

Enhancing Goal-oriented Proactive Dialogue Systems via Consistency Reflection and Correction

Didi Zhang*, Yaxin Fan*, Peifeng Li†, and Qiaoming Zhu

School of Computer Science and Technology, Soochow University, Suzhou, China
{ddzhang2023, yxfansuda}@stu.suda.edu.cn, {pfli, qmzhu}@suda.edu.cn

Abstract

Goal-oriented proactive dialogue systems are designed to guide user conversations seamlessly towards specific objectives by planning a goal-oriented path. However, previous research has focused predominantly on optimizing these paths while neglecting the inconsistencies that may arise between generated responses and dialogue contexts, including user profiles, dialogue history, domain knowledge, and subgoals. To address this issue, we introduce a model-agnostic two-stage Consistency Reflection and Correction (CRC) framework. Specifically, in the consistency reflection stage, the model is prompted to reflect on the discrepancies between generated responses and dialogue contexts, identifying inconsistencies and suggesting possible corrections. In the consistency correction stage, the model generates responses that are more consistent with the dialogue context based on these reflection results. We conducted experiments on various model architectures with different parameter sizes, including encoder-decoder models (BART, T5) and decoder-only models (GPT-2, DialoGPT, Phi3, Mistral and LLaMA3), and the experimental results on three datasets demonstrate that our CRC framework significantly improves the consistency between generated responses and dialogue contexts. Our code is publicly available at: <https://github.com/zhyidi/CRC>.

1 Introduction

The Goal-oriented Proactive Dialogue System (GPDS) focuses on achieving specific objectives by actively guiding and anticipating user needs (Liu et al., 2020; Wang et al., 2024a, 2023b). Unlike traditional dialogue systems that passively respond to user requests (Touvron et al., 2023; Achiam et al., 2024), GPDS strategically steers the conversation along a goal-oriented path, ensuring that a goal

is naturally achieved while maintaining a positive user experience. GPDS has a wide range of applications in various domains, such as recommendation systems (Fu et al., 2020; Liu et al., 2020, 2021) and medical consultations (Xu et al., 2024b).

Figure 1 presents an example of GPDS, which generates responses (e.g., S_i) based on the dialogue context, such as the user profile, dialogue history, domain knowledge, and subgoals within a goal-oriented path. GPDS can be divided into two primary sub-tasks: goal-oriented path planning and response generation. Initially, the system plans a goal-oriented path where each step is represented by an <action, topic> pair (e.g., “Q&A | Jimmy Lin’s constellation” → “Chat about the Star | Jimmy Lin” → “Movie recommendation | Grandpa’s Love” in Figure 1). Following this, the system generates responses aligned with the planned path, thereby guiding the conversation proactively and naturally toward achieving the final target (recommending the movie *Grandpa’s Love*).

Most previous studies on GPDS have primarily focused on planning goal-oriented paths using techniques such as CNN-based classifiers (Liu et al., 2020), Seq2seq paradigms (Deng et al., 2023; Wang et al., 2023c, 2024b,a), and graph interaction methods (Zhang et al., 2024). However, these approaches often overlook the inconsistencies that arise between generated responses and dialogue contexts. These inconsistencies manifest in several ways. First, there is an inconsistency with the dialogue history. As shown in Figure 1, the system asserts “of course he knows Lin” in S_2 , even though the dialogue history does not inquire about his acquaintance with Lin. Second, there is an inconsistency with the subgoal. Although S_4 ’s action is to recommend a movie, it fails to address the topic of *Grandpa’s Love*, resulting in an invalid recommendation. Third, there is an inconsistency with the domain knowledge. S_5 states that Yanping Zhu is the star of *Grandpa’s Love*, whereas the do-

*Equal contribution

†Corresponding author

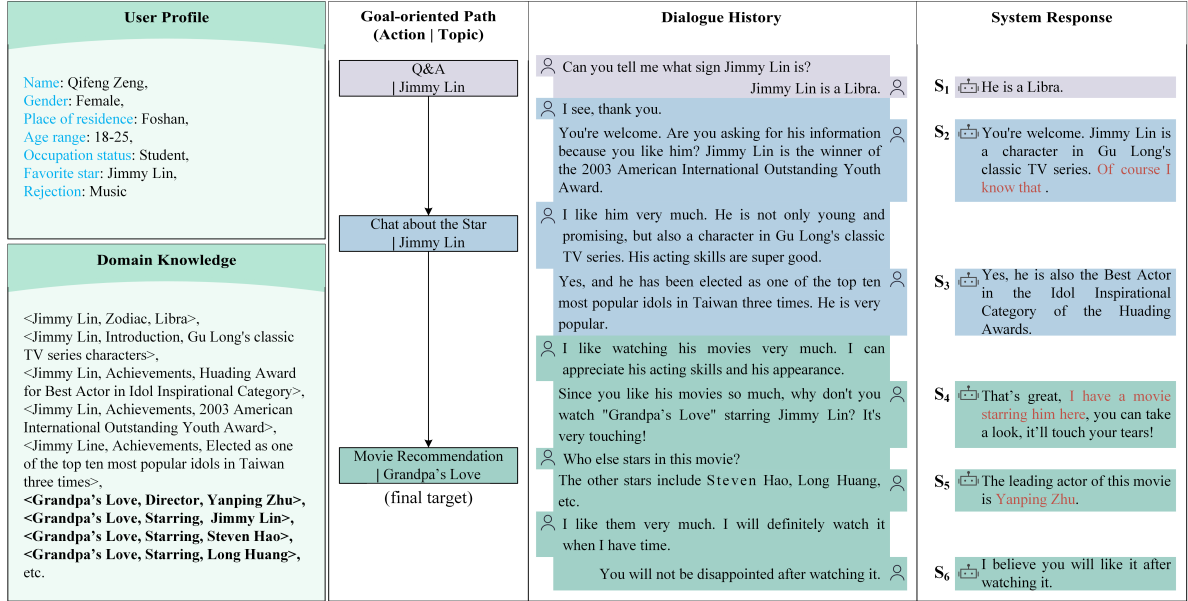


Figure 1: An example of GPDS from the DuRecDial (Liu et al., 2020) dataset, where the system responses are generated by TPNet (Wang et al., 2024a).

main knowledge indicates that Zhu is the director of the movie. Lastly, there is an inconsistency with the user profile. The system might generate a response that does not align with the user’s profile, as illustrated in Appendix A. For instance, the user’s profile shows a preference for news about Nicholas Tse. However, the system recommends unrelated social news. Hence, these inconsistencies can lead to a poor user experience in real-world scenarios, causing conversations to abruptly break down and failing to achieve the intended targets.

To address these inconsistencies, we propose a Consistency Reflection and Correction (CRC) framework, drawing inspiration from the reflective practice theory (Checkoway and Schön, 1985) of human cognition that emphasizes the systematic reflection on experiences to identify and improve areas of weakness. Specifically, in the **consistency reflection** stage, we guide the model to reflect its experience, i.e., whether its generated responses are consistent with the elements of dialogue contexts. If any inconsistencies are identified, the model is prompted to categorize the types of discrepancies and suggest potential corrections. In the **consistency correction** stage, the model is instructed to regenerate the responses more consistent with dialogue contexts based on the insights gained from the reflection stage.

To validate the effectiveness of our framework, we conducted extensive experiments on three

widely-used datasets: DuRecDial (Liu et al., 2020), DuRecDial 2.0 (Liu et al., 2021) and TopDial (Wang et al., 2023b). Since our framework is model-agnostic, we tested it on different model architectures and various parameter sizes, including encoder-decoder models (BART and T5) and decoder-only models (GPT-2, DialoGPT, Phi3-3.8B, Mistral-7B and LLaMA3-8B). The experimental results demonstrate that our CRC framework significantly improves the consistency between generated responses and dialogue context.

2 Related Work

2.1 Goal-oriented Proactive Dialogue System

Previous studies on GPDS typically began by planning a sequence of subgoals, followed by generating responses based on the subgoals to guide the conversation toward specific objectives. Most of them concentrated on goal-oriented path planning, employing the techniques such as CNN-base classifier (Liu et al., 2020), target-driven method (Wang et al., 2022, 2024a), Brownian bridge (Wang et al., 2023c), prompt-based method (Deng et al., 2023), graph-grounded planning (Liu et al., 2023), graph-interaction planning (Zhang et al., 2024), and bidirectional planning (Wang et al., 2024b).

However, these efforts often emphasize the importance of planning a goal-oriented path, overlooking the consistency between generated responses and dialogue contexts. This paper primarily fo-

cuses on response generation, aiming to improve GPDS by enhancing the consistency between generated responses and dialogue contexts.

2.2 Discourse Consistency in Dialogue

Discourse consistency in dialogue refers to the logical coherence and uniformity of information and themes, which is essential for fostering understanding and effective communication among participants. Previous research (Song et al., 2021; Wang et al., 2023a; Chen et al., 2023; Zhou et al., 2023) has frequently employed Natural Language Inference (NLI) models to assess and enhance discourse consistency between generated responses and dialogue contexts. However, these NLI models often depend on additional training data, which can hinder their generalizability. In an innovative approach, we propose leveraging the model’s inherent reflective capabilities to enhance the consistency between generated responses and dialogue contexts, thereby improving its generalizability.

3 Task Definition

Given a dataset $D = \{U^i, K^i, H^i, G^i\}_{i=1}^N$, where N is the size of the dataset. $U^i = \{u_j^i\}_{j=1}^{N_u}$ is the i -th user profile, where each item u_j^i is a key-value pair, representing the user’s personal information (e.g., name and gender). $K^i = \{k_j^i\}_{j=1}^{N_k}$ is the domain knowledge related to the i -th conversation of D , where each item k_j^i is a triple $\langle \text{head}, \text{relation}, \text{tail} \rangle$. $H^i = \{h_m^i\}_{m=1}^M$ is the content of the i -th conversation, consisting of M turns. $G^i = \{g_m^i\}_{m=1}^M$ is the goal-oriented path for the dialogue H^i , where each g_m^i consists of a dialogue action a_m^i and a dialogue topic t_m^i . The final goal of the dialogue is represented by g_M^i . GPDS can be divided into two primary sub-tasks: goal-oriented path planning and response generation.

Goal-oriented Path Planning Goal-oriented path planning aims to plan a sequence of subgoals to proactively guide the conversation to achieve the final target g_M^i . Each subgoal $g_m^i (1 \leq m \leq M)$ which is formulated as follows.

$$g_m^i = GPP(U^i, K^i, H_{\leq m}^i, G_{< m}^i) \quad (1)$$

where GPP is a path prediction model, which has attracted the attention of most previous work. In this paper, we adopted the same path prediction model as Wang et al. (2024a) and mainly focus on generating responses that are consistent with dialogue contexts.

Response Generation A generative model is employed to generate a response that aligns with the action and topic in g_m^i , thereby actively steering the conversation towards the final target. This process is represented as follows.

$$r_m^i = RG(U^i, K^i, H_{\leq m}^i, g_m^i) \quad (2)$$

where r_m^i is the generated response, and RG denotes an autoregressive model. Specifically, RG autoregressively generates r_m^i conditioned on the concatenation of the dialogue context, and it is optimized by minimizing the negative log-likelihood as follows.

$$\mathcal{L}(\theta) = - \sum_{i=1}^N \sum_{t=1}^T \log P(r_{m,t}^i | r_{m,<t}^i, U^i, K^i, H_{\leq m}^i, g_m^i; \theta) \quad (3)$$

where θ represents the trainable parameters, $r_{m,t}^i$ and $r_{m,<t}^i$ are the t -th token and the previous $t-1$ tokens of the response r_m^i , respectively, and T is the length of r_m^i . In this paper, we focus on improving RG to enhance the consistency between generated responses and dialogue contexts.

4 CRC Framework

Motivation Most of the Response Generation (RG) models in GPDS cannot produce responses that align with dialogue context. This is likely because their learning experience primarily involves imitation learning, which lacks deep reflection and correction mechanisms. According to the theory of reflective practice (Checkoway and Schön, 1985), the process of learning and growth comes from a cycle of experience, i.e., reflection and correction. Without this reflective practice, most models struggle to learn from their own experiences and summarize effectively, making it challenging for them to consistently generate responses that align with the dialogue context. Therefore, establishing a framework that can involve the reflection and correction mechanism to improve the consistency between responses and dialogue contexts is crucial.

In this paper, we proposed a model-agnostic, two stage CRC framework including consistency reflection and consistency correction, as shown in Figure 2. In the consistency reflection stage, we first guide a RG model to reflect on its experience, i.e., reflecting on the types of inconsistency between generated responses and dialogue contexts, and then suggest possible corrections. In the correction

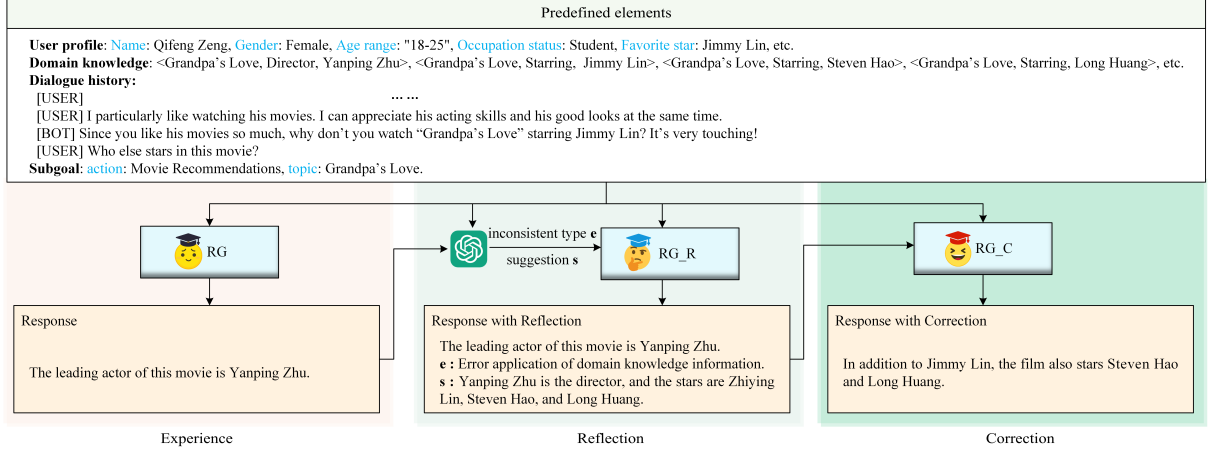


Figure 2: Overview of our CRC framework.

stage, we further guide the RG model to regenerate responses that are consistent with dialogue contexts on the reflection results.

Consistency Reflection As introduced in Section 1, the responses generated by the RG model may exhibit inconsistencies with dialogue context. These inconsistencies primarily pertain to the user profile U , the domain knowledge K , the dialogue history H , and the subgoal g . To address this issue, we prompt the model to reflect on and consider ways to correct these inconsistencies between dialogue responses and dialogue contexts, providing suggestions for improvement. This can be formalized as follows.

$$r_m^i, e_m^i, s_m^i = RG_R(U^i, K^i, H_{\leq m}^i, g_m^i) \quad (4)$$

where RG_R represents an autoregressive model with reflective ability, e_m^i denotes the inconsistent type of the reflection results on the response r_m^i , and s_m^i is a correction suggestion on the inconsistent type e_m^i . Therefore, the key lies in how to obtain the high-quality inconsistent type e and the correction suggestion s to stimulate the model's reflective capabilities.

While manually annotating inconsistency types and correction suggestions is an ideal approach, the time-consuming and costly nature of manual annotation hinders its practical application. Thanks to the powerful understanding capabilities of Large Language Models (LLMs) like ChatGPT, which have already achieved success in data annotation across various fields (Wang et al., 2023d; Xu et al.,

2024a), we utilize ChatGPT¹ to act as an annotator for the annotations of inconsistency types and correction suggestions. The prompt used is illustrated in Appendix B.

We feed the dialogue context and the generated response to ChatGPT and let it evaluate the consistencies. ChatGPT first requires to identify the inconsistency types and then provide correction suggestions. An example is shown in Figure 2, where the inconsistency type is related to the domain knowledge, and the correction suggestion is that "Yanping Zhu is the director, and the stars are Zhiying Lin, Steven Hao, and Long Huang.". After obtaining the annotations e and s , we continue to fine-tune the RG model to obtain a new RG_R model with reflective capabilities. Let the concatenation of r_m^i , e_m^i , and s_m^i be denoted as c_m^i , the optimization of RG_R is as follows.

$$\mathcal{L}_{cr}(\theta) = - \sum_{i=1}^N \sum_{t=1}^T \log P(c_{m,t}^i | c_{m,<t}^i, U^i, K^i, H_{\leq m}^i, g_m^i; \theta) \quad (5)$$

where θ represents the trainable parameters, N is the data size and T is the token length of c_m^i . During the learning process, the RG_R model needs to generate not only the response r but also both the reflective results e and s regarding r .

Consistency Correction During the consistency correction phase, we continue to train RG to generate the response $r_m^{i'}$ that is more consistent with the dialogue context on the reflective results from RG_R. This can be formalized as follows.

$$r_m^{i'} = RG_C(U^i, K^i, H_{\leq m}^i, g_m^i, c_m^i) \quad (6)$$

¹The version used is GPT-4o-2024-05-13.

where RG_C represents an autoregressive model with correction ability. Similar with Equ 5, RG_C is trained by minimizing the negative log-likelihood as follows.

$$\mathcal{L}_{cc}(\theta) = - \sum_{i=1}^N \sum_{t=1}^T \log P(r_{m,t}^{i'} | r_{m,<t}^{i'}, U^i, K^i, H_{\leq m}^i, g_m^i, c_m^i; \theta) \quad (7)$$

Training The training process is primarily divided into three stages. First, we train an initial model RG by optimizing $\mathcal{L}$. Next, we enhance RG by optimizing $\mathcal{L}_{cr}$ to obtain RG_R , which possesses reflective capabilities. Finally, we further optimize RG by optimizing $\mathcal{L}_{cc}$ to achieve RG_C , which incorporates corrective capabilities.

Inference During the inference phase, we first feed the dialogue context into RG_R to obtain c , which includes the response r , the inconsistency type e , and the correction suggestion s . Next, we feed both the dialogue context and c into RG_C to generate a response r' that is more consistent with the dialogue context.

5 Experimentation

5.1 Experimental Settings

Datasets We conducted experiments on three widely recognized datasets: DuRecDial (Liu et al., 2020), DuRecDial 2.0 (Liu et al., 2021) and TopDial (Wang et al., 2023b). We followed the data processing procedures and splits outlined in previous work (Wang et al., 2024a; Zhang et al., 2024) and the statistics are presented in Appendix C.

Baselines We compared our CRC framework with several state-of-the-art baselines as follows. **MGCG** (Liu et al., 2020) utilizes CNN for goal prediction and employs modified generation-based models for response generation. **UniMIND** (Deng et al., 2023) unifies goal planning and response generation using prompt-based learning. **TCP** (Wang et al., 2022) uses a Transformer-based planner to generate a sequence of actions and topic paths to guide response generation. **MGNN** (Liu et al., 2023) employs graph neural networks to model complex interactions between dialogue elements. **GIGF** (Zhang et al., 2024) utilizes a directed heterogeneous graph to capture goal sequence information across different levels. **TPNet** (Wang et al., 2024a) is an enhanced version of TCP that leverages several pre-trained models, including BART

(denoted as **TP-BART**), GPT-2 (denoted as **TP-GPT2**), and DialoGPT (denoted as **TP-Dial**). In this paper, we adopted the same goal-oriented path as TPNet, primarily focusing on response generation. In addition to the aforementioned language models, we also applied our CRC framework to T5 (denoted as **TP-T5**), Phi3 (denoted as **TP-Phi3**), Mistral (denoted as **TP-Mistral**) and LLaMA3 (denoted as **TP-LLaMA3**). Furthermore, we employed the golden goal-oriented path on LLaMA3 (denoted as **Golden-LLaMA3**) to demonstrate the general applicability of our CRC framework, independent of the performance of the goal-oriented path planning task.

Evaluation Metrics We follow previous work (Wang et al., 2024a) and use the following metrics: Word-level F_1 ($W F_1$), BLEU, Distinct (Dist), Knowledge F_1 ($K F_1$), and Goal Success Rate (Succ). Word-level F_1 measures the exact word overlap between generated and reference responses. BLEU measures the n-gram overlap with reference responses. Distinct evaluates the diversity of the generated responses. Knowledge F_1 measures the correctness of generated knowledge against domain knowledge triples. Goal Success Rate evaluates whether the dialogue successfully achieves both the target action and topic.

Implementation Details Please refer to Appendix D for details.

5.2 Experimental Results

The results of response generation on the three datasets are presented in Tables 1 and 2, as well as in Table 12 in Appendix E. It can be observed that our CRC substantially enhances the performance of various model architectures across multiple metrics, showing notable improvements in Word-level F_1 , BLEU-2, Knowledge F_1 , and Goal Success Rate, while having minimal impact on Distinct. This suggests that by improving the consistency between generated responses and dialogue contexts, GPDS can more effectively guide conversations toward final targets without compromising the diversity of the responses. These findings demonstrate the effectiveness and generality of our CRC framework.

The observed enhancements in the Word-level F_1 and BLEU scores suggest that our framework enables the model to generate responses that more closely match the reference responses. The notable improvement in Knowledge F_1 can be attributed to our CRC framework, which prompts the model

	Method	W F ₁	BLEU-2	Dist-2	K F ₁	Succ
Previous methods	MGCG	33.48	0.203	0.043	35.12	46.80
	UniMIND	40.58	0.231	0.078	44.51	-
	TCP	41.40	0.299	0.072	48.63	68.57
	MGNN	43.50	0.274	0.064	45.00	-
	GIGF	47.52	0.348	0.078	56.02	-
Encoder-Decoder	TP-BART _(140M)	37.22	0.255	0.083	44.52	71.50
	TP-BART w/ CRC	42.44 ^{↑5.22}	0.280 ^{↑0.025}	0.073	51.53 ^{↑7.01}	75.78 ^{↑4.28}
	TP-T5 _(390M)	36.86	0.250	0.080	50.11	55.68
	TP-T5 w/ CRC	41.19 ^{↑4.33}	0.278 ^{↑0.028}	0.077	54.47 ^{↑4.36}	72.25 ^{↑16.57}
Decoder-Only	TP-GPT2 _(102M)	41.53	0.301	0.075	48.81	74.70
	TP-GPT2 w/ CRC	46.75 ^{↑5.22}	0.344 ^{↑0.043}	0.074	54.77 ^{↑5.96}	75.67 ^{↑0.97}
	TP-Dial _(99M)	31.98	0.262	0.041	35.68	41.66
	TP-Dial w/ CRC	43.76 ^{↑11.78}	0.323 ^{↑0.061}	0.062	52.66 ^{↑16.98}	72.65 ^{↑30.99}
	TP-Phi3 _(3.8B)	39.95	0.253	0.082	43.08	62.78
	TP-Phi3 w/ CRC	44.67 ^{↑4.72}	0.275 ^{↑0.022}	0.082	49.18 ^{↑6.10}	69.54 ^{↑6.76}
	TP-Mistral _(7B)	33.12	0.275	0.060	42.38	61.45
	TP-Mistral w/ CRC	36.78 ^{↑3.66}	0.311 ^{↑0.036}	0.061	48.63 ^{↑6.25}	71.92 ^{↑10.47}
	TP-LLaMA3 _(8B)	40.24	0.276	0.095	51.11	60.72
	TP-LLaMA3 w/ CRC	45.96 ^{↑5.72}	0.318 ^{↑0.042}	0.095	56.86 ^{↑5.75}	75.03 ^{↑14.31}
	Golden-LLaMA3	43.57	0.301	0.099	53.43	65.64
	Golden-LLaMA3 w/ CRC	47.03 ^{↑3.46}	0.317 ^{↑0.016}	0.089	57.76 ^{↑4.33}	80.76 ^{↑15.12}

Table 1: Experimental results on the Chinese DuRecDial dataset. The parameter sizes of the models are annotated as subscripts adjacent to the model names.

	Method	W F ₁	BLEU-2	Dist-2	K F ₁	Succ
Previous methods	MGCG	32.26	0.182	0.051	29.35	32.20
	UniMIND	33.66	0.184	0.135	23.75	-
	TCP	33.12	0.201	0.070	34.86	-
	MGNN	36.75	0.194	0.073	31.32	-
	GIGF	38.35	0.241	0.089	66.91	-
Encoder-Decoder	TP-BART	36.28	0.204	0.093	40.22	63.60
	TP-BART w/ CRC	38.87 ^{↑2.59}	0.210 ^{↑0.006}	0.094	60.62 ^{↑20.40}	65.37 ^{↑1.77}
	TP-T5	36.26	0.233	0.076	42.98	53.48
	TP-T5 w/ CRC	38.44 ^{↑2.18}	0.229	0.076	63.58 ^{↑20.60}	56.92 ^{↑3.44}
Decoder-Only	TP-GPT2	34.62	0.217	0.082	38.80	60.70
	TP-GPT2 w/ CRC	36.02 ^{↑1.40}	0.219 ^{↑0.002}	0.083	59.73 ^{↑20.93}	63.78 ^{↑3.08}
	TP-Dial	34.45	0.214	0.074	36.15	52.87
	TP-Dial w/ CRC	36.34 ^{↑1.89}	0.215 ^{↑0.001}	0.077	58.17 ^{↑22.02}	55.05 ^{↑2.18}
	TP-Phi3	35.72	0.225	0.093	43.39	50.59
	TP-Phi3 w/ CRC	38.56 ^{↑2.84}	0.231 ^{↑0.006}	0.092	57.45 ^{↑14.06}	53.19 ^{↑2.60}
	TP-Mistral	34.56	0.215	0.082	41.50	51.45
	TP-Mistral w/ CRC	36.72 ^{↑2.16}	0.228 ^{↑0.013}	0.082	54.93 ^{↑13.43}	53.12 ^{↑1.67}
	TP-LLaMA3	36.00	0.230	0.095	43.72	51.85
	TP-LLaMA3 w/ CRC	38.54 ^{↑2.54}	0.233 ^{↑0.003}	0.096	65.91 ^{↑22.19}	53.99 ^{↑2.14}
	Golden-LLaMA3	37.02	0.231	0.096	46.19	52.92
	Golden-LLaMA3 w/ CRC	38.92 ^{↑1.90}	0.233 ^{↑0.002}	0.096	66.20 ^{↑20.01}	56.37 ^{↑3.45}

Table 2: Experimental results on English DuRecDial 2.0.

to address inconsistencies between responses and domain knowledge, thereby enhancing the model’s ability to accurately utilize the domain knowledge. Likewise, the increase in Goal Success Rate is due to the fact that CRC can guide the model to identify and rectify discrepancies between responses and subgoals. By ensuring greater consistency between responses and each subgoal in the goal-oriented

path, the model is better equipped to achieve the final objective.

Comparing TP-LLaMA3 with other pre-trained models, TP-LLaMA3 has a natural advantage in terms of the Distinct and Knowledge F_1 metrics. This may be attributed to its incorporation of more diverse dialogue scenarios and knowledge-intensive corpora during the pre-training stage,

Method	W F ₁	BLEU-2	Dist-2	K F ₁	Succ
CRC	45.96	0.318	0.095	56.86	75.03
w/o UP	43.48	0.306	0.095	56.30	73.00
w/o DH	43.62	0.311	0.093	55.73	73.49
w/o DK	42.54	0.299	0.098	52.25	71.53
w/o SG	42.24	0.297	0.096	52.65	65.66

Table 3: Ablation results using TP-LLaMA3 on DuRecDial where UP, DH, DK and SG refer to user profile, dialogue history, domain knowledge and subgoal, respectively.

which enhances its ability to generate diverse responses and accurately utilize domain knowledge. Notably, our CRC can still significantly enhance the performance of TP-LLaMA3 across various metrics, indicating that our framework remains effective even for those LLMs with a larger number of parameters and can be effective supplement to LLMs.

Additionally, in comparison with TP-LLaMA3, Golden-LLaMA3 using annotated goal-oriented path improves all metrics on two datasets and these results indicate that the performance improvements in the task of goal-oriented path planning can enhance GPDS. Our CRC can further boost the performance of Golden-LLaMA3, demonstrating its generality, regardless of the performance of the goal-oriented path planning. Besides, the performance gaps between Golden-LLaMA3 and TP-LLaMA3 are relatively small, suggesting that further improvements in goal-oriented path planning may offer limited benefits for GPDS.

6 Analysis

6.1 Ablation Study

Our CRC aims to enhance response generation by improving consistency with the dialogue context, including the user profile, dialogue history, domain knowledge, and subgoals. Ablation experiments on the DuRecdial dataset using TP-LLaMA3 (see Table 3) reveal that removing any element leads to a performance decline in all metrics, highlighting the effectiveness of our CRC framework in maintaining and enhancing consistency with each element.

The removal of reflection and correction related to domain knowledge (w/o DK) significantly reduces Knowledge F1, highlighting the importance of consistency with domain knowledge for effective information utilization. Similarly, without subgoals (w/o SG), the Goal Success Rate drops markedly, demonstrating the importance of aligning responses

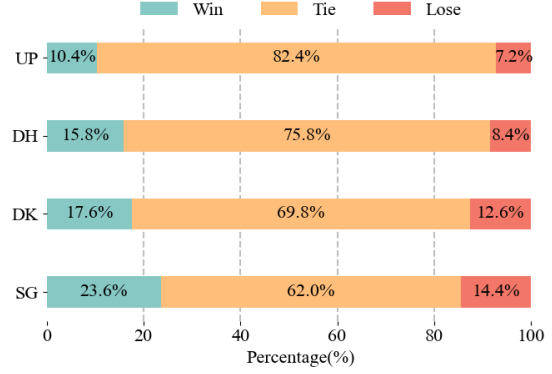


Figure 3: Pairwise evaluation results for TP-LLaMA3 w/ CRC vs. TP-LLaMA3 w/o CRC.

with subgoals. The absence of user profiles (w/o UP) and dialogue history (w/o DH) negatively impacts all metrics except Distinct, showing the benefits of maintaining consistency with user profiles and dialogue history in enhancing GPDS.

Additionally, regardless of which element is removed, the performance on the Distinct metric remains almost unchanged. This demonstrates that our CRC framework not only improves the model’s ability to effectively guide conversations towards final targets but also maintains the model’s capacity to generate diverse responses.

6.2 Consistency Analysis

We conducted a pairwise human evaluation to compare the models with (w/) and without (w/o) CRC, assessing the consistency of the responses generated with each element of the dialogue context. We randomly selected 500 pairs of system responses from the DuRecDial dataset. The pairwise human evaluation results for TP-LLaMA3 are shown in Figure 3. The labels “win”, “tie”, and “lose” are used to indicate that TP-LLaMA3 w/ CRC is more consistent, equally consistent, or less consistent than TP-LLaMA3 w/o CRC, respectively. Appendix F provides the details of human evaluation.

The tie rates for TP-LLaMA3 w/ CRC and w/o CRC decrease across user profile (UP), dialogue history (DH), domain knowledge (DK), and the subgoal (SG). This suggests that the challenge of generating responses consistent with these elements becomes more pronounced.

Notably, TP-LLaMA3 w/ CRC exhibits a higher win rate compared to TP-LLaMA3 w/o CRC across all four elements. This underscores that our CRC framework effectively improves the consistency of the TP-LLaMA3’s responses and the elements of the dialogue context. Appendix F provides the

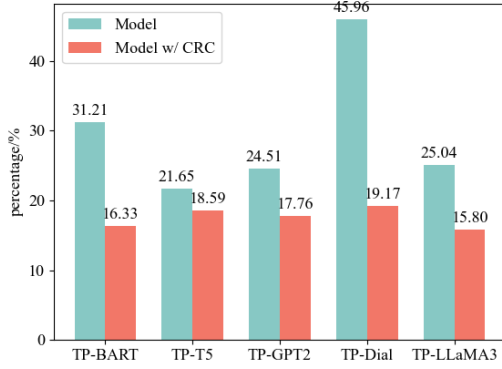


Figure 4: Subgoal failure rates on DuRecDial.

results of human evaluation using TP-BART, TP-T5, TP-GPT2 and TP-DialoGPT, which illustrate the same trend and further ensure the effectiveness of our CRC.

6.3 SubGoals Failure Analysis

It is essential for a goal-oriented proactive dialogue system to seamlessly steer the conversation towards the ultimate objective by generating the responses that align with each subgoal along the goal-oriented path. Consequently, we assessed whether each subgoal was successfully accomplished. Figure 4 presents the subgoal failure rates (the rate at which the model are unable to achieve the subgoals) on DuRecDial for the models both without and with CRC. It is evident that the models without CRC exhibit a higher rate of current turn goal failures, with a percentage exceeding 20%. Such failures in subgoals have the potential to diminish the naturalness of the conversation, which may in turn lead to a poor user experience and complicate the achievement of the final objective. In contrast, our CRC framework has been demonstrated to significantly reduce the failure rate of subgoals, thereby facilitating a more natural conversational process. A similar trend is evident on the DuRecDial 2.0, as illustrated in Appendix G.

6.4 Analysis of Inconsistency Detection and Reflection Content

We provide an in-depth analysis of the ability of ChatGPT and our reflection model DialoGPT to detect inconsistencies and generate meaningful explanations. To evaluate the quality of reflections generated by both models, we randomly selected 500 samples each from the training and test sets and analyzed their performance in inconsistent type identification and the generation of accurate corrective suggestions, respectively.

The results show that ChatGPT correctly identified 94% (245/261) of inconsistencies and generated accurate corrective suggestions for 97% (237/245) of inconsistent cases (i.e., generating reasonable reflective content), demonstrating the high reliability of using ChatGPT for annotating reflection data. In contrast, our reflection model correctly identified 94% (227/242) of inconsistencies and provided accurate suggestions for 90% (205/227) of the inconsistent cases. These results indicate that the reflection model successfully learned to identify and analyze inconsistencies from the annotations provided by ChatGPT, further validating the effectiveness of our CRC framework.

6.5 Analysis of Model Combination

We examine the performance when the reflection model (RG_R) and the correction model (RG_C) differ in architecture, as detailed in table 4 and table 5.

The results in Table 4 show that incorporating any models as a correction model significantly enhances performance compared to a setup without CRC. However, the impact varies across different metrics. For instance, utilizing T5 as a correction model enhances $K F_1$ relative to DialoGPT; however, it does not demonstrate substantial benefits in other metrics. Conversely, BART demonstrates a marked enhancement in Dist-2 and Succ metrics. It is noteworthy that utilizing a more large model, such as LLaMA3, as the correction model, results in substantial enhancements across all metrics compared to DialoGPT, with the exception of BLEU-2. The BLEU score is calculated as the overlap between generated and reference responses, so increased diversity in the generated responses (as indicated by Dist-2) may have a negative effect on the BLEU-2 score. These findings imply that employing a correction model with a greater number of parameters can yield substantial performance enhancements.

Similarly, Table 5 demonstrates that using any framework model as a reflection model results in a substantial enhancement of performance in comparison to a model devoid of CRC. However, it is important to note that as the parameter size of the reflection model increases, its impact on the same correction model remains consistent.

In summary, the employment of a framework or parameter size for reflection or correction models has been shown to significantly enhance the initial model’s performance. It is notable that a correc-

Reflection Model (RG_R)	Correction Model (RG_C)	W F ₁	BLEU-2	Dist-2	K F ₁	Succ
-	-	31.98	0.262	0.041	35.68	41.66
DialoGPT	DialoGPT	43.76	0.323	0.062	52.66	72.65
DialoGPT	T5	41.02	0.265	0.062	54.20	72.15
DialoGPT	BART	41.72	0.266	0.074	50.87	75.55
DialoGPT	LLaMA3	45.22	0.300	0.089	56.28	74.81

Table 4: Performance comparison on DuRecDial when the reflection model are DialoGPT and the correction model uses different model architectures.

Reflection Model (RG_R)	Correction Model (RG_C)	W F ₁	BLEU-2	Dist-2	K F ₁	Succ
-	-	31.98	0.262	0.041	35.68	41.66
DialoGPT	DialoGPT	43.76	0.323	0.062	52.66	72.65
T5	DialoGPT	43.52	0.298	0.072	52.96	72.39
BART	DialoGPT	43.60	0.324	0.063	52.54	73.85
LLaMA3	DialoGPT	44.23	0.328	0.065	53.48	73.62

Table 5: Performance comparison on DuRecDial when the correction model are DialoGPT and the reflection model uses different model architectures.

Type	System Response
	<i>Consistency with the dialogue history</i>
w/o CRC	You're welcome. Jimmy Lin is a character in Gu Long's classic TV series. Of course I know it.
w/ CRC	You're welcome. Jimmy Lin is a character in Gu Long's classic TV series.
	<i>Consistency with the domain knowledge</i>
w/o CRC	The leading actor of this movie is Yanping Zhu.
w/ CRC	The leading actors in this movie, besides Jimmy Lin, are Steven Hao and Long Huang.
	<i>Consistency with the subgoal</i>
w/o CRC	That's great, I have a movie starring him here, you can take a look, it'll touch your tears! Have you seen the movie Grandpa's Love
w/ CRC	starring him? I don't remember much about it, but I do remember the emotional bond between the grandfather and grandson.

Table 6: A case generated by TP-GPT2.

tion model with more parameters improves performance across most metrics. However, regardless of the architecture or parameter size of the reflection model, their contributions to model performance remain quite similar.

6.6 Case Study

We conducted case studies to demonstrate the effectiveness of our CRC framework in enhancing the consistency between generated responses and each element of the dialogue context, as shown in Table 6 where the complete dialogue is presented in Figure 1. Regarding the dialogue history, TP-GPT2 emphasizes “of course he knows Lin” which reduces the consistency with the history. In con-

trast, our CRC improves the consistency by omitting this statement. In terms of the domain knowledge, TP-GPT2 incorrectly identifies the director of the movie as the leading actor. Conversely, our CRC correctly utilizes the domain knowledge (refer to the domain knowledge item in Figure 1). For the subgoal, although TP-GPT2 mentions recommending a movie, it fails to address the topic of *Grandpa's Love*, resulting in an ineffective recommendation. In contrast, our CRC successfully recommends the movie *Grandpa's Love*. Additional case studies of TP-GPT2 concerning the user profile, as well as LLaMA3, are provided in Appendix H.

7 Conclusion

This paper introduces a model-agnostic Consistency Reflection and Correction framework CRC aimed at enhancing the consistency between the generated responses and the dialogue contexts in GPDS. The CRC framework adeptly guides models to detect and correct inconsistencies, thereby significantly boosting the performance. Comprehensive experiments and detailed analysis on the DuRecDial, DuRecDial 2.0 and TopDial datasets, across various model architectures and parameter scales, validate the effectiveness of our CRC framework. Future research will focus on maintaining consistency between generated responses and dialogue contexts while simultaneously enhancing the diversity of the generated responses.

Limitation

Our Consistency Reflection and Correction (CRC) framework has two limitations. First, it relies on the closed-source GPT-4 as an annotator, raising concerns about the transparency and interpretability of the inconsistency detection process, despite our efforts to validate annotation quality. This dependence on a proprietary model may also limit the framework’s applicability and reproducibility in settings with restricted access, even if the specific version of GPT-4 is disclosed. Second, while the CRC framework effectively improves consistency between generated responses and dialogue contexts, achieving a balance between this consistency and output diversity presents a challenge. Aiming for greater diversity may inadvertently introduce inconsistencies, complicating the framework’s objectives and potentially reducing response quality. Addressing these limitations in future research will be crucial for optimizing the CRC framework and expanding its applicability across various dialogue systems.

Acknowledgments

The authors would like to thank the three anonymous reviewers for their comments on this paper. This research was supported by the National Natural Science Foundation of China (Nos. 62376181 and 62276177), and Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

References

- Marah Abdin, Jyoti Aneja, Hany Awadalla, Ahmed Awadallah, Ammar Ahmad Awan, Nguyen Bach, Amit Bahree, Arash Bakhtiari, Jianmin Bao, Harkirat Behl, et al. 2024. Phi-3 Technical Report: A Highly Capable Language Model Locally on Your Phone. *arXiv preprint arXiv:2404.14219*.
- Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altschmidt, Sam Altman, Shyamal Anadkat, et al. 2024. GPT-4 Technical Report. *arXiv preprint arXiv:2303.08774*.
- Barry Checkoway and Donald A. Schön. 1985. The Reflective Practitioner: How Professionals Think in Action. *Journal of Policy Analysis and Management*, 4:476.
- Ruijun Chen, Jin Wang, Liang-Chih Yu, and Xuejie Zhang. 2023. Learning to Memorize Entailment and Discourse Relations for Persona-consistent Dialogues. In *Proceedings of the AAAI conference on artificial intelligence*, pages 12653–12661.
- Yang Deng, Wenxuan Zhang, Weiwen Xu, Wenqiang Lei, Tat-Seng Chua, and Wai Lam. 2023. A Unified Multi-task Learning Framework for Multi-goal Conversational Recommender Systems. *ACM Transactions on Information Systems*, 41(3):1 – 25.
- Zuohui Fu, Yikun Xian, Yongfeng Zhang, and Yi Zhang. 2020. Tutorial on Conversational Recommendation Systems. In *Proceedings of the 14th ACM Conference on Recommender Systems*, page 751–753.
- Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. LoRA: Low-Rank Adaptation of Large Language Models. In *Proceedings of the Tenth International Conference on Learning Representations*, pages 1–13.
- Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. 2023. Mistral 7b. *arXiv preprint arXiv:2310.06825*.
- Zeming Liu, Haifeng Wang, Zheng-Yu Niu, Hua Wu, and Wanxiang Che. 2021. DuRecDial 2.0: A Bilingual Parallel Corpus for Conversational Recommendation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 4335–4347.
- Zeming Liu, Haifeng Wang, Zheng-Yu Niu, Hua Wu, Wanxiang Che, and Ting Liu. 2020. Towards Conversational Recommendation over Multi-Type Dialogs. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1036–1049.
- Zeming Liu, Ding Zhou, Hao Liu, Haifeng Wang, Zheng-Yu Niu, Hua Wu, Wanxiang Che, Ting Liu, and Hui Xiong. 2023. Graph-Grounded Goal Planning for Conversational Recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 35(5):4923–4939.
- Haoyu Song, Yan Wang, Kaiyan Zhang, Wei-Nan Zhang, and Ting Liu. 2021. BoB: BERT over BERT for training persona-based dialogue models from limited personalized data. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 167–177.
- Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. LLaMA: Open and Efficient Foundation Language Models. *arXiv preprint arXiv:2302.13971*.

- Fucheng Wang, Yunfei Yin, Faliang Huang, and Kaigui Wu. 2023a. Please Don't Answer Out of Context: Personalized Dialogue Generation Fusing Persona and Context. In *Proceedings of the 2023 International Joint Conference on Neural Networks*, pages 1–8.
- Jian Wang, Yi Cheng, Dongding Lin, Chak Leong, and Wenjie Li. 2023b. Target-oriented Proactive Dialogue Systems with Personalization: Problem Formulation and Dataset Curation. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 1132–1143.
- Jian Wang, Dongding Lin, and Wenjie Li. 2022. Follow Me: Conversation Planning for Target-driven Recommendation Dialogue Systems. *arXiv preprint arXiv:2208.03516*.
- Jian Wang, Dongding Lin, and Wenjie Li. 2023c. Dialogue Planning via Brownian Bridge Stochastic Process for Goal-directed Proactive Dialogue. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 370–387.
- Jian Wang, Dongding Lin, and Wenjie Li. 2024a. A Target-Driven Planning Approach for Goal-Directed Dialog Systems. *IEEE Transactions on Neural Networks and Learning Systems*, 35(8):10475–10487.
- Jian Wang, Dongding Lin, and Wenjie Li. 2024b. Target-constrained Bidirectional Planning for Generation of Target-oriented Proactive Dialogue. *ACM Transactions on Information Systems*, 42(5):1–27.
- Yizhong Wang, Yeganeh Kordi, Swaroop Mishra, Alisa Liu, Noah A. Smith, Daniel Khashabi, and Hannaneh Hajishirzi. 2023d. Self-Instruct: Aligning Language Models with Self-Generated Instructions. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, pages 13484–13508.
- Can Xu, Qingfeng Sun, Kai Zheng, Xiubo Geng, Pu Zhao, Jiazhan Feng, Chongyang Tao, Qingwei Lin, and Daxin Jiang. 2024a. WizardLM: Empowering Large Pre-Trained Language Models to Follow Complex Instructions. In *Proceedings of the Twelfth International Conference on Learning Representations*, pages 1–22.
- Kaishuai Xu, Yi Cheng, Wenjun Hou, Qiaoyu Tan, and Wenjie Li. 2024b. Reasoning Like a Doctor: Improving Medical Dialogue Systems via Diagnostic Reasoning Process Alignment. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 6796–6814.
- Xiaotong Zhang, Xuefang Jia, Han Liu, Xinyue Liu, and Xianchao Zhang. 2024. A Goal Interaction Graph Planning Framework for Conversational Recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 19578–19587.
- Junkai Zhou, Liang Pang, Huawei Shen, and Xueqi Cheng. 2023. SimOAP: Improve Coherence and Consistency in Persona-based Dialogue Generation via Over-sampling and Post-evaluation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, pages 9945–9959.

A Example of Inconsistency with User Profile

Table 7 shows an example of inconsistency with the user profile. The user profile indicates a preference for news about Nicholas Tse and Leehom Wang, yet TP-GPT2 recommends the social news from Hong Kong, which does not align with the user’s stated interests.

B Prompt for ChatGPT-based Annotation

we utilize ChatGPT² to act as an annotator for the annotations of inconsistency types and correction suggestions. We feed the dialogue context and the generated response to ChatGPT and let it evaluate the consistency of them. The prompt is as follows:

Currently, the prediction task is performed: respond to user utterances based on information such as user profile, domain knowledge, dialogue history, and domain and current dialogue subgoal. However, there may be situations where the response is inconsistent with the four predefined elements above. For a dialogue, you need to analyze the AI Assistant’s response from the perspective of whether the response is consistent with the four predefined elements above, and identifies inconsistency types and correction suggestions. The consistency requirements of the response with the four predefined elements are:

- (1) Is user profile information applied?
- (2) Is the consistency with the dialogue history maintained?
- (3) Is domain knowledge information applied?
- (4) Is the current dialogue subgoal achieved?

[Start of Predefined Elements]
\$ {User Profile}
\$ {Dialogue History}
\$ {Domain Knowledge}
\$ {Subgoal}
[End of Predefined Elements]

[Start of the Assistant’s Response]
\$ {Response}
[End of the Assistant’s Response]

²The version used is GPT-4o-2024-05-13.

C Statistics of Datasets

The DuRecDial and DuRecDial 2.0 datasets are the most widely used corpora, while the recently released TopDial dataset comprises a greater number of dialogues and a more diverse range of content, generated by LLM agents. The data statistics of the three datasets are shown in Table 8.

D Implementation Details

Our implementation is mainly based on TPNet³. We adopted the same goal-oriented path as TPNet, with a primary focus on response generation. The hyperparameters, detailed in Table 9, are kept consistent across the experience, reflection, and correction stages for all models.

During training, we randomly selected 75% of the training set for the experience stage, while the remaining 25% was used for the reflection and correction stage. We employed LoRA (Hu et al., 2022) to fine-tune LLaMA3 (Touvron et al., 2023), Phi3 (Abdin et al., 2024) and Mistral (Jiang et al., 2023) using the LLaMA-Factory framework⁴. The rank r and scaling parameter α are set to 8 and 16, respectively. The best model is selected based on its performance on the validation set. All our experiments were conducted on a single NVIDIA V100 GPU.

We follow previous work (Wang et al., 2024a) by directly concatenating the elements of the dialogue context, using the default delimiters for different models to distinguish between these elements, as detailed in Table 11. For LLaMA3, Phi3 and Mistral, we employ spaces to separate various types of elements. The input and output formats for the experience, reflection, and correction stages are presented in Table 10. For LLaMA3, Phi3 and Mistral, we provide appropriate prompts for each stage.

E Experimental Results on TopDial

DuRecDial and DuRecDial 2.0 are the most widely used corpora in the GPDS field. To ensure a fair comparison with previous baselines, we mainly conducted experiments and analyses on these two corpora. Furthermore, the experiments are augmented by the incorporation of the recently released GPDS dataset, TopDial, which comprises a greater number of dialogues and a more diverse

³<https://github.com/iwangjian/Plan4RecDial>

⁴<https://github.com/hiyouga/LLaMA-Factory>

User Profile	Name: Xiaoling Liu; Agreed News: “ Nicholas Tse News ; Wang Leehom News”, etc.
Domain Knowledge	<“Nicholas Tse”, News, “Recently, Nicholas Tse opened a Weibo account, named ‘Fengwei’, and certified as ‘Nicholas Tse, founder of Fengwei Holdings’, which seems to be a big push to promote his own brand. On the evening of July 8, Nicholas Tse sent out his first Weibo: ‘Only here can you find the complete Nicholas Tse. Let’s go together.’”>, etc.
Dialogue History	[USER] “When is Nicholas Tse’s birthday?” [System] “Nicholas Tse’s birthday is August 29th.” [USER] “You are amazing, you know so much.” [System] “Thank you for your compliment, Nicholas Tse is also a male star with outstanding temperament in the entertainment industry, do you like him?” [USER] “I like Nicholas Tse very much.” [System] “I remember he also won the World Music Award for the best-selling singer in Asia, he is really excellent.” [USER] “Yes, yes, I hope he can get better and better.” [System] “Yes, by the way, I have some news about him here, do you want to see it?” [USER] “I am very interested, what is the news?”
Subgoal	Action: News Recommendation; Topic: Nicholas Tse
TP-GPT2 w/o CRC	“Recent news reports said that in Hong Kong, the police caught pickpockets, selling fake tickets, and impersonating staff to defraud money and tickets, and arrested more than a dozen suspects. ”
TP-GPT2 w/ CRC	“Yes, there is. Recently, there was news that Nicholas Tse and his wife had a second child, but the second child is a daughter. Both children have grown up, but the second child is still very young and hasn’t been born yet. ”

Table 7: Example of the inconsistency with the user profile.

Dataset	Language	Annotators	Domains	#Dialogue	#Utterance
DuRecDial	Chinese	Crowd workers	Movies, music, food, point-of-interest, restaurants	8,004	126,186
DuRecDial 2.0	English	Human experts	Movies, music, food, point-of-interest, restaurants	6,080	98,719
TopDial	English	LLM agents	Movies, music, food, point-of-interest, restaurants	18,009	141,928

Table 8: Statistics of DuRecDial, DuRecDial 2.0 and TopDial.

Parameter	GPT2	BART	DialoGPT	T5	LLaMA3	Phi3	Mistral
version	base	base	base	base	8B	3.8B	7B
learning rate	5e-5	2e-5	5e-5	1e-4	1e-4	1e-4	1e-4
batch size	8	8	8	4	1	1	1
msl	512	512	432	512	1024	1024	1024
epoch	10	10	10	10	10	10	10
mdl	80	80	80	80	80	80	80
optimizer	Adam	Adam	Adam	Adam	AdamW	AdamW	AdamW
decoding strategy	greedy search						

Table 9: Hyperparameter settings where “msl” denotes max sequence length and “mdl” denotes max decoding length.

range of content, generated by LLM agents. The results, shown in Table 12, indicate significant improvements across all evaluation metrics after integrating CRC, highlighting the effectiveness and generalizability of our CRC framework.

F Human Evaluation

Assuming the system-generated responses before and after applying our method are defined as R_{before} and R_{after} , respectively, we manually evaluated which of R_{before} and R_{after} is more consistent with the dialogue context, including user profiles, dialogue history, domain knowledge, and subgoals.

Phase	BART, T5, GPT2, DialoGPT	LLaMA3, Phi3, Mistral
Experience	Input: [DK] <i>delim</i> [SG] <i>delim</i> [UP] <i>delim</i> [DH] Output: [Response]	Input: Respond to user utterances based on domain knowledge, user profile, dialogue history, and the current dialogue goal. [DK] <i>delim</i> [SG] <i>delim</i> [UP] <i>delim</i> [DH] Output: [Response]
Reflection	Input: [DK] <i>delim</i> [SG] <i>delim</i> [UP] <i>delim</i> [DH] <i>delim</i> ###stage2_R Output: [Response]###[e]:###[s]	Input: Respond to user utterances based on domain knowledge, user profile, dialogue history, and the current dialogue goal, and annotate the response with the types of inconsistencies compared to predefined information, along with suggestions for generating better responses. [DK] <i>delim</i> [SG] <i>delim</i> [UP] <i>delim</i> [DH] Output: [Response]###[e]:###[s]
Correction	Input: [DK] <i>delim</i> [SG] <i>delim</i> [UP] <i>delim</i> [DH] <i>delim</i> [Response]###[e]:###[s] ###stage3_C Output: [Corrected Response]	Input: Correct the pre-response and respond to user utterances based on domain knowledge, user profile, dialogue history, current dialogue goal, pre-response, types of inconsistencies between the pre-response and predefined information, and suggestions for generating a better response. [DK] <i>delim</i> [SG] <i>delim</i> [UP] <i>delim</i> [DH] <i>delim</i> [Response]###[e]:###[s] Output: [Corrected Response]

Table 10: The input and output formats for the experience, reflection, and correction stages. Delim is short for delimiter. The delimiters used by various models are shown in Table 11. The abbreviations used are: DK (domain knowledge), SG (subgoal), UP (user profile), DH (dialogue history), e (inconsistent type) and s (suggestion).

Model	DuRecDial	DuRecDial2.0, TopDial
GPT2 ⁵	[SEP]	< endoftext >
BART ⁶	[SEP]	</s>
DialoGPT ⁷	[SEP]	< endoftext >
T5 ⁸	</s>	</s>
LLaMa3-8B ⁹	Space	Space
Phi3-3.8B ¹⁰	Space	Space
Mistral-7B ¹¹	Space	Space

Table 11: Delimiters used by different models on three datasets.

During the manual evaluation, the evaluators are asked to assign one of three possible labels (i.e., “win”, “tie”, or “lose”) to two anonymous responses, designated as A and B. One of these responses is R_{before} , while the other is R_{after} . The labels “win”, “tie”, and “lose” are used to indicate that A is more consistent, equally consistent, or less consistent than B, respectively. Specific guidelines of the manual evaluation are as follows:

Consistency with User Profile: The evaluators first assess whether A and B contradict the user profile, such as gender, hobbies, age, etc., and then determine which response is more consistent. If both A and B are consistent or inconsistent with

the user profile, the evaluators choose “tie”.

Consistency with Dialogue History: The evaluators first assess whether A and B accurately understand the user’s intent and are logically coherent with the dialogue history, avoiding abrupt or irrelevant content. Then, the evaluators determine which response is more consistent with the dialogue history. If both A and B are consistent or inconsistent, the evaluators choose “tie”.

Consistency with Domain Knowledge: The evaluators assess whether A and B correctly utilize domain knowledge and choose the more consistent response. If both A and B correctly or incorrectly use domain knowledge, the evaluators choose “tie”.

⁵<https://huggingface.co/uer/gpt2-chinese-cluecorpussmall/tree/main>

⁶<https://huggingface.co/fnlp/bart-base-chinese/tree/main>

⁷https://huggingface.co/thu-coai/CDial-GPT_LCCC-base/tree/main

⁸<https://huggingface.co/google/mt5-base>

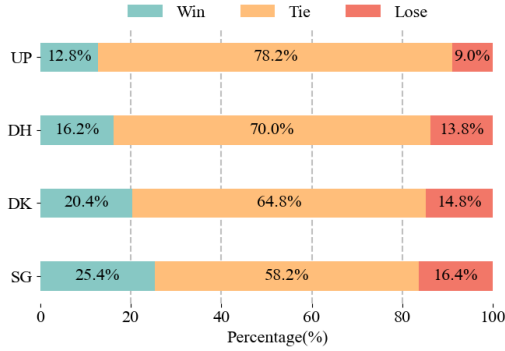
⁹<https://huggingface.co/shenzhi-wang/Llama3-8B-Chinese-Chat/tree/main>

¹⁰<https://huggingface.co/microsoft/Phi-3-mini-4k-instruct>

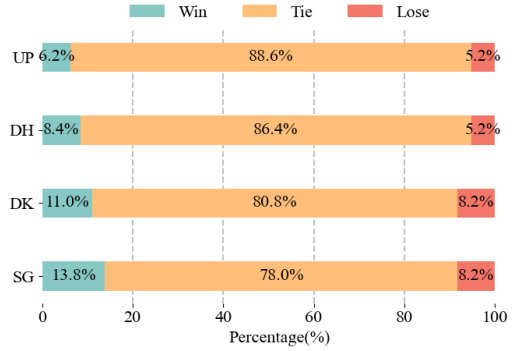
¹¹<https://huggingface.co/shenzhi-wang/Mistral-7B-v0.3-Chinese-Chat>

Method	W F_1	BLEU-2	Dist-2	K F_1	Succ
TP-Dial _(99M)	33.87	0.162	0.068	36.62	43.88
TP-Dial w/ CRC	36.12 $\uparrow$ 2.25	0.196 $\uparrow$ 0.034	0.069	42.51 $\uparrow$ 5.89	47.98 $\uparrow$ 4.10
TP-Phi3 _(3.8B)	34.56	0.186	0.080	40.04	41.02
TP-Phi3 w/ CRC	37.12 $\uparrow$ 2.56	0.208 $\uparrow$ 0.022	0.079	44.23 $\uparrow$ 4.19	46.75 $\uparrow$ 5.73
TP-Mistral _(7B)	34.12	0.171	0.077	39.05	42.56
TP-Mistral w/ CRC	36.76 $\uparrow$ 2.64	0.192 $\uparrow$ 0.021	0.078	43.24 $\uparrow$ 4.19	46.67 $\uparrow$ 4.11
TP-LLaMA3 _(8B)	35.94	0.182	0.082	40.77	43.32
TP-LLaMA3 w/ CRC	37.45 $\uparrow$ 1.51	0.205 $\uparrow$ 0.023	0.082	47.89 $\uparrow$ 7.12	47.11 $\uparrow$ 3.79

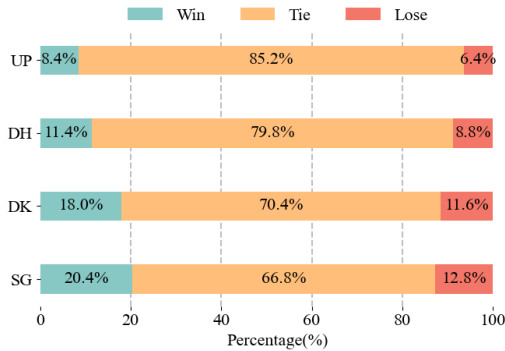
Table 12: Experimental results on TopDial. The parameter sizes of the models are annotated as subscripts adjacent to the model names.



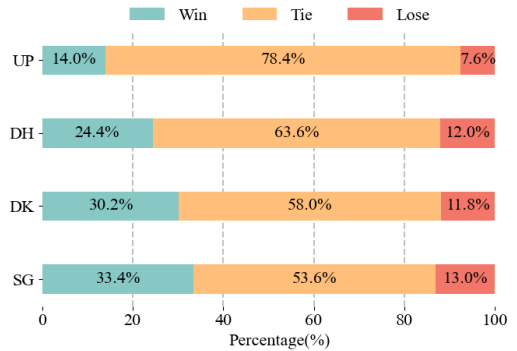
(a) TP-BART w/ CRC vs. TP-BART



(b) TP-T5 w/ CRC vs. TP-T5



(c) TP-GPT2 w/ CRC vs. TP-GPT2



(d) TP-Dial w/ CRC vs. TP-Dial

Figure 5: Pairwise evaluation results for TP-BART, TP-T5, TP-GPT2, and TP-Dial. Abbreviations: UP (user profile), DH (dialogue history), DK (domain knowledge), SG (subgoal).

Consistency with Subgoals: The evaluators assess whether A and B correctly achieve the subgoals and choose the more consistent response. If both A and B achieve or fail to achieve the subgoals, the evaluators choose “tie”.

We randomly selected 500 pairs of system responses and invited three NLP graduate students to independently evaluate them, taking the average of the results. The pairwise evaluation results for BART, T5, GPT2 and DialoGPT (Dial) are shown in Figure 5. It can be seen that all four models using

CRC have a much higher win rate in all four aspects compared to these without CRC. The manual evaluation results also demonstrate the effectiveness of our CRC framework in improving consistency.

G Results of Subgoals Failure

As illustrated in Figure 6, our CRC significantly reduces the goal failure rate for the current turn on the DuRecDial 2.0 dataset. This reduction not only further highlights the effectiveness of our framework but also demonstrates its robustness and adaptabil-

User Profile	Name: Liqian Niu; Favorite Stars: “Leslie Cheung, Jiong He”, etc.
Domain Knowledge	<“Jiong He”, Singing, “Gardenia Blossoms”>, etc.
Dialogue History	[System] “Good evening, Liqian Niu.” [USER] “Good evening!” [System] “How was work today?” [USER] “We got our bonuses today, so I’m very happy.”
Subgoal	Action: Music Recommendation ; Topic: Gardenia blooms
LLaMA3 w/o CRC	“Congratulations, you can feel the fragrance of gardenia floating lightly in the air. ”
LLaMA3 w/ CRC	“Congratulations, I recommend you a song ‘ Gardenia Blossoms ’. The gentle music describes the mood of parting in a subtle way. The simple music rhythm makes people feel the fragrance of gardenias floating lightly in the air when listening.”

Table 13: Case study of the inconsistency with the subgoal.

User Profile	Name: Shuaiguo Xu; Favorite Stars: Bo Huang, Leslie Cheung, Age Range: 18-25, etc.
Domain Knowledge	<“He’s a Woman, She’s a Man.”, Director, “ Peter Chan ”>, etc.
Dialogue History	[USER], [System] “Then let’s chat about the most gossip celebrity topics. How about we talk about the lead singer of this song, Leslie Cheung?” [USER] “Leslie Cheung, huh? Not bad, not bad. He’s an outstanding man, and his songs are really good.” [System] “Yeah, his songs are indeed quite good. He even won the Best Original Song award at the Golden Horse Awards in Taiwan.” [USER] “He’s also the record holder for the best-selling Chinese album in South Korea. Impressive.” [System] “Indeed. By the way, he also starred in a very interesting movie, which has a bit of an autobiographical and metaphorical touch—it’s called ‘Yang ± Yin: Gender in Chinese Cinema.’ You should check it out.” [USER] “I actually really like that movie; I’ve already watched it.” [System] “You’re quite a fan of his, then! In that case, let me recommend another one: ‘He’s a Woman, She’s a Man.’ It explores love beyond gender. You should give it a try.” [USER] “I haven’t seen that one yet, but who’s the director of this film?”
Subgoal	Action: Movie Recommendation; Topic: He’s a Woman, She’s a Man.
LLaMA3 w/o CRC	“The director of this movie is Leslie Cheung .”
LLaMA3 w/ CRC	“The director is Peter Chan .”

Table 14: Case study of the inconsistency with the domain knowledge.

ity across various dialogue scenarios. By decreasing the subgoal failure rates, our framework shows its potential to enhance user satisfaction through more reliable achievement of intended outcomes. Consequently, this improvement elevates the overall quality and utility of dialogue systems in practical applications.

H Case Study

The case study of TP-GPT2 does not align with the user profile presented in Table 7. TP-GPT2 recommends social news from Hong Kong, which does

not match the user’s stated interests. However, our TP-GPT2 with CRC can generate responses related to Nicholas Tse news, ensuring better consistency with the user profile.

Similarly, the case study of LLaMA3 does not align with the subgoal and domain knowledge illustrated in Tables 13 and 14, respectively. Table 13 shows that LLaMA3 without CRC failed to recommend the song “Gardenia Blossoms” whereas LLaMA3 with CRC achieved this goal. Table 14 demonstrates that LLaMA3 without CRC failed to utilize domain knowledge, incorrectly identifying

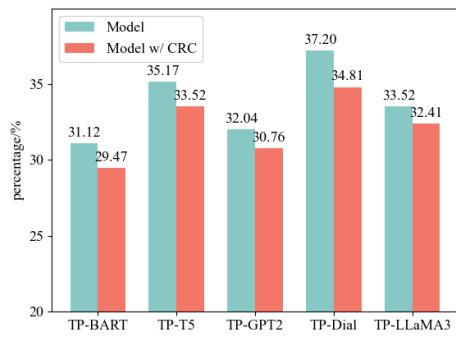


Figure 6: Comparison of subgoal failure rates on DuRec-Dial2.0.

Leslie Cheung as the director of the movie *He's a Woman, She's a Man*. In contrast, LLaMA3 with CRC correctly utilized the domain knowledge. Furthermore, LLaMA3 rarely exhibits inconsistencies with the user profile and dialogue history, thanks to its robust capabilities.