. Specifically, leveraging debates from the DDO corpus [129], the authors initially generate argument structures for these debates using the pre-trained model introduced by [123], followed by employing BERT to extract argument structure features. Subsequently, two LSTM-based models encode the sequence of turns from opposing sides (i.e., PRO versus CON). The experimental findings suggest that argument structure features encapsulate crucial persuasion strategies, underscoring their significance in facilitating an effective persuasion process.

4 Datasets, evaluation metrics, and performance analysis for CogAgent

4.1 Datasets for CogAgent

Massive data is undeniably indispensable for training high-quality CogAgent. To foster advancement in this field, numerous large-scale and high-quality datasets have been released. In this section, we categorize existing datasets by application scenarios, including psychological counseling, debate, price negotiation, persuasion for donation, and product recommendation, summarized as Table 6.

4.1.1 Datasets for psychological counseling

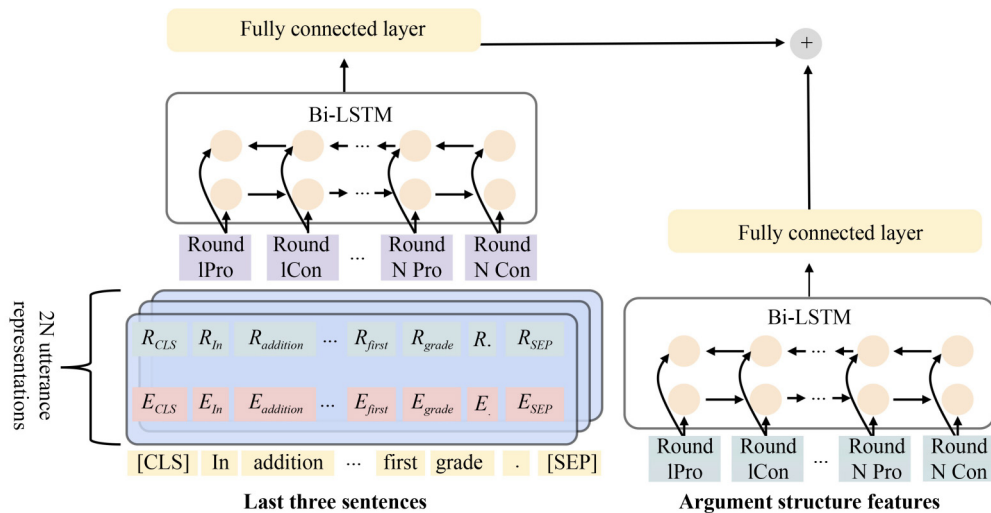Psychological counseling is a typical field of persuasive dialogue, where CogAgent reduces users' psychological

Fig. 6   The overall framework to investigate the influence of discourse structure on the persuasiveness of arguments extracted from online debates [124]

Table 6   A review of available datasets for CogAgent

Scenario	Dataset	Description
Psychological counseling	ESConv [49]	The first dataset for psychological counseling, annotated with persuasion strategies.
	AUGESC [130]	The enhanced dataset from ESConv using LLMs with a broader range of topics.
	PsyQA [131]	A Chinese mental health support dataset featuring annotated persuasion strategies.
Debate	IAC [132], IAC 2 [133]	Argumentative dialog dataset with curated threads, posts, and annotations.
	Winning arguments [134]	A metadata-rich subset of r/ChangeMyView subreddit conversations includes data on the success of user utterances in persuading the poster.
	Debate Sum [135]	A dataset for the competitive formal debate with corresponding argument and extractive summaries.
Negotiation	CraigslistBargain [57]	A human-human dialogue dataset for price negotiation where the buyer and seller are encouraged to reach an agreement to get a better deal.
	NegotiationCoach [136]	An additional negotiation coach based on CraigslistBargain, which monitors the exchange between two annotators and provides real-time negotiation strategy.
Social good	PersuasionForGood [8]	A collection of online conversations where one participant (the persuader) tries to convince the other (the persuadee) to donate to a charity.
	EPP4G and ETP4G [71]	Datasets extending Persuasion For Good by annotating it with the emotion and politeness-strategy labels.
	FaceAct [137]	A dataset extending Persuasion For Good by adding the utterance-level annotations that change the positive and/or the negative face of the participants in a conversation.
Recommendation	REDIAL [138]	A dataset comprising recommendation conversations about movies.
	TG-ReDial [11]	A dataset consisting of dialogues between a seeker and a recommender.
	DuRecDial [139]	A human-to-human Chinese dialog dataset, which contains multiple sequential dialogues for every pair of a recommendation seeker and a recommender.
	INSPIRED [140]	A movie recommendation dataset, consisting of human-human dialogues with an annotation scheme for persuasion strategies.

anxiety and encourages positive emotions through the persuasive dialogue process. Researchers have released several datasets for psychological counseling.

ESConv ESConv [49] is a well-designed and rich, effective corpora for psychological counseling, consisting of 1,053 dialogue pairs and a total of 31,410 sentences. Each dialogue pair includes information about the initial emotional state of the seeker, the persuasion strategies employed by the supporter during each interaction, and the content of the conversation. The dataset encompasses seven distinct emotional states and eight supportive strategies, with the labeling of these strategies being inspired by Hill's Helping Skills Theory [101].

AUGESC The limitations imposed by crowdsourcing platforms on data themes and collection methods, along with the substantial regulatory costs, have hindered the extension of downstream dialogue models to open-domain topics. In response, Zheng et al. augment ESConv to AUGESC [130] using LLMs, which comprises 65,000 dialogue sessions and a total of 1,738,000 utterances. It substantially expands the scale of ESConv and encompasses a broader range of topics.

PsyQA PsyQA [131] is a Chinese mental health support dataset collected from a Chinese mental health service platform, including 22,000 questions and 56,000 lengthy, well-structured answers. In line with psychological counseling theory, PsyQA annotates some of the answer texts with persuasion strategies and further conducts in-depth analyses of the lexical features and strategic patterns within counseling responses.

4.1.2 Datasets for debate

Debates are typically persuasive scenarios in which each party of the debate organizes arguments to persuade the other party to accept his or her side's viewpoints. Existing datasets for debate are listed as follows.

Internet argument corpus (IAC) IAC [132] is a scriptless argumentative dialog dataset, comprising 390,704 posts extracted from 11,800 discussions on the online debate platform. Within this corpus, a manually curated subset of 2,866 threads and 130,206 posts is formed, categorized based on discussion topics. Extended from IAC, IAC 2 [133] is a corpus for research in political debate on Internet forums, consisting of three data sets: 4forums (414k posts), ConvinceMe (65k posts), and a sample from CreateDebate (3k posts).

Winning arguments To delve deeper into the mechanisms of changing others' viewpoints in social interactions, Tan et al. [134] introduce the Wining Arguments (ChangeMyView) Corpus. Wining ArgumentsCorpus is a metadata-rich subset of conversations made in the *r/ChangeMyview* subreddit between 1 Jan 2013 - 7 May 2015, with information on the delta (success) of a user's utterance in convincing the poster. There are 34,911 Speakers, 293,297 Utterances, and 3,051 Conversations.

DebateSum DebateSum [135] is a dataset for the competitive formal debate, including 187,386 unique pieces of evidence with corresponding argument and extractive summaries. The argument data is collected from the National Speech and Debate Association over 7 years.

4.1.3 Datasets for negotiation

Price negotiation is an everyday persuasive scenario where buyers and sellers reach their desired price through the persuasive dialog process. Datasets for price negotiation are summarized as follows.

CraigslistBargain CraigslistBargain [57] is a human-human dialogue dataset for price negotiation, which consists of 6,682 dialogues, collected using Amazon Mechanical Turk (AMT) in a negotiation setting where two workers were assigned the roles of buyer and seller, respectively. The buyer is additionally given a target price and both parties are encouraged to reach an agreement while each of the workers

tries to get a better deal.

NegotiationCoach NegotiationCoach [136] introduces an additional negotiation coach based on CraigslistBargain, which monitors the exchange between two annotators and provides real-time negotiation strategy recommendations to the seller for achieving better deals.

4.1.4   Datasets for social good

Persuasion for donation is very common in life, where the persuader persuades others to donate property or labor to charities for a public good purpose. Datasets for persuasion for donation are listed as follows.

PersuasionForGood PersuasionForGood [8] is a collection of online conversations generated by AMT workers, where one participant (the persuader) tries to convince the other (the persuadee) to donate to a charity. This dataset contains 1,017 conversations, along with demographic data and responses to psychological surveys from users. 300 conversations also have per-sentence human annotations of dialogue acts that pertain to the persuasion setting, and sentiment.

EPP4G and ETP4G EPP4G and ETP4G [71] extend PersuasionForGood by annotating it with the emotion and politeness-strategy labels.

FaceAct FaceAct [137] further extend PersuasionForGood by adding the utterance-level annotations that change the positive and/or the negative face of the participants in a conversation. A face act can either raise or attack the positive face or negative face of either the speaker or the listener in the conversation.

4.1.5   Datasets for recommendation

Product recommendation intends to induce the recommended person to accept or buy a particular product through persuasive dialogues. Datasets for product recommendation are listed as follows.

REDIAL REDIAL [138] comprises over 10,000 conversations centered around conversations about movies, where one party in the conversation is seeking recommendations and the other party is providing recommendations.

TG-ReDial TG-ReDial [11] consists of 10,000 two-party dialogues between a seeker and a recommender in the movie domain.

DuRecDial DuRecDial [139] is a human-to-human Chinese dialog dataset (about 10k dialogs, 156k utterances), which contains multiple sequential dialogues for every pair of a recommendation seeker (user) and a recommender (bot). In each dialogue, the recommender proactively leads a multi-type dialogue to approach recommendation targets and then makes multiple recommendations with rich interaction behavior.

INSPIRED INSPIRED [140] is a movie recommendation dataset, consisting of 1,001 human-human dialogues with an annotation scheme for persuasion strategies based on social science theories.

4.2   Evaluation metrics toward CogAgent

The reasonable evaluation of the quality of CogAgent is a challenging dilemma. Different from the open-domain dialog system, the evaluation of CogAgent needs to be performed under different persuasion scenarios and multifaceted persuasive goals. This requires judging the quality of dialogue response while emphasizing the persuasive effects in specific persuasive contexts and assessing the adaptability and persuasiveness of the system's cognitive strategies in different domains. Up to now, there is no unified theory on how to effectively evaluate CogAgent, and researchers predominantly employ automatic evaluation and human evaluation. The evaluation of CogAgent based on LLMs has also gradually garnered widespread attention. We summarize typical evaluation methods in Table 7.

4.2.1   Automatic evaluation metrics

Automatic evaluation metrics evaluate the performance of CogAgent by calculating the similarity between the responses generated by CogAgent and ground truths. There are typical categories of automatic evaluation metrics: overlap-based methods, embedding-based methods, and learning-based techniques.

Overlap-based metrics Overlap-based methods measure the degree of text overlap between generated responses and golden responses, with particular emphasis on the number of the same n-grams. These methods quantify the similarity of the text, especially the local structural similarity, to measure the quality of generated responses. Classical Overlap-based methods include **BLEU** [91], **ROUGE** [93], **METEOR** [141], and **CIDEr** [142]. Among these, BLEU evaluates response quality by comparing the harmonic mean of n-gram overlaps between generated responses and the golden ones. BLEU is a straightforward and intuitive metric, yet it is constrained by surface features and may exhibit a weak capture of semantic relevance. ROUGE calculates the length of the longest common subsequences between generated and golden

Table 7   Evaluation metrics for CogAgent

Evaluation method	Category	Description	Metrics
Automatic evaluation	Overlap-based	Measuring the degree of text overlap between generated responses and golden responses.	BLEU [91], ROUGE [93], METEOR [141], CIDEr [142]
	Embedding-based	Evaluating the semantic similarity of embedding vectors between generated responses and reference ones.	Greedy Matching [143], Embedding averaging [144], Vector Extreme [145]
	Learning-based	Employing machine learning models to predict the quality scores of generated responses, relying not only on given references.	ADEM [146]
Human evaluation		Scoring by human annotators to evaluate the quality of the generated responses with subjective judgment.	Naturalness, Appropriateness, Informativeness, Diversity, Humanness
LLMs-based evaluation		Scoring by utilizing one or multiple LLMs as expert evaluators for evaluating the performance of CogAgent.	Engagingness, Persuasiveness, Strategy selection proficiency

responses and considers the precision and recall to evaluate the quality. METEOR integrates multiple aspects of information, including precision, recall, and syntactic structure, providing a more comprehensive evaluation. CIDEr evaluates the semantic similarity between generated responses and ground truths using n-gram level cosine similarity. These metrics have been widely applied in the evaluation of open-domain dialog systems, but they focus mainly on surface features of the response and may not capture semantic relevance. In addition, relying solely on n-gram overlap to measure similarity may not always accurately evaluate the quality of long texts.

Embedding-based metrics Embedding-based metrics evaluate the semantic similarity of embedding vectors between generated responses and reference ones. These methods utilize pre-trained word embedding models (e.g., BERT [147]) to map textual responses into embedding vectors, thus capturing the semantic relationships between the texts more accurately. Specifically, **Greedy matching** [143] computes the cosine similarity of word embeddings between each word in generated response and golden ones. **Embedding averaging** [144] averages all words in the sentence to calculate the sentence-level similarity. **Vector extrema** [145] takes the most extreme value in the embedding vector to represent the response to be evaluated. In essence, embedding-based metrics emphasize the semantic quality of CogAgent more than overlap-based metrics and better capture the semantic correlations between generated responses and references.

Learning-based metrics Learning-based metrics employ machine learning models to predict the quality scores of generated responses, relying not only on given references but aiming to better correlate with human judgment. ADEM [146] is a deep model-based evaluation metric for dialogue systems. A hierarchical RNN model is trained in a semi-supervised manner to capture semantic information and contextual associations and align with the human preferences for dialogue responses.

In summary, automatic evaluation metrics offer advantages in terms of efficiency and consistency. However, they face challenges in terms of semantic understanding, manual annotation costs, and model complexity. When selecting and applying automated evaluation metrics, it is important to balance their advantages and disadvantages according to specific persuasive tasks and scenarios.

4.2.2 Human evaluation

Human evaluation entails the subjective analysis and scoring by human annotators to assess the quality of the generated responses of CogAgent. This process often involves domain experts and crowd workers who judge the responses against specific criteria and task requirements, emphasizing the subjective nature, emotional depth, and persuasion strategies employed by CogAgent. Unlike automated metrics, human evaluation offers a nuanced appreciation of these aspects, making it an indispensable tool for a thorough and nuanced assessment of CogAgent's performance.

Human evaluation metrics for conventional dialogue systems typically include the following aspects:

- Naturalness: assessing if the responses are grammatically and syntactically fluent, accurate, and coherent.
- Appropriateness: evaluating the relevance of responses to the ongoing dialogue context.
- Informativeness: determining whether the responses enrich the conversation with practical insights.
- Diversity: evaluating the variety in language expression with little repetition in responses.
- Humanness: judging whether the responses appear as if authored by a human.

These metrics, while foundational, overlook the potential influence of cognitive strategies on the evaluation of CogAgent. For a precise evaluation of cognitive strategies' role in persuasion, it's pivotal to design evaluation criteria based on the principles of cognitive psychology in persuasive dialogue. We propose several promising human evaluation metrics that incorporate cognitive psychology insights and existing research to elucidate cognitive strategies' impact on CogAgent:

- Strategy Utilization: measuring the adept incorporation of appropriate cognitive strategies in the responses.
- Cognitive Engagement: evaluating the capacity to stimulate the persuadee's cognitive processes, encourage them with critical reflection and alternative perspectives, and potentially alter their pre-existing beliefs or attitudes.
- Emotional Resonance: assessing the efficacy of CogAgent in harnessing cognitive strategies to modify users' emotional states, thereby bolstering the persuasiveness of conversations.
- Persuasiveness: evaluating the CogAgent's effectiveness in shifting the persuadee's attitudes, beliefs, or behaviors towards the desired direction through the adoption of cognitive strategies.

In essence, human evaluation stands as the most reliable and reasonable method to evaluate CogAgent, necessitating further exploration into the integration of cognitive strategies to advance the evaluation of CogAgent.

4.2.3 LLMs-based evaluation

Due to the complexity and diversity within human language, accurately evaluating the quality of dialogue systems has long posed a significant challenge in dialogue system research [148,149]. Automatic evaluation metrics mentioned above (e.g., BLEU, METEOR) often exhibit a limited correlation with human judgments and necessitate substantial effort to collect reliable reference labels for metric computation. LLMs with exceptional natural language understanding and generation capabilities have been widely applied across various scenarios [17,18,150,151]. Employing LLMs to evaluate the quality of dialogue systems and CogAgent is very promising and has garnered considerable attention from researchers.

Initially, researchers [152,153] directly employ LLMs to score candidate outputs based on their generation

probabilities, without relying on reference responses. For instance, Wang et al. [152] first regard ChatGPT as a human evaluator, providing task-specific (e.g., text summarization, story generation) and aspect-specific (e.g., relevance, fluency) instructions to prompt evaluations of natural language generation models. While these methods showed a higher correlation with human evaluations than traditional metrics, they still fell short compared to medium-sized neural evaluators [154,155] in terms of human correspondence.

Subsequent frameworks for evaluating dialogue systems effectively and reliably have been proposed to address these limitations. For example, Liu et al. [156] introduce G-EVAL, a framework utilizing LLMs (GPT-3.5 and GPT-4) with chain-of-thoughts (CoT) [157] to generate detailed evaluation steps based on prompts with task introductions (e.g., text summarization, dialogue generation) and evaluation criteria. This framework introduces the criteria of "Engagingness'' to evaluate the quality of dialogue systems, which could be extended to evaluate the persuasiveness, fluency of persuasion processes, and other criteria of CogAgent. To enhance the interpretability of LLMs-based metrics and provide explicit explanations of their evaluations, INSTRUCTSCORE [158], a fine-grained explainable evaluation metric for text generation, is proposed. This approach extracts latent evaluation knowledge from an instruction-following LLM (e.g., GPT-4) and aligns LLaMA-based evaluation mode with human evaluations through diagnostic reports.

While single-agent-based approaches have demonstrated significant effectiveness in evaluation, further advancements are necessary to achieve human-level evaluation quality. The multi-agent-based approach facilitates the collaboration of a group of LLMs with an assortment of intelligent counterparts, leveraging their distinct capabilities and expertise to augment efficiency and effectiveness in handling intricate tasks. ChatEval [159] introduces a multi-agent referee team designed to autonomously discuss and evaluate the quality of generated dialogue responses from diverse models. Each debater agent within ChatEval possesses a unique persona, and various communication strategies are meticulously crafted to ensure that each agent can efficiently communicate with other agents to evaluate results.

The use of LLMs for evaluating dialogue systems is recognized as an effective approach for achieving reliable human-like evaluations and introduces new perspectives to the evaluation of CogAgent. Through the design of tailored evaluation criteria, such as persuasiveness [160], strategy selection proficiency [161], and fluency of persuasion process [162], LLMs have the potential to deeply comprehend the persuasion process and provide precise, human-like evaluations for CogAgent. Additionally, by creating distinct agents accompanied by specific evaluation criteria, the persuasion process can be analyzed from various aspects, leading to a more detailed understanding and more accurate evaluation of CogAgent.

4.3 Performance analysis of CogAgent

In light of the rapid advancement of CogAgent, researchers have undertaken extensive investigations across diverse scenarios (e.g., persuasion for social good, persuasion for psychological counseling, persuasion for debate). To gain a more intuitive understanding of the evolution of current CogAgent systems, this section conducts a comprehensive comparison and analysis of their performance in several typical application scenarios.

4.3.1 Persuasion for Psychological Counseling

The global rise in the prevalence of mental disorders and diseases is having a profound impact on both society and healthcare systems [163,164]. Despite emotional distress being a common part of the human experience, only a small fraction of individuals seek professional help. Persuasion for psychological counseling aims to reduce emotional distress and assist individuals in understanding and addressing the challenges they face. Alleviating the psychological burden of the persuadee through conversation is of great importance for CogAgent. The Emotion Support Conversation dataset (ESConv) [49] serves as a benchmark for persuasion in psychological counseling, and numerous studies based on it have explored the potential for CogAgent to provide psychological counseling.

To evaluate the performance of CogAgent on the ESConv benchmark, researchers typically adopt several evaluation metrics, including perplexity (PPL), BLEU-1/2/3/4 (B-1/2/3/4), ROUGE-L (R-L), METEOR (MET), and CIDEr. The performance of current state-of-the-art methods is summarized in Table 8. Research on persuasion for psychological counseling focuses on two main aspects to strengthen the persuasiveness of CogAgent. One aspect is the accurate persuasion strategy prediction. Psychological research [101,169] emphasizes the importance of modeling long-term conversational procedures and employing appropriate persuasion strategies at different conversation stages to effectively and naturally support emotions. Several notable works include:

- DialoGPT-Joint and BlenderBot-Joint [49], based on DialoGPT [170] and BlenderBot [171], respectively, predict persuasion strategies using a simple multi-classifier and concatenate them with dialogue context to generate persuasive responses.
- MISC [97] introduces a mixed strategy probability distribution to generate persuasive responses guided by a mixture of multiple persuasion strategies instead of predicting a single strategy.
- FADO [167] introduces a dual-level feedback strategy selector to encourage or penalize strategies during the strategy selection process.
- MultiESC [30] develops lookahead heuristics to estimate the expectation of future user feedback, aiding in selecting appropriate persuasion strategies to effectively achieve long-term persuasive goals.

Another research aspect focuses on understanding and modeling the evolving emotional states (e.g., emotion causes, emotion transitions) of persuadees during conversations, which is crucial for CogAgent to empathize with persuadees. Some notable studies include:

Table 8   Performance of various CogAgent models on the ESConv dataset. The best results are highlighted in bold

Models	PPL↓	B-1↑	B-2↑	B-3↑	B-4↑	R-L↑	MET↑	CIDEr↑
DialoGPT-Joint [49]	15.71	17.39	5.59	2.03	1.18	16.93	7.55	11.86
BlenderBot-Joint [49]	16.79	17.62	6.91	2.81	1.66	17.94	7.54	18.04
MISC [97]	16.16	−	7.31	−	2.20	17.91	**11.05**	−
TransESC [165]	15.85	17.92	7.64	4.01	2.43	17.51	−	−
GLHG [166]	15.67	19.66	7.57	3.74	2.13	16.37	−	−
FADO [167]	15.72	−	8.00	4.00	2.32	17.53	−	−
KEMI [168]	15.92	−	8.31	−	2.51	17.05	−	−
MultiESC [30]	15.41	21.65	9.18	4.99	3.09	20.41	8.84	29.98
MODERN [96]	**14.99**	**23.19**	**10.13**	**5.53**	**3.39**	**20.86**	9.26	**30.08**

- TransESC [165] considers turn-level state transitions in emotional support conversations from three perspectives: semantics, strategy, and emotion transitions, aiming for a smooth and natural conversation flow.
- GLHG [166] uses a graph neural network to model relationships between the user's emotional causes, intentions, and dialogue history to generate persuasive dialogues.
- KEMI [168] is a knowledge acquisition module to retrieve emotional support knowledge from a large-scale knowledge graph on mental health dialogues. The retrieved knowledge serves as additional input to generate persuasive responses.
- MODERN [96] utilizes a knowledge-enriched dialogue context encoding to perceive dynamic emotional changes during the conversation and selects context-related concepts from ConceptNet for response generation.

4.3.2   Persuasion for recommendation

Engaging in dialogue-based recommendations for movies, products, or other items represents a profoundly practical application of a CogAgent [172,173]. CogAgent integrates persuasion strategies to gradually direct the user's interest toward the recommended target, thereby facilitating the persuasion process. Two typical benchmarks, ReDial [138] and INSPIRED [140] have been instrumental in exploring the potential of CogAgent in persuasive recommendation, leading to extensive research in this area.

The performance evaluation of CogAgent on the ReDial and INSPIRED benchmarks typically focuses on two aspects: recommendation performance and response generation performance. For recommendation performance, the metric $H@K$ ($K = 1, 10, 50$) [174,175] is widely utilized to assess whether the top-K items recommended by the system include the ground truth (i.e., true answers). Regarding response

generation performance, metrics like Distinct n-gram ($n = 3, 4$) [175,176] and *HitX* are popular for evaluating the diversity of generated responses at a sentence level and whether a generated recommendation response contains explanations with any features of the target item, respectively. Table 9 summarizes the performance of current state-of-the-art methods, detailed as follows:

- KGSF [174] incorporates both word-oriented and entity-oriented knowledge graphs (KG) to enhance data representations. It proposes a KG-enhanced recommender component and a KG-enhanced dialogue component to make accurate recommendations and generate corresponding informative responses with keywords or entities.
- CR-Walker [177] performs tree-structured reasoning on a knowledge graph and generates informative dialogue acts to guide persuasive dialogue generation.
- RevCore [178] takes reviews of target items as additional inputs to enrich item information and assist in generating coherent and informative responses.
- C²-CRS [179] is a coarse-to-fine contrastive learning framework that enhances data semantic fusion for persuasive recommendation. It extracts and represents multi-grained semantic units from different data signals and aligns them to generate persuasive dialogue content.
- UniCRS [176] unifies the recommendation and conversation subtasks into the prompt learning paradigm and utilizes knowledge-enhanced prompts based on fixed LLMs to fulfill both subtasks in a unified approach.
- LATTE [180] pre-trains each core module in persuasive recommendation systems (i.e., the recommendation and conversation module) through abundant external data. The two pre-trained modules enable LATTE as a domain expert for persuasive recommendation systems.

Table 9   Performance of various CogAgent models on the ReDial and INSPIRED benchmarks. The best results are highlighted in bold

Methods	ReDial						INSPIRED					
	H@1↑	H@10↑	H@50↑	Dist-3↑	Dist-4↑	HitX↑	H@1↑	H@10↑	H@50↑	Dist-3↑	Dist-4↑	HitX↑
KGSF [174]	3.55	18.04	37.89	0.40	0.50	2.40	6.05	15.88	27.75	2.40	3.20	11.40
CR-Walker [177]	3.70	17.60	37.10	3.00	3.70	3.20	−	−	−	−	−	−
RevCore [178]	4.05	21.21	41.02	0.90	1.50	8.30	7.20	16.53	33.47	6.00	6.80	30.70
C²-CRS [179]	4.32	21.50	41.20	1.70	1.90	2.90	6.38	21.28	36.60	7.50	9.00	6.60
UniCRS [176]	5.03	21.82	41.62	0.70	0.90	0.70	9.01	25.00	41.23	5.50	6.50	0.10
LATTE [180]	**5.61**	**23.93**	**44.01**	**5.10**	**6.00**	**97.00**	**9.79**	**27.23**	**44.26**	**9.50**	**10.10**	**95.40**

5   Open issues and future trends

Though researchers have made considerable efforts to address the above challenges in CogAgent, there are still open issues to be resolved. In this section, we present some open issues and future development trends for CogAgent to promote the advancement of the research community.

5.1   Comprehensive modeling of cognitive psychology theory for CogAgent

Although we have summarized some of the cognitive psychology theories, a comprehensive investigation of the cognitive mechanisms of persuasive dialogues from a cognitive psychology perspective is essential for understanding users' cognitive weaknesses and generating engaging persuasive dialogues. Many researchers have demonstrated the indispensability of employing specific strategies to achieve persuasive effects based on different cognitive psychology theories. Utilizing cognitive strategies, CogAgent can avoid cognitive dissonance in users and efficiently persuade them to accept specific viewpoints [8,55,99]. Prakken et al. [181] argue that psychological dissonance occurs when individuals are confronted with multiple conflicting cognitions. To alleviate this dissonance, three approaches can be used: changing cognitively relevant factors in the environment, introducing new cognitive elements, and changing cognitive elements in behavior. CogAgent should be aware of cognitive dissonance to mitigate the obstacles it creates in persuasion. In addition, researchers utilize the dual process theory of persuasion and guide the persuasive process with the Elaboration Likelihood Model (ELM) [28], a theory that focuses on cognitive and affective appeals in persuasion. Another noteworthy aspect is modeling the user's cognition. Proposing agreements or making concessions promptly facilitates the perception of the user's cognitive state, enabling CogAgent to adapt to changes in the user's cognition on time and avoiding the failure of the persuasive process [51].

Besides using data analysis to study the mechanisms of persuasive dialog, we can also explore this phenomenon from the perspective of the cognitive functions of the human brain. Advances in neuroscience have provided valuable methods for studying the cognitive mechanisms of persuasive dialogue. As Poldrack et al. state [182], the use of electroencephalography (EEG), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and other brain-imaging tools can deepen our understanding of how the human brain produces social behavior. Arapakis et al. [183] use brainwave recordings to measure users' interest in news articles, and the experimental results suggest that frontal asymmetry (FFA) can objectively assess users' receptive preferences for content. Exploring the changes in neural signals in the brain of the persuadee during persuasive conversations to model which persuasive factors are effective in being accepted by users and convincing them to adopt persuasive targets is a promising research direction.

5.2   Enriched cognitive strategy mining for CogAgent

In this paper, we have ventured into the intricate realm of persuasive dialogue systems, categorizing cognitive strategies derived from cognitive psychology theory into three types: persuasion strategy, topic path planning strategy, and argument structure prediction strategy. This classification, inspired by both foundational theories and a comprehensive review of existing works, aims to summarize the broad spectrum of research in this domain. Notably, theories such as Pre-suasion, Principle of Consistency, Theory of Mind, and Aristotle's rhetoric principles, provide a robust theoretical underpinning for the design and effectiveness of CogAgent.

While these categories offer a structured approach to understanding the dynamics of persuasion in dialogue systems, the complexity and nuance of human persuasion reveal a vast landscape of unexplored cognitive mechanisms. These mechanisms hold the potential to revolutionize the design and efficacy of CogAgent. The challenge and opportunity lie in mining novel cognitive strategies that can navigate and leverage the intricate psychological processes involved in persuasion. The dynamic and multifaceted nature of human cognition suggests the vast potential for innovative combinations and applications of these strategies.

The integration of cognitive dissonance theory, for example, could enhance the understanding of how individuals reconcile conflicting beliefs and information during persuasion, offering new pathways for strategy development. Furthermore, incorporating insights from dual-process theories, such as the Elaboration Likelihood Model (ELM) [3] and the Heuristic-Systematic Model (HSM) [184], could lead to more sophisticated approaches that adapt to the persuadee's level of cognitive engagement and processing route preference. These theories suggest that persuasion can be achieved through either systematic processing of arguments or heuristic cues, depending on the individual's motivation and ability to process the message.

Moreover, exploring the role of narrative persuasion—how stories and narratives can influence beliefs and attitudes —could unlock new dimensions in persuasive dialogue systems. Narratives are a powerful tool for persuasion because they allow individuals to experience vicarious emotions and to process information in a more engaged and less resistant manner [185]. This approach could be particularly effective in conjunction with existing strategies, offering a more immersive and emotionally resonant persuasion experience. The intersection of social identity theory with persuasion strategies presents another promising area. Understanding how individuals' identification with various social groups influences their reception to persuasive messages could guide the development of dialogue systems that leverage shared identity and values to foster persuasion [186].

To explore these cognitive strategies further, research must delve into interdisciplinary studies, combining findings from psychology, neuroscience, and computational linguistics. Mining cognitive strategies and modeling persuasive processes from a more comprehensive cognitive psychological perspective is a promising direction to promote CogAgent development.

5.3   Model adaptivity/generality of CogAgent

Equipping CogAgent with cross-domain understanding and

generation capabilities is a promising research direction. Existing CogAgent usually focuses on one specific persuasion scenario, such as persuasion for social good, bargaining, and debating. However, it is crucial to develop the ability of CogAgent to understand and transfer through multiple domains, which enables CogAgent to dynamically optimize cognitive strategies based on different persuasive targets and efficiently perform persuasive tasks. For example, Wolf et al. [187] utilize transfer learning to jointly fine-tune multiple unsupervised response prediction tasks. They demonstrate the effectiveness of language model transfer learning on the PERSONA-CHAT dataset, especially on the dialogue response generation task. Qian et al. [188] propose a meta-learning-based approach to domain adaptive dialogue generation that learns from multiple resource-rich tasks. They utilize multiple resource-rich single-domain dialog datasets to train the dialogue system so that it can adapt to new domains with minimal training samples. Therefore, improving the transferability of CogAgent across different domains using transfer learning and other advanced approaches is an important step towards the universal CogAgent.

5.4 Multi-party CogAgent

Existing research of CogAgent has demonstrated remarkable performance in two-party conversational scenarios. However, in real world, multi-party conversations (MPCs) are more prevalent and require CogAgent to persuade multiple participants simultaneously. Unlike existing persuasive dialog systems, multi-party dialogue scenarios require the collaboration of multiple CogAgent to efficiently achieve persuasion targets [189–191]. Specifically, a single CogAgent is prone to be overly purposeful when interacting with users, which can cause the users' resentment and resistance and hinder the realization of persuasion targets. In contrast, multiple CogAgent can assume different persuasive roles, cooperate, and persuade from different perspectives, thus winning users' trust and realizing persuasion targets more effectively. Existing studies have explored MPCs in open-domain dialogue systems. For instance, Ito et al. [192] construct a multi-modal and multi-party model based on GRU to predict the persuasiveness of multiple members within a group during multi-party conversations, thereby providing a model paradigm for the study of multi-party dialogues. Gu et al. [193] propose a Speaker-Aware BERT (SABERT) model to select appropriate speaking targets from multiple users based on dialogue contexts. Gu et al. [194] explore the problem of "who says what to whom" in MPCs and propose a plug-and-play graphically-induced fine-tuning (GIFT) module for tuning a variety of PLMs for generalized multi-party conversation understanding. Inspired by multi-party dialogue research, it is promising to utilize multiple CogAgent to collaborate on persuasive tasks to enhance the credibility and efficiency of the persuasion process. Multiple CogAgent utilize persuasive roles with complementary capabilities, strategies, and trust-building to enhance persuasion and effectiveness, thereby facilitating more persuasive and successful persuasion results.

5.5 Interpretability of persuasive process

Interpretability of models can improve their credibility.

Improving the interpretability of the persuasion process is essential to ensure that persuasive dialogue contents produced by CogAgent are accepted and adopted. In recent years, the field of Natural Language Processing (NLP) has increasingly focused on improving the interpretability of deep models [195–197]. For example, Gaur et al. [198] argue that domain-specific knowledge helps to understand how deep models work. They demonstrate the utility of incorporating knowledge-infused learning in knowledge graph format into complex neural networks to achieve model interpretability. Similarly, Yasunaga et al. [199] demonstrate model interpretability and structure inference by combining a pre-trained language model a knowledge graph, and a quality assurance context into a unified graph. Currently, research on the interpretability of the persuasion process still lacks an overall framework. For the interpretability of the persuasion process, the effectiveness of the persuasion strategy can be verified from the cognitive theory, combined with the knowledge graph reasoning, and the persuasion behavior can be analyzed interactively.

5.6 Multimodal CogAgent

Multimodal perception and comprehension capabilities are essential for human beings in daily conversations. By understanding the multimodal surroundings around them, including visual, textual, auditory, and other modal information, we humans can produce engaging dialogues to communicate messages, emotions, and attitudes with others [200,201]. Despite the outstanding natural language understanding and generation capabilities, perceiving and understanding multimodal context information is essential for natural and harmonious human-machine conversation systems [202,203]. To persuade people to change their thoughts, opinions, or attitudes, it is crucial to understand the multimodal surroundings of users. Different environments may lead users to develop different attitudes towards things. Combining multimodal contextual information, persuasive dialogue systems can comprehensively understand users' mental states to generate more specific persuasive dialogue content. There has been extensive research on multimodal dialogue systems that enable the understanding of image or video content through dialogue [204–206]. For example, Murahari et al. adapt ViLBERT [207] to achieve multi-turn image-based dialogue, which understands the image information through image-text pre-trained on multimodal datasets. Visual ChatGPT [206] integrates ChatGPT with visual foundation models to achieve visual dialogue. Different kinds of visual information, such as images, depth images, and mask matrices, are converted into language formats based on visual foundation models and the prompt manager. Then ChatGPT takes the information from visual and textual modalities to generate dialogue responses. These efforts have laid a solid foundation for multimodal persuasive dialogue systems. The integration of multimodal information to generate more persuasive conversational content is a highly promising research direction.

5.7 Data and model co-optimization for CogAgent

The huge impact of LLMs (e.g., ChatGPT) in the field of

dialog systems has sparked the enthusiasm of researchers and has been widely used in many domains [208–210]. For example, Liang et al. [209] rewrite the policy code for controlling a robot using LLMs. The policy code can receive and understand commands and then outputs the execution code to the API to achieve coherent control of the robot's actions through the classical chain logic. Similarly, Wen et al. [210] combine the common-sense knowledge implicit in LLMs with the domain-specific knowledge of mobile applications to realize hands-free speech-based interaction between users and smartphones. LLM can be surprisingly useful in a variety of domains. To develop a high-quality CogAgent, we can utilize LLMs to generate large-scale persuasive dialogue data to quickly validate the algorithm at an early stage. Since the capability of LLMs stems from massive amounts of data, retraining this data is hugely expensive. Therefore, the persuasion process also needs to be modeled to efficiently and accurately perform the persuasion task. The combination of data-driven LLMs and model-driven persuasion process is the most efficient way to develop intelligent CogAgent. Future research directions for combining LLMs and model-driven persuasion processes include issues such as when to employ the generation abilities of LLMs, when model constraints are needed, and the rules and timing of collaboration between LLMs and the persuasion process.

5.8   Construction of standardized datasets and benchmarks

Despite the significant progress researchers have made in CogAgent, datasets, and benchmarks for the study of CogAgent are still scarce. The relatively small size of many existing datasets (e.g., Persuasion for good [8]) limits the performance of the model in a wider range of applications. The limited amount of data hinders the ability to capture the full complexity and diversity of persuasive dialogue. Moreover, the lack of detailed annotations about cognitive strategies in existing datasets creates challenges for training persuasive dialogue agents. Building large-scale, high-quality datasets of persuasive dialogues with rich cognitive strategy annotations is indispensable for the development of CogAgent. Combining the superior text generation capabilities of LLMs [211,212] is a potential way to build large-scale and high-quality datasets for CogAgent.

6   Conclusion

Persuasion is an essential ability in human social communication, and people often skillfully persuade others to accept their standpoints, views, or perspectives for various purposes. Consequently, persuasive dialogue systems have become an engaging research direction. In this paper, we have made a systematic survey of CogAgent. We first present some representative cognitive psychology theories to guide the design of CogAgent at the principle level and formalize the necessary cognitive strategies for generating highly persuasive dialogue contents, including the persuasion strategy, the topic path planning strategy, and the argument structure prediction strategy. Based on the formalized definition and generic architecture of CogAgent, we comprehensively investigate

representative works by categorizing cognitive strategies. The available datasets and evaluation metrics for CogAgent are also summarized. Despite significant progress, the research of CogAgent is still in the early stage and massive open issues and prospective future trends to be explored, such as model adaptivity/generality of CogAgent, multi-party CogAgent, and multimodal CogAgent.

References

1.   Guo B, Wang H, Ding Y, Wu W, Hao S, Sun Y, Yu Z. Conditional text generation for harmonious human-machine interaction. ACM Transactions on Intelligent Systems and Technology, 2021, 12(2): 14

2.   Huang M, Zhu X, Gao J. Challenges in building intelligent open-domain dialog systems. ACM Transactions on Information Systems, 2020, 38(3): 21

3.   Petty R E, Cacioppo J T. The elaboration likelihood model of persuasion. In: Petty R E, Cacioppo J T, eds. Central and Peripheral Routes to Attitude Change. New York: Springer, 1986

4.   Fogg B J. Persuasive technology: using computers to change what we think and do. Ubiquity, 2002, 2002: 5

5.   IJsselsteijn W, De Kort Y, Midden C, Eggen B, van den Hoven E. Persuasive technology for human well-being: setting the scene. In: Proceedings of the 1st International Conference on Persuasive Technology for Human Well-Being. 2006, 1−5

6.   Fogg B J. Mass interpersonal persuasion: an early view of a new phenomenon. In: Proceedings of the 3rd International Conference on Persuasive Technology. 2008, 23−34

7.   Wood W. Attitude change: persuasion and social influence. Annual Review of Psychology, 2000, 51: 539−570

8.   Wang X, Shi W, Kim R, Oh Y, Yang S, Zhang J, Yu Z. 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. 2019, 5635−5649

9.   Slonim N, Bilu Y, Alzate C, Bar-Haim R, Bogin B, et al. An autonomous debating system. Nature, 2021, 591(7850): 379−384

10. Kolenik T, Gams M. Intelligent cognitive assistants for attitude and behavior change support in mental health: state-of-the-art technical review. Electronics, 2021, 10(11): 1250

11. Zhou K, Zhou Y, Zhao W X, Wang X, Wen J R. Towards topic-guided

conversational recommender system. In: Proceedings of the 28th International Conference on Computational Linguistics, 2020. 4128−4139

12. Torning K, Oinas-Kukkonen H. Persuasive system design: state of the art and future directions. In: Proceedings of the 4th International Conference on Persuasive Technology. 2009, 30

13. Eagly A H, Chaiken S. Cognitive theories of persuasion. Advances in Experimental Social Psychology, 1984, 17: 267−359

14. Shi W, Wang X, Oh Y J, Zhang J, Sahay S, Yu Z. Effects of persuasive dialogues: testing bot identities and inquiry strategies. In: Proceedings of 2020 CHI Conference on Human Factors in Computing Systems. 2020, 1−13

15. Joshi R, Balachandran V, Vashishth S, Black A W, Tsvetkov Y. Dialograph: incorporating interpretable strategy-graph networks into negotiation dialogues. In: Proceedings of the 9th International Conference on Learning Representations. 2021

16. Min B, Ross H, Sulem E, Veyseh A P B, Nguyen T H, Sainz O, Agirre E, Heintz I, Roth D. Recent advances in natural language processing via large pre-trained language models: a survey. ACM Computing Surveys, 2023, 56(2): 30

17. Zhao W X, Zhou K, Li J, Tang T, Wang X, et al. A survey of large language models. 2023, arXiv preprint arXiv: 2303.18223

18. Touvron H, Martin L, Stone K, Albert P, Almahairi A, et al. Llama 2: open foundation and fine-tuned chat models. 2023, arXiv preprint arXiv: 2307.09288

19. Zhou C, Li Q, Li C, Yu J, Liu Y, et al. A comprehensive survey on pretrained foundation models: a history from BERT to chatGPT. 2023, arXiv preprint arXiv: 2302.09419

20. Ray P P. ChatGPT: a comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet of Things and Cyber-Physical Systems, 2023, 3: 121−154

21. Li J, Han D, Guo Z, Qiao B, Wu G. Generating empathetic responses through emotion tracking and constraint guidance. Frontiers of Computer Science, 2024, 18(2): 182330

22. Wang W, Feng S, Song K, Wang D, Li S. Informative and diverse emotional conversation generation with variational recurrent pointer-generator. Frontiers of Computer Science, 2022, 16(5): 165326

23. Breckler S J, Wiggins E C. Cognitive responses in persuasion: affective and evaluative determinants. Journal of Experimental Social Psychology, 1991, 27(2): 180−200

24. Johnson B T, Eagly A H. Effects of involvement on persuasion: a meta-analysis. Psychological Bulletin, 1989, 106(2): 290−314

25. Friestad M, Wright P. The persuasion knowledge model: how people cope with persuasion attempts. Journal of Consumer Research, 1994, 21(1): 1−31

26. Petty R, Ostrom T M, Brock T C. Cognitive responses in persuasion. New York: Psychology Press, 2014

27. Bless H, Bohner G, Schwarz N, Strack F. Mood and persuasion: a cognitive response analysis. Personality and Social Psychology Bulletin, 1990, 16(2): 331−345

28. Petty R E, Briñol P. Emotion and persuasion: cognitive and meta-cognitive processes impact attitudes. Cognition and Emotion, 2015, 29(1): 1−26

29. Shevchuk N, Degirmenci K, Oinas-Kukkonen H. Adoption of gamified persuasive systems to encourage sustainable behaviors: Interplay between perceived persuasiveness and cognitive absorption. In: Proceedings of International Conference on Information Systems. 2019

30. Cheng Y, Liu W, Li W, Wang J, Zhao R, Liu B, Liang X, Zheng Y. Improving multi-turn emotional support dialogue generation with lookahead strategy planning. In: Proceedings of 2022 Conference on Empirical Methods in Natural Language Processing. 2022, 3014−3026

31. Qin J, Ye Z, Tang J, Liang X. Dynamic knowledge routing network for target-guided open-domain conversation. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence. 2020, 8657−8664

32. Zhong P, Liu Y, Wang H, Miao C. Keyword-guided neural conversational model. In: Proceedings of the 35th AAAI Conference on Artificial Intelligence. 2021, 14568−14576

33. Wang J, Lin D, Li W. A target-driven planning approach for goal-directed dialog systems. IEEE Transactions on Neural Networks and Learning Systems, 2023, doi: 10.1109/TNNLS.2023.3242071

34. Prakken H. A persuasive chatbot using a crowd-sourced argument graph and concerns. Computational Models of Argument, 2020, 326: 9

35. Tran N, Litman D. Multi-task learning in argument mining for persuasive online discussions. In: Proceedings of the 8th Workshop on Argument Mining. 2021, 148−153

36. Srivastava P, Bhatnagar P, Goel A. Argument mining using BERT and self-attention based embeddings. In: Proceedings of the 4th International Conference on Advances in Computing, Communication Control and Networking. 2022, 1536−1540

37. Dijkstra A. The psychology of tailoring-ingredients in computer-tailored persuasion. Social and Personality Psychology Compass, 2008, 2(2): 765−784

38. Chen M, Shi W, Yan F, Hou R, Zhang J, Sahay S, Yu Z. 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. 2022, 399−413

39. Duerr S, Gloor P A. Persuasive natural language generation−a literature review. 2021, arXiv preprint arXiv: 2101.05786

40. Zhan H, Wang Y, Feng T, Hua Y, Sharma S, Li Z, Qu L, Haffari G. Let's negotiate! A survey of negotiation dialogue systems. 2022, arXiv preprint arXiv: 2212.09072

41. Deng Y, Lei W, Lam W, Chua T S. A survey on proactive dialogue systems: problems, methods, and prospects. In: Proceedings of the 32nd International Joint Conference on Artificial Intelligence. 2023, 6583−6591

42. Cialdini R. Pre-suasion: A Revolutionary Way to Influence and Persuade. New York: Simon & Schuster, 2016

43. Bilu Y, Gera A, Hershcovich D, Sznajder B, Lahav D, Moshkowich G, Malet A, Gavron A, Slonim N. Argument invention from first principles. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019, 1013−1026

44. Premack D, Woodruff G. Does the chimpanzee have a theory of mind? Behavioral and Brain Sciences, 1978, 1(4): 515−526

45. Wu J, Chen Z, Deng J, Sabour S, Huang M. COKE: a cognitive knowledge graph for machine theory of mind. 2023, arXiv preprint arXiv: 2305.05390

46. Sap M, Le Bras R, Fried D, Choi Y. Neural theory-of-mind? On the limits of social intelligence in large LMs. In: Proceedings of 2022 Conference on Empirical Methods in Natural Language Processing. 2022, 3762−3780

47. Roman H R, Bisk Y, Thomason J, Celikyilmaz A, Gao J. RMM: a recursive mental model for dialogue navigation. In: Proceedings of Findings of the Association for Computational Linguistics. 2020, 1732−1745

48. Campbell G, Bitzer L F. The Philosophy of Rhetoric. Illinois: Southern Illinois University Press, 1988

49. Liu S, Zheng C, Demasi O, Sabour S, Li Y, Yu Z, Jiang Y, Huang M. 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. 2021, 3469−3483

50. Lopes J D, Hastie H. The language of persuasion, negotiation and trust. In: Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue. 2021, 1−12

51. Thimm M. Strategic argumentation in multi-agent systems. KI-Künstliche Intelligenz, 2014, 28(3): 159−168

52. Maher M L, Balachandran M B, Zhang D M. Case-based Reasoning in Design. New York: Psychology Press, 2014

53. Asch S E. Opinions and social pressure. Scientific American, 1955, 193(5): 31−35

54. Xu F, Warkentin M. Integrating elaboration likelihood model and herd theory in information security message persuasiveness. Computers & Security, 2020, 98: 102009

55. Chen Y, Deng S, Kwak D H, Elnoshokaty A, Wu J. A multi-appeal model of persuasion for online petition success: a linguistic cue-based approach. Journal of the Association for Information Systems, 2019. 20(2): 105−131

56. McGuire W J. The effectiveness of supportive and refutational defenses in immunizing and restoring beliefs against persuasion. Sociometry, 1961, 24(2): 184−197

57. He H, Chen D, Balakrishnan A, Liang P. Decoupling strategy and generation in negotiation dialogues. In: Proceedings of 2018 Conference on Empirical Methods in Natural Language Processing, 2018. 2333−2343

58. Cacioppo J T, Petty R E. Effects of message repetition and position on cognitive response, recall, and persuasion. Journal of Personality and Social Psychology, 1979, 37(1): 97−109

59. Cialdini R B. Influence: The Psychology of Persuasion. New York: Collins Business, 2007

60. Ni J, Pandelea V, Young T, Zhou H, Cambria E. HiTKG: towards goal-oriented conversations via multi-hierarchy learning. In: Proceedings of the 36th AAAI Conference on Artificial Intelligence. 2022, 11112−11120

61. Tang Z H, Yeh M Y. EAGLE: enhance target-oriented dialogs by global planning and topic flow integration. In: Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. 2023, 2402−2411

62. Petty R E, Cacioppo J T. Communication and Persuasion: central and Peripheral Routes to Attitude Change. New York: Springer, 2012

63. Swanson R, Ecker B, Walker M. Argument mining: extracting arguments from online dialogue. In: Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue. 2015, 217−226

64. Chakrabarty T, Hidey C, Muresan S, McKeown K, Hwang A. AMPERSAND: argument mining for persuasive online discussions. In: Proceedings of 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. 2019, 2933−2943

65. Rach N, Schindler C, Feustel I, Daxenberger J, Minker W, Ultes S. From argument search to argumentative dialogue: a topic-independent approach to argument acquisition for dialogue systems. In: Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue. 2021, 368−379

66. Wambsganss T, Kueng T, Soellner M, Leimeister J M. ArgueTutor: an adaptive dialog-based learning system for argumentation skills. In: Proceedings of 2021 CHI Conference on Human Factors in Computing Systems. 2021, 683

67. Ni J, Young T, Pandelea V, Xue F, Cambria E. Recent advances in deep learning based dialogue systems: a systematic survey. Artificial Intelligence Review, 2023, 56(4): 3055−3155

68. Bubeck S, Chandrasekaran V, Eldan R, Gehrke J, Horvitz E, Kamar E,

Lee P, Lee Y T, Li Y, Lundberg S, Nori H, Palangi H, Ribeiro M T, Zhang Y. Sparks of artificial general intelligence: early experiments with GPT-4. 2023, arXiv preprint arXiv: 2303.12712

69. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997, 9(8): 1735−1780

70. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł, Polosukhin I. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017, 6000−6010

71. Mishra K, Samad A M, Totala P, Ekbal A. PEPDS: a polite and empathetic persuasive dialogue system for charity donation. In: Proceedings of the 29th International Conference on Computational Linguistics. 2022, 424−440

72. Walker E R, McGee R E, Druss B G. Mortality in mental disorders and global disease burden implications: a systematic review and meta-analysis. JAMA Psychiatry, 2015, 72(4): 334−341

73. Xu B, Zhuang Z. Survey on psychotherapy chatbots. Concurrency and Computation: Practice and Experience, 2022, 34(7): e6170

74. Liang Y, Liu L, Ji Y, Huangfu L, Zeng D D. Identifying emotional causes of mental disorders from social media for effective intervention. Information Processing & Management, 2023, 60(4): 103407

75. Zhou J, Zheng C, Wang B, Zhang Z, Huang M. CASE: aligning coarse-to-fine cognition and affection for empathetic response generation. In: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. 2023, 8223−8237

76. Bosselut A, Rashkin H, Sap M, Malaviya C, Celikyilmaz A, Choi Y. COMET: commonsense transformers for automatic knowledge graph construction. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019, 4762−4779

77. Speer R, Chin J, Havasi C. ConceptNet 5.5: an open multilingual graph of general knowledge. In: Proceedings of the 31st AAAI Conference on Artificial Intelligence. 2017, 4444−4451

78. Nortio E, Jasinskaja-Lahti I, Hämäläinen M, Pakkasvirta J. Fear of the Russian bear? Negotiating finnish national identity online. Nations and Nationalism, 2022, 28(3): 861−876

79. Sakai K, Higashinaka R, Yoshikawa Y, Ishiguro H, Tomita J. Hierarchical argumentation structure for persuasive argumentative dialogue generation. IEICE Transactions on Information and Systems, 2020, E103(2): 424−434

80. Rach N, Minker W, Ultes S. Increasing the naturalness of an argumentative dialogue system through argument chains. Computational Models of Argument (COMMA 2020), 2020. 331−338

81. Gupta P, Jhamtani H, Bigham J. Target-guided dialogue response generation using commonsense and data augmentation. In: Proceedings of Findings of the Association for Computational Linguistics. 2022, 1301−1317

82. Hua W, Li L, Xu S, Chen L, Zhang Y. Tutorial on large language models for recommendation. In: Proceedings of the 17th ACM Conference on Recommender Systems. 2023, 1281−1283

83. Wu L, Zheng Z, Qiu Z, Wang H, Gu H, Shen T, Qin C, Zhu C, Zhu H, Liu Q, Xiong H, Chen E. A survey on large language models for recommendation. 2023, arXiv preprint arXiv: 2305.19860

84. Harte J, Zorgdrager W, Louridas P, Katsifodimos A, Jannach D, Fragkoulis M. Leveraging large language models for sequential recommendation. In: Proceedings of the 17th ACM Conference on Recommender Systems, 2023. 1096−1102

85. Mondal P. A unifying perspective on perception and cognition through linguistic representations of emotion. Frontiers in Psychology, 2022, 13: 768170

86. Shettleworth S J. Cognition, Evolution, and Behavior. Oxford: Oxford University Press, 2010

87. Nguyen H, Masthoff J. Designing persuasive dialogue systems: using

argumentation with care. In: Proceedings of the 3rd International Conference on Persuasive Technology. 2008, 201−212

88. Orji R. Why are persuasive strategies effective? Exploring the strengths and weaknesses of socially-oriented persuasive strategies. In: Proceedings of the 12th International Conference on Persuasive Technology, 2017. 253−266

89. Ham J, Bokhorst R, Cuijpers R, van der Pol D, Cabibihan J J. Making robots persuasive: the influence of combining persuasive strategies (gazing and gestures) by a storytelling robot on its persuasive power. In: Proceedings of the 3rd International Conference on Social Robotics, 2011. 71−83

90. Samad A M, Mishra K, Firdaus M, Ekbal A. Empathetic persuasion: reinforcing empathy and persuasiveness in dialogue systems. In: Proceedings of Findings of the Association for Computational Linguistics. 2022, 844−856

91. Papineni K, Roukos S, Ward T, Zhu W J. Bleu: a method for automatic evaluation of machine translation. In: Proceedings of the 40th annual meeting of the Association for Computational Linguistics. 2002, 311−318

92. Banerjee S, Lavie A. METEOR: an automatic metric for MT evaluation with improved correlation with human judgments. In: Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization. 2005, 65−72

93. Lin C Y. ROUGE: a package for automatic evaluation of summaries. In: Proceedings of Text Summarization Branches Out. 2004, 74−81

94. Yu X, Chen M, Yu Z. Prompt-based Monte-Carlo tree search for goal-oriented dialogue policy planning. In: Proceedings of 2023 Conference on Empirical Methods in Natural Language Processing. 2023, 7101−7125

95. Zhou Y, Tsvetkov Y, Black A W, Yu Z. Augmenting non-collaborative dialog systems with explicit semantic and strategic dialog history. In: Proceedings of the 8th International Conference on Learning Representations. 2020

96. Jia M, Chen Q, Jing L, Fu D, Li R. Knowledge-enhanced memory model for emotional support conversation. 2023, arXiv preprint arXiv: 2310.07700

97. Tu Q, Li Y, Cui J, Wang B, Wen J R, Yan R. MISC: a mixed strategy-aware model integrating COMET for emotional support conversation. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. 2022, 308−319

98. Liu C, Gao C, Yuan Y, Bai C, Luo L, Du X, Shi X, Luo H, Jin D, Li Y. Modeling persuasion factor of user decision for recommendation. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2022, 3366−3376

99. Yang R, Chen J, Narasimhan K. 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. 2021, 681−693

100. Greene J O, Burleson B R. Handbook of Communication and Social Interaction Skills. Mahwah: L. Erlbaum Associates, 2003

101. Hill C E. Helping Skills: Facilitating Exploration, Insight, and Action. 3rd ed. Washington: American Psychological Association, 2009

102. Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks. In: Proceedings of the 5th International Conference on Learning Representations. 2017

103. Velickovic P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y. Graph attention networks. Stat, 2017, 1050(20): 10-48550

104. Wu Z, Pan S, Chen F, Long G, Zhang C, Yu P S. A comprehensive survey on graph neural networks. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(1): 4−24

105. Wu L, Chen Y, Shen K, Guo X, Gao H, Li S, Pei J, Long B. Graph neural networks for natural language processing: a survey. Foundations and Trends® in Machine Learning, 2023, 16(2): 119−328

106. Wang H, Guo B, Liu J, Ding Y, Yu Z. Towards informative and diverse dialogue systems over hierarchical crowd intelligence knowledge graph. ACM Transactions on Knowledge Discovery from Data, 2023, 17(7): 105

107. Zheng C, Liu Y, Chen W, Leng Y, Huang M. CoMAE: a multi-factor hierarchical framework for empathetic response generation. In: Proceedings of Findings of the Association for Computational Linguistics. 2021, 813−824

108. Zheng Z, Liao L, Deng Y, Nie L. Building emotional support chatbots in the era of LLMs. 2023, arXiv preprint arXiv: 2308.11584

109. Xu J, Wang H, Niu Z, Wu H, Che W. Knowledge graph grounded goal planning for open-domain conversation generation. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence. 2020, 9338−9345

110. Liu J, Pan F, Luo L. GoChat: goal-oriented chatbots with hierarchical reinforcement learning. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2020, 1793−1796

111. Lei W, Zhang Y, Song F, Liang H, Mao J, Lv J, Yang Z, Chua T S. Interacting with non-cooperative user: a new paradigm for proactive dialogue policy. In: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2022, 212−222

112. Zou Y, Liu Z, Hu X, Zhang Q. Thinking clearly, talking fast: concept-guided non-autoregressive generation for open-domain dialogue systems. In: Proceedings of 2021 Conference on Empirical Methods in Natural Language Processing. 2021, 2215−2226

113. Wang J, Lin D, Li W. Dialogue planning via Brownian bridge stochastic process for goal-directed proactive dialogue. In: Proceedings of Findings of the Association for Computational Linguistics. 2023, 370−387

114. Ren X, Yin H, Chen T,Wang H, Zheng K. Learning to ask appropriate questions in conversational recommendation. In: Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval. 2021, 808−817

115. Yang Z, Wang B, Zhou J, Tan Y, Zhao D, Huang K, He R, Hou Y. TopKG: Target-oriented dialog via global planning on knowledge graph. In: Proceedings of the 29th International Conference on Computational Linguistics. 2022, 745−755

116. Mnih V, Badia A P, Mirza M, Graves A, Harley T, Lillicrap T P, Silver D, Kavukcuoglu K. Asynchronous methods for deep reinforcement learning. In: Proceedings of the 33rd International Conference on Machine Learning. 2016, 1928−1937

117. Wang H, Guo B, Wu W, Liu S, Yu Z. Towards information-rich, logical dialogue systems with knowledge-enhanced neural models. Neurocomputing, 2021, 465: 248−264

118. .Wu S, Li Y, Zhang D, Zhou Y, Wu Z. Diverse and informative dialogue generation with context-specific commonsense knowledge awareness. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020, 5811−5820

119. Tang J, Zhao T, Xiong C, Liang X, Xing E, Hu Z. Target-guided open-domain conversation. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019, 5624−5634

120. Vecchi E M, Falk N, Jundi I, Lapesa G. Towards argument mining for social good: a survey. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. 2021, 1338−1352

121. Al-Khatib K, Wachsmuth H, Hagen M, Köhler J, Stein B. Cross-domain mining of argumentative text through distant supervision. In:

Proceedings of 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016, 1395−1404

122. Hua X, Hu Z, Wang L. Argument generation with retrieval, planning, and realization. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019, 2661−2672

123. Niculae V, Park J, Cardie C. Argument mining with structured SVMs and RNNs. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017, 985−995

124. .Li J, Durmus E, Cardie C. Exploring the role of argument structure in online debate persuasion. In: Proceedings of 2020 Conference on Empirical Methods in Natural Language Processing, 2020, 8905−8912

125. Cheng L, Bing L, He R, Yu Q, Zhang Y, Si L. IAM: a comprehensive and large-scale dataset for integrated argument mining tasks. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. 2022, 2277−2287

126. Wang S, Yin Z, Zhang W, Zheng D, Li X. Two stage learning for argument pairs extraction. In: Proceedings of the 10th CCF International Conference on Natural Language Processing and Chinese Computing. 2021, 538−547

127. Sun J, Zhu Q, Bao J, Wu J, Yang C, Wang R, Xu R. A hierarchical sequence labeling model for argument pair extraction. In: Proceedings of the 10th CCF International Conference on Natural Language Processing and Chinese Computing. 2021, 472−483

128. Chawla N V, Bowyer K W, Hall L O, Kegelmeyer W P. SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 2002, 16: 321−357

129. Durmus E, Cardie C. A corpus for modeling user and language effects in argumentation on online debating. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019. 602−607

130. Zheng C, Sabour S, Wen J, Zhang Z, Huang M. AugESC: dialogue augmentation with large language models for emotional support conversation. In: Proceedings of Findings of the Association for Computational Linguistics. 2023, 1552−1568

131. Sun H, Lin Z, Zheng C, Liu S, Huang M. PsyQA: a Chinese dataset for generating long counseling text for mental health support. In: Proceedings of Findings of the Association for Computational Linguistics. 2021, 1489−1503

132. Walker M A, Tree J E F, Anand P, Abbott R, King J. A corpus for research on deliberation and debate. In: Proceedings of the 8th International Conference on Language Resources and Evaluation, 2012. 812−817

133. Abbott R, Ecker B, Anand P, Walker M. Internet argument corpus 2.0: an SQL schema for dialogic social media and the corpora to go with it. In: Proceedings of the 10th International Conference on Language Resources and Evaluation. 2016, 4445−4452

134. Tan C, Niculae V, Danescu-Niculescu-Mizil C, Lee L. Winning arguments: interaction dynamics and persuasion strategies in good-faith online discussions. In: Proceedings of the 25th International Conference on World Wide Web. 2016, 613−624

135. Roush A, Balaji A. DebateSum: a large-scale argument mining and summarization dataset. In: Proceedings of the 7th Workshop on Argument Mining. 2020, 1−7

136. Zhou Y, He H, Black A W, Tsvetkov Y. A dynamic strategy coach for effective negotiation. In: Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue. 2019, 367−378

137. Dutt R, Joshi R, Rose C. Keeping up appearances: computational modeling of face acts in persuasion oriented discussions. In: Proceedings of 2020 Conference on Empirical Methods in Natural Language Processing. 2020, 7473−7485

138. Li R, Kahou S, Schulz H, Michalski V, Charlin L, Pal C. Towards deep conversational recommendations. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. 2018, 9748−9758

139. Liu Z, Wang H, Niu Z Y, Wu H, Che W, Liu T. Towards conversational recommendation over multi-type dialogs. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020, 1036−1049

140. Hayati S A, Kang D, Zhu Q, Shi W, Yu Z. INSPIRED: toward sociable recommendation dialog systems. In: Proceedings of 2020 Conference on Empirical Methods in Natural Language Processing. 2020, 8142−8152

141. Lavie A, Agarwal A. Meteor: an automatic metric for MT evaluation with high levels of correlation with human judgments. In: Proceedings of the 2nd Workshop on Statistical Machine Translation. 2007, 228−231

142. .Vedantam R, Lawrence Zitnick C, Parikh D. CIDEr: consensus-based image description evaluation. In: Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. 2015, 4566−4575

143. Rus V, Lintean M. An optimal assessment of natural language student input using word-to-word similarity metrics. In: Proceedings of the 11th International Conference on Intelligent Tutoring Systems. 2012, 675−676

144. Wieting J, Bansal M, Gimpel K, Livescu K. Towards universal paraphrastic sentence embeddings. In: Proceedings of the 4th International Conference on Learning Representations. 2016

145. Forgues G, Pineau J, Larchevêque J, Tremblay R. Bootstrapping dialog systems with word embeddings. In: Proceedings of NIPS, Modern Machine Learning and Natural Language Processing Workshop. 2014, 2: 168

146. Lowe R, Noseworthy M, Serban I V, Angelard-Gontier N, Bengio Y, Pineau J. Towards an automatic turing test: learning to evaluate dialogue responses. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017, 1116−1126

147. Devlin J, Chang M W, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019, 4171−4186

148. Liu C W, Lowe R, Serban I, Noseworthy M, Charlin L, Pineau J. How not to evaluate your dialogue system: an empirical study of unsupervised evaluation metrics for dialogue response generation. In: Proceedings of 2016 Conference on Empirical Methods in Natural Language Processing. 2016, 2122−2132

149. Deriu J, Rodrigo A, Otegi A, Echegoyen G, Rosset S, Agirre E, Cieliebak M. Survey on evaluation methods for dialogue systems. Artificial Intelligence Review, 2021, 54(1): 755−810

150. Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, et al. GPT-4 technical report. 2024, arXiv preprint arXiv: 2303.08774

151. Du Z, Qian Y, Liu X, Ding M, Qiu J, Yang Z, Tang J. GLM: general language model pretraining with autoregressive blank infilling. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. 2022, 320−335

152. Wang J, Liang Y, Meng F, Sun Z, Shi H, Li Z, Xu J, Qu J, Zhou J. Is chatGPT a good NLG evaluator? A preliminary study. In: Proceedings of the 4th New Frontiers in Summarization Workshop. 2023, 1−11

153. Fu J, Ng S K, Jiang Z, Liu P. GPTScore: evaluate as you desire. 2023, arXiv preprint arXiv: 2302.04166

154. Zhong M, Liu Y, Yin D, Mao Y, Jiao Y, Liu P, Zhu C, Ji H, Han J. Towards a unified multi-dimensional evaluator for text generation. In: Proceedings of 2022 Conference on Empirical Methods in Natural Language Processing. 2022, 2023−2038

155. Chiang C H, Lee H Y. Can large language models be an alternative to human evaluations? In: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. 2023, 15607−15631

156. Liu Y, Iter D, Xu Y, Wang S, Xu R, Zhu C. G-Eval: NLG evaluation using gpt-4 with better human alignment. In: Proceedings of 2023 Conference on Empirical Methods in Natural Language Processing, 2023. 2511−2522

157. Wei J, Wang X, Schuurmans D, Bosma M, Ichter B, Xia F, Chi E H, Le Q V, Zhou D. Chain-of-thought prompting elicits reasoning in large language models. In: Proceedings of the 36th International Conference on Neural Information Processing Systems. 2022, 1800

158. Xu W, Wang D, Pan L, Song Z, Freitag M, Wang W, Li L. INSTRUCTSCORE: towards explainable text generation evaluation with automatic feedback. In: Proceedings of 2023 Conference on Empirical Methods in Natural Language Processing. 2023, 5967−5994

159. Chan C M, Chen W, Su Y, Yu J, Xue W, Zhang S, Fu J, Liu Z. ChatEval: towards better LLM-based evaluators through multi-agent debate. 2023, arXiv preprint arXiv: 2308.07201

160. Al Khatib K, Völske M, Syed S, Kolyada N, Stein B. Exploiting personal characteristics of debaters for predicting persuasiveness. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020, 7067−7072

161. Kang D, Kim S, Kwon T, Moon S, Cho H, Yu Y, Lee D, Yeo J. Can large language models be good emotional supporter? Mitigating preference bias on emotional support conversation. 2024, arXiv preprint arXiv: 2402.13211

162. Bullock O M, Shulman H C, Huskey R. Narratives are persuasive because they are easier to understand: examining processing fluency as a mechanism of narrative persuasion. Frontiers in Communication, 2021, 6: 719615

163. Patel V, Saxena S, Lund C, Thornicroft G, Baingana F, et al. The *lancet* commission on global mental health and sustainable development. The Lancet, 2018, 392(10157): 1553−1598

164. GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990−2019: a systematic analysis for the global burden of disease study 2019. The Lancet Psychiatry, 2022, 9(2): 137−150

165. Zhao W, Zhao Y, Wang S, Qin B. TransESC: smoothing emotional support conversation via turn-level state transition. In: Proceedings of Findings of the Association for Computational Linguistics. 2023, 6725−6739

166. Peng W, Hu Y, Xing L, Xie Y, Sun Y, Li Y. Control globally, understand locally: a global-to-local hierarchical graph network for emotional support conversation. In: Proceedings of the 31st International Joint Conference on Artificial Intelligence. 2022, 4324−4330

167. Peng W, Qin Z, Hu Y, Xie Y, Li Y. FADO: feedback-aware double controlling network for emotional support conversation. Knowledge-Based Systems, 2023, 264: 110340

168. Deng Y, Zhang W, Yuan Y, Lam W. Knowledge-enhanced mixed-initiative dialogue system for emotional support conversations. In: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. 2023, 4079−4095

169. Burleson B R. Emotional support skills. In: Greene J O, Burleson B R, eds. Handbook of Communication and Social Interaction Skills. Mahwah: Lawrence Erlbaum Associates Publishers, 2003. 569−612

170. Zhang Y, Sun S, Galley M, Chen Y C, Brockett C, Gao X, Gao J, Liu J, Dolan B. DIALOGPT: large-scale generative pre-training for conversational response generation. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations. 2020, 270−278

171. Roller S, Dinan E, Goyal N, Ju D, Williamson M, Liu Y, Xu J, Ott M, Smith E M, Boureau Y L, Boureau Y L, Weston J. Recipes for building an open-domain chatbot. In: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics. 2021, 300−325

172. Gao C, Lei W, He X, de Rijke M, Chua T S. Advances and challenges in conversational recommender systems: a survey. AI Open, 2021, 2: 100−126

173. Jannach D, Manzoor A, Cai W, Chen L. A survey on conversational recommender systems. ACM Computing Surveys, 2021, 54(5): 105

174. Zhou K, Zhao W X, Bian S, Zhou Y, Wen J R, Yu J. Improving conversational recommender systems via knowledge graph based semantic fusion. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020, 1006−1014

175. Chen Q, Lin J, Zhang Y, Ding M, Cen Y, Yang H, Tang J. Towards knowledge-based recommender dialog system. In: Proceedings of 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. 2019, 1803−1813

176. Wang X, Zhou K, Wen J R, Zhao W X. Towards unified conversational recommender systems via knowledge-enhanced prompt learning. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2022, 1929−1937

177. Ma W, Takanobu R, Huang M. CR-walker: tree-structured graph reasoning and dialog acts for conversational recommendation. In: Proceedings of 2021 Conference on Empirical Methods in Natural Language Processing. 2021, 1839−185

178. Lu Y, Bao J, Song Y, Ma Z, Cui S, Wu Y, He X. RevCore: review-augmented conversational recommendation. In: Proceedings of Findings of the Association for Computational Linguistics. 2021, 1161−117

179. Zhou Y, Zhou K, Zhao W X, Wang C, Jiang P, Hu H. C^2-CRS: coarse-to-fine contrastive learning for conversational recommender system. In: Proceedings of the 15th ACM International Conference on Web Search and Data Mining. 2022, 1488−1496

180. Kim T, Yu J, Shin W Y, Lee H, Im J H, Kim S W. LATTE: a framework for learning item-features to make a domain-expert for effective conversational recommendation. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023, 1144−1153

181. Wen T J, Kim E, Wu L, Dodoo N A. Activating persuasion knowledge in native advertising: the influence of cognitive load and disclosure language. International Journal of Advertising, 2020, 39(1): 74−93

182. Poldrack R A, Farah M J. Progress and challenges in probing the human brain. Nature, 2015, 526(7573): 371−379

183. Arapakis I, Barreda-Ángeles M, Pereda-Baños A. Interest as a proxy of engagement in news reading: spectral and entropy analyses of EEG activity patterns. IEEE Transactions on Affective Computing, 2019, 10(1): 100−114

184. Chaiken S. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. Journal of Personality and Social Psychology, 1980, 39(5): 752−766

185. Green M C, Brock T C. The role of transportation in the persuasiveness of public narratives. Journal of Personality and Social Psychology, 2000, 79(5): 701−721

186. Tajfel H. An integrative theory of intergroup conflict. The social psychology of intergroup relations. 1979, 33: 33-47

187. Wolf T, Sanh V, Chaumond J, Delangue C. TransferTransfo: a transfer learning approach for neural network based conversational agents. 2019, arXiv preprint arXiv: 1901.08149

188. Qian K, Yu Z. Domain adaptive dialog generation via meta learning. In: Proceedings of the 57th Annual Meeting of the Association for

Computational Linguistics, 2019. 2639−2649

189. Shi Z, Huang M. A deep sequential model for discourse parsing on multi-party dialogues. In: Proceedings of the 33rd AAAI Conference on Artificial Intelligence, 2019. 7007−7014

190. Ju D, Feng S, Lv P, Wang D, Zhang Y. Learning to improve persona consistency in multi-party dialogue generation via text knowledge enhancement. In: Proceedings of the 29th International Conference on Computational Linguistics. 2022, 298−309

191. Yuan L, Chen F, Zhang Z, Yu Y. Communication-robust multi-agent learning by adaptable auxiliary multi-agent adversary generation. Frontiers of Computer Science, 2024, 18(6): 186331

192. Ito A, Nakano Y I, Nihei F, Sakato T, Ishii R, Fukayama A, Nakamura T. Predicting persuasiveness of participants in multiparty conversations. In: Proceedings of the 27th International Conference on Intelligent User Interfaces. 2022, 85−88

193. Gu J C, Li T, Liu Q, Ling Z H, Su Z, Wei S, Zhu X. Speaker-aware BERT for multi-turn response selection in retrieval-based chatbots. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management. 2020, 2041−2044

194. Gu J C, Ling Z, Liu Q, Liu C, Hu G. GIFT: graph-induced fine-tuning for multi-party conversation understanding. In: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. 2023, 11645−11658

195. Belinkov Y, Gehrmann S, Pavlick E. Interpretability and analysis in neural NLP. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts. 2020, 1−5

196. Jacovi A, Goldberg Y. Towards faithfully interpretable NLP systems: how should we define and evaluate faithfulness? In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020. 4198−4205

197. Zhang Q, Guo B, Liu S, Liu J, Yu Z. CrowdDesigner: information-rich and personalized product description generation. Frontiers of Computer Science, 2022, 16(6): 166339

198. Gaur M, Faldu K, Sheth A. Semantics of the black-box: can knowledge graphs help make deep learning systems more interpretable and explainable? IEEE Internet Computing, 2021, 25(1): 51−59

199. Yasunaga M, Ren H, Bosselut A, Liang P, Leskovec J. QA-GNN: reasoning with language models and knowledge graphs for question answering. In: Proceedings of 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2021, 535−546

200. Quek F, McNeill D, Bryll R, Duncan S, Ma X F, Kirbas C, McCullough K E, Ansari R. Multimodal human discourse: gesture and speech. ACM Transactions on Computer-Human Interaction, 2002, 9(3): 171−193

201. Turk M. Multimodal interaction: a review. Pattern Recognition Letters, 2014, 36: 189−195

202. Jaimes A, Sebe N. Multimodal human−computer interaction: a survey. Computer Vision and Image Understanding, 2007, 108(1-2): 116−134

203. Baltrušaitis T, Ahuja C, Morency L P. Multimodal machine learning: a survey and taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(2): 423−443

204. Qi J, Niu Y, Huang J, Zhang H. Two causal principles for improving visual dialog. In: Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020, 10857−10866

205. Alamri H, Cartillier V, Das A, Wang J, Cherian A, Essa I, Batra D, Marks T K, Hori C, Anderson P, Lee S, Parikh D. Audio visual scene-aware dialog. In: Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019, 7550−7559

206. Wu C, Yin S, Qi W, Wang X, Tang Z, Duan N. Visual chatGPT: talking, drawing and editing with visual foundation models. 2023, arXiv preprint arXiv: 2303.04671

207. Lu J, Batra D, Parikh D, Lee S. ViLBERT: pretraining task-agnostic visiolinguistic representations for vision-and-language tasks. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems. 2019, 2

208. Pal S, Bhattacharya M, Lee S S, Chakraborty C. A domain-specific next-generation large language model (LLM) or chatGPT is required for biomedical engineering and research. Annals of Biomedical Engineering, 2024, 52(3): 451−454

209. Liang J, Huang W, Xia F, Xu P, Hausman K, Ichter B, Florence P, Zeng A. Code as policies: language model programs for embodied control. In: Proceedings of 2023 IEEE International Conference on Robotics and Automation. 2023, 9493−9500

210. Wen H, Li Y, Liu G, Zhao S, Yu T, Li T J J, Jiang S, Liu Y, Zhang Y, Liu Y. Empowering LLM to use smartphone for intelligent task automation. 2023, arXiv preprint arXiv: 2308.15272

211. Kim H, Hessel J, Jiang L, West P, Lu X, Yu Y, Zhou P, Le Bras R, Alikhani M, Kim G, Sap M, Choi Y. SODA: million-scale dialogue distillation with social commonsense contextualization. In: Proceedings of 2023 Conference on Empirical Methods in Natural Language Processing. 2023, 12930−12949

212. Zheng C, Sabour S, Wen J, Huang M. AugESC: large-scale data augmentation for emotional support conversation with pre-trained language models. 2022, arXiv preprint arXiv: 2202.13047

Mengqi Chen received her master's degree in digital textiles from Xi'an Polytechnical University (XPU), China in 2022. She is currently working toward a PhD degree at Northwestern Polytechnical University (NWPU), China. Her current research interests include natural language processing, dialog systems, and large language models.

Bin Guo is a PhD professor and PhD supervisor at Northwestern Polytechnical University (NWPU), China. He is a senior member of the China Computer Federation. His main research interests include ubiquitous computing, social and community intelligence, urban big data mining, mobile crowdsensing, and human-computer interaction.

Hao Wang received his BE degree in computer science and technology from Northwestern Polytechnical University (NWPU), China in 2019. He is currently working toward a PhD degree at NWPU. His current research interests include natural language processing, dialog systems, and large language models.
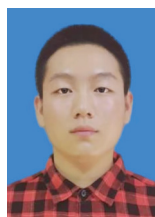
Haoyu Li received his BE degree in computer science and technology from Northwestern Polytechnical University (NWPU), China in 2023. He is currently working toward a master's degree at NWPU. His current research interests include natural language processing, large language models, and robot dynamic obstacle avoidance.

Qian Zhao received her BE degree in Internet of Things engineering from Tianjin University of Technology (TUT), China in 2023. She is currently working toward a master's degree at Northwestern Polytechnical University (NWPU), China. Her current research interests include multimodal dialogue, large language models, and visual human-computer interaction.

Jingqi Liu entered Northwestern Polytechnical University(NWPU) to study for a bachelor's degree in information and computing science in 2020. Her current research interests include natural language processing, dialogue systems, and large language models.

Yasan Ding received his BE degree in computer science and technology from Northwestern Polytechnical University (NWPU), China in 2018. He is currently working toward a PhD degree at NWPU. His current research interests include fake news detection and natural language processing.

Yan Pan is a lecturer at the Science and Technology on Information Systems Engineering Laboratory. He respectively received the BS degree in 2013 and the PhD degree in 2020 from Northwestern Polytechnical University (NWPU), China. His research interests include Big Data, Machine Learning, and Crowd Intelligence.

Zhiwen Yu is a PhD professor and PhD supervisor. He is a senior member of the China Computer Federation. His main research interests include mobile internet, ubiquitous computing, social and community intelligence, urban big data mining, mobile crowdsensing, and human-computer interaction.