Enhancing User-Oriented Proactivity in Open-Domain Dialogues with Critic Guidance

Yufeng Wang[1,2,3*], **Jinwu Hu**[1,3*], **Ziteng Huang**[1], **Kunyang Lin**[4], **Zitian Zhang**[1], **Peihao Chen**[5], **Yu Hu**[6], **Qianyue Wang**[1], **Zhuliang Yu**[1], **Bin Sun**[7], **Xiaofen Xing**[1†], **Qingfang Zheng**[2†], **Mingkui Tan**[1,3]

[1]South China University of Technology
[2]Peng Cheng Laboratory
[3]Pazhou Laboratory
[4]Tencent AI Lab
[5]Tencent Robotics X Lab
[6]Hong Kong Polytechnic University
[7]Hunan University
yufeng6568@gmail.com, xfxing@scut.edu.cn, zhengqf01@pcl.ac.cn

Abstract

Open-domain dialogue systems aim to generate natural and engaging conversations, providing significant practical value in real applications such as social robotics and personal assistants. The advent of large language models (LLMs) has greatly advanced this field by improving context understanding and conversational fluency. However, existing LLM-based dialogue systems often fall short in proactively understanding the user's chatting preferences and guiding conversations toward user-centered topics. This lack of user-oriented proactivity can lead users to feel unappreciated, reducing their satisfaction and willingness to continue the conversation in human-computer interactions. To address this issue, we propose a **U**ser-oriented **P**roactive **C**hatbot (**UPC**) to enhance the user-oriented proactivity. Specifically, we first construct a critic to evaluate this proactivity inspired by the LLM-as-a-judge strategy. Given the scarcity of high-quality training data, we then employ the critic to guide dialogues between the chatbot and user agents, generating a corpus with enhanced user-oriented proactivity. To ensure the diversity of the user backgrounds, we introduce the *ISCO-800*, a diverse user background dataset for constructing user agents. Moreover, considering the communication difficulty varies among users, we propose an iterative curriculum learning method that trains the chatbot from easy-to-communicate users to more challenging ones, thereby gradually enhancing its performance. Experiments demonstrate that our proposed training method is applicable to different LLMs, improving user-oriented proactivity and attractiveness in open-domain dialogues. Code and appendix are available at github.com/wang678/LLM-UPC.

1 Introduction

Humans are naturally more engaged in conversations relevant to their interests [Oertel *et al.*, 2020]. To meet humans' chatting preferences, the open-domain dialogue task aims to build chatbots that produce engaging and meaningful natural language responses [Kann *et al.*, 2022]. It yields practical value in various human-computer interaction applications, such as personal assistants [Luo *et al.*, 2022], education [Rodrigues *et al.*, 2022] and social robotics [Grassi *et al.*, 2022]. Recently, the rise of large language models (LLMs) [Zhao *et al.*, 2023; Hu *et al.*, 2024] has led to unprecedented popularity in dialogue systems. LLM-based dialogue systems have demonstrated extraordinary context understanding and coherent response generation capabilities [Deng *et al.*, 2023a], significantly advancing the development of open-domain dialogue.

Despite excellent dialogue capabilities of LLMs, they still have limitations due to the inability of *proactivity*. The proactive chatbot is designed to go beyond passive response to user inputs, but actively explore the user's chatting interests and lead the topics to the user's preferences. Since LLM-based methods heavily rely on the existing training conversations and knowledge to provide suggestions and complete tasks [Achiam *et al.*, 2023], they typically generate responses to user queries in a passive manner [Liao *et al.*, 2023]. They tend to help solve problems rather than proactively learn about the background and chatting interests of the user (Figure 1, left). Existing methods enhance proactivity by topic planning and shifting, or topic-aware response generation [Tang *et al.*, 2019; Deng *et al.*, 2023b; Deng *et al.*, 2023c; Wang *et al.*, 2024]. Although they can actively lead the chatting topic, these target-oriented methods focus more on target-centered topic switching, resulting in a lack of proactive attention toward the users themselves (Figure 1, medium). This may not fully align with the original intention of open-domain chatbots, which is to enhance the user's chat experience and create engaging dialogues. To address this, we **delve into the proactivity of exploring user's**

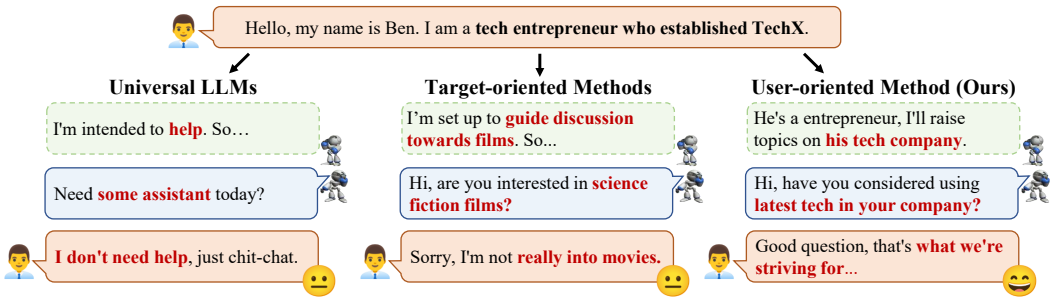

Figure 1: Illustration of different open-domain dialogue methods. Left: universal LLMs. Medium: target-oriented methods. Right: our user-oriented method. The green dotted box means the intention of the chatbot.

background and chatting interests and take proactivity to lead the conversation toward user-centered topics, namely the *User-oriented Proactivity* (Figure 1, right). It enables the chatbot to deliver responses that are not only relevant but also resonate with the user's chatting interests and background, thereby fostering more satisfying interaction.

In this paper, we propose a **U**ser-oriented **P**roactive **C**hatbot (**UPC**) to address the lack of user-oriented proactivity. We build a chatbot that explores the user's background and chatting interests, and actively leads the conversation towards user-centered topics. Specifically, since assessing proactivity by humans is costly, LLM can serve as a judge that aligns with humans [Zheng *et al.*, 2024]. We appoint an LLM to be a critic and design evaluation metrics for this proactivity. Besides, the lack of a high-quality training corpus for a proactive chatbot puts us in a dilemma of making bricks without straw. To this end, we introduce a critic-guided dialogue corpus generation paradigm. We employ the critic to guide dialogues between the chatbot and different user agents, generating a dialogue corpus with enhanced user-oriented proactivity. Since communication difficulty varies across users, the chatbot may struggle with users who have uncommon backgrounds or preferences in the early training stages. Therefore, we propose a communication difficulty-aware iterative curriculum learning approach. It adapts the chatbot to easy-to-communicate users in early training iterations and incrementally learns the chatbot with more challenging users, achieving steady performance improvement. Finally, we construct the *ISCO-800*, a dataset with 800 user backgrounds, to create diverse user agents. Our main contributions are as follows:

1) Design of a critic to assess user-oriented proactivity. To address the lack of user-oriented proactivity in the chatbot, we develop a critic to quantify and assess this proactivity. It also guides high-quality corpus generation for training. The real user evaluation verifies its alignment with humans.

2) Iterative curriculum learning is proposed for chatbot training. Recognizing that communication difficulty varies among different users, we define adaptation difficulty and introduce an iterative curriculum learning method. It first adapts the chatbot to easy-to-communicate users in early training iterations and progressively includes more challenging ones. Experiments show that this paradigm improves average performance by 8.2% compared to the original LLM.

3) Construction of a user background dataset *ISCO-800*. To train a chatbot capable of interacting with users from various backgrounds. We construct a user background dataset that includes 800 types of user background information. It provides realistic background information, including occupation, hobby, education, and personality, enabling the chatbot to interact with a wide range of users.

2 Related Works

Significant efforts have been made to enhance the human-like performance of open-domain dialogue systems. Based on key issues, existing studies relevant to our work can be broadly categorized into three areas: coherent dialogue, personalized dialogue, and target-oriented dialogue.

Coherent dialogue aims to improve contextual coherence during conversations. Memochat [Lu *et al.*, 2023] uses memorization-retrieval-response cycles to teach the LLMs to memorize and retrieve past dialogues with structured memos, leading to enhanced consistency. Sun *et al.* [Sun *et al.*, 2023] propose a hybrid latent variable method. It combines discrete and continuous latent variables to improve both semantic correlation and diversity. Although these methods improve coherence, they do not prioritize the proactive, user-centric dialogue that is central to our work.

Personalized dialogue focuses on building a chatbot endowed with a persona [Tu *et al.*, 2023; Chen *et al.*, 2023], or modeling the persona of the other party [Chen *et al.*, 2024; Zhong *et al.*, 2024]. For example, ORIG [Chen *et al.*, 2023] proposes a model-agnostic framework to fine-tune the persona dialogue model. It enables the dialogue models to learn robust representations and improve the consistency of response generation. In addition to improving the chatbot's role-playing abilities, Memorybank [Zhong *et al.*, 2024] adds long-term memory functionality for LLMs by summarizing past interactions to capture user personalities. However, unlike the static personality, the user's chatting interests evolve dynamically during a conversation. These methods often struggle to actively explore and engage with these changing interests, limiting their ability to generate topics that align with what the user is currently interested in.

Target-oriented dialogue improve proactivity by planning and steering the dialogue to pre-set topics [Tang *et al.*, 2019; Deng *et al.*, 2023b; Deng *et al.*, 2023c; Wang *et al.*, 2024]. For example, TRIP [Wang *et al.*, 2024] generates the dialogue path with a bidirectional planning method to drive the conver-

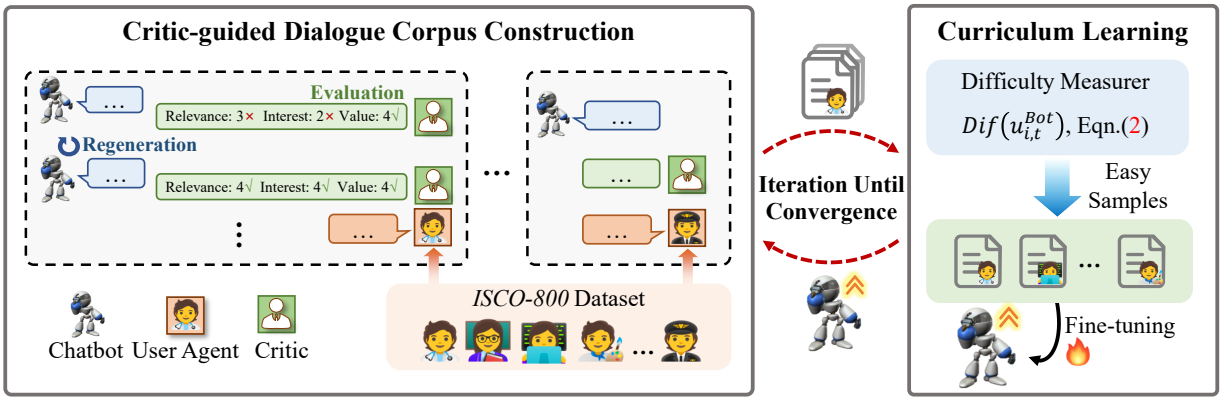

Figure 2: General diagram of proposed UPC. We train UPC in an iterative process. In each iteration, the chatbot engages in dialogue with user agents from the *ISCO-800* dataset, guided by the critic to generate a corpus with enhanced proactivity. In the iterative curriculum learning process, we measure the communication difficulty and fine-tune the chatbot with the corpus corresponding to easy-to-communicate users. The fine-tuned chatbot is then used to generate the corpus for the next iteration. This iterative process repeats until convergence.

sation to the goal. ProCoT [Deng *et al.*, 2023c] proposes the proactive chain-of-thought prompting scheme, which augments LLMs with the goal planning capability. Although they actively lead the conversation towards system-side goals, they lack the ability to understand the user's chatting interests, resulting in low engagement and a diminished willingness to continue the conversation. In contrast, our method emphasizes user-oriented proactivity by focusing on topics that are centered around the user's background and interests.

3 Problem Definition

Given a chatbot C_θ parameterized by θ, we consider the task of having a multi-turn open-domain conversation between the chatbot and each user $\mathcal{U}_i$ from a user set $\{\mathcal{U}_i\}_{i=1}^N$. The resulting dialogue corpus is denoted as $\{\mathcal{D}_i\}_{i=1}^N$ with the total dialogue number of N. $\mathcal{D}_i = \left\{\left(u_{i,t}^{Bot}, u_{i,t}^{User}\right)_{t=1}^{T}\right\}$ is the one conversation with T turns, and $u_{i,t}^{Bot}$, $u_{i,t}^{User}$ are the utterances of the chatbot and user at turn t, respectively.

Existing methods often passively respond to the user queries or focus on target-oriented topics. This lack of user-oriented proactivity can make users feel unappreciated, reducing their satisfaction and willingness to continue the conversation in human-computer interactions. Therefore, our goal is to develop a chatbot that proactively explores the user's chatting interests and guides conversations toward user-centered topics. To evaluate this kind of proactivity, the primary evaluation metric is whether the user finds the chatbot's response interesting. In addition, background relevance plays a crucial role, as the user typically focuses on topics related to their context. Finally, the value of the chatbot's response is an important factor, as it directly influences the quality of the conversation. To this end, We aim to build a chatbot with user-oriented proactivity from the perspective of the user's chatting interests $S_{int.}^{i,t}$, background relevance $S_{rel.}^{i,t}$ and response value $S_{val.}^{i,t}$, respectively. These metrics gauge the chatbot's ability to proactively explore the user's background, engage with the user's desired topics, and provide responses that are both interesting and valuable.

4 User-oriented Proactive Chatbot

The overall framework for UPC training is shown in Figure 2. We first construct a critic to evaluate user-oriented proactivity, then train the chatbot in an iterative curriculum learning process. In each iteration, the chatbot engages in a dialogue with different user agents, each with a distinct background from our *ISCO-800* dataset. During the conversation, the chatbot receives evaluations from the critic, which guides the generation of a high-quality corpus with improved proactivity. Since the communication difficulty varies across users, we assess the difficulty based on the quality of the generated corpus, so as to fine-tune the chatbot with the corpus corresponding to easy users. Then, the fine-tuned chatbot is used to start a new iteration of dialogue corpus generation. We repeat this iterative process to improve the performance until convergence.

4.1 Evaluation of User-oriented Proactivity

To evaluate the user-oriented proactivity $\mathbf{S}_{i,t}$ from the three perspectives of user's interest level, background relevance, and response value, we follow the LLM-as-a-judge method [Zheng *et al.*, 2024] and construct the critic $\mathcal{J}$ by prompting the ChatGPT to score the abilities. Specifically, we let the critic obtain the user's background and the dialogue history $\mathcal{D}_{i,t-1}$ between the user and the chatbot. Then we use the 5-point scoring system to evaluate the current response $u_{i,t}^{Bot}$ of the chatbot. It is formalized:

$$\mathbf{S}_{i,t} = \left[S_{int.}^{i,t}, S_{rel.}^{i,t}, S_{val.}^{i,t}\right] = \mathcal{J}\left(u_{i,t}^{Bot}, \mathcal{U}_i, \mathcal{D}_{i,t-1}\right), \quad (1)$$

where the dialogue history $\mathcal{D}_{i,t-1} = \left\{\left(u_{i,j}^{Bot}, u_{i,j}^{User}\right)_{j=1}^{t-1}\right\}$. Figure 3 illustrates the prompt of critic. We provide descriptions for scores of 1, 3, and 5 to improve the rating accuracy.

4.2 Critic-guided Dialogue Corpus Generation

We employ the critic to guide dialogues between the chatbot and user agents, generating a dialogue corpus with enhanced user-oriented proactivity. We use Qwen1.5-72B-Chat,

Prompt of the critic

You are now playing a character named {name}, and you are chatting with another person named Bot. You need to evaluate and rate the Bot's response based on its performance.
This is your background information: \n{background}
The historical dialogue between you and Bot is: \n{history}
Bot's next response is: \n{response}
Using the description below, combined with the historical conversation records, evaluate and rate the response of the Bot with score 1 to 5:
• **Relevance to your background of Bot's next response.** Score: 1, 2, 3, 4, 5: [1: <The content only has a little connection to my stated interests or background>, 3: <The content generally relates to my interests or background, with a fair level of relevance throughout the conversation>, 5: <The content is highly tailored to my specific interests and background. It may also expand on topics that are relevant to me>], because...
• **Value of Bot's next response.** Score: 1, 2, 3, 4, 5: [1: <The content only has a little substantive information. It may include meaningless repetition, incorrect information>, 3: <The content can somewhat meet my needs, although it may not be completely comprehensive or in-depth>, 5: <The content is insightful or deeply analytical, greatly facilitating my understanding>], because...
• **Your level of interest in the Bot's next response.** Score: 1, 2, 3, 4, 5: [1: <I only have a little interest in the response>, 3: <The response is reasonably interesting, though it may not be my favorite subject>, 5: <I am deeply fascinated by this response. It is highly engaging and something I look forward to>], because...
Be cautious in giving a score of 4 or above. Only use a score of 5 when the content is truly exceptional and exceeds your expectations.
You can only output three sentences, each following the format below:
Relevance to your background of Bot's next response: (score), because...
Value of Bot's next response: (score), because...
Level of interest in the Bot's next response: (score), because...

Figure 3: The prompt of the critic.

an LLM with strong role-playing abilities [LMSYS, 2024], to act as user agents. Each user agent is assigned a unique background from our *ISCO-800* dataset, described in Section 4.4. The system prompt of the user agent is shown in Appendix E. We then put the chatbot and each user agent into multiple rounds of open-domain conversation. Notably, the chatbot does not have prior access to the user's background but is instead prompted to actively learn about the background of the user during the conversation, and find topics of potential interest for the user. The system prompt is shown in Appendix E. In the first round, we prompt the chatbot to greet first and let the user give a brief self-introduction. Then they engaged in open-ended conversation. The critic evaluates each response from the chatbot, excluding the initial greeting. Each response is scored on three dimensions: user interest, background relevance, and response value. Scores range from 1 to 5, with corresponding reasons provided. For responses scoring below 4, the chatbot is prompted to regenerate based on the critic's feedback, as shown in Appendix E. The regeneration continues until all scores are 4 or higher, or until the maximum number of regenerations is reached. Our critic-guided approach ensures that the chatbot generates high-quality responses by trial and error like humans [Young, 2009], progressively enhancing user-oriented proactivity. Besides, the proportion of high-quality responses increases with the evolution of the chatbot. The corpus generation pipeline in each training iteration k is detailed in Algorithm 1.

Algorithm 1 Dialogue corpus generation in iteration k.

Require: The chatbot C_{θ_k} in iteration k, the user agent $\mathcal{U}_i$ from *ISCO-800* dataset $\{\mathcal{U}_i\}_{i=1}^{N}$. The critic $\mathcal{J}$. The maximum re-gen attempts R, dialogue turns T and score buffer $\mathcal{S}_i$.

for $i = 1, ..., N$ **do**
  for $t = 1, ..., T$ **do**
    if $t == 1$ **then**
      The chatbot C_{θ_k} gives an greeting utterance $u_{i,t}^{Bot}$.
      The user agent $\mathcal{U}_i$ gives brief self-introduction $u_{i,t}^{User}$.
      Add $(u_{i,t}^{Bot}, u_{i,t}^{User})$ to the dialogue history $\mathcal{D}_{i,t}$.
    else
      The chatbot C_{θ_k} gives response $u_{i,t}^{Bot}$.
      Compute scores $\mathbf{S}_{i,t}$ for current $u_{i,t}^{Bot}$ via Eqn. (1).
      Cache scores, utterances: $\mathbf{S}'_{i,t} \leftarrow \mathbf{S}_{i,t}$, $u'^{Bot}_{i,t} \leftarrow u_{i,t}^{Bot}$.
      Initialize $r = 0$.
      while ($S_{int.}^{i,t}$, $S_{rel.}^{i,t}$ or $S_{rel.}^{i,t} < 4$) *and* $r <= R$ **do**
        Re-generate utterance $u_{i,t}^{Bot}$, $r = r + 1$.
        Compute scores $\mathbf{S}_{i,t}$ for $u_{i,t}^{Bot}$ via Eqn. (1).
      end while
      The user agent $\mathcal{U}_i$ gives response $u_{i,t}^{User}$.
      Add $(u'^{Bot}_{i,t}, u_{i,t}^{Bot}, u_{i,t}^{User})$ to $\mathcal{D}_i$, Add $(\mathbf{S}'_{i,t}, \mathbf{S}_{i,t})$ to $\mathcal{S}_i$.
    end if
  end for
end for
return Dialogue corpus $\{\mathcal{D}_i\}_{i=1}^{N}$ and scores $\{\mathcal{S}_i\}_{i=1}^{N}$ in iter. k.

4.3 Communication Difficulty-Aware Iterative Curriculum Learning

Since the training data is generated through interactions between the chatbot and the user, timely improvement of the chatbot helps enhance the generated data and final performance. Therefore, it is crucial to conduct iterative training that uses the model from the previous iteration to generate data for the next iteration. Besides, since the communication difficulty varies among different users [Lindqvist, 2015], similar to human learning, fine-tuning models benefit from an easy-to-difficult curriculum during model training [Gao *et al.*, 2024; Hu *et al.*, 2025]. Therefore, we propose a communication difficulty-aware iterative curriculum learning framework that consists of a Communication Difficulty Measurer and Training Scheduler.

Communication Difficulty Measurer. Our difficulty measurer captures two aspects: how well the chatbot performs and whether it can improve. The first checks if the chatbot can produce acceptable responses for the user. The second evaluates whether the chatbot can make meaningful progress from the feedback. Therefore, we use the scores described in Section 4.1 and their improvement with the critic as the Difficulty Measurer $Dif(u_{i,t}^{Bot})$:

$$Dif(u_{i,t}^{Bot}) = True \text{ if } P \text{ else } False, \tag{2}$$

the condition P is formalized:

$$P : \exists p \in \{S_{int.}^{i,t}, S_{rel.}^{i,t}, S_{val.}^{i,t}\}\ s.t.\ p<\alpha \ or \sum_{c \in \{\mathbf{S}_{i,t} - \mathbf{S}'_{i,t}\}} \mathbb{I}(c>0) < \beta, \tag{3}$$

Algorithm 2 Iterative Curriculum Learning.

Require: The chatbot C_θ with parameter θ. User set $\{\mathcal{U}_i\}_{i=1}^N$. The maximum number of iterations K and dialogue turns T.

for $k = 1, \ldots, K$ **do**
  Collect a corpus $\{\mathcal{D}_i\}_{i=1}^N$ according to Algorithm 1.
  Initialize $\mathcal{D}^* = \emptyset$.
  for $i = 1, \ldots, N$ **do**
    $is\_easy \leftarrow$ **true**
    for $t = 1, \ldots, T$ **do**
      Calculate $Dif(u_{i,t}^{Bot})$ via Eqn. (2).
      if $Dif(u_{i,t}^{Bot})$ == **true then**
        $is\_easy \leftarrow$ **false**
        break
      end if
    end for
    Add $\mathcal{D}_i$ to $\mathcal{D}^*$ **if** is_easy.
  end for
  Fine-tuning the model C_{θ_k} with $\mathcal{D}^*$.
end for
return Chatbot C_{θ_K}.

where $True$ represents difficult and the reverse is easy. α and β are hyperparameters, specified in Appendix B. A sample is considered difficult if any metric is below α or the number of boosted metrics is fewer than β. These criteria identify user samples that are either hard to generate high-quality responses or hard to improve.

Training Scheduler. We adjust the generated training corpus at each training iteration k with the Difficulty Measurer for the collected multi-round dialogue corpus. Specifically, for the dialogue corpus $\mathcal{D}$ of different users, we follow Eqn. (2) to measure the difficulty of the corpus $\{\mathcal{D}_i\}_{i=1}^N$. The easy corpus constitutes the training dataset $\mathcal{D}^*$, while the difficult corpus will be discarded and its corresponding users will be engaged in dialogues in the next round. Notably, as the chatbot is updated, its ability to adapt to a wider range of users improves, leading to an increase in the amount of easy corpus. The training pipeline is shown in the Algorithm. 2.

4.4 User Background Dataset *ISCO-800*

To train a chatbot capable of interacting with users from diverse backgrounds, we construct a dataset called *ISCO-800*, containing 800 pieces of user background information. Specifically, to ensure the dataset is representative of global occupations, we refer to the ISCO-08 classification by the International Labour Organization [Organization, 2008]. We randomly select 40 out of 43 sub-major occupation groups from ISCO-08 as the source data. The details of these occupation groups are listed in Appendix C. We designed a prompt in Appendix C to guide GPT-4 in generating realistic user backgrounds for each occupation. These backgrounds include names, career histories, educations, personalities, and hobbies. Each user's background consists of 50 to 100 words. To ensure realism, not all users have positive personalities or smooth careers. We generate 20 different backgrounds for each occupation group, resulting in 800 user backgrounds. Statistics of *ISCO-800* and comparison with other related datasets in terms of the user background are in Appendix C.

5 Experiment

5.1 Experimental Settings

Implementation Details. We implemented UPC with Qwen1.5-32B-Chat [Bai *et al.*, 2023] as the base LLM. We use user backgrounds from *ISCO-800* to form the user agents and each engages in 5-turn dialogues with the chatbot. The 800 user agents are divided into training, validation, and test sets (500, 100, and 200 users, respectively) for dialogue generation. They are not duplicated in the occupation groups. This ensures the model is always evaluated on unseen domains during testing. See Appendix B and C for more details.

Compared Methods. We compare UPC with existing open-domain dialogue methods, including the universal LLMs, recent target-oriented open-domain dialogue methods (ProCoT [Deng *et al.*, 2023c] and TRIP [Wang *et al.*, 2024]) and other LLM-based methods (BlenderBot 3 [Shuster *et al.*, 2022], MemoChat [Lu *et al.*, 2023] and MemoryBank [Zhong *et al.*, 2024]). For a fair comparison, we adapt them to the user-oriented dialogue tasks by fine-tuning the models or using prompt strategies. See Appendix B for details.

Metrics. We use our designed scoring system (including the background relevance $Rel.$, user's interest level $Int.$, and response value $Val.$) to evaluate the chatbot's user-oriented proactivity. Notably, the critic used for evaluation includes gpt-3.5-turbo-0125 or gpt-4-turbo-2024-04-09. In addition, we also adopt the widely used perplexity (PPL) [Jelinek *et al.*, 1977] to measure the quality of the generated corpus.

Evaluation. In the test phase, we obtain the user backgrounds of the test set from *ISCO-800* and construct different user agents. Then we let the chatbot start 5 turns of open-domain dialogue with each user agent. The chatbot does not obtain the user's background information in advance. We use the critic to evaluate metrics of user-oriented proactivity of each turn of the chatbot's response for the generated dialogue corpus and obtain the averaged scores.

5.2 Comparison Experiments

We compare our UPC with other open-domain dialogue methods and the results are shown in Table 1. Our proposed UPC, with only 32B parameters, outperforms all other methods in terms of $Rel.$, $Int.$, $Val.$, and PPL, demonstrating its effectiveness. The detailed analyses are as follows.

UPC outperforms universal LLMs with fewer parameters. With GPT-3.5 as the critic, our UPC outperforms the strongest baseline Llama-3-70B-Instruct by 3.32%, 4.42%, and 4.03% on $Rel.$, $Int.$ and $Val.$. With GPT-4 as the critic, our UPC outperforms the strongest baseline Qwen1.5-72B-Chat by 1.39%, 1.33% and 3.69% on $Rel.$, $Int.$ and $Val.$. Our method also reduces the PPL by 21.64% compared with Llama-3-70B-Instruct and 29.36% compared with Qwen1.5-72B-Chat. It indicates that our UPC improves the performance even with half of the model parameters. We attribute the improvement to the critic-guided corpus construction paradigm that facilitates the LLM to generate user-interested dialogue data. Besides, the iterative curriculum learning method helps adapt to different users from easy to hard, enhancing the chatbot's user-oriented proactivity. Moreover, the

Category	Methods	Param.	Critic: GPT-3.5			Critic: GPT-4			PPL ↓
			Rel. ↑	*Int.* ↑	*Val.* ↑	*Rel.* ↑	*Int.* ↑	*Val.* ↑	
Universal LLMs	Llama-3-70B-Instruct [AI@Meta, 2024]	70B	4.702	3.776	3.773	4.619	3.628	3.734	10.153
	Qwen1.5-72B-Chat [Bai *et al.*, 2023]	72B	4.553	3.670	3.664	4.752	3.672	3.764	11.263
	Qwen1.5-32B-Chat [Bai *et al.*, 2023]	32B	4.534	3.672	3.639	4.690	3.640	3.746	11.038
	Vicuna-33b-v1.3 [Chiang *et al.*, 2023]	33B	4.508	3.626	3.609	4.507	3.595	3.644	10.821
	GLM-4 [ZHIPUAI, 2024]	NA	4.401	3.581	3.486	4.640	3.657	3.668	12.326
	GPT-3.5 Turbo [OpenAI, 2022]	≈175B	4.625	3.719	3.701	4.760	3.579	3.688	12.691
	GPT-4o [OpenAI, 2024]	NA	4.591	3.546	3.578	4.599	3.656	3.644	17.624
Target-Oriented Methods	ProCoT [Deng *et al.*, 2023c]	≈175B	4.363	3.526	3.496	4.279	3.001	3.099	19.895
	TRIP [Wang *et al.*, 2024]	32B	4.521	3.634	3.626	4.341	3.461	3.455	9.327
Other LLM-Based Methods	BlenderBot 3 [Shuster *et al.*, 2022]	30B	3.414	3.188	3.056	3.231	3.124	3.008	13.287
	MemoChat [Lu *et al.*, 2023]	33B	4.609	3.713	3.771	4.203	3.438	3.619	8.773
	MemoryBank [Zhong *et al.*, 2024]	6B	3.798	3.220	3.253	3.312	2.676	2.996	12.564
	Ours (Qwen1.5-32B-Chat)	32B	**4.858**	**3.943**	**3.925**	**4.818**	**3.721**	**3.903**	**7.956**

Table 1: Comparsion experimental results on the *ISCO-800*.

SFT	CDC	IFT	CL	*Rel.* ↑	*Int.* ↑	*Val.* ↑
				4.522	3.629	3.620
✓				4.423	3.547	3.525
✓	✓			4.534	3.672	3.639
	✓	✓		4.578	3.738	3.758
	✓	✓	✓	**4.858**	**3.943**	**3.925**

Table 2: Ablation experiments on *ISCO-800*. The row without ticks means the original Qwen1.5-32B-Chat.

GPT-4-based critic exhibits similar scoring trends to the GPT-3.5-based critic. It suggests that although we use GPT-3.5-based critic during training, we still obtain a consistent judgment with the better-performing GPT-4 in the test. Therefore, we use GPT-3.5-based critic in ablation studies.

UPC outperforms the target-oriented methods. UPC outperforms the strongest baseline TRIP by 7.45%, 8.50%, and 8.25% on *Rel*., *Int*. and *Val*. (Critic: GPT-3.5) and 10.99%, 7.51%, and 12.97% (Critic: GPT-4). UPC also reduces the PPL by 14.70%. We attribute this to the fact that although it has proactivity in target topic switching, the topic is not necessarily the one that the user is interested in, leading to lower performance of user-oriented proactivity.

UPC outperforms other LLM-based methods. UPC outperforms the strongest MemoChat by 5.40%, 6.19%, and 4.08% on *Rel*., *Int*. and *Val*. (Critic: GPT-3.5) and 14.63%, 8.23%, and 7.85% (Critic: GPT-4). Our method also reduces the PPL by 9.31%. We attribute this to the reason that MemoChat's focus on maintaining consistency offers limited improvement in open-domain conversations where topics shift constantly, whereas our trial-and-error strategy yields better performance across a diverse range of topics.

5.3 Ablation Studies

We conducted ablation experiments with the GPT-3.5-based critic to validate the effectiveness of our proposed iterative curriculum learning. We compared different training strategies in our iterative curriculum learning, described as follows:

SFT: An initial LLM generates a dialogue corpus through multiple dialogues with each user agent for supervised fine-tuning, without the critic's involvement. **SFT+CDC (Critic-guided Dialogue Corpus Collection)**: Based on SFT, a critic evaluates the chatbot's responses and requests regenerations for low-rated responses, producing a high-quality corpus for fine-tuning. We repeat the above two strategies for the same number of iterations as UPC to ensure a consistent amount of training data. **CDC+IFT (Iterative Fine-Tuning)**: The chatbot engages in dialogue with each user agent, incorporating critic feedback. The collected corpus is used for initial fine-tuning. Then the fine-tuned chatbot re-engages with each user agent and collects the corpus. This process of corpus collection and fine-tuning undergoes the same iterative rounds as UPC, gradually enhancing the performance of the chatbot. **CDC+IFT+CL (Curriculum Learning)**: Building on CDC+IFT, we additionally use our Communication Difficulty Measurer to identify easy samples after each dialogue generation, and we only use easy samples for fine-tuning.

Effectiveness of CDC. As in Table 2, UPC (SFT+CDC) outperforms UPC (SFT) in all metrics, as the critic feedback improves the dialogue corpus quality, resulting in better fine-tuning and enhanced LLM performance [Zhou *et al.*, 2024]. Besides, UPC (SFT) performs even inferior to that of the original LLM. We attribute this to the poor-quality corpus generated during training data collection without feedback.

Effectiveness of IFT. As in Table 2, compared to only CDC, training with CDC and IFT improves the user-oriented proactivity of chatbots by 0.97%, 1.80%, and 3.27% on *Rel*., *Int*., and *Val*., respectively. It indicates that re-collecting the dialogue corpus data using the fine-tuned model helps to improve the quality of the generated data and improves the performance of the model in the next round of fine-tuning.

Effectiveness of CL. As in Table 2, UPC has significant improvement compared to UPC (w/o CL). Specifically, there are 6.12%, 5.48% and 4.44% increases on *Rel*., *Int*., and *Val*. respectively. In conclusion, curriculum learning helps the model learn the essential features of the task more quickly and improves the generalization ability [Wang *et al.*, 2021].

5.4 More Discussions

Online Chatbot Interaction with Real Users. To illustrate the practicality of UPC, we deploy UPC and one of the strongest baselines Llama-3-70B-Instruct on the Internet and

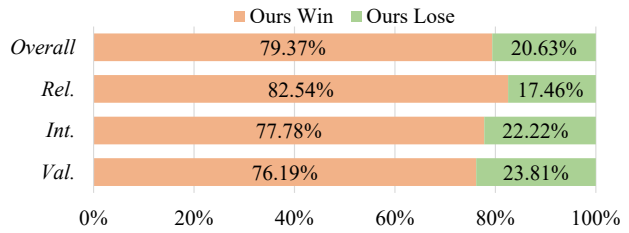

Figure 4: Real user evaluation results between our UPC and Llama-3-70B-Instruct baseline.

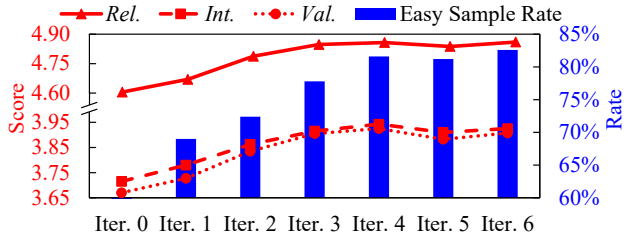

Figure 5: The performance during the iterative curriculum learning. The curves represent the metrics of user-oriented proactivity. The bars represent the rate of the easy sample.

recruited 62 participants to have free chat with each of the two chatbots. Participants are unknown to the names of the chatbots. The participants are first advised to chat with the two chatbots separately with any identity and topics. Then they are required to compare the performance of the two chatbots. See Appendix D for detailed settings. As in Figure 4, over 75% of participants preferred UPC in the overall performance as well as the *Rel.*, *Int.* and *Val.*. It indicates our UPC has enhanced performance and aligns with human preferences. Most Participants consider that UPC pays more attention to their background and chatting interests, gives more valuable responses, and leads to a better chatting performance. It reveals the practical value in real chatting scenarios.

Performance evolution during iterative curriculum learning. To illustrate the performance improvement process during iterative curriculum learning, we present our designed metrics on the validation set and the rate of easy samples in the training data at each iteration in Figure 5. It shows that performance and the rate of easy samples gradually improve with each iteration, converging at the 4^{th} iteration. This suggests that iterative curriculum learning progressively enhances model performance by training the chatbot to adapt to users with increasing communication difficulty. Additionally, the maximum number of iterations is set to 4 due to convergence. We further calculate the regeneration rate and average number of regenerations during the training process. As in Table 3, both the regeneration rate and average regenerations decrease as the iteration proceeds. It indicates that through iterative curriculum learning, the cost of regeneration is reduced as the iteration progresses. Besides, the chatbot increasingly delivers satisfactory responses on the first attempt.

Effects of LLM category and size. To verify the applicability of UPC, we conduct experiments on LLMs of different categories and sizes. In Table 4, for various sizes in Qwen1.5 series, UPC consistently improves performance, with gains of 5%∼10%. For LLMs of similar size but different types, e.g., Qwen1.5 and Vicuna, UPC also yields improvements. Besides, stronger base LLMs like original Qwen1.5-14B and Qwen1.5-32B show greater gains than Vicuna-33B. In summary, UPC works well across LLM categories and sizes, and brings larger gains on stronger base LLMs.

Training Process	Iter. 1	Iter. 2	Iter. 3	Iter. 4
Regeneration Rate	37.7%	37.9%	26.0%	23.2%
Average Number of Regenerations	0.785	0.784	0.552	0.474

Table 3: Regeneration rate and average number of regenerations during the iterative curriculum learning.

LLMs	*Rel.* ↑	*Int.* ↑	*Val.* ↑
Original Qwen1.5-14B-Chat	4.356	3.550	3.490
Ours (Qwen1.5-14B-Chat)	4.653	3.811	3.789
Original Vicuna-33B-Chat	4.323	3.514	3.479
Ours (Vicuna-33B-Chat)	4.363	3.568	3.479
Original Qwen1.5-32B-Chat	4.522	3.629	3.620
Ours (Qwen1.5-32B-Chat)	**4.858**	**3.943**	**3.925**

Table 4: Results of different size and types of LLMs on *ISCO-800*.

5.5 Case Study

In Appendix F, we provide test set samples to compare the performance of UPC with GPT-3.5-Turbo baseline. After the introduction of the user, UPC ask specific questions about the user's background, like "Innovative supply chain management solutions". When the user mentions blockchain technology, UPC asks about its implementation. The responses are highly interactive and relevant to the user's interests, keeping the user engaged in the discussion. In contrast, the dialogues of GPT-3.5 are more generic, with the chatbot relying on simple pleasantries like "How can I assist you today?" and failing to address the user's specific background. The responses are shallow, do not effectively explore the user's interests, and result in superficial dialogues that neither engage the user nor encourage further discussion. Thus, our UPC exhibits better user-oriented proactivity. Another case between our UPC and target-oriented method TRIP is provided in Appendix F.

6 Conclusion

We propose a User-oriented Proactive Chatbot to address the lack of user-oriented proactivity in open-domain dialogue. We first construct a critic to evaluate the user-oriented proactivity. Then we use the critic to guide dialogues between the chatbot and user agents, generating a corpus with enhanced user-oriented proactivity. We introduce a user background dataset *ISCO-800* to ensure the diversity of user backgrounds. Finally, we train the chatbot with an iterative curriculum learning strategy to adapt to different users from easy to hard. Experiments demonstrate that our proposed training method is applicable to different LLMs, improving user-oriented proactivity in open-domain dialogues. We hope to bring new insight into the improvement of dialogue quality for LLMs and human-computer interaction experience.

Acknowledgments

This work was supported in part by the National Key R&D Program of China under Grant No. 2022YFB4500600, in part by the Joint Funds of the National Natural Science Foundation of China under Grant No. U24A20327, in part by the Guangdong Provincial Key Laboratory of Human Digital Twin under Grant No. 2022B1212010004, and in part by Major Key Project of Peng Cheng Laboratory (PCL) under Grant No. PCL2023A08.

Contribution Statement

This work was a collaborative effort by all contributing authors. Yufeng Wang and Jinwu Hu made equal contributions to this study and are designated as co-first authors. Xiaofen Xing and Qingfang Zheng, serving as the corresponding authors, are responsible for all communications related to this manuscript.

References

[Achiam *et al.*, 2023] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

[AI@Meta, 2024] AI@Meta. Llama 3 model card. 2024.

[Bai *et al.*, 2023] Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, Binyuan Hui, Luo Ji, Mei Li, Junyang Lin, Runji Lin, Dayiheng Liu, Gao Liu, Chengqiang Lu, Keming Lu, Jianxin Ma, Rui Men, Xingzhang Ren, Xuancheng Ren, Chuanqi Tan, Sinan Tan, Jianhong Tu, Peng Wang, Shijie Wang, Wei Wang, Shengguang Wu, Benfeng Xu, Jin Xu, An Yang, Hao Yang, Jian Yang, Shusheng Yang, Yang Yao, Bowen Yu, Hongyi Yuan, Zheng Yuan, Jianwei Zhang, Xingxuan Zhang, Yichang Zhang, Zhenru Zhang, Chang Zhou, Jingren Zhou, Xiaohuan Zhou, and Tianhang Zhu. Qwen technical report. *arXiv preprint arXiv:2309.16609*, 2023.

[Chen *et al.*, 2023] Liang Chen, Hongru Wang, Yang Deng, Wai Chung Kwan, Zezhong Wang, and Kam-Fai Wong. Towards robust personalized dialogue generation via order-insensitive representation regularization. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 7337–7345, 2023.

[Chen *et al.*, 2024] Yi-Pei Chen, Noriki Nishida, Hideki Nakayama, and Yuji Matsumoto. Recent trends in personalized dialogue generation: A review of datasets, methodologies, and evaluations. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 13650–13665, 2024.

[Chiang *et al.*, 2023] Wei-Lin Chiang, Zhuohan Li, Zi Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E Gonzalez, et al. Vicuna: An open-source chatbot impressing gpt-4 with 90%* chatgpt quality. *See https://vicuna. lmsys. org (accessed 14 April 2023)*, 2(3):6, 2023.

[Deng *et al.*, 2023a] Yang Deng, Wenqiang Lei, Minlie Huang, and Tat-Seng Chua. Rethinking conversational agents in the era of llms: Proactivity, non-collaborativity, and beyond. In *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region*, pages 298–301, 2023.

[Deng *et al.*, 2023b] Yang Deng, Wenqiang Lei, Wai Lam, and Tat-Seng Chua. A survey on proactive dialogue systems: problems, methods, and prospects. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 6583–6591, 2023.

[Deng *et al.*, 2023c] Yang Deng, Lizi Liao, Liang Chen, Hongru Wang, Wenqiang Lei, and Tat-Seng Chua. Prompting and evaluating large language models for proactive dialogues: Clarification, target-guided, and non-collaboration. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 10602–10621, 2023.

[Gao *et al.*, 2024] Shen Gao, Zhengliang Shi, Minghang Zhu, Bowen Fang, Xin Xin, Pengjie Ren, Zhumin Chen, Jun Ma, and Zhaochun Ren. Confucius: Iterative tool learning from introspection feedback by easy-to-difficult curriculum. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 18030–18038, 2024.

[Grassi *et al.*, 2022] Lucrezia Grassi, Carmine Tommaso Recchiuto, and Antonio Sgorbissa. Knowledge-grounded dialogue flow management for social robots and conversational agents. *International Journal of Social Robotics*, 14(5):1273–1293, 2022.

[Hu *et al.*, 2024] Jinwu Hu, Yufeng Wang, Shuhai Zhang, Kai Zhou, Guohao Chen, Yu Hu, Bin Xiao, and Mingkui Tan. Dynamic ensemble reasoning for llm experts. *arXiv preprint arXiv:2412.07448*, 2024.

[Hu *et al.*, 2025] Jinwu Hu, Wei Zhang, Yufeng Wang, Yu Hu, Bin Xiao, Mingkui Tan, and Qing Du. Dynamic compressing prompts for efficient inference of large language models. *arXiv preprint arXiv:2504.11004*, 2025.

[Jelinek *et al.*, 1977] Fred Jelinek, Robert L Mercer, Lalit R Bahl, and James K Baker. Perplexity—a measure of the difficulty of speech recognition tasks. *The Journal of the Acoustical Society of America*, 62(S1):S63–S63, 1977.

[Kann *et al.*, 2022] Katharina Kann, Abteen Ebrahimi, Joewie Koh, Shiran Dudy, and Alessandro Roncone. Open-domain dialogue generation: What we can do, cannot do, and should do next. In *Proceedings of the 4th Workshop on NLP for Conversational AI*, pages 148–165, 2022.

[Liao *et al.*, 2023] Lizi Liao, Grace Hui Yang, and Chirag Shah. Proactive conversational agents in the post-chatgpt world. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 3452–3455, 2023.

[Lindqvist, 2015] Susanne Lindqvist. Interprofessional communication and its challenges. *Clinical communication in medicine*, pages 157–167, 2015.

[LMSYS, 2024] LMSYS. Chatbot Arena, 2024. https://openlm.ai/chatbot-arena/.

[Lu *et al.*, 2023] Junru Lu, Siyu An, Mingbao Lin, Gabriele Pergola, Yulan He, Di Yin, Xing Sun, and Yunsheng Wu. Memochat: Tuning llms to use memos for consistent long-range open-domain conversation. *arXiv preprint arXiv:2308.08239*, 2023.

[Luo *et al.*, 2022] Bei Luo, Raymond YK Lau, Chunping Li, and Yain-Whar Si. A critical review of state-of-the-art chatbot designs and applications. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 12(1):e1434, 2022.

[Oertel *et al.*, 2020] Catharine Oertel, Ginevra Castellano, Mohamed Chetouani, Jauwairia Nasir, Mohammad Obaid, Catherine Pelachaud, and Christopher Peters. Engagement in human-agent interaction: An overview. *Frontiers in Robotics and AI*, 7:92, 2020.

[OpenAI, 2022] OpenAI. ChatGPT, 2022. https://openai.com/blog/chatgpt/.

[OpenAI, 2024] OpenAI. GPT-4o System Card, 2024. https://openai.com/index/gpt-4o-system-card/.

[Organization, 2008] International Labour Organization. International Standard Classification of Occupations 2008, 2008. www.ilo.org/publications/international-standard-classification-occupations-2008-isco-08-structure.

[Rodrigues *et al.*, 2022] Carlos Rodrigues, Arsénio Reis, Rodrigo Pereira, Paulo Martins, José Sousa, and Tiago Pinto. A review of conversational agents in education. In *International Conference on Technology and Innovation in Learning, Teaching and Education*, pages 461–467. Springer, 2022.

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

[Sun *et al.*, 2023] Bin Sun, Yitong Li, Fei Mi, Weichao Wang, Yiwei Li, and Kan Li. Towards diverse, relevant and coherent open-domain dialogue generation via hybrid latent variables. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 13600–13608, 2023.

[Tang *et al.*, 2019] Jianheng Tang, Tiancheng Zhao, Chenyan Xiong, Xiaodan Liang, Eric Xing, and Zhiting Hu. Target-guided open-domain conversation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5624–5634, 2019.

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

[Wang *et al.*, 2021] Xin Wang, Yudong Chen, and Wenwu Zhu. A survey on curriculum learning. *IEEE transactions on pattern analysis and machine intelligence*, 44(9):4555–4576, 2021.

[Wang *et al.*, 2024] Jian Wang, Dongding Lin, and Wenjie Li. Target-constrained bidirectional planning for generation of target-oriented proactive dialogue. *ACM Transactions on Information Systems*, 42(5):1–27, 2024.

[Young, 2009] H Peyton Young. Learning by trial and error. *Games and economic behavior*, 65(2):626–643, 2009.

[Zhao *et al.*, 2023] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. A survey of large language models. *arXiv preprint arXiv:2303.18223*, 1(2), 2023.

[Zheng *et al.*, 2024] Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing, et al. Judging llm-as-a-judge with mt-bench and chatbot arena. *Advances in Neural Information Processing Systems*, 36, 2024.

[ZHIPUAI, 2024] ZHIPUAI. GLM-4, 2024. https://chatglm.cn/.

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

[Zhou *et al.*, 2024] Chunting Zhou, Pengfei Liu, Puxin Xu, Srinivasan Iyer, Jiao Sun, Yuning Mao, Xuezhe Ma, Avia Efrat, Ping Yu, Lili Yu, et al. Lima: Less is more for alignment. *Advances in Neural Information Processing Systems*, 36, 2024.