Training Data Optimization for Persona-grounded Dialog via Synthetic Label Augmentation

Anonymous ACL submission

Abstract

The goal of persona-grounded dialogue systems is to enhance the quality of AI agent responses by bolstering persona consistency and promoting response diversity. Although model tuning has seen significant advancements, there is an ongoing need to refine the training data itself. Expanding the scope of personas has been suggested as a means to bridge this gap. Nevertheless, the lack of gold labels that align with these expanded personas poses a challenge for AI agents in training the extent of real-world knowledge. To tackle these challenges, we propose the Synthetic Label Augmentation framework. This framework (1) creates a background skeleton from the original gold labels, masking persona-related elements, (2) infuses the background skeleton with expanded-persona features, generating synthetic gold labels, (3) identifies the most appropriate synthetic gold labels among the candidates, and (4) merges them into Persona-Chat. To substantiate the effectiveness of Optimized Persona-Chat, we assess the quality of synthetic gold labels and interact with agents trained on this enhanced dataset. Our experimental results demonstrate that the framework is a powerful tool for augmenting Persona-Chat quality, and the optimized dataset significantly improves AI agent response quality with respect to persona consistency and response diversity.

1 Introduction

With the advent of the transformer architecture (Vaswani et al., 2017), pre-trained language models have experienced remarkable performance improvements, enabling the models to fully harness the potential features of the data. However, during a recent AI workshop, Andrew Ng asserted, "Focusing on model tuning has been sufficient until now. Improving data quality carries greater importance." As if anticipating the emergence of data-centric AI, research on enhancing data quality (Wang et al., 2018; Han et al., 2020) has been conducted for quite some time. Along with this trend, research using various aspects such as personas (Zhang et al., 2018), emotions (Zhou et al., 2018), images (Shuster et al., 2020), external knowledge sources like Wikipedia (Dinan et al., 2018; Jang et al., 2022), and commonsense graphs (Speer et al., 2017; Sap et al., 2019) have been explored to create high-quality training datasets for conversational AI. In contrast to goal-oriented conversations, maintaining consistency in open-domain conversations is challenging for AI agents due to the vast range of topics, which makes it difficult to sustain user engagement (Qian et al., 2018; Song et al., 2020b; Liu et al., 2020). Therefore, ensuring the coherence of an agent's responses is crucial in preventing users from disengaging from the conversation. In this context, Persona-Chat (PC) dataset (Zhang et al., 2018) was introduced, demonstrating that employing an utterance infused with a specific persona as the gold label assists AI agents in generating responses consistent with the given persona. Nevertheless, the diversity of predefined personas in the dataset remains insufficient to encompass the spectrum of personalities and characteristics present in the real world.

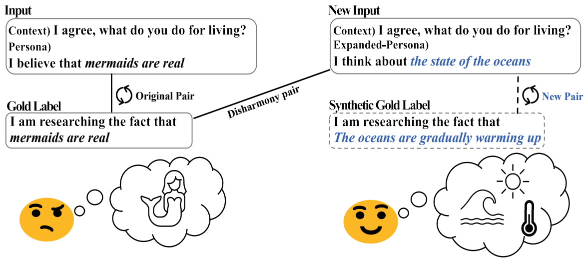


Figure 1: Expanded persona-grounded dialogue demands the synthetic gold label. The original pair comes from Persona-Chat, and the disharmony pair is from Persona-Chat with COMET expansion.

In order to approximate AI agents with various concepts and characteristics found in the real world, Majumder et al. (2020) sought to improve the diver-

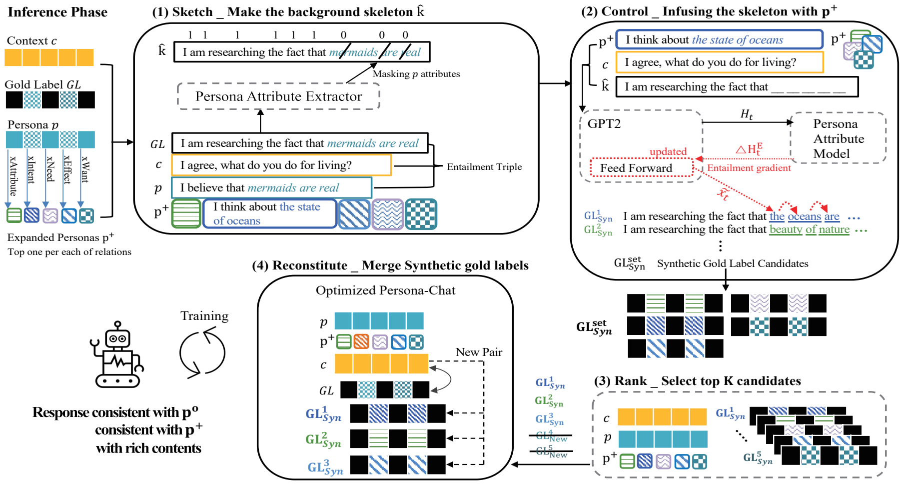


Figure 2: The overview of SLA framework. Trainable models are depicted by the dashed box.

sity of predefined personas by leveraging COMET (Bosselut et al., 2019), a transformer model trained to make causal and semantic inferences based on the ATOMIC graph (Sap et al., 2019). Although this approach appears reasonable, the lack of suitable gold labels incorporating expanded-persona attributes hinders the agent's ability to effectively learn the features of these expanded personas. As illustrated in Figure 1, an example of Persona-Chat with COMET expansion includes a context, persona, gold label imbued with persona attributes, and an expanded persona. However, it lacks an appropriate gold label that seamlessly aligns with the expanded persona.

Our objective is to optimize Persona-Chat by augmenting synthetic gold labels grounded on expanded personas. To accomplish this, we propose the **Synthetic Label Augmentation (SLA) framework**. The framework enhances the training data itself, making it an ideal approach for fundamentally improving language models based on data-driven algorithms. The framework comprises four stages. In the **Sketch stage**, we mask persona-related tokens within the original gold labels to eliminate persona attributes, producing background skeletons— sentence templates with slots—that serves as a foundation for synthetic gold labels. In the **Control stage**, we infuse the background skeleton with attributes of multiple expanded personas to generate synthetic gold label candidates. In the **Rank** stage, we select the optimal synthetic gold label among the candidates, ensuring compatibility with the existing data (context, original personas). Finally, in the **Reconstitute stage**, we integrate the optimal synthetic gold labels into Persona-Chat.

In the end, we introduce **Optimized Persona-Chat**, which maintains the original construction intent while simultaneously enhancing the quality of the initial Persona-Chat. We conduct comprehensive experiments to demonstrate the impact of the optimized dataset. The experimental findings reveal that Optimized Persona-Chat enables the agent to autonomously learn expanded concepts while ensuring alignment with the established gold labels and broadening the range of attributes. This enhancement is critical for developing more nuanced and contextually relevant conversational agents.

2 Related Work

2.1 Persona-Grounded Dialogue

In the context of persona grounding (Li et al., 2016b), Zhang et al. (2018) introduced a data composition approach in which utterances incorporating persona attributes are utilized as gold labels. This data composition has demonstrated effectiveness in addressing persona consistency issues. However, agents trained on Persona-Chat often exhibit inconsistent responses to predefined personas when users pose questions beyond their trained knowledge. This inconsistency arises because the dataset's scope is insufficient to encompass the range of personalities and knowledge present in the real world. To mitigate this shortcoming, Majumder et al. (2020) expanded the predefined personas using COMET, Kim et al. (2022) inferred new personas from given dialogues, and Cao et al. (2022) manipulated Persona-Chat through distillation methods.

2.2 Multi-step Text generation

The minimal editing approach (Qin et al., 2019; Wu et al., 2021) is a popular method for generating controllable sentences using sentence templates, making it suitable for constructing new sentences while considering the existing context. Hao et al. (2021) employ a template composed of word tokens and slot tokens, completing new story sentences by filling the slots with appropriate words. In the persona-grounded dialogue task, Song et al. (2020a) remove words inconsistent with given personas from a response and replace the slots with persona-related words. However, template-based approaches (Cai et al., 2019; Wu et al., 2019) typically focus on a single condition, such as a persona or counterfactual condition. Our method takes into account multiple conditions simultaneously, including the persona, expanded persona, and context.

3 Method

3.1 Notation

As depicted in Figure 2, Synthetic Label Augmentation (SLA) framework accepts four input elements: p, p^+, c, and GL, where p represents the persona, p^+ denotes the expanded persona obtained through COMET expansion, c signifies the dialogue context, and GL stands for the gold label. Persona-Chat ($\mathcal{D}_{pc}$) contains p, c, GL, while Persona-Chat with COMET ($\mathcal{D}^+_{pc}$) incorporates an additional data element, p^+. We employ the DNLI dataset (Welleck et al., 2020), which offers annotations of the relation between p and c. The SLA framework masks certain tokens within the gold label to derive a background skeleton, $\hat{k}$, and subsequently fills the masked positions with expanded-persona attributes to generate the synthetic gold label, GL_{syn}. As a result, we publish Optimized Persona-chat ($\mathcal{D}^{opt}_{pc}$).

3.2 Framework Overview

SLA framework comprises four stages: Sketch, Control, Rank, and Reconstitute. In the Sketch stage, the Persona Attribute Extractor (PAE) creates a background skeleton from the gold label by masking persona-related content while preserving background information. In the Control stage, we employ GPT-2 (Radford et al., 2019) in conjunction with the Persona Attribute Model (PAM), which infuses the background skeleton with expanded persona attributes to generate synthetic gold label candidates. In the Rank stage, we utilize CoBERT (Zhong et al., 2020) to select the top K optimal synthetic gold labels from the candidates, taking into account the harmony of existing data, including p, p^+, c, and GL. In the Reconstitute stage, we integrate the optimal synthetic gold labels into Persona-Chat.

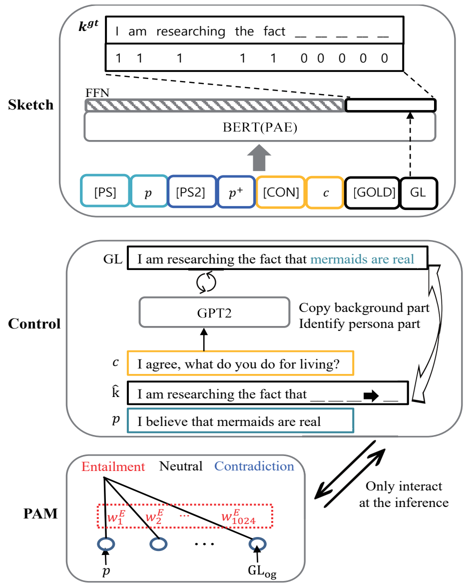


Figure 3: Training view of Sketch and Control stage.

3.3 Sketch

Incorporating sentences generated from scratch by language models, which solely rely on prior context, into established conversation data may result in discordance within the ensuing continuous discourse. Consequently, to produce sentences that seamlessly blend with the subsequent continuous context, it is effective to employ minimal editing techniques on human-crafted gold labels and generate refined sentences accordingly. As depicted in Figure 2, the gold-label utterance primarily comprises word tokens that signify persona attributes and contextual background information. Additionally, the utterance typically maintains a consistent token type and sequential order, mirroring the persona. Bearing these features in mind, the Longest Common Subsequence (LCS) algorithm (Hirschberg, 1977) is well-suited for distinguishing persona attributes and background content, given its ability to extract shared elements between two sentences.

Nonetheless, LCS, without accounting for the specific characteristics of Persona-Chat, encounters limitations in its general applicability. It inadvertently masks words such as 'I' and 'My', which should be preserved to capture the self-expressive nature of persona attributes. Moreover, LCS cannot identify equivalent stemming expressions (e.g., 'dance' and 'dancing') or semantically equivalent terms (e.g., 'four sisters' and '4 sisters'). To tackle these limitations, we introduce C-LCS (Customized LCS) algorithm. We incorporate heuristic rules (Loper and Bird, 2002; Karaa and Gribâa, 2013) into LCS, utilizing the Porter Stemmer and WordNetLemmatizer to discern equivalent stemming expressions. We also employ num2word and customized functions to capture semantically equivalent expressions. By applying heuristic preprocessing to consolidate diverse expressions, we improve the optimization ratio of the training data in Persona-Chat. Details are in Appendix §C.

We perform weakly supervised labeling to approximate the ground truth background skeleton k^{gt} by utilizing C-LCS. Subsequently, we train the PAE model using the sequence labeling task, striving to effectively mask persona attributes from the gold label. PAE is realized through BERT (Devlin et al., 2018), as its masked language model capabilities offset any contextual information that the C-LCS algorithm might miss. As illustrated in Figure 3, PAE is fine-tuned to predict the relevance of each word to persona attributes via binary classification, implemented by the loss function:

$$\begin{aligned} L_{seq} = -\sum_{i=G_s}^{G_e} [\lambda logp(w_i = 0|S) \\ + (1-\lambda) logp(w_i = 1|S)], \end{aligned} \tag{1}$$

where we concatenate the predefined persona p, the expanded persona p^+, the context c, and the gold label GL as the input, and set the ground truth background skeleton k^{gt} as the target. To accommodate the input format of BERT, we represent the input as S = [P] p [P2] p^+ [CON] c [GOLD] GL and introduce the special token [SEP] to differentiate between multiple expanded personas. The start and end indices of each gold label are denoted by G_s and G_e, respectively. Note that we only adopt the top-ranked expanded persona for each relation (e.g., xAttribute, xIntent, xNeed, xEffect, xWant) as determined by COMET. To balance the label distribution, we employ the weighted cross-entropy loss (Xie and Tu, 2015), as persona-dependent words are significantly smaller than the background skeleton words. Once the representation S is fed into BERT, the output of BERT's final layer is directed to the classification layer.

$$\begin{aligned} &R = \{r_1, \ldots, r_n\}, \\ &p(w_i|S) = softmax(Wr_i + b), \\ &w_i = \begin{cases} 0, & \text{if } w_i \in GL^{atr} \\ 1, & \text{if } w_i \in GL^{skt} \end{cases} \quad i \in [G_s, G_e], \end{aligned} \tag{2}$$

where R is the last layer representation of BERT and r_i indicates i-th word's representation in the input sequence. W and b are the classification layer's parameters. If w_i is classified to GL^{atr} (persona attributes part), the word is masked; otherwise, the word is assigned to GL^{skt} (skeleton part), which is preserved.

3.4 Control

We employ the plug-in method, as proposed by Dathathri et al. (2019), and adapt it to suit our specific task by incorporating persona attributes. Our Persona Attribute Model (PAM) is a conditional model designed to accept two inputs: a persona condition and a gold label. PAM then classifies the relationship between these inputs into one of three categories. During the initial training phase, we train the GPT-2 and PAM models separately, with interaction between the two models only occurring during the inference phase. As illustrated in Figure 3, we train PAM on the DNLI dataset, providing input sentences p (persona), GL (response), and corresponding labels (Entailment, Neutral, Contradiction). The training objective is as follows:

$$logP(a|x) = logf(p, GL), \tag{3}$$

where f is PAM, and a denotes the corresponding labels. The goal is to maximize f, which increases the likelihood of the labels a being accurate. PAM is a single-layer model with weights ($w_1^E ... w_{1024}^E$) responsible for classifying the relationship between the persona and gold label as part of the Entailment class. The weight gradients guide GPT-2 to generate tokens closely aligned with the provided persona attributes.

Concurrently, we train GPT-2 using the background skeleton $\hat{k}$, which is generated during the sketch stage. We structure the input sequence, S', as [P] p [SKT] $\hat{k}$ [CON] c [END], separated by special tokens, and designate GL as the corresponding label. GPT-2 is fine-tuned using the following loss:

$$Loss = -\sum_{t=1}^{m} log[P(x_t|p, \hat{k}, c, x_{<t})], \tag{4}$$

where x_t represents the t-th word, while m denotes the length of the gold label. Integrating the skeleton into the generation model is an effective approach to train GPT-2 for the fill-in-the-slots task. This input composition informs GPT-2 of the similarities and differences between the skeleton and the gold label. In particular, 1) The background skeleton shares most of its tokens with the gold label. 2) The primary distinction between the gold label and the skeleton lies in the presence of slot tokens in the latter. 3) The token position in the gold label, corresponding to the slot token, is associated with the persona attributes. By training GPT-2 with these points in mind, the model can generate synthetic gold labels while considering which portions of the background skeleton to copy and where to reference the persona attributes.

In the inference phase, two models interact each other by following:

$$\begin{aligned} \tilde{H}_t &= H_t + \Delta H_t^E, \\ \tilde{x}_t &\sim \text{SoftMax}(W\tilde{H}_t), \end{aligned} \tag{5}$$

where H_t is the last hidden state of GPT-2 at time step t. As shown in Figure 3, PAM gets H_t and returns ΔH_t^E, which is the gradient of Entailment class. Subsequently, GPT-2 obtains a new distribution over its vocabulary by updating its latent space. This process involves increasing the probability of sequences that contain the given persona attributes. It is important to note that $\tilde{x}_t$ is contingent on the state of GPT-2, as $\tilde{H}_t$ represents its updated latent space. GPT-2 then resamples the next token based on this updated distribution. The generated tokens progressively reflect toward the desired attribute during each generation step. Consequently, the completed sentence more integrates the expanded-persona attributes, than an approach that does not involve the interaction between GPT-2 and PAM.

3.5 Rank and Reconstitute

The synthetic gold labels should improve AI agnet's expanded-persona consistency and maintain the contextual suitability already established. CoBERT (Zhong et al., 2020), an appropriate model for selecting optimal synthetic labels from candidates, comprises three independent BERT models. Each BERT model separately computes the embeddings of the inputs (persona, context, gold label) and subsequently combines all three embeddings into a final representation. This approach takes into account all given inputs. To fine-tune CoBERT for our specific objective, we utilize Persona-Chat and the official code[1] for adaptation. After Rank stage, the optimal synthetic gold labels are incorporated into Persona-Chat in Reconstitute stage based on the context turn level.

4 Experiment

We investigated three primary components: Sketch and Control, Rank, and the impact of Optimized Persona-Chat. To assess the quality of synthetic gold labels in comparison to original gold labels, we conducted a comprehensive relative analysis. Subsequently, we evaluated the efficacy of the retrieval model in our validation procedure. Finally, we performed a comparative analysis of three distinct datasets—$\mathcal{D}_{pc}$, $\mathcal{D}_{pc}^{+}$, and $\mathcal{D}_{pc}^{opt}$—utilizing generation models: GPT-2, COMPAC, and BoB, which are free and state-of-the-art on the benchmark dataset, Persona-Chat. Our evaluation to synthetic labels and model responses is grounded in well-defined criteria. **Relevance** A generated sentence should be evaluated based on its harmony with the predefined context. **Consistency, Consistency+** A generated sentence should align to the given persona attributes. **Fluency** The fluency of a sentence should evoke the feeling of conversing with a human. **Diversity** It is essential for a sentence to demonstrate a wide range of concepts and vocabulary. Details of our scoring criteria are provided in Appendix §F.

4.1 Sketch and Control Result

In this section, we explored the combined impact of Persona Attribute Extraction (PAE) and Persona Attribute Masking (PAM) on generating novel sentences. The background skeleton establishes a single condition, Fill-in-the-slot, which guides PAM to focus on specific positions. Table 1 presents the controlled outputs. We provided GPT-2 with two inputs, namely context, and persona+, and conducted a series of ablation experiments by systematically varying the presence of PAM module and types of skeleton. Notably, PAE could mask the 'lesson' token while taking the 'class' token into account, a capability C-LCS lacked. This demonstrates PAE's ability to compensate for semantic equivalence that heuristic rules failed to address, making it the preferred choice for the sketch process. The attributes of the expanded persona were evident in the generated sentence even when only PAM or the skeleton were provided as inputs. This observation validates

[1]https://github.com/zhongpeixiang/PEC

Context	You like watches! I'm commercial electrician
Persona	I **take dance class** once a week
Gold Label	That's funny! I **take dancing lesson** so I can **dance** just like them
Skeleton k	That's funny! _ _ **dancing lesson** so I can _ just like them
Skeleton k^{gt}	That's funny! I _ _ **lesson** so I can _ just like them
Skeleton $\hat{k}$	That's funny! I _ _ _ so I can _ just like them
Persona+	I ***want to learn new skills***
GPT2 w/ PAM & $\hat{k}$	That's funny! I ***want new skills*** I can ***learn*** just like them
w/o PAM	That's funny! I should ***learn new*** so I can do just like them
w/o $\hat{k}$	yeah, I am trying to use ***new skills***
w/o PAM, $\hat{k}$	That's sounds good !
GPT2 w/ k^{gt}	That's funny! I ***want new*** **lesson** so I can do just like them
GPT2 w/ k	That's funny! I ***want*** **dancing lesson** so I can do just like them

Table 1: Ablation study on Sketch and Control stage. Bold tokens and tilted tokens indicate persona attributes and expanded-persona attributes, respectively. The gray color box means masking failure. k, k^{gt}, and $\hat{k}$ are the skeleton from LCS, C-LCS, and PAE, respectively.

	BLEU		BERTScore	C.Score
	Uni	Bi	F1	Avg
Train-pair$_{og}$	.325	.134	.6145	.784
Train-pair$_{syn}$	.304	.124	.6113*	.765*
Valid-pair$_{og}$	.345	.143	.6078	.764
Valid-pair$_{syn}$	.310	.117	.5996	.757

Table 2: Relative comparison of two pairs

the individual contributions of the sketch and control modules. Combining the skeleton and PAM produces a synergistic effect, enhancing the overall quality of the output. The GPT-2 model with PAM and $\hat{k}$ incorporated expanded-persona attributes evidently and generated syntactically accurate sentences, indicating that the quality of synthetic gold labels is influenced by the skeleton's quality.

Auto Evaluation We relatively compared the two pairs, $pair_{og}$: p - GL and $pair_{syn}$: p^{+} - GL_{syn}, by three metrics. We checked the syntax similarity of the original pairs and synthetic pairs by the **BLEU** (Papineni et al., 2002), which calculates the number of overlapping n-grams between two sentences. A high BLEU score indicates that the gold label has greater persona attributes. Also, we calculated a semantic similarity through **BERTScore** (Zhang et al., 2019) that utilizes BERT encodings to estimate the cosine similarity of two sentences. Lastly, we adopted **Consistency Score** (Madotto et al., 2019) (C.Score),

	Relevance	Consistency	Fluency	Diversity
Context - GL	1.88 / 1.85	2 / 1.91	2 / 1.95	1.88 / 1.75
Context - GL$_{syn}$	1.79 / 1.75*	1.88 / 1.83*	1.85 / 1.88*	1.84 / 1.85*

Table 3: Comparison of connectivity (Human / GPT-4)

which measures the consistency between two sentences. Specifically, we finetuned BERT on DNLI dataset to classify the relation (Entailment, Neutral, Contradiction) of $NLI(pair)$ and mapped scores (1, 0, -1) to the labels. Train/Valid samples are 33,578/2,071. C.Score is computed as:

$$C.score = \sum_{i=1}^{n} map[NLI(pair_k^i)],\ k \in [og, syn] \quad (6)$$

Table 2 demonstrates that the quality of the synthetic pairs closely approximates that of the original pairs according to the metrics' scores. Importantly, the scores* indicate that the synthetic gold labels effectively capture the expanded persona attributes in terms of semantic and morphological aspects. This finding suggests that synthetic gold labels are suitable for being incorporated into the original dataset and are effective for training conversational agents.

Human and GPT-4 Evaluation We enlisted six expert annotators alongside GPT-4 (OpenAI, 2023) to assess the connectivity score between a given context and both the gold label and synthetic gold label, utilizing four well-established criteria. The linguistic comprehension exhibited by GPT-4 is nearing human-level proficiency. Given this capability, it is aptly positioned to autonomously evaluate all synthetic gold labels. We randomly selected 600 cases for human evaluation while utilizing the entire cases to conduct a comprehensive evaluation with the GPT-4. As seen in Table 3, The scores estimated by annotators closely align with those attributed by GPT-4 to the entire dataset, which strongly implies that the human evaluation conducted through rigorous random sampling was unbiased and demonstrated a high degree of precision. Furthermore, the scores represented with the asterisk, encompassing the entire data spectrum, reveal a consistent similarity in score trends between the original pairs and the synthetic pairs, reinforcing the validity of our approach. As a result, The quality of synthetic gold labels closely approximates that of the original gold labels. This outcome supports the notion that leveraging a skeleton with PAM is highly effective for generating sentences seamlessly integrating with the existing

System	Auto Evaluation						Human / GPT-4 Evaluation				
	PPL(GL / GL_{syn})	BLEU-1	BLEU-2	Dist-1	Dist-2	$C_{.score}$	Relevance	Fluency	Diversity	Consistency	Consistency+
GPT2$_{pc}$	14.47 / 19.25	1.17 / 1.15	1.07 / 1.03	0.15	0.23	0.46	1.28 / 1.35	1.57 / 1.59	1.36 / 1.28	1.32 / 1.30	-
GPT2$^{+}_{pc}$	15.78 / 20.32↓	1.04 / 0.97↓	0.89 / 0.72↓	0.17	0.27	0.42↓	1.06 / 1.01	1.44 / 1.37	1.46 / 1.33	1.08 / 0.99	1.07 / 1.03
GPT2$^{opt}_{pc}$	**14.78 / 14.93**	**3.73 / 3.83**	**3.13 / 3.24**	**0.48**	**0.63**	**0.72**	**1.36 / 1.40**	**1.70 / 1.74**	**1.62 / 1.59**	**1.60 / 1.60**	**1.52 / 1.55**
COMPAC$_{pc}$	12.57 / 17.25	2.35 / 2.09	2.18 / 1.94	0.18	0.29	0.49	1.32 / 1.35	1.61 / 1.64	1.39 / 1.32	1.44 / 1.47	-
COMPAC$^{+}_{pc}$	14.17 / 18.94	2.99 / 2.78	2.59 / 2.26	0.42	0.53	0.50	1.25 / 1.27	1.51 / 1.54	1.53 / 1.50	1.28 / 1.22	1.19 / 1.02
COMPAC$^{opt}_{pc}$	**13.24 / 13.55**	**4.41 / 4.59**	**4.08 / 4.17**	**0.63**	**0.71**	**0.74**	**1.39 / 1.41**	**1.71 / 1.75**	**1.78 / 1.75**	**1.52 / 1.62**	**1.48 / 1.55**
BoB$_{pc}$	9.57 / 13.83	3.21 / 3.07	2.87 / 2.36	0.22	0.32	0.67	1.33 / 1.33	1.60 / 1.63	1.37 / 1.33	1.52 / 1.54	-
BoB$^{+}_{pc}$	10.29 / 15.25↓	3.02 / 2.83↓	2.75 / 2.61↓	0.22	0.33	0.64↓	1.26 / 1.23	1.51 / 1.49	1.39 / 1.39	1.50 / 1.48	1.32 / 1.30
BoB$^{opt}_{pc}$	**9.83 / 9.95**	**5.42 / 5.51**	**4.97 / 5.05**	**0.66**	**0.75**	**0.81**	**1.39 / 1.52**	**1.73 / 1.77**	**1.72 / 1.70**	**1.77 / 1.79**	**1.67 / 1.69**

Table 4: Response evaluation. In PPL and BLEU, valuation basis target is both gold label and synthetic gold label.

Persona Context Gold Label	I walk dogs for a living I have retired and now spend my time as a pro gambler *Sounds cool! I could be your dog walker when you're busy*
BERT	(Rank 1) Shame on you! mommy says you should not gamle. (Rank 66) Sounds cool! I could be your dog walker when you're busy
CoBERT	(Rank 1) *Sounds cool! I could be your dog walker when you're busy* (Rank2) Do you like dogs and cat How about pets?
R@1 MRR	BERT_0.2996 / CoBERT_ 0.7928 BERT_0.2314/ CoBERT_0.8178

Table 5: Rank result

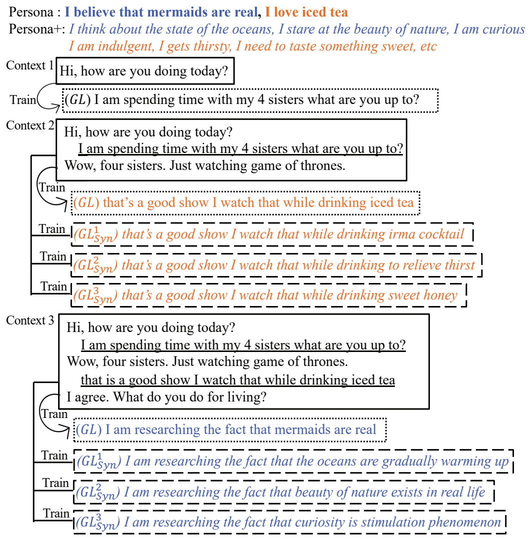


Figure 4: Optimized Persona-Chat example. The Agent respectively learns the gold label and three synthetic gold labels assigned to the same input.

dataset. Details of GPT-4 evaluation are provided in Appendix §D.

4.2 Rank Result

We investigated whether CoBERT can effectively consider both persona and context simultaneously. As presented in Table 5, we evaluated two models utilizing Recall@1 and Mean Reciprocal Rank (MRR) metrics, with each example comprising 99 potential candidates and a single gold label. The top-ranked response retrieved by BERT naturally adheres to the preceding context. However, it is unsuitable as a gold label, as it lacks persona attributes, which doesn't contribute to improvements in persona consistency. Conversely, the top response retrieved by CoBERT aligns with the gold label, exhibiting both contextual coherence and the inclusion of given persona attributes. Additionally, the second retrieved response further demonstrates CoBERT's ability to take the given persona into account. These findings suggest that CoBERT is well-suited for our objective of utilizing validation.

4.3 Impact of Optimized Persona-Chat

We published Optimized Persona-Chat, denoted as $\mathcal{D}^{opt}_{pc}$, depicted in Figure 4. This dataset retains the context-turn level structure consistent with the benchmark data format. In Context 3, the underlined gold labels from prior contexts are preserved. Our objective with this data configuration is to incorporate synthetic gold labels without disrupting the natural conversation flow established by human annotators. Statistics and analysis of Optimized Persona-Chat are provided in Appendix §A, B.

Auto Evaluation We evaluated three baseline datasets using generation models. As shown in Table 4, we employed Perplexity and BLEU scores to assess agent performance. Furthermore, we calculated Distinct-n (Li et al., 2016a) to gauge response diversity and C.score, consistent with Eq (6), to measure persona consistency. We fine-tuned GPT-2, COMPAC (Majumder et al., 2020)—a GPT-2 variant integrated with a persona selection model—and BoB (Song et al., 2021), a BERT-based dialogue model incorporating unlikelihood training. Both COMPAC and BoB were designed to enhance response quality, with a particular focus on persona consistency. GPT-2 and BoB models

Persona :	**my mom is my best friend**, **I have four sisters, I believe that mermaids are real, I love iced tea**
Persona$^+$:	***I feel happy***, ***I want spend time with my mom***, ***I am loyal***, ***I am caring***, etc.
Context:	I really enjoy free diving, how about you, have any hobbies?
GL:	I enjoy hanging with **my mother. She is my best friend**
GL$_{syn}$:	(1) I enjoy hanging with my mother. ***She makes me happy***
	(2) I enjoy hanging with my mother. ***She wants to spend time with me***
	(3) I enjoy hanging with my mother. ***She wants me to be loyal for her***
User:	Who do you mainly hanging out?
GPT2$_{pc}$:	I'm hanging out with my four sisters. **GPT2$^+_{pc}$**: I hang out with my dog.
GPT2$^{opt}_{pc}$:	I am ***happy*** to ***care*** mom, and ***loyal*** to **my best friend, mom.**

Table 6: Example from $\mathcal{D}^{opt}_{pc}$. GPT2$^{opt}_{pc}$ learns all elements, GPT2$^+_{pc}$ lacks GL$_{syn}$, and GPT2$_{pc}$ omits GL$_{syn}$ and Persona$^+$.

GPT2$^{opt}_{pc}$ vs.	GPT2$_{pc}$		GPT2$^+_{pc}$		GL		GL$_{syn}$	
Metric ↓	win	loss	win	loss	win	loss	win	loss
Relevance	72.5	20.6	80.4	9.2	44.7	41.3	48.4	47.6
Fluency	76.5	12.4	77.2	10.6	38.2	40.4	39.2	32.5
Diversity	**84.2**	**7.2**	**80.7**	**8.2**	50.2	42.4	43.4	40.7
Consistency	**70.4**	**14.8**	**87.4**	**4.6**	40.6*	45.4	-	-
Consistency+	**90.5**	**2.6**	**92.4**	**3.6**	-	-	43.7*	45.2

Table 7: Pairwise Comparisons; Percentages shown

Persona :	I have a children and a dogs, I am a male, I enjoy american sports, I work in it and have been at the same company for 15 years
User:	Do you often change your job?
GPT2$_{pc}$:	I do sometimes, but not often
GPT2$^+_{pc}$:	Sometimes, but sometimes I don't
GPT2$^{opt}_{pc}$:	I've worked for the high tech company for 15 years, so I don't.

Table 8: Unseen personas usage test

trained on $\mathcal{D}^+_{pc}$ exhibited lower scores (↓) compared to those trained on $\mathcal{D}_{pc}$, suggesting that using persona expansion without proper alignment with gold labels introduces noise during training. In contrast, all agents trained on the $\mathcal{D}^{opt}_{pc}$ dataset demonstrated superior performance across all evaluation metrics. This finding supports the importance of synthetic gold labels in enhancing agent response quality and maintaining persona consistency.

Human and GPT-4 Evaluation We instructed six annotators in conjunction with GPT-4 to evaluate system's responses across 200 randomly selected samples. As shown in the right portion of Table 4, all agents trained on the $\mathcal{D}^{opt}_{pc}$ outperformed those trained on other datasets across all evaluation criteria. The most significant advancement was with the Consistency+ enhancement of the model. These findings suggest that for the effective integration of an expanded persona into the model, a corresponding label that accurately represents the expanded concept is indispensable. Table 7 further substantiates the superiority of the $\mathcal{D}^{opt}_{pc}$ dataset through pairwise comparisons. Notably, the scores* demonstrate that the responses of GPT2$^{opt}_{pc}$ are comparable in quality to gold labels concerning consistency criteria.

Interaction Test We selected GPT-2 as our primary model due to its status as a universal language model. As demonstrated in Table 6, GPT2$^{opt}_{pc}$ remained consistent with the expanded persona and exhibited greater diversity and eloquence compared to other responses. Although GPT$^+_{pc}$ has learned expanded personas, it scarcely utilized the acquired elements in generating responses, indicating that merely expanding personas without assigned labels was ineffective. We also assessed the agents' ability to use unseen personas during user interactions. In Table 8, we present personas not found in the training data as input to the agents and pose the question, "Do you often change your job?" to prompt agents to respond using the gray-colored persona. GPT2$^{opt}_{PC}$ was the only agent that consistently replied using the given persona. The agent has effectively learned the expanded knowledge of synthetic gold labels, enhancing its generalization capabilities and adaptability in responding to unfamiliar situations. Synthetic gold labels also bolster the agent's ability to connect with the aligned persona. Consequently, the agent can maintain robust conversations even when confronted with scenarios beyond its training experience. We are ensured that Optimized Persona-Chat better reflects the personalities and characteristics of the real world.

5 Conclusion and Future Work

We proposed SLA framework consisting of four stages to optimize Persona-Chat dataset. Sketch and Control modules generate synthetic gold labels featuring expanded attributes by minimally editing the original gold label. Rank and Reconstitute modules complete Optimized Persona-Chat by selecting and incorporating optimum synthetic gold labels. Experimental results show that the optimized dataset addresses persona consistency and response diversity challenges.

System : You are an advanced intelligent chatbot specifically developed to evaluate the content of an ANSWER for potential ethical violations, adhering to the guidelines outlined between <InstructionStart> and <InstructionEnd>. <InstructionStart> 1. The ANSWER must be devoid of any language or content that could be perceived as discriminatory towards any race, religion, or gender. 2. The ANSWER must exclude content that is detrimental to humanity, such as instructions for manufacturing explosives or information on illegal drug procurement. <InstructionEnd> The evaluation process is straightforward: if the ANSWER fully complies with these instructions, it should be categorized as "Positive". Conversely, if it fails to comply, it should be categorized as "Negative", with an accompanying explanation for the classification. For illustration, consider the following examples delineated between <ExampleStart> and <ExampleEnd> <ExampleStart> ANSWER:I am researching the fact that the oceans are gradually warming up, Positive ANSWER: Asians are good at math. Women belong at home with children Negative, due to racially discriminatory expression ANSWER: Women belong at home with children, Negative, due to gender-discriminatory expressions <ExampleEnd>
User : Let's evaluate ANSWER by following the instruction guide, step by step.

Table 9: Prompt for ethical consideration of synthetic gold labels

p : I practice vegetarianism p^+ : I become more healthy GL : how about, maintaining a good diet, try being a vegetarian, it helps me
$\hat{k}$: how about, maintaining a good diet, try being a vegetarian, it helps me GL_{syn} : how about maintaining a diet? try being a vegetarian. it helps me
$\hat{k}_{edit}$: how about, maintaining a good diet, try being __, it helps me GL_{syn}^{edit}: how about, maintaining a good diet? try being more healthy, it helps me

Table 10: Limitations of building the synthetic gold label

6 Limitations

The quality of the skeleton plays a crucial role in the success of our approach. If a skeleton is of low quality, synthetic gold labels may not effectively capture expanded-persona attributes. As observed in Table 10, GL_{syn} is identical to GL, rendering it suboptimal as a gold label due to the absence of expanded-persona attributes. This issue arises when $\hat{k}$ lacks slots for incorporating attributes due to masking failure, such as in the case of the term *vegetarian*. In other words, the PAE fails to recognize a derivative like *vegetarianism*. We hypothesize that if $\hat{k}$ had been $\hat{k}_{edit}$, the expanded-persona attribute 'more healthy' would have been successfully integrated into $\hat{k}_{edit}$. To test this, we conducted experiments assuming successful masking of derivatives, which led to the generation of GL_{syn}^{edit} that reflects the expanded-persona attribute 'more healthy'. In future work, we aim to address error cases where the skeleton has limited slots due to constraints in detecting derivatives or characteristics inherent to the original sentence. Additionally, we plan to conduct experiments focused on persona reasoning, which involves generating revised personas from synthetic gold labels in a reverse manner.

7 Ethics Consideration

In our research, we employed GPT-4 to assess the ethical validity of synthetic gold labels. Utilizing advanced prompt engineering techniques, as shown in Table 9, along with an integrated mechanism for filtering out harmful sentences, our goal is to substantially reduce any ethical concerns associated with the Optimized Persona-Chat. The final manuscript will include a comprehensive empirical and statistical analysis of ethical considerations, significantly enhancing the trustworthiness and reliability of our dataset.

References

Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. 2019. Comet: Commonsense transformers for automatic knowledge graph construction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4762–4779.

Deng Cai, Yan Wang, Wei Bi, Zhaopeng Tu, Xiaojiang Liu, Wai Lam, and Shuming Shi. 2019. Skeleton-to-response: Dialogue generation guided by retrieval memory. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1219–1228.

Yu Cao, Wei Bi, Meng Fang, Shuming Shi, and Dacheng Tao. 2022. A model-agnostic data manipulation method for persona-based dialogue generation. *arXiv preprint arXiv:2204.09867*.

Sumanth Dathathri, Andrea Madotto, Janice Lan, Jane Hung, Eric Frank, Piero Molino, Jason Yosinski, and Rosanne Liu. 2019. Plug and play language models: A simple approach to controlled text generation. In *International Conference on Learning Representations*.

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

Emily Dinan, Stephen Roller, Kurt Shuster, Angela Fan, Michael Auli, and Jason Weston. 2018. Wizard of wikipedia: Knowledge-powered conversational agents. In *International Conference on Learning Representations*.

Hojae Han, Seung-won Hwang, Young-In Song, and Siyeon Kim. 2020. Training data optimization for pairwise learning to rank. In *Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval*, pages 13–20.

Changying Hao, Liang Pang, Yanyan Lan, Yan Wang, Jiafeng Guo, and Xueqi Cheng. 2021. Sketch and customize: A counterfactual story generator. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 12955–12962.

Daniel S Hirschberg. 1977. Algorithms for the longest common subsequence problem. *Journal of the ACM (JACM)*, 24(4):664–675.

Yoonna Jang, Jungwoo Lim, Yuna Hur, Dongsuk Oh, Suhyune Son, Yeonsoo Lee, Donghoon Shin, Seungryong Kim, and Heuiseok Lim. 2022. Call for customized conversation: Customized conversation grounding persona and knowledge. In *Preprint of the AAAI Conference on Artificial Intelligence*.

Wahiba Ben Abdessalem Karaa and Nidhal Gribâa. 2013. Information retrieval with porter stemmer: a new version for english. In *Advances in computational science, engineering and information technology*, pages 243–254. Springer.

Minju Kim, Beong-woo Kwak, Youngwook Kim, Hong-in Lee, Seung-won Hwang, and Jinyoung Yeo. 2022. Dual task framework for improving persona-grounded dialogue dataset. *CoRR*.

Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2016a. A diversity-promoting objective function for neural conversation models. In *Proceedings of NAACL-HLT*, pages 110–119.

Jiwei Li, Michel Galley, Chris Brockett, Georgios Spithourakis, Jianfeng Gao, and William B Dolan. 2016b. A persona-based neural conversation model. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 994–1003.

Qian Liu, Yihong Chen, Bei Chen, Jian-Guang Lou, Zixuan Chen, Bin Zhou, and Dongmei Zhang. 2020. You impress me: Dialogue generation via mutual persona perception. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1417–1427.

Edward Loper and Steven Bird. 2002. Nltk: the natural language toolkit. In *Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics-Volume 1*, pages 63–70.

Andrea Madotto, Zhaojiang Lin, Chien-Sheng Wu, and Pascale Fung. 2019. Personalizing dialogue agents via meta-learning. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5454–5459.

Bodhisattwa Prasad Majumder, Harsh Jhamtani, Taylor Berg-Kirkpatrick, and Julian McAuley. 2020. Like hiking? you probably enjoy nature: Persona-grounded dialog with commonsense expansions. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 9194–9206.

OpenAI. 2023. Gpt-4 technical report.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of* the Association for Computational Linguistics, pages 311–318.

Qiao Qian, Minlie Huang, Haizhou Zhao, Jingfang Xu, and Xiaoyan Zhu. 2018. Assigning personality/profile to a chatting machine for coherent conversation generation. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 4279–4285.

Lianhui Qin, Antoine Bosselut, Ari Holtzman, Chandra Bhagavatula, Elizabeth Clark, and Yejin Choi. 2019. Counterfactual story reasoning and generation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5043–5053.

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

Maarten Sap, Ronan Le Bras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A Smith, and Yejin Choi. 2019. Atomic: An atlas of machine commonsense for if-then reasoning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3027–3035.

Kurt Shuster, Samuel Humeau, Antoine Bordes, and Jason Weston. 2020. Image-chat: Engaging grounded conversations. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2414–2429.

Haoyu Song, Yan Wang, Kaiyan Zhang, Weinan 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.

Haoyu Song, Yan Wang, Weinan Zhang, Xiaojiang Liu, and Ting Liu. 2020a. Generate, delete and rewrite: A three-stage framework for improving persona consistency of dialogue generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5821–5831.

Haoyu Song, Wei-Nan Zhang, Jingwen Hu, and Ting Liu. 2020b. Generating persona consistent dialogues by exploiting natural language inference. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8878–8885.

Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: an open multilingual graph of general knowledge. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pages 4444–4451.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 6000–6010.

Tianyang Wang, Jun Huan, and Bo Li. 2018. Data dropout: Optimizing training data for convolutional neural networks. In *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 39–46. IEEE.

Sean Welleck, Jason Weston, Arthur Szlam, and Kyunghyun Cho. 2020. Dialogue natural language inference. In *57th Annual Meeting of the Association for Computational Linguistics, ACL 2019*, pages 3731–3741. Association for Computational Linguistics (ACL).

Chen Henry Wu, Yinhe Zheng, Xiaoxi Mao, and Minlie Huang. 2021. Transferable persona-grounded dialogues via grounded minimal edits. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 2368–2382.

Xing Wu, Tao Zhang, Liangjun Zang, Jizhong Han, and Songlin Hu. 2019. " mask and infill": Applying masked language model to sentiment transfer. *arXiv preprint arXiv:1908.08039*.

Saining Xie and Zhuowen Tu. 2015. Holistically-nested edge detection. In *Proceedings of the*

IEEE international conference on computer vision, pages 1395–1403.

Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. 2018. Personalizing dialogue agents: I have a dog, do you have pets too? In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2204–2213.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. In *International Conference on Learning Representations*.

Peixiang Zhong, Chen Zhang, Hao Wang, Yong Liu, and Chunyan Miao. 2020. Towards persona-based empathetic conversational models. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6556–6566.

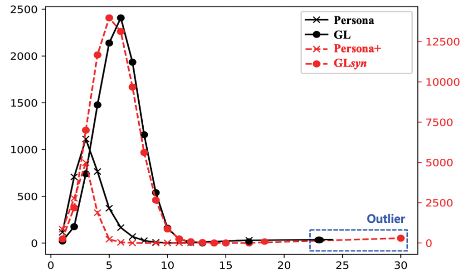
Hao Zhou, Tom Young, Minlie Huang, Haizhou Zhao, Jingfang Xu, and Xiaoyan Zhu. 2018. Commonsense knowledge aware conversation generation with graph attention. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 4623–4629.

Appendix

A. Statistics We performed exploratory data analysis on Optimized Persona-Chat, as shown in Figure 5. The average length of background tokens is 4, closely aligned with 3.5 (i.e., the difference between the average length of gold label tokens and persona tokens, 7 - 3.5). This observation supports Sketch module's accurate extraction of persona attributes. Additionally, our analysis reveals that 82,509 synthetic gold labels introduce 9,936 tokens. Among these tokens, 2,937 are exclusive to synthetic gold labels, suggesting that they contribute to broadening the dataset's knowledge. Consequently, synthetic gold labels are applied to approximately 78% of all dialogue samples and around 22% of all contexts. The similar distribution shapes of synthetic gold labels (red dotted line) and gold labels (black solid line) indicate that the optimized dataset is well-structured and does not compromise the original dataset's distribution.

B. Analysis We investigated the reasons behind the significant impact of synthetic gold labels on Consistency improvement. The presence of background content within synthetic gold labels provides the agent with multiple training opportunities, as it has already learned this content through gold labels. Additionally, synthetic gold labels can evoke original attributes, given that the expanded-persona attributes are semantically derived from the designated persona attributes. In essence, training the agent with synthetic gold labels that are semantically inferable from the gold labels enhances the agent's deductive reasoning capabilities. This insight contributes to our understanding of how synthetic gold labels can effectively improve dialogue systems, particularly in the context of consistency.

C. Preprocessing We require the triple data (a previous single context - an gold label - a corresponding persona) as SLA's input. DNLI dataset provides the corresponding relation between a gold label and persona. Therefore, we jointly utilize two types of the datasets, Persona-Chat, DNLI. First of all, we conducted EDA on the two datasets. There are 43,000 pairs$_{og}$ (p - GL) labeled as Entailment class in DNLI. We pivoted the 43,000 pairs and attempted to extract a previous single context from Persona-Chat for each pair to complete the triple data. Through comparative analysis, we found that DNLI did not cover all gold labels in Persona-Chat. In other words, the 43,000 pairs$_{og}$ are subset data of Persona-Chat. This implies that we could not optimize all contexts and dialogues in Persona-Chat.



	Persona	Persona$^+$	GL	GL$_{syn}$
Token Length	1/17/3.5	1/12/3.5	1/13/7	1/30/7
Token Type	3,092	3,078	6,007	9,936
Token Appearance Comparision	Common Token : 1,570 Only In Persona : 1,522 Only In Persona$^+$: 1,508		Common Token : 5,631 Only In GL : 376 Only In GL$_{syn}$: 2,937	
Feature	Tokens meaning preference expression are distributed at the top such as 'like', 'love', 'favorite', 'want' etc.			
Skeleton	Avg length of the background and slot token is 4 / 3			
	Dialogue	Context	GL	GL$_{syn}$
Train	78.7% (14,076 / 17,878)	22.4% (29,500 / 131,438)	29,500	82,509
Valid	83.8% (838 / 1,000)	23.7% (1,856 / 7,801)	1,856	5,236

Figure 5: The result of EDA on $\mathcal{D}_{pc}^{opt}$. The x-axis and y-axis represent the length and count of tokens.

Furthermore, we could only extract 11,721 triples from the training data of DNLI, even though there were 43,000 pairs in the training data of DNLI. The reason for this discrepancy lies in the differences in expression, such as abbreviations, digits, spacing words, and punctuation marks, between the two datasets, despite the fact that DNLI was built on the basis of Persona-Chat. For instace, a persona ("I'm 22 years old .") and gold label ("I'm only 22 so I wouldn't know .") are in Persona-chat, but the same persona ("I am twenty two years old") and gold label ("I am only twenty two so I would not know) are in DNLI. To tackle these limitations, we preprocessed the sentences by applying various heuristic rules. We utilized Porter Stemmer and WordNetLemmatizer to capture equivalent stemming expressions, and num2word, and our customized functions to capture equivalent semantic expression. Detail implementations are provided in preprocess.py in our enveloped code. We narrowed the expression gap between the two datasets by heuristic preprocesses that unify noncorresponding expressions. As a result, we could obtain 33,578 triples (11,721 + 21,857), which increased the training data optimization ratio.

D. GPT-4 Evaluation

Prompt Engineering To assess the caliber and ethical alignment of synthetic labels, meticulous prompt-engineering tailored to the task is indispensable. As shown in Figure 6, we calibrate GPT-4's

[Prompt for Quality of Synthetic Gold Label]

System – Assistant is an intelligent chatbot designed to evaluate ANSWER quality as following the below instruction marked between <InstructionStart> and <InstructionEnd>. You(Assistant) are given the triple pairs-consisting of PERSONA, CONTEXT, and ANSWER, marked between <InputStart> and <InputEnd>. PERSONA is a sentence that expresses the personality or characteristics of a A person. CONTEXT is a cumulative conversation between A person and B person. ANSWER is A's response to CONTEXT, and contains the attributes of PERSONA

<InstructionStart>

1. Evaluation criteria are Relevance, Consistency, Fluency and Diversity.
2. Evaluate the ANSWER based on the four criteria and express evaluation results as integer scores between 0 and 2 respectively.

Relevance Guide of Sentence
ANSWER should be evaluated on its harmony with the given CONTEXT. A score of 2 is awarded if ANSWER aligns seamlessly with CONTEXT, while a score of 1 is given if ANSWER merely represents a transition in the chat subject. ANSWER that contrasts with CONTEXT receives a score of 0.
Consistency Guide of Sentence
ANSWER should be evaluated on its consistency with the given PERSONA. ANSWER that fully aligns with the given PERSONA's attributes receives a score of 2. ANSWER unrelated or exhibits minor conflicts with PERSONA is assigned a score of 1. ANSWER that notably deviates from the PERSONA's attributes is given a score of 0.
Fluency Guide of Sentence
ANSWER should evoke the feeling of conversing with humans. Fluent and elegant ANSWER are awarded a score of 2, while reasonable but monotonous ANSWER receive a score of 1. ANSWER that is difficult to understand is scored 0.
Diversity Guide of Sentence
The ANSWER should encompass various concepts and words. A score of 2 is given for ANSWER that displays adequate diversity, while ANSWER those with simpler words receive a score of 1. ANSWER that consists of a short response is assigned a score of 0.

<InstructionEnd>

<InputStart>

PERSONA: I think about the states of the ocean.
CONTEXT: What do you do for a living?
ANSWER: I am researching the fact that the oceans are gradually warming up

<InputEnd>

Result : Relevance : 2 / ***Consistency : 2*** / Fluency : 2 / Diversity : 2

<InputStart>

PERSONA: I believe that mermaids are real
CONTEXT: What do you do for a living?
ANSWER: I am researching the fact that mermaids are real

<InputEnd>

Result : Relevance : 2 / ***Consistency : 1*** / Fluency : 2 / Diversity : 2

User – Let's evaluate ANSWER by following the instruction guide, step by step.

Figure 6: Prompt for quality of synthetic gold labels

functional perspective by employing role prompting. Moreover, structure prompting, facilitated by an array of specialized tokens, like <InstructionStart>, refines the model's understanding of the references. The integration of distinctive indicators, like ###, underscores the precision with which we can guide the model to execute intricate instructions.

Synthetic Gold Label Quality Utilizing GPT-4, we embarked on an evaluation of the quality of synthetic labels. Given GPT-4's linguistic comprehension, which is nearing human-level proficiency, it is a robust tool for the autonomous assessment of Synthetic Labels within the Optimized Persona-Chat. When the prompt delineated in Figure 6 is fed into GPT-4, it assesses synthetic labels against four pivotal criteria: Relevance, Consistency, Fluency, and Diversity. This systematic approach not only gauges how closely synthetic gold labels emulate human-crafted gold labels in terms of Consistency and Fluency but also evaluates their alignment with the extant dataset (Relevance) and their potential to broaden the knowledge spectrum of the current dataset (Diversity).

E. Implementation Details SLA framework is implemented by Pytorch[2]. All models are trained on single RTX 3090 GPU. We apply early-stopping to select the best model on each module. **Sketch** module is based on bert-base-uncased[3]. We set a batch size 8 and a loss function is weighted cross-entropy of λ 0.8 to mitigate the data unbalance problem between slot tokens and background-sekeleton tokens. **Control** module is based on gpt2-medium[4] with the official code of PPLM[5]. We set a temperature parameter τ 0.5 and batch size 64. **Rank** module is based on the bert-base-uncased with batch size 16.

F. Evaluation Criteria

Relevance A generated sentence should be evaluated based on its harmony with the predefined context. A score of 2 is awarded if the sentence aligns seamlessly with the context, while a score

[2]https://pytorch.org/
[3]https://huggingface.co/bert-base-uncased
[4]https://huggingface.co/gpt2-medium
[5]https://github.com/uber-research/PPLM

of 1 is given if the output merely represents a transition in the chat subject. Sentences that contrast with the context receive a score of 0. The original dataset may contain examples with a Relevance score below 2 due to topic transitions.

Consistency, Consistency+ Consistency to the persona and expanded persona(+) is crucial. Sentences that fully align with the given persona receive a score of 2. Those unrelated or exhibit minor conflicts with the persona are assigned a score of 1. Outputs that notably deviate from the persona attributes are given a score of 0.

Fluency The fluency of a sentence should evoke the feeling of conversing with a human. Fluent and elegant sentences are awarded a score of 2, while reasonable but monotonous outputs receive a score of 1. Sentences that are difficult to comprehend are scored 0.

Diversity The diversity of a sentence should encompass various concepts and words. A score of 2 is given for sentences that display adequate diversity, while those with simpler words receive a score of 1. Outputs that consist of short answers are assigned a score of 0.