

Partner Personas Generation for Dialogue Response Generation

Hongyuan Lu, Wai Lam, Hong Cheng, Helen M. Meng
The Chinese University of Hong Kong
{hylu, wlam, hcheng, hmmeng}@se.cuhk.edu.hk

Abstract

Incorporating personas information allows diverse and engaging responses in dialogue response generation. Unfortunately, prior works have primarily focused on self personas and have overlooked the value of partner personas. Moreover, in practical applications, the availability of the gold partner personas is often not the case. This paper attempts to tackle these issues by offering a novel framework that leverages automatic partner personas generation to enhance the succeeding dialogue response generation. Our framework employs reinforcement learning with a dedicatedly designed critic network for reward judgement. Experimental results from automatic and human evaluations indicate that our framework is capable of generating relevant, interesting, coherent and informative partner personas, even compared to the ground truth partner personas. This enhances the succeeding dialogue response generation, which surpasses our competitive baselines that condition on the ground truth partner personas.

1 Introduction

Building informative and engaging dialogue agents (Zhang et al., 2020; Roller et al., 2021) is a popular research direction within the area of natural language processing. For the sake of engagement, diverse and consistent responses (Song et al., 2020, 2021) are important factors, and personas information (Zhang et al., 2018) gives rise to both. There are two types of personas, namely self persona and partner persona. The former refers to a self profile consisting of several sentences representing the dialogue agents. Such a persona allows producing consistent responses rather than solely relying on the personas that are randomly learned and embedded in the model parameters (Kim et al., 2020). The latter refers to a profile that represents the users. Leveraging such partner personas has been empirically shown to be helpful for dialogue response selection (Gu et al., 2021).

Unfortunately, the existence of partner personas suffers from the cold start (Schein et al., 2002; Zhang et al., 2014; Li et al., 2021) at the beginning of the conversation. Most of the works, if not all, (Li et al., 2016b; Mazaré et al., 2018; Song et al., 2019; Gu et al., 2019; Zhao et al., 2019; Madotto et al., 2019; Liu et al., 2020; Majumder et al., 2020; Wu et al., 2020a; Song et al., 2020) have been either overlooking partner personas or simply focusing on the impractical situation where partner personas guarantee to exist. In contrast, our work does not suffer from the practical issue when partner personas are missing during inference, and our proposed framework surpasses the baseline that conditions on the ground truth partner personas.

To our knowledge, this is the first attempt to formulate partner personas generation for improved performance on the downstream dialogue response generation. Our work is motivated by the underlying hypothesis that partner personas generation is plausible given the self personas and dialogue context. Automatic and human evaluation results support the hypothesis and indicate that generated personas are even more interesting than the ground truth, which improves the downstream dialogue response generation. This paper thus paves the way to exploit partner personas generation (PPG) for dialogue response generation (DRG).

We propose a novel framework composed of three major components, namely a personas generator, a dialogue response generator and a critic network. The personas generator generates partner personas, which the dialogue response generator conditions on. We employ reinforcement learning with a critic network that propagates the reward back to the generators for joint training.

Prior works have investigated partner persona retrieval (Zhang et al., 2018; Song et al., 2019). The human-constructed ground truth personas serve as the upper bound for such retrieval-based systems, and we argue that the ground truth is not coherent

and diverse enough. Interestingly, we observe that the generative counterpart proposed in our framework generates relevant, informative and coherent partner personas, which further improves the succeeding dialogue response generation. It follows another advantage that our framework does not need an external database to retrieve from (Madotto et al., 2020; Xu et al., 2021).

One close work to ours is a multi-task framework for meta-learning (Lee et al., 2021) that uses personas reconstruction as an auxiliary task to improve response consistency. The differences are that theirs does not differentiate between self personas and partner personas, while ours does. Theirs indicates an improvement over personality consistency, while ours report improvements for the overall quality. We conduct an empirical comparison with their model by reconstructing the partner personas. Experimental results indicate that such a multi-task model does not work well in our problem setting. Very recently, Zhou et al. (2021) formulates personas generation as a Seq2Seq task for improved downstream response generation via multi-task learning. In contrast, our work leverages reinforcement learning to jointly train the partner personas generator and the response generator.

Automatic and human evaluation results indicate that our framework can generate partner personas that are more diverse and interesting than the ground truth partner personas and generate more diverse and engaging responses than the baseline conditioned on ground truth partner personas.¹

2 Related Work

2.1 Personalized Dialogue Generation

Conditioning on personas helps to produce informative and engaging responses. The most well-known multi-turn dialogue dataset conditioned on personal profiles is PERSONACHAT (Zhang et al., 2018), in which two crowdsourcers converse and find more about each other. The community has proposed many methods to better utilize self personas. Mazaré et al. (2018) employs a pre-training stage based on dedicatedly extracted large-scale persona-based dialogues. Zhao et al. (2019) fuses information in personas and dialogue context into individual contextualized representations by attending to different parts of both. Gu et al. (2019) exploits the interaction between personas, dialogue

context and response to improve retrieval-based dialogue agents. Madotto et al. (2019) leverages meta-learning with several dialogues of the current speakers to enhance response personality. Welleck et al. (2019) releases a dataset for measuring dialogue consistency. Song et al. (2020) employs a multi-stage pipeline to improve response personality by response rewriting. Lee et al. (2021) uses multi-task learning for improved personality consistency in the meta-learning scenario. Gu et al. (2021) employs four different strategies for personas fusing to leverage both self persona. However, most of these works focus on exploiting self personas rather than partner personas, and they assume the existence of the gold partner personas.

2.2 User Profile Extraction

Li et al. (2014) leverages distant supervision to classify the spouse, education and job information from user twitters. Wu et al. (2020b) proposes a two-staged profile extractor that extracts attributes before extracting the underlying relationship. Wang et al. (2021) proposes to categorize the profile extraction task into two different difficulties, namely ‘extraction’ and ‘inference’, and they leverage a GPT-based generator to extract user profiles. These works have formulated user profile extraction as a classification task that conditions on an input sentence, and they aim at better profile extraction. In contrast, we propose to formulate personas generation to be conditioned dialogue input to be jointly trained with response generation. While ground truth personas serve as the upper bound for these user profile extractors, we empirically demonstrate that our reinforcement learning algorithm surpasses the response model conditioned on the ground truth partner personas. As supported by our human evaluation, we believe the underlying reason is that our model can leverage pre-trained generators to generate coherent and relevant partner personas.

2.3 Reinforcement Learning

Reinforcement learning (RL), or specifically, policy gradient methods (Williams, 1992), have been frequently adopted to both task-oriented dialogue agents (Roman Roman et al., 2020; Deng et al., 2021) or open-domain chitchat agents (Li et al., 2016c; Saleh et al., 2020). It can either propagate non-differentiable loss (Cai et al., 2019a) or optimize an expert reward such as ease of answering (Li et al., 2016c). It also adopts a scenario where a user simulator and a dialogue agent interact, and an

¹Related resources can be found at <https://github.com/HongyuanLuke/PPG>.

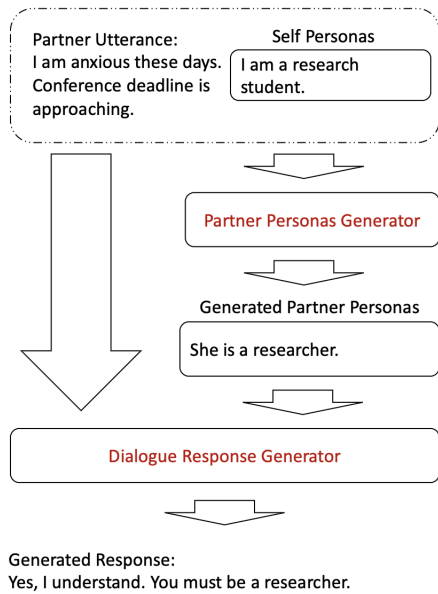


Figure 1: An example of the inference flow that shows the generated partner personas and the incorporation of partner personas generation into response generation.

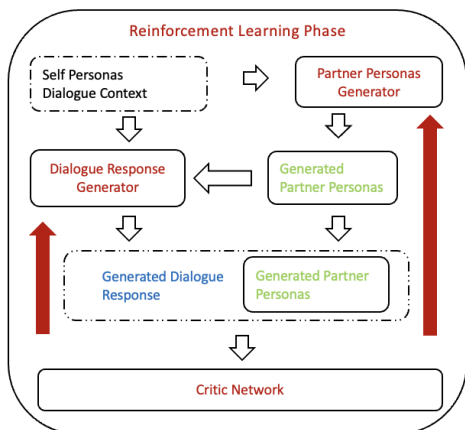


Figure 2: The illustrated reinforcement learning strategy that directly backpropagates the response-related rewards from the critic network to the partner personas generator and the dialogue response generator.

expert reward function can be defined to assign the goodness to each response generated (Roman Roman et al., 2020).

3 Proposed Framework

We propose a novel framework composed of three major components, namely a partner personas generator, a dialogue response generator and a reinforcement learning component with a critic network. Figure 1 depicts the inference flow of our setting. The input dialogue context with self persona is first fed into the partner personas generator.

The generated partner personas output is then concatenated with the dialogue context and the self personas as the input into the dialogue response generator. In the beginning, we train our partner personas generator and dialogue response generator under supervised learning. In the training stage, we use the ground truth partner personas to train the dialogue response generator, and we replace it with generated partner personas in the inference stage. After the supervised learning stage, the second stage is a reinforcement learning stage which jointly optimizes both partner personas generator and dialogue response generator as depicted in Figure 2 to train the partner personas generator under the reward signal that is relevant to dialogue response generation as well as fine-tuning dialogue response generator trained on the generated partner personas.² Particularly, we employ a dedicatedly designed critic network that receives generated partner personas and generated dialogue responses as the input and output a reward that measures the relevance between the generated personas and responses and propagates back to the generators.

3.1 Partner Personas Generation (PPG)

A Seq2Seq neural network (Sutskever et al., 2014) is adopted as our partner personas generator for the task of partner personas generation (PPG). The concatenation of dialogue context \mathbf{c} and self personas \mathbf{s} is fed as an input into the partner personas generator. The personas generator then outputs an approximated partner personas $\hat{\mathbf{p}}$ conditioned on the input that maximises the following likelihood:

$$P(\hat{\mathbf{p}} | \mathbf{s}, \mathbf{c}) = \prod_{t=1}^T P(\hat{p}_t | \hat{p}_1, \dots, \hat{p}_{t-1}, \mathbf{s}, \mathbf{c}),$$

where T represents the length of the generated partner personas and \hat{p}_t represents the word at the position t that has been inferred.

For training, the ground truth partner personas \mathbf{p} is used and we train our generator to maximise the likelihood $P(\mathbf{p} | \mathbf{s}, \mathbf{c})$. We generate the complete partner personas profiles in an one-off shot for all the dialogue samples.

3.2 Dialogue Response Generation (DRG)

We also adopt a Seq2Seq neural network for the task of dialogue response generation (DRG). During inference, the concatenation of dialogue context

²Section 5.7 presents an ablation study on reinforcement learning that demonstrates the effectiveness of this approach.

\mathbf{c} , self personas \mathbf{s} , and generated partner personas $\hat{\mathbf{p}}$ is fed as an input into the dialogue response generator. The response generator then outputs a dialogue response $\hat{\mathbf{r}}$ conditioned on the input, which maximises the conditional likelihood: $P(\hat{\mathbf{r}} | \mathbf{s}, \hat{\mathbf{p}}, \mathbf{c})$.

For training, the ground truth partner personas \mathbf{p} and the ground truth dialogue responses \mathbf{r} are used.

3.3 Reinforcement Learning (RL)

We employ a critic network to compute the reinforcement learning rewards for our generators. We use a binary classifier as critic by extracting training instances $(\mathbf{s}, \mathbf{r}, \mathbb{L}=1)$,³ $(\mathbf{s}^A, \mathbf{r}^A, \mathbb{L}=1)$ and $(\mathbf{s}^B, \mathbf{r}^B, \mathbb{L}=1)$. Then we can derive two negative samples as: $(\mathbf{s}^A, \mathbf{r}^B, \mathbb{L}=0)$ and $(\mathbf{s}^B, \mathbf{r}^A, \mathbb{L}=0)$. Thereafter, we fine-tune on a binary classifier to be used as our critic in RL on the training partition by minimizing the binary cross-entropy loss:

$$-\mathbb{L} \log(P(\mathbb{L} | \mathbf{s}, \mathbf{r})) - (1-\mathbb{L}) \log(1 - P(\mathbb{L} | \mathbf{s}, \mathbf{r})),$$

where the binary label \mathbb{L} indicates whether the response is relevant to the personas.

We then use this classifier acting as a critic network that outputs $\hat{\mathbb{L}}$, conditioned on the generated partner personas $\hat{\mathbf{p}}$ and generated response $\hat{\mathbf{r}}$. The predicted binary label $\hat{\mathbb{L}}$ is then converted to a reward \mathbb{R} . \mathbb{R} is a positive reward when $\hat{\mathbb{L}}=1$, and \mathbb{R} is a negative reward when $\hat{\mathbb{L}}=0$. We empirically set the reward \mathbb{R} for RL to $\{1, -1\}$ for both PPG and DRG. We then update our RL agents with the following gradients:

$$\Delta\theta_{\text{PPG}} = -\mathbb{R} \nabla_{\theta_{\text{PPG}}} \log P(\hat{\mathbf{p}} | \mathbf{s}, \mathbf{c})$$

for the partner personas generator (PPG), and for the dialogue response generator (DRG):

$$\Delta\theta_{\text{DRG}} = -\mathbb{R} \nabla_{\theta_{\text{DRG}}} \log P(\hat{\mathbf{r}} | \mathbf{s}, \hat{\mathbf{p}}, \mathbf{c})$$

By formulating a reward that measures the relevance between generated partner personas and generated dialogue response, we are motivated by the following objectives:

- Further fine-tune the partner personas generator to generate personas that benefits the downstream dialogue response generation.

³Our critic reports a test accuracy of about 75%. We empirically choose to use \mathbf{s} instead of \mathbf{p} . The latter reports a test accuracy of about 60%. This indicates that it seems people talk more seldomly about their partner during the conversation, but they still do, and it is useful to exploit partner personas.

- Further fine-tune the dialogue response generator trained with ground-truth partner personas to adapt to noisy partner personas generated by the partner personas generator.

As mentioned in Section 3.1, the first motivation is that we are generating the complete personas profile. However, some of them can be irrelevant and unhelpful for the next-turn dialogue response generation. It could be challenging for the partner personas generator alone to identify which personas could be helpful. Therefore, we design such a reward to train the personas generator to learn to generate a set of personas that is more helpful for the downstream dialogue response generation. Our second motivation is that the dialogue response generator has not been exposed to the generated partner personas. We would like to fine-tune the response generator to mitigate the potential training-inference discrepancy. Experimental results indicate that our design empirically works well.

The previous work from Cai et al. (2019a) employed critic network for RL loss backpropagation. The major difference is that their critic is trained in an adversarial manner (Li et al., 2018) to pick up the gold response among other negative candidates. Also, their critic network conditions only on the dialogue response but not on the generated skeleton. In contrast, we aim for improved response generation with a classifier conditioning on both the generated personas and the generated response.

3.4 Evaluation Metrics

For both PPG and DRG, perplexity (PPL) is reported to measure the intrinsic performance with the ground truth output (Roller et al., 2021). We adopt well-known sequence evaluation metrics weighted BLEU (Papineni et al., 2002) and F-measure for ROUGE-L (Lin, 2004) as the extrinsic evaluations. For PPG, we also report Distinct-N with $N=\{1,2\}$ to measure the response diversity (Li et al., 2016a; Cai et al., 2019b; Gao et al., 2019) with the ratio of distinct unigrams/bigrams against total number of unigrams/bigrams generated.

4 Experimental Setup

4.1 Dataset

We conduct experiments on the PERSONACHAT (Zhang et al., 2018), the most well-known multi-turn dialogue dataset conditioned on personas. We follow the train/valid/test split from the PARLAI

Model	PERSONACHAT-ORI			PERSONACHAT-REV		
	Perplexity↓	BLEU	ROUGE	Perplexity↓	BLEU	ROUGE
E2E w/o Partner Personas	14.70	0.7502	16.87	14.81	0.6772	16.79
E2E w/ Partner Personas in Training	14.76	0.7109	16.56	14.44	0.7083	16.43
E2E w/ Partner Personas in Training and Inference	14.00	0.8105	17.52	14.17	0.6631	16.64
GPT-2	17.53	0.8031	17.31	15.11	0.6095	16.59
TRANSFERTRANSFO (Wolf et al., 2019)	15.55	0.7346	17.02	18.32	0.5534	16.40
PERCVAE (Song et al., 2019)	41.43	0.2400	13.90	42.80	0.2385	13.51
PAML (Madotto et al., 2019)	41.86	0.4448	13.83	—	—	—
MTL w/ Personas Reconstruction (Lee et al., 2021)	232.1	0.0570	9.937	244.1	0.0504	9.916
Ours w/o Reinforcement Learning	13.91	0.8068	17.40	13.88	0.6583	17.01
Ours w/ Reinforcement Learning on PPG&DRG	13.05	1.0862	18.11	13.85	0.7380	17.47

Table 1: Automatic evaluation results on PERSONACHAT-ORI and PERSONACHAT-REV. Perplexity (PPL) attains a better quality with lower values and the remaining metrics attain a better quality with higher values.

platform (Miller et al., 2017) that contains about 65,000/7,800/7,500 instances respectively. Each instance contains about 8 utterances on average and about 4 traits for each of the self and partner personas. We denote the dataset with this original personas as PERSONACHAT-ORI. Later the original personas have been manually scrutinized by rephrasing, generalizing or specializing, which we denote as PERSONACHAT-REV. We apply the same preprocessing operation to both datasets. To train the critic for RL, we collected about 130,000 instances from the train split with equally distributed positive and negative samples.

4.2 Baselines and Comparison Models

E2E w/o Partner Personas This is an end-to-end (E2E) response generator without partner persona.

E2E w/ Partner Personas in Training With partial ground truth partner personas for training only.

E2E w/ Partner Personas in Training and Inference With ground truth partner personas for training and inference. During early experiments, we found that feeding all traits yields lower performance. Retrieving Top-3 relevant partner personas using BM25 (Robertson and Walker, 1994) yields the best performance on the original personas.

GPT-2 This is a comparison model fine-tuned on GPT-2 (Radford et al., 2019). We build the same three E2E systems described above, and the best model is selected, the third one.

TRANSFERTRANSFO A comparison model built with a Transformer-based model pre-trained on gen-

eral domain corpus, which is then fine-tuned on PERSONACHAT (Wolf et al., 2019).

PERCVAE This is a comparison model that employs a memory-augmented architecture incorporated with conditional variational autoencoder that exploits persona information (Song et al., 2019).

PAML This is a comparison model that leverages several dialogues collected from the same speaker to enhance response personality via meta-learning (Madotto et al., 2019). As the authors did not conduct experiments on the PERSONACHAT-REV and no preprocessing scripts are provided for the revised personas, we only report the results of their model on the PERSONACHAT-ORI only.

MTL w/ Personas Reconstruction This is a multi-task learning (MTL) comparison model (Lee et al., 2021) trained to maximise the objective:

$$\alpha \mathcal{L}_{\text{PPG}} + (1 - \alpha) \mathcal{L}_{\text{DRG}},$$

where \mathcal{L}_{PPG} represents the auxiliary PPG likelihood, and \mathcal{L}_{DRG} represents the DRG likelihood. α is weight tuned over the validation set, and both tasks condition on dialogue context and self personas and share the same model parameters.

5 Results

5.1 Dialogue Response Generation Results

We build our baselines, the partner personas generator and the dialogue response generator based on a state-of-the-art pre-trained dialogue model DIALOGPT (Zhang et al., 2020) for parameters

Dialogue Context	PARTNER PERSONAS		DIALOGUE RESPONSE		
	Gold Partner	Generated Partner	Ours	E2E w/ Full	Human
Hey ! Do you like music? I like metallica.	My favourite band is metallica.	My favourite band is metallica.	Hi! I do! I love metallica!	I do.	I like country music mostly.
I know, I grew up in a horse ranch that is so large.	I also enjoy cooking meals with food from our garden.	I like to cook with the food I grow in my garden.	I have a small garden in my home.	I am sorry, I am not very tall.	I wish my childhood was as exciting.
I meditate. It helps with my anger, I can be pretty violent.	I am in the army.	I am an army ranger.	I am a general in the army.	I do not have time for things like that. I am a busy person.	Do you like to work for long hour? Do you have any pet?
Yes, I have 8 dogs, 7 cats and 17 birds.	–	I have a dog and two cats.	I have a dog and a cat.	I have a dog.	That is a lot! I have 3 cats and 2 dogs.

Table 2: Case studies that compare our framework against the baseline with the complete partner personas as well as the human response. We present the preceding partner utterance as dialogue context, and we give the most salient ground truth partner personas (**Gold Partner**) and generated partner personas (**Generated Partner**) for clarity.

initialization. More implementation details can be found in Appendix B.

The dialogue response generation results are presented in Table 1. Our framework with reinforcement learning attains the best over all the metrics on both PERSONACHAT-ORI and PERSONACHAT-REV. This supports the usefulness of our framework, which generates reasonable personas and effectively enhances the succeeding dialogue response generation, through the use of RL.

Although TRANSFERTRANSFO attains a better score on the PPL than the fine-tuned GPT-2, GPT-2 have better extrinsic scores than TRANSFERTRANSFO. GPT-2 also has better overall scores than the E2E baselines without the complete partner personas. However, it is surpassed by the E2E baseline with the complete partner personas during training and inference.

The E2E baseline with the complete ground truth partner personas attains better scores on all of the metrics than our remaining baselines. Our framework with RL succeeds the performance of such a competitive baseline for both PERSONACHAT-ORI and PERSONACHAT-REV, indicating our proposed framework’s robustness against paraphrasal.

The multi-task learning comparison model (Lee et al., 2021) produces less promising results. Concretely, we postulate that the nature of PPG and DRG largely differs. The textual format of partner personas always initiates with first-person sentence starters, while dialogue responses are more general, ranging from greetings to goodbyes. Therefore, it could be hard to capture both in a single model.

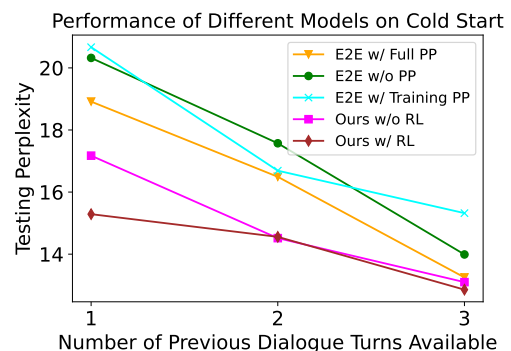


Figure 3: Analysis for the cold start problem with limited dialogue turns available. Note that all of these baselines are fine-tuned on DIALOGPT.

5.2 Cold Start

Cold start is a common problem in recommender systems (Schein et al., 2002; Zhang et al., 2014; Li et al., 2021). This also applies to dialogue systems, as the partner personas are commonly missing in early turns. We conduct an analysis on the baselines and our framework when N turns are available where $N=\{1,2,3\}$, using PERSONACHAT-ORI. As demonstrated in Figure 3, all the methods attain a better PPL when N increases, which indicates the existence of the cold start. This is also the case for the baseline with ground truth personas, and we postulate that it fails to learn how to use partner personas during cold start due to the lack of clues. Our framework effectively mitigates the cold start problem and attains the best among them for all N .

5.3 Case Study on Dialogue Response Generation

Table 2 depicts the case study for response generation using PERSONACHAT-ORI. In the first case,

Generated Partner Personas	Ground Truth Partner Personas
I am an army ranger. I secretly love my long deployments, because it gets me away from conventional life. I have a wife and two kids back in the states. I would be honored to give my life for my country. I am not afraid to die.	I am in the army. I am serving in South Korea. I was born in puerto rico. I am a violent person.
I drink protein powder with nothing but water. I like to watch mma. My prized possession is a bowie knife. I life weights, but I never do squats.	I am happy being single and alone. I only drink water. I go to the gym a days a week. I do not want children. I work in labor and delivery.
I like to watch football. My friends like watching it too. Its great fun. We drink beer and eat food.	I love watching football on Sunday. I have three dogs. My favroutie food is cheese piazza. I am a hair stylist.

Table 3: Case studies to show that our generated personas are relevant, informative and coherent.

Model	PPL↓	B	R	D-1	D-2
Gold	-	-	-	0.003	0.008
Our PPG	56.2	2.99	22.5	0.012	0.042
Gold	-	-	-	0.004	0.009
Our PPG	111	1.34	20.8	0.013	0.042

Table 4: Results for PPG. Upper for PERSONACHAT-ORI and lower for PERSONACHAT-REV. **B**, **R** and **D** represents BLEU, ROUGE and Distinct respectively.

our framework successfully recognizes that the partner is asking specifically for metallica. It then conditions on the generated personas to generate a much more entailed response than the baseline. The human response expresses negatively and thus seems less engaging. In the second case, our framework recognizes that the partner has a garden. It then talks about the garden rather than the irrelevant response from the baseline that we postulate is misled by the ‘large’ adjective in the dialogue context. The human response is potentially sarcastic if the partner is not joking, while our generation does not have such issue. For the third case, the baseline produces a response that could be potentially offensive, which could be biased by the word ‘violent’ in the dialogue context. In contrast, our framework recognizes the identity of the partner to generate a response without such an issue. The human response tends to raise a new topic and is less relevant. For the fourth case, we observe that the annotator sometimes converses based on the partner profile rather than his own traits. In this case, the annotator (**Dialogue Context**) said that he has many pets, which is not in his own traits (**Gold Partner**). Rather, his conversation partner expressed his passion for animals in previous dialogue contexts. We postulate that the annotator attempted to engage the conversation by conditioning his partner personas and telling a relevant joke. Our PPG can recognize this, which further tweaks the model output to talk about dogs and cats rather than

the dog only. These cases validate that leveraging partner personas is beneficial, and our framework can generate reasonable partner personas, which is not even in the ground truth.

5.4 Partner Personas Generation Results

Table 4 presents the quality measurements of the generated partner personas from our PPG with no RL. We observe that our models have much higher Distinct-N scores as the number of unique words in the generated output is much higher than the ground truth test personas. Compared to the ground truth personas that are limited sets of traits, our generator can leverage the power of pre-trained models for better diversity. The remaining metrics also report reasonable scores, suggesting the plausibility to formulate personas generation as a Seq2Seq task.

5.5 Case Study on Partner Personas Generation

Table 3 presents generated partner personas using PERSONACHAT-ORI. As depicted, our PPG can generate reasonable partner personas which are relevant to the ground truth partner personas. It sometimes gives a reasonable generation which is even not in the ground truth partner personas. In the first case, the generator successfully identifies the partner as being an army ranger. It then becomes rather positive than a violent person as given in the ground truth personas. Conditioning on such positive contents can give a positive response. In the second case, it recognizes the partner as a gym person, and imagines that the partner drinks protein and life weights, which is not in the ground truth personas. In the third case, the generator generates coherent personas, saying that the partner would drink beer and eat food while watching football, which is also not in the ground truth. We postulate that personas could be semantically closer to each other when they frequently co-occur in the training set. Our PPG then tends to generate more coherent

Criteria	E2E w/ Full PP	Our Framework
Appropriateness	42	58 [‡]
Informativeness	41	59 [‡]
Engagingness	40	60 [‡]
Human-likeness	42	58 [‡]

Table 5: Human evaluation results of dialogue response generation in winning percentages. [‡] indicates the results as passing a two-tailed binomial significance test with $p < 0.001$.

Criteria	Ground Truth	Our Framework
Coherence	41	59 [‡]
Interestingness	39	61 [‡]
Engagingness	44	56 [†]
Human-likeness	47	53

Table 6: Human evaluation results of PPG in winning percentages. [†] and [‡] indicates the results as passing a two-tailed binomial significance test with $p < 0.05$ and $p < 0.001$ respectively.

personas by learning such semantical relationship.

Since our generated personas are relevant and coherent, we postulate it as the underlying reason why our method gives a better generalization to DRG. In contrast, as demonstrated by Table 3, ground truth personas tend to be more like discrete collections of traits. This could be the reason why some of our generated partner personas could beat the ground truth, which is also supported by our human evaluation in Section 5.6. This is a potential benefit of our approach compared to sentence-level user profile extraction (Li et al., 2014; Wu et al., 2020b; Wang et al., 2021) that is upper bounded by the discrete ground truth. We present more examples in Table 8 in the Appendix.

5.6 Human Evaluation

We hired experienced annotators who have degrees relevant to English Linguistics to conduct evaluation on PERSONACHAT-ORI. For both DRG and PPG, we present a questionnaire composed of 800 questions with randomly sampled 200 test instances to three annotators who compare model outputs under A/B testing. As in Zou et al. (2021) and ACUTE-Evals (Li et al., 2020), annotators follow the criteria which we present in Appendix D.

Table 5 presents the human evaluation results on dialogue response generation. Our framework

Model	PPL↓	BLEU	ROUGE
Ours w/ PPG	13.80	0.8293	17.42
Ours w/ DRG	13.17	0.8391	17.36
Ours w/ PPG&DRG	13.05	1.0862	18.11

Table 7: Ablation study for our proposed framework on PERSONACHAT-ORI.

trained under RL surpasses the E2E model that leverages both training and inference ground truth partner personas from all the aspects.

Table 6 presents the human evaluation results on PPG. We observe that our PPG generates personas that are more coherent and interesting than the ground truth, which align with the facts observed in Section 5.4 and Section 5.5 indicating that our generated partner personas are more coherent and diverse.

5.7 Ablation Study

We conduct an ablation study on PERSONACHAT-ORI as reported in Table 7 to present the performance of our framework when one of the components is frozen during RL. The result indicates that our proposed framework yields the best result when both of the components are actively trained under RL. We also notice that scaling the RL reward for either PPG or DRG by 10 leads to minor decrease in the performance. Further scaling deteriorates the quality of response generation.

6 Conclusion

Our novel framework incorporates partner personas generation into dialogue response generation. It effectively mitigates the problem that partner personas are not available in practical applications as well as the cold start problem during early conversation. The experimental results with both automatic and human evaluation demonstrate that our framework generates coherent, diverse, interesting and engaging partner personas, even compared to the ground truth partner personas. We employ reinforcement learning with a dedicatedly designed critic network that boosts the response generation by conditioning on the generated personas. Automatic and human evaluation results indicate that our response generator surpasses our competitive baselines that condition on the ground truth partner personas. Extensive case studies demonstrate that our framework can generate satisfying dialogue responses and partner personas.

Acknowledgments

This research/paper was supported by the Center for Perceptual and Interactive Intelligence (CPII) Ltd under the Innovation and Technology Commission's InnoHK scheme.

References

- Ashutosh Baheti, Maarten Sap, Alan Ritter, and Mark Riedl. 2021. [Just say no: Analyzing the stance of neural dialogue generation in offensive contexts](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 4846–4862, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Deng Cai, Yan Wang, Wei Bi, Zhaopeng Tu, Xiaojiang Liu, Wai Lam, and Shuming Shi. 2019a. [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, Minneapolis, Minnesota. Association for Computational Linguistics.
- Deng Cai, Yan Wang, Wei Bi, Zhaopeng Tu, Xiaojiang Liu, and Shuming Shi. 2019b. [Retrieval-guided dialogue response generation via a matching-to-generation framework](#). 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 1866–1875, Hong Kong, China. Association for Computational Linguistics.
- Yang Deng, Yaliang Li, Fei Sun, Bolin Ding, and Wai Lam. 2021. [Unified conversational recommendation policy learning via graph-based reinforcement learning](#). In *SIGIR*, pages 1431–1441.
- Jun Gao, Wei Bi, Xiaojiang Liu, Junhui Li, and Shuming Shi. 2019. [Generating multiple diverse responses for short-text conversation](#). *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):6383–6390.
- Jia-Chen Gu, Zhen-Hua Ling, Xiaodan Zhu, and Quan Liu. 2019. [Dually interactive matching network for personalized response selection in retrieval-based chatbots](#). 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 1845–1854, Hong Kong, China. Association for Computational Linguistics.
- Jia-Chen Gu, Hui Liu, Zhen-Hua Ling, Quan Liu, Zhigang Chen, and Xiaodan Zhu. 2021. [Partner matters! an empirical study on fusing personas for personalized response selection in retrieval-based chatbots](#). In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR 2021, page 565–574, New York, NY, USA. Association for Computing Machinery.
- Hyunwoo Kim, Byeongchang Kim, and Gunhee Kim. 2020. [Will I sound like me? improving persona consistency in dialogues through pragmatic self-consciousness](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 904–916, Online. Association for Computational Linguistics.
- Diederik P. Kingma and Jimmy Ba. 2015. [Adam: A method for stochastic optimization](#). In *ICLR (Poster)*.
- Jing Yang Lee, Kong Aik Lee, and Woon Seng Gan. 2021. [Generating personalized dialogue via multi-task meta-learning](#). In *Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*, Potsdam, Germany. SEMDIAL.
- Dianqi Li, Qiuyuan Huang, Xiaodong He, Lei Zhang, and Ming-Ting Sun. 2018. [Generating Diverse and Accurate Visual Captions by Comparative Adversarial Learning](#). *CoRR*, abs/1804.00861.
- Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2016a. [A diversity-promoting objective function for neural conversation models](#). In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 110–119, San Diego, California. Association for Computational Linguistics.
- Jiwei Li, Michel Galley, Chris Brockett, Georgios Spithourakis, Jianfeng Gao, and Bill 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, Berlin, Germany. Association for Computational Linguistics.
- Jiwei Li, Will Monroe, Alan Ritter, Dan Jurafsky, Michel Galley, and Jianfeng Gao. 2016c. [Deep reinforcement learning for dialogue generation](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1192–1202, Austin, Texas. Association for Computational Linguistics.
- Jiwei Li, Alan Ritter, and Eduard Hovy. 2014. [Weakly supervised user profile extraction from twitter](#). volume 1, pages 165–174.
- Margaret Li, Stephen Roller, Ilia Kulikov, Sean Welleck, Y-Lan Boureau, Kyunghyun Cho, and Jason Weston. 2020. [Don't say that! making inconsistent dialogue unlikely with unlikelihood training](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4715–4728, Online. Association for Computational Linguistics.
- Margaret Li, Jason Weston, and Stephen Roller. 2019. [ACUTE-EVAL: Improved Dialogue Evaluation with Optimized Questions and Multi-turn Comparisons](#). *CoRR*, abs/1909.03087.

- Shijun Li, Wenqiang Lei, Qingyun Wu, Xiangnan He, Peng Jiang, and Tat-Seng Chua. 2021. [Seamlessly unifying attributes and items: Conversational recommendation for cold-start users](#). *ACM Trans. Inf. Syst.*, 39(4).
- Chin-Yew Lin. 2004. [ROUGE: A package for automatic evaluation of summaries](#). In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.
- 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, Online. Association for Computational Linguistics.
- Andrea Madotto, Samuel Cahyawijaya, Genta Indra Winata, Yan Xu, Zihan Liu, Zhaojiang Lin, and Pascale Fung. 2020. [Learning knowledge bases with parameters for task-oriented dialogue systems](#). In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2372–2394, Online. Association for Computational Linguistics.
- 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, Florence, Italy. Association for Computational Linguistics.
- 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, Online. Association for Computational Linguistics.
- Pierre-Emmanuel Mazaré, Samuel Humeau, Martin Raison, and Antoine Bordes. 2018. [Training millions of personalized dialogue agents](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2775–2779, Brussels, Belgium. Association for Computational Linguistics.
- Alexander Miller, Will Feng, Dhruv Batra, Antoine Bordes, Adam Fisch, Jiasen Lu, Devi Parikh, and Jason Weston. 2017. [ParLA: A dialog research software platform](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 79–84, Copenhagen, Denmark. Association for Computational Linguistics.
- 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, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.
- Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners.
- Stephen Robertson and Steve Walker. 1994. [Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval](#). pages 232–241.
- Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Eric Michael Smith, Y-Lan Boureau, and Jason Weston. 2021. [Recipes for building an open-domain chatbot](#). In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 300–325, Online. Association for Computational Linguistics.
- Homero Roman Roman, Yonatan Bisk, Jesse Thomason, Asli Celikyilmaz, and Jianfeng Gao. 2020. [RMM: A recursive mental model for dialogue navigation](#). In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1732–1745, Online. Association for Computational Linguistics.
- Abdelrhman Saleh, Natasha Jaques, Asma Ghandeharion, Judy Shen, and Rosalind Picard. 2020. [Hierarchical reinforcement learning for open-domain dialog](#). *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05):8741–8748.
- Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. [DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter](#). *CoRR*, abs/1910.01108.
- Andrew I. Schein, Alexandrin Popescul, Lyle H. Ungar, and David M. Pennock. 2002. [Methods and metrics for cold-start recommendations](#). In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '02*, page 253–260, New York, NY, USA. Association for Computing Machinery.
- 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, Online. Association for Computational Linguistics.
- Haoyu Song, Yan Wang, Wei-Nan Zhang, Xiaojiang Liu, and Ting Liu. 2020. [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, Online. Association for Computational Linguistics.
- Haoyu Song, Wei-Nan Zhang, Yiming Cui, Dong Wang, and Ting Liu. 2019. [Exploiting persona information for diverse generation of conversational responses](#). In *Proceedings of the Twenty-Eighth International*

- Joint Conference on Artificial Intelligence, IJCAI-19*, pages 5190–5196. International Joint Conferences on Artificial Intelligence Organization.
- Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2, NIPS'14*, page 3104–3112, Cambridge, MA, USA. MIT Press.
- Jack Urbanek, Angela Fan, Siddharth Karamcheti, Saachi Jain, Samuel Humeau, Emily Dinan, Tim Rocktäschel, Douwe Kiela, Arthur Szlam, and Jason Weston. 2019. [Learning to Speak and Act in a Fantasy Text Adventure Game](#). *CoRR*, abs/1903.03094.
- Zhilin Wang, Xuhui Zhou, Rik Koncel-Kedziorski, Alex Marin, and Fei Xia. 2021. [Extracting and Inferring Personal Attributes from Dialogue](#). *CoRR*, abs/2109.12702.
- Sean Welleck, Jason Weston, Arthur Szlam, and Kyunghyun Cho. 2019. [Dialogue natural language inference](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3731–3741, Florence, Italy. Association for Computational Linguistics.
- Ronald J. Williams. 1992. [Simple statistical gradient-following algorithms for connectionist reinforcement learning](#). *Mach. Learn.*, 8(3–4):229–256.
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. [Transformers: State-of-the-art natural language processing](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.
- Thomas Wolf, Victor Sanh, Julien Chaumond, and Clement Delangue. 2019. [TransferTransfo: A Transfer Learning Approach for Neural Network Based Conversational Agents](#). *CoRR*, abs/1901.08149.
- Bowen Wu, MengYuan Li, Zongsheng Wang, Yifu Chen, Derek F. Wong, Qihang Feng, Junhong Huang, and Baoxun Wang. 2020a. [Guiding variational response generator to exploit persona](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 53–65, Online. Association for Computational Linguistics.
- Chien-Sheng Wu, Andrea Madotto, Zhaojiang Lin, Peng Xu, and Pascale Fung. 2020b. [Getting to know you: User attribute extraction from dialogues](#). In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 581–589, Marseille, France. European Language Resources Association.
- Yan Xu, Etsuko Ishii, Zihan Liu, Genta Indra Winata, Dan Su, Andrea Madotto, and Pascale Fung. 2021. [Retrieval-Free Knowledge-Grounded Dialogue Response Generation with Adapters](#). *CoRR*, abs/2105.06232.
- Mi Zhang, Jie Tang, Xuchen Zhang, and Xiangyang Xue. 2014. [Addressing cold start in recommender systems: A semi-supervised co-training algorithm](#). In *Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '14*, page 73–82, New York, NY, USA. Association for Computing Machinery.
- 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, Australia. Association for Computational Linguistics.
- Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan. 2020. [DIALOGPT : Large-scale generative pre-training for conversational response generation](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 270–278, Online. Association for Computational Linguistics.
- Xueliang Zhao, Chongyang Tao, Wei Wu, Can Xu, Dongyan Zhao, and Rui Yan. 2019. [A document-grounded matching network for response selection in retrieval-based chatbots](#). In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 5443–5449. International Joint Conferences on Artificial Intelligence Organization.
- Yinhe Zheng, Guanyi Chen, Minlie Huang, Song Liu, and Xuan Zhu. 2019. [Personalized Dialogue Generation with Diversified Traits](#). *CoRR*, abs/1901.09672.
- 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, Online. Association for Computational Linguistics.
- Wangchunshu Zhou, Qifei Li, and Chenle Li. 2021. [Learning to Predict Persona Information for Dialogue Personalization without Explicit Persona Description](#). *CoRR*, abs/2111.15093.
- Yicheng Zou, Zhihua Liu, Xingwu Hu, and Qi Zhang. 2021. [Thinking clearly, talking fast: Concept-guided non-autoregressive generation for open-domain dialogue systems](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 2215–2226, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

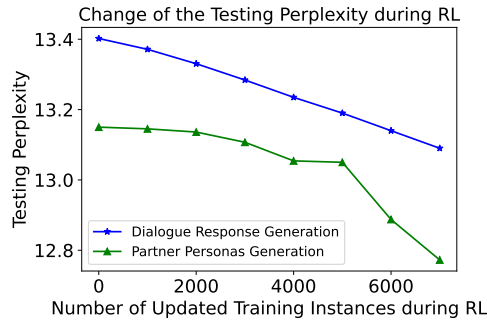


Figure 4: Change of the testing PPL during RL.

A Ethics Statement

The PERSONACHAT dataset used in this work is well-known and widely used. In our view, there is no known ethical issue with its usage. Large-scale pre-trained models are also employed, but they are widely known to be subject to potential problems such as generating offensiveness context. With its use, our partner personas generator could generate unseen personas, which are also subject to potential offensive generation. An offensiveness check can be incorporated to alleviate this problem for actual usage (Baheti et al., 2021).

B Implementation Details

For supervised phase, we set Adam (Kingma and Ba, 2015) as our optimizer, with hyperparameters $\eta = 5e-4$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1e-8$. The models are fine-tuned for 2 epochs. For RL phase, we set Adam as our optimizer, with $\eta = 5e-6$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1e-8$. We update the model parameters every 20 training instances and validate the model performance every 50 updates. DistilBERT (Sanh et al., 2019) is used to initialize the model parameters for the critic network. We set Adam as our optimizer, with hyperparameters $\eta = 5e-6$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1e-8$. We fine-tune the critic for 1 epoch, and we freeze it empirically during RL. All the experiments are conducted based on the TRANSFORMERS library from HUGGINGFACE (Wolf et al., 2020).

C Analysis on Dialogue Reponse Generation

We present the progressive change of the testing perplexity for DRG and PPG on PERSONACHAT-ORI in Figure 4.⁴ We observe that they improve

⁴The result is scaled for the sake of space and clarity.

simultaneously, which supports our motivation to use RL for joint training.

D Human Evaluation Criteria

- **(Appropriateness):** "Who is more appropriate given the previous dialogue context?"
- **(Informativeness):** "Who is more diverse instead of null answers such as I do not know?"
- **(Engagingness):** "Who would you prefer to talk with for a long conversation?"
- **(Human-likeness):** "Which speaker do you think sounds more like a real person?"
- **(Coherence):** "Which persona contains traits that are more coherent to each other?"
- **(Interestingness):** "Which persona is more interesting and diverse?"

The first four are from the existing work (Li et al., 2019; Zou et al., 2021) and we propose the last two for evaluating PPG. We report the first four for DRG, and we report the last four for PPG.

E Dataset Limitations

Our work uses an off-the-shelf persona-based conversational dataset PERSONACHAT (Zhang et al., 2018), which is collected and built by crowdsourcing to converse based on a fake set of discrete traits. There is no personal information and hence no ethics concern, but this might result in limited usefulness as there could be discrepancies between the collected samples and real-life conversation. It is also more expensive to collect real data. However, PERSONACHAT has been widely used by the community as a standard dataset. Many well-known persona-based datasets suffer from the same problem (Urbanek et al., 2019) as widely known.

Although Mazaré et al. (2018) proposed a useful method to collect large-scale persona-based dialogue datasets by extracting persona from user comments with classifiers trained on revised personas from PERSONACHAT which can improve the model performance on PERSONACHAT. For legal reasons, they did not release this dataset at the time of writing. Similarly, Zheng et al. (2019) proposed a persona-based dialogue dataset with diversified traits, but it is not currently online readily available.

Zhong et al. (2020) has followed the approach suggested by Mazaré et al. (2018) to build an empathetic conversation dataset based on personas.

Generated Partner Personas

Personas A: I am a shy person, but I love to sing. Until recently, I ve never been able to sing in front of anyone. Anyways, I decided to give it a try and participaed in an audition for a talent show . My shyness made me panick and I didn t show up.

Personas B: I play the violin. I am married with 5 kids. I am nurse. I met my husband when I was a freshman in college .

Personas C: I am a soccer player. I am a goalie. My number is 42. Nike cleats are my favorite. I joined a new team last month.

Personas D: I have two kids, ages 2 and 6. I am from sterling heights, michigan . My favorite movie is titanic . I work part time at aldis. My husband owns a small auto repair shop.

Personas E: I am a retired computer programmer. I have one grandson and one daughter. I just turned 77. I love animals. I like watching british tv shows and movies .

Personas F: I like to go hunting. I like to remodel homes. I like to shoot a bow. My favorite holiday is halloween . I like to go shopping with my daughters.

Personas G: I have a large cd collection. I collect stamps. Favorite band is the beetles . I like vintage furniture .

Personas H: I like to drink wine. I enjoy reading history books. I am a teacher. I love to write stories while sitting in the grass in my back yard. I grew up in new hampshire.

Personas I: I am retired. I stay active. I have eight grandchildren. I have good health.

Personas J: I m a student. I like to go out to eat. I like listening to other rap music too. One of my favorite artists is drake . A hobby of mine is the drums. I also enjoy cooking .

Personas K: I have two children. I like to go on walks. I am from mexico. I used to be a chef, but I am a teacher now. I like to bake.

Personas L: I like to do all my shopping at walmart . I m deathly terrified of heights. I prefer to live where the weather s cold. Winter s my favorite time of the year. I m really excited to see how game of thrones ends.

Personas M: I like to cook. I am a foodie. I love to chat with my friends. I love kids and dogs. I like to go shopping with my daughters.

Personas N: My family hates my fiance. We will be traveling to niagra falls for our honeymoon. We are getting married in a park. My dog is the ring bearer.

Personas O: My favorite color is red. I have 2 dogs as pets. I leave the dogs home when I visit my parents. I love dogs. I work as a veterinarian s assistant .

Personas P: I am a musician. I wish I could spend more time at home. I like to write my own songs. I have taken formal music lessons since i was 5.

Personas Q: I love comics. I love reading. I ve started creating my own comics and presenting them to publishers. I decided to publish my creations on internet. I ve been rejected several times and thought of giving up with this .

Table 8: More generated personas. We highlight in pink for informativeness and in yellow for coherence.

However, their main focus is to investigate the impact of personas on empathetic dialogue generation. Therefore, we choose to follow the community to investigate our method on the most well-known dataset, PERSONACHAT.

F Computing Infrastructure

We run all our experiments on a single NVIDIA TITAN RTX with 24GB GPU memory. Fine-tuning

the generators for 2 epochs as we have done on our preprocessed PERSONACHAT train split consumes about 3-4 hours. Fine-tuning our critic classifier for 1 epoch consumes about 1 hour. Our RL phase consumes about 15 hours to achieve the best validation loss before being early stopped. We report averaged results from 3 runs for our dialogue response generation and partner personas generation results reported in Table 1, Table 4 and Table 7.