MCL: Meta-Collaborative Learning for Generalized Personalized Dialogue Generation

Anonymous ACL submission

Abstract

Large language models (LLMs) have greatly advanced conversational systems, making Personalized Dialogue Generation (PDG) a core direction for tailoring responses to individual users. However, retrieval-augmented and user-parameterization approaches typically rely on dense user-level persona information and consistent domain-level persona distributions, leading to two key challenges for scalable PDG: sparse personas of unseen users hinder personalization, while domain-bound user parameters require costly retraining across dialogue scenes. Therefore, we argue that these limitations originate from an over-reliance on static data fitting, as opposed to mastering atomic personalization capabilities. To this end, we propose Meta-Collaborative Learning (MCL), a novel learning paradigm that proactively supplements cues via *Neighbor Identifier* and *Booster*, and dynamically aligns responses to preferences by *Generator*. By decomposing PDG into synergistic modules with iterative joint optimization, MCL internalizes universal personalization skills for acquiring and leveraging contextual signals across diverse users and domains. Extensive experiments validate that MCL delivers robust personalization under practical unseen-user and cross-domain settings.

1 Introduction

Substantial advances in Large Language Models (LLMs) have significantly enhanced the quality and naturalness of modern conversational systems. However, generic responses lack tailored adaptation to distinct users, making personalization a core demand for boosting user engagement and retention (Kirk et al., 2023; Zhang et al., 2024b). To this aim, Personalized Dialogue Generation (PDG) has emerged as a critical direction, with current approaches falling into two categories: RAG-based and user-parameterization. RAG-based methods (Salemi et al., 2024; Huang et al., 2023, 2024; Tang et al., 2024) learn to personalize generation on related histories or personas, implicitly requiring dense data to model user preferences. Parameterization methods (Zhong et al., 2021; Zhuang et al., 2024; Zhang et al., 2025) encode individuals in specific parameters or vectors which can be loaded based on their IDs, hinging on abundant labeled data and consistent train-deploy distributions.

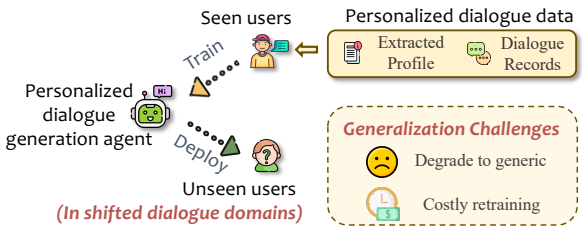

Figure 1: The generalization issues of personalized dialogue generation for scalable deployments.

Despite impressive progress, existing models exhibit inherent bottlenecks that severely limit their scalability and hinder real-world deployment. In practice, as illustrated in Figure 1, it is impractical to collect comprehensive interaction data for every potential user in advance. As a result, deployed PDG systems must handle unseen users who are not included during training and provide minimal examples at test time. Such data sparsity prevents conventional methods from adapting learned representations to unseen users, thus resulting in a fallback to generic responses. Moreover, the efficacy of user-parameterization is inherently tied to the specific domains they are trained on. Consequently, maintaining personalization across domains with significant distributional gaps requires extensive, costly retraining. These inherent constraints become prominently evident when PDG models are deployed or licensed to distinct applications, services or client organizations.

We argue that the above generalization challenges of existing methods stem from an over-

reliance on a rigid data-driven paradigm, which reduces PDG to aligning static patterns. The dynamic and autonomous nature of real-world personalized interactions, however, necessitates PDG systems capable of scalable, cost-effective adaptation. This dilemma implies that effective PDG requires learning atomic skills of *how to personalize*, rather than memorizing static patterns of what to personalize. Thereby, we make the first attempt to ***elevate the PDG paradigm from data-fitting to skill-mastering***, so as to ensure models themselves learn to acquire and exploit contextual signals in case of data sparsity or distribution shifts.

To achieve generalized PDG, models are expected to supplement personalized cues and dynamically align responses to user preferences. To this end, we propose a meta-learning paradigm that explicitly learns universal personalization skills. Inspired by advances in LLM agents (Chowa et al., 2026), we introduce **M**eta-**C**ollaborative **L**earning **(MCL)**, a novel learning framework that inherits the strengths of the existing approaches while mitigating their practical limitations.

The key insight of MCL is to address the generalization challenges via dynamic collaboration. Concretely, it is instantiated with three cooperative modules that directly map to the challenges above: (1) a *Neighbor Identifier* that collects relevant user–context pairs to serve as collaborative signals; (2) a *Booster* that refines contexts to distill user preferences; and (3) a *Generator* that produces coherent and personalized responses conditioned on the enhanced context. We devise an iterative optimization scheme to optimize the Booster and Generator jointly: the Booster learns to improve the utility of contextual personalization cues by exploring refinements that enhance downstream response quality, while the Generator adapts its generation behavior to the evolving context representations. This interplay encourages models to discover personalization mechanisms that generalize beyond specific users and domains, rather than overfitting to memorized patterns. At deployment, MCL operates with a lightweight, single-pass inference pipeline: once neighbors are retrieved, the enhanced context is forwarded sequentially through the pre-trained modules. Critically, no test-time optimization or user-specific adapters retraining is needed, ensuring inference latency comparable to standard generation pipelines.

We characterize the generalized PDG task from a domain-centric perspective and consider two practical settings. In the intra-domain generalization testbed, both training and testing data originate from the same dialogue dataset, including unseen users with no prior knowledge during training, representing the common generalization challenge in deployed systems. In the cross-domain generalization testbed, training and testing users come from different dialogue domains, introducing additional shifts in interaction styles and data distributions. We benchmark MCL across these settings, where intra-domain experiments assess sparsity resilience and cross-domain experiments measure distributional robustness. Extensive results demonstrate that MCL consistently outperforms baselines, validating its ability to deliver robust personalization across the full spectrum of real-world deployment challenges. Our key contributions are threefold:

- We reveal the core generalization challenges in personalized dialogue generation, and highlight the necessity of elevating the PDG paradigm from data-fitting to skill-mastering.
- We propose Meta-Collaborative Learning, a novel learning paradigm that learns atomic personalization skills via cooperative modules optimized with an iterative scheme.
- Comprehensive experiments validate that MCL enables generalized personalization across users and domains without retraining, offering valuable insights for scalable adaptation.

2 Related Work

2.1 Personal Knowledge Acquisition

Personal knowledge acquisition focuses on perceiving and representing user-related information from diverse sources, such as external corpora or ongoing conversations. This stage is fundamental for enabling dialogue models to capture user context and produce coherent, personalized responses. Existing methods can be broadly categorized into retrieval-based and representation-learning approaches. Retrieval-based methods explicitly retrieve user-relevant records to supplement the dialogue context, using lexical matching or semantic similarity search to select relevant information, which is then incorporated into prompts (Salemi et al., 2024; Song et al., 2021). Representation-learning approaches, in contrast, encode personal attributes into learnable latent structures. For example, CLV (Tang et al., 2023)

employs contrastive latent variables combining sparse and dense persona embeddings; SPT (Huang et al., 2024) distills persona attributes into soft prompts selectively activated by context; and MORPHEUS (Tang et al., 2024) maintains role representations compactly in a persona codebook.

2.2 Personalized Generation Alignment

Building upon the acquired user knowledge, personalized generation alignment focuses on tailoring model outputs to user-specific characteristics. Recent methods primarily differ in whether the tuning modules are shared across users or individually assigned. Shared-module approaches train a single fine-tuning component to encode aggregated user information (Woźniak et al., 2024; Zhang et al., 2024a), yielding efficient yet static personalization, which struggles to capture individual nuances. Individual-module approaches assign each user a dedicated lightweight component: LaMP PEFT-RAG (Salemi and Zamani, 2025) integrates train-time fine-tuning with test-time RAG, while PLoRA (Zhang et al., 2024a) extends this design with the combination of PKI and LoRA to inject personalized information into LLMs. Although some recent work has explored adaptation to users with masked predefined profiles (Tang et al., 2024; Liu et al., 2022), these methods operate under overly simplified assumptions that rarely hold in real-world commercial deployments, that the training and deployment user domains are identically distributed and that extensive dialogue histories are always available, highlighting the need for approaches capable of robust generalization to user domains with shifted distributions.

2.3 Domain Generalization

Domain Generalization (DG) focuses on training models using data from one or multiple related but distinct source domains, with the goal of performing well on target domains without access to target-domain data during training (Zhou et al., 2022; Wang et al., 2021). Existing DG methodologies generally fall into three categories: (1) Data manipulation, including augmentation or synthetic sample generation, to increase the diversity and coverage of source domains (Shankar et al., 2018; Zhou et al., 2021); (2) Disentangled representation learning, which seeks to separate domain-invariant and domain-specific features to improve transferability (Piratla et al., 2020; Ilse et al., 2019); and (3) Learning strategies, encompassing meta-learning to simulate domain shifts, ensemble techniques, and reinforcement learning approaches (Hansen and Wang, 2021; Cha et al., 2021; Lee et al., 2020). Building upon these DG principles, our work aims to endow PDG models with a generalization capability, enabling them to handle complex and compound user scenarios that involve both data sparsity and domain shifts.

A concurrent submission (Anonymous, 2026) proposes a personalized steering method to disentangle preference from contextual bias, while our work orthogonally focuses on cross-user/domain generalization via training-time meta-learning.

3 Problem Formulation

We define Generalized Personalized Dialogue Generation as the task of producing coherent and user-tailored responses for target users under realistic generalization conditions, including unseen users and domain-level distribution shifts.

3.1 Preliminaries and Notation

Let $\mathcal{D}$ denote a domain, corresponding to an independent application or service with unique interaction patterns and user data characteristics. Each user $u \in \mathcal{U}$ possesses a predefined persona P_u in text form and engages in multi-turn dialogues with other users, forming a conversation context $H_t = \{(q_1, y_1), \ldots, (q_{t-1}, y_{t-1}), q_t\}$. The goal of personalized dialogue generation is to generate a response y_t that aligns with u's persona and the ongoing conversational context. We denote the sets of seen and unseen users as $\mathcal{U}_{\text{seen}}$ and $\mathcal{U}_{\text{unseen}}$, sampled from the source and target domains $\mathcal{D}_S$ and $\mathcal{D}_T$, respectively, without overlapping user identities: $\mathcal{U}_{\text{seen}} \cap \mathcal{U}_{\text{unseen}} = \emptyset$, $\mathcal{U}_{\text{seen}} \sim \mathcal{D}_S$, $\mathcal{U}_{\text{unseen}} \sim \mathcal{D}_T$. The personalized dialogue model is parameterized by Θ and trained on $\mathcal{U}_{\text{Seen}}$, while being deployed on $\mathcal{U}_{\text{Unseen}}$.

3.2 Generalized PDG Tasks

To facilitate systematic study and evaluation, we organize generalized scenarios from a domain-centric perspective[1].

Task 1: Intra-domain Generalization (IDG) Both training and target users originate from the same domain without overlaps, representing the

[1]Henceforth, unless otherwise specified, we suppress the turn subscript t to avoid ambiguity and use H, P, and y for the dialogue history, profile, and the target response of user u.

prevalent data-sparsity challenge in deployed systems: $y = \Theta(H, P), \quad \mathcal{D}_T = \mathcal{D}_S$. *Challenge:* Generalize to unseen users within same domain.

Task 2: Cross-domain Generalization (CDG) Target users are drawn from a different domain than the training data, introducing inherent distribution shifts in interaction styles, topic distributions, and personalization patterns across domains: $y = \Theta(H, P), \quad \mathcal{D}_T \neq \mathcal{D}_S$. *Challenge:* Maintain consistent personalization under domain-level distributional gaps.

Together, these two tasks capture the core practical barriers to scalable PDG deployment highlighted in real-world scenarios: personalization under unseen users and robustness to cross-domain shifts. They motivate learning personalization mechanisms that generalize beyond specific users or domains, rather than memorizing static personalization patterns.

4 Methodology

4.1 Motivation

Personalized dialogue models deployed in real-world scenarios face two fundamental challenges: (1) generating coherent and user-aware responses for unseen users whose preferences or profiles are unavailable during training; and (2) maintaining consistent personalization performance under cross-domain distribution shifts. Addressing these challenges calls for a paradigm that can leverage transferable personalization knowledge from similar users and dynamically enrich user-specific contextual cues for the target user and domain context.

To this end, we propose **Meta-Collaborative Learning (MCL)**, a novel learning paradigm that enables adaptive and robust personalization. As illustrated in Figure 2, MCL decomposes personalization process into three synergistic components:

- Collaborative context enrichment: Identify and aggregate informative signals from similar users to form enriched contextual representations for unseen users.
- Dynamic persona refinement: Adapt and filter retrieved persona information to ensure consistency and faithfulness with respect to the target user and dialogue context.
- Tailored response generation: Generate personalized responses conditioned on the refined collaborative context.

MCL embodies a core Domain Generalization principle: aligning posterior distributions to learn domain-invariant features (Zhou et al., 2022). Instead of minimizing KL-divergence between posteriors (Lin et al., 2022), we achieve this implicitly by aligning persona–context→response distributions across users and domains. MCL learns personalization skills robust to cross-domain distribution shifts, establishing a principled DG foundation for our approach and extending alignment concepts to personalized dialogue—demonstrating how MCL naturally instantiates alignment for generalized PDG.

4.2 Architecture and Pipeline

Neighbor Identifier In real-world deployments, a dialogue model typically accesses only sparse user records within a given domain, without prior knowledge of unseen users. To overcome this limitation, the **Neighbor Identifier** Θ_N is introduced to figure out users u' whose personal data most closely resemble that of the target user u (with $u, u' \in \mathcal{U}$). These retrieved neighbors act as collaborative references, providing auxiliary signals that enrich the target user's context H_u and capture dialogue patterns within the same domain. Formally, the resulting collaborative neighborhood is defined as $\mathcal{N}_u = \Theta_N(H_u, P_u; \mathcal{U})$, serving as a foundation for improved generalization to unseen users and cross-domain distribution shifts.

Booster & Generator Simply incorporating neighbor personas, however, can be suboptimal since static profiles may lack sufficient contextual alignment with the target user's ongoing dialogue content. To address this, **Booster** Θ_B is designed to refine the retrieved neighborhood $\mathcal{N}_u$ into an enhanced collaborative context C_u, which serves as a dynamic and task-aware input augmentation.

Since explicit supervision for the Booster is unavailable in the dataset, we establish an implicit learning signal through feedback from the downstream **Generator** Θ_G. Specifically, for each query, the Booster performs J stochastic trials to produce multiple enhanced context candidates:

$$C_u^j = \Theta_B(H_u, P_u; \mathcal{N}_u), \quad j = 1, \ldots, J. \quad (1)$$

Each C_u^j consists of the refined persona and dialogue history of neighbor users and is subsequently passed to the Generator, which outputs a set of corresponding response candidates:

$$y^j = \Theta_G(H_u, P_u; C_u^j), \quad j = 1, \ldots, J. \quad (2)$$

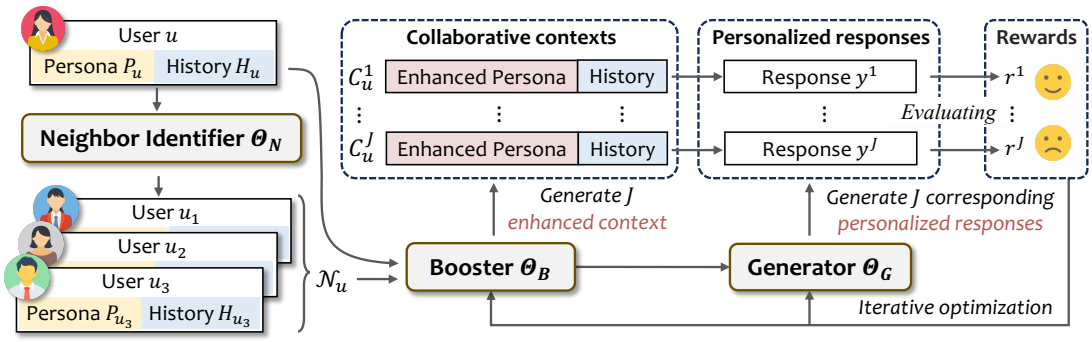

Figure 2: Overview of the proposed Meta-Collaborative Learning paradigm for generalized PDG.

Rather than treating the ground-truth response as a unique target, we use it only as a reference for preference construction, allowing MCL to favor candidates that better preserve coherence, diversity, and persona consistency instead of imitating a single lexical form. Therefore, a reward function $\mathcal{R}(\cdot)$ is employed to evaluate the generated responses along multiple dimensions—such as coherence, diversity, and persona consistency—yielding a scalar reward that reflects the overall response quality:

$$r^j = \mathcal{R}(y^j). \tag{3}$$

These reward signals are then used to infer relative preferences among candidate contexts and responses, providing the supervision necessary for preference-based learning.

These two modules are optimized using a preference alignment framework inspired by Kahneman–Tversky Optimization (KTO) (Ethayarajh et al., 2024), which converts relative rewards into likelihood-based updates rather than external heuristic ranking. For each user u, we identify the candidate with the highest scalar reward r^j as the positive sample (C_u^+, y^+) and the one with the lowest scalar reward as the negative sample (C_u^-, y^-). The loss functions are defined as follows:

$$\mathcal{L}_M = -\frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}}\Big[\lambda_M^+ \log\sigma(\beta_M\Delta_M^+ - z_M) + \lambda_M^- \log\sigma(z_M - \beta_M\Delta_M^-)\Big], M \in \{\Theta_B, \Theta_G\}. \tag{4}$$

Here, $(x_B, o_B) = (H_u, P_u, \mathcal{N}_u; C_u)$ and $(x_G, o_G) = (H_u, P_u, C_u; y)$, where o_B is an enhanced context and o_G is a response. We define $\Delta_M^\pm = \log\pi_M(o_M^\pm|x_M) - \log\pi_M^{\text{ref}}(o_M^\pm|x_M)$, where $\log\pi_M(o_M|x_M) = |o_M|^{-1}\sum_t \log\pi_M(o_{M,t}|x_M, o_{M,<t})$ is the length-normalized sequence log-probability. $o_M^\pm$ denotes the desirable/undesirable context or response selected by $\mathcal{R}(\cdot)$. π_M^{ref} is initialized from the same backbone and kept frozen, while β_M, z_M, and $\lambda_M^\pm$ denote the update scale, KL baseline, and sample-type weights, respectively.

Iterative Co-training Optimization The Booster and Generator are trained alternately under an iterative optimization scheme. At each iteration step k, the procedure unfolds as follows:

Booster Update. Freeze the current Generator $\Theta_G^{(k)}$. The Booster $\Theta_B^{(k)}$ produces a set of context candidates $\{C^j\}$, which are evaluated through $\Theta_G^{(k)}$'s outputs and the reward function $\mathcal{R}(\cdot)$. The Booster is then updated as:

$$\Theta_B^{(k+1)} = \arg\min_{\Theta_B} \mathcal{L}_{\Theta_B}(\Theta_B^{(k)}; \Theta_G^{(k)}). \tag{5}$$

Generator Update. Freeze the optimized Booster $\Theta_B^{(k+1)}$. The Generator $\Theta_G^{(k)}$ generates response candidates $\{y^j\}$ conditioned on $\Theta_B^{(k+1)}$'s enhanced context. Upon receiving the reward signals, the Generator is updated as:

$$\Theta_G^{(k+1)} = \arg\min_{\Theta_G} \mathcal{L}_{\Theta_G}(\Theta_G^{(k)}; \Theta_B^{(k+1)}). \tag{6}$$

This iterative optimization process, summarized in Algorithm 1, enables the model to progressively internalize transferable knowledge from similar users, refine contextual representations, and produce coherent, user-tailored responses, thus yielding robust generalization across unseen users and cross-domain deployments.

5 Experiments

5.1 Experimental Setup

Datasets. ConvAI2 (Dinan et al., 2019) is a high-quality dialogue benchmark built on the

PERSONA-CHAT dataset (Zhang et al., 2018). Synthetic-Persona-Chat (SPC) (Jandaghi et al., 2023) is created using the Generator-Critic framework and an instruction-tuned LLM. As standard experimental settings for evaluating user-aware conversational models, both datasets contain conversations between crowd workers who are each assigned a distinct persona, described by at least five self-reported sentences.

Baselines. We compare our proposed MCL-based method against several popular state-of-the-art approaches for persona-based personalized dialogue generation. **CLV** clusters textual persona into sparse categories and integrates them with history via a conditional variational autoencoder for generation. **SPT** maintains a pool of soft prompts and uses a learned retriever to dynamically select appropriate ones for each conversational context, trained via contrastive and fusion losses to encourage diversity and avoid repetitive generic responses. **MORPHEUS** learns a persona codebook in latent space for role representations and constructs a posterior distribution of role embeddings from dialogue history, enabling response personalization for roles with masked personas. **LaMP PEFT-RAG** integrates retrieval models with a personalized LLM to perform RAG personalization with PEFT. **PLoRA** combines the PKI and LoRA to inject personalized information into LLMs. Evaluation metrics and implementation details are included in Appendix B.

5.2 Main Results

Table 1 summarizes the performance of all methods under both intra-domain and cross-domain generalization settings. Overall, our MCL-based method achieves the most consistent results across coherence, diversity, and persona consistency. Averaged over all baselines and task settings, MCL brings an overall relative improvement of 47.9%, with gains of 41.56% under intra-domain generalization and 54.24% under cross-domain generalization. These results demonstrate its ability to produce fluent and personalized responses for both unseen users and shifted domains.

In the intra-domain setting, baseline methods still suffer from user-level data sparsity, as their personalization mechanisms largely rely on static profiles, retrieved histories, or user-specific parameters. By contrast, MCL dynamically incorporates collaborative signals from similar users and refines them into task-aware contexts, enabling more reliable personalization without additional user-specific training.

In the cross-domain setting, the advantage of MCL becomes even more evident. Existing methods tend to rely on domain-dependent personalization patterns and are therefore less robust under distribution shifts. Comparing the CDG results with the intra-domain setting in Table 1, baseline methods experience severe performance degradation due to domain shifts, whereas MCL maintains strong and stable personalization performance.

These results demonstrate that MCL successfully mitigates cross-domain discrepancies by learning atomic personalization skills, enabling robust deployment without domain-specific retraining. We conduct a human evaluation to complement the automated metrics, which is included in Appendix C.

5.3 Ablation and In-depth Analysis

Component ablation. We first examine the contribution of each component in MCL. As shown in Table 2, MCL w/o Booster exhibits notable declines in coherence and diversity under domain shift, indicating that the Booster plays a central role in refining and reweighting contextual personalization cues to ensure transferability across domains. These results align with our motivation that generalized PDG requires learning *how to exploit contextual signals*, rather than relying on domain- or user-specific memorization. w/o Generator[2] leads to pronounced degradation in absolute dialogue quality, particularly in coherence. This confirms that the Generator is indispensable for grounding responses in fluent and persona-consistent language. In contrast, the full MCL paradigm achieves superior performance in **92%** of metric-task combinations, demonstrating synergistic effects. The sole exception occurs in CDG's DIST-1, suggesting the Generator's persona-filtering may occasionally over-constrain lexical diversity during domain transfer. Overall, the Booster is crucial for acquiring and distilling atomic personalization skills that generalize across users and domains, while the Generator ensures coherent and persona-aligned realization of these skills.

Effect of Neighbor Identifier. To examine the robustness of our paradigm with respect to the retrieval backbone used in the Neighbor Identifier, we

[2] In the "w/o Generator" ablation, we directly incorporate the Booster's enhanced context into a vanilla LLM for generation, *i.e.*, degrading to generic generation.

Source → Target	Method	Coherence		Diversity		Persona-Consistency	AVG$_\uparrow$
		BLEU-1	ROUGE-L	DIST-1	DIST-2	P-Co	
Intra-domain Generalization (IDG)							
ConvAI2 → ConvAI2	CLV	11.85	11.30	27.42	68.71	14.67	37.00%
	MORPHEUS	12.67	15.18	20.47	64.42	18.64	31.25%
	SPT	4.42	6.56	20.56	49.05	12.09	131.32%
	LaMP PEFT-RAG	9.99	13.84	30.83	71.83	17.70	27.98%
	PLoRA	8.66	14.72	30.90	73.19	17.89	31.06%
	MCL	**17.13**	**16.98**	**34.71**	**79.01**	**21.80**	-
SPC → SPC	CLV	13.69	15.44	19.53	53.59	12.45	46.42%
	MORPHEUS	12.08	15.69	24.48	**67.32**	14.39	34.05%
	SPT	16.11	25.84	21.99	51.37	10.86	29.80%
	LaMP PEFT-RAG	11.52	21.96	21.66	50.08	14.82	34.56%
	PLoRA	15.49	26.70	22.74	56.11	14.63	18.18%
	MCL	**17.54**	**27.97**	**30.33**	66.23	**17.92**	-
Cross-domain Generalization (CDG)							
ConvAI2 → SPC	CLV	10.99	10.27	25.57	68.02	11.37	51.35%
	MORPHEUS	7.97	8.28	22.83	59.38	12.17	77.42%
	SPT	6.06	10.15	24.84	50.68	9.20	92.18%
	LaMP PEFT-RAG	9.87	12.11	27.30	68.02	17.04	35.78%
	PLoRA	11.40	17.70	28.92	**68.78**	17.50	17.11%
	MCL	**16.98**	**23.10**	**29.52**	67.93	**18.42**	-
SPC → ConvAI2	CLV	11.35	7.52	18.90	52.06	11.93	72.88%
	MORPHEUS	8.06	13.13	17.68	62.29	12.37	60.14%
	SPT	11.97	**22.52**	24.84	50.68	11.20	32.46%
	LaMP PEFT-RAG	6.53	13.53	26.46	56.22	13.76	56.74%
	PLoRA	6.75	14.38	29.15	69.52	13.66	46.36%
	MCL	**16.95**	18.95	**31.89**	**73.47**	**18.28**	-

Table 1: Performance comparison of baselines and Our MCL-based (blue) method on two tasks. Best results are in **bold**, sub-optimal ones are underlined. **AVG$_\uparrow$** denotes the average relative improvement of MCL over each corresponding baseline across the five metrics.

compare three representative methods: a lexical-based retriever (BM25), a sparse neural retriever (Splade-Index), and a dense embedding model (GTE). As reported in Table 2, across both settings, all variants achieve comparable performance across metrics. Notably, although retrieval quality affects the informativeness of collaborative signals, the relative downstream differences among retrievers remain marginal, and no single method consistently dominates across all evaluation dimensions. This suggests that MCL is not merely driven by raw neighbor retrieval quality; instead, its collaborative learning and dynamic refinement components can mitigate moderate retrieval noise, enabling stable performance across sparse, lexical, and dense retrieval backbones.

Effect of Reward Function. To further verify the effectiveness of our multi-objective reward design in the iterative co-training optimization process, we conduct an ablation study by isolating each reward component. Specifically, we compare three single-objective variants, *i.e.*, optimizing only for Coherence, Diversity, or Persona-Consistency, against our full combination that jointly balances all three. As reported in Table 2, single-objective optimization often leads to trade-offs across dimensions. For example, optimizing only coherence improves BLEU and ROUGE scores but yields less diverse or less persona-aligned responses. Similarly, diversity-oriented optimization enhances lexical variation at the cost of coherence and consistency. In contrast, our unified reward formulation consistently achieves a better balance across all evaluation aspects, with notable gains in overall fluency, richness, and personalization. These findings highlight the necessity of a holistic reward signal to guide the model towards well-rounded improvements. By jointly considering several aspects, our iterative training framework effectively harmonizes competing objectives, resulting in more natural and user-aligned dialogue generation under generalized PDG challenges.

Robustness to imperfect neighbor information. We further evaluate MCL under information flaws, including incomplete, noisy, and misleading neighbor signals. Results in Appendix D.2 show that

Source → Target	Variant	Coherence		Diversity		Persona-Consistency
		BLEU-1	ROUGE-L	DIST-1	DIST-2	P-Co
Component Ablation						
ConvAI2 → ConvAI2	MCL	**17.13**	**16.98**	**34.71**	**79.01**	**21.80**
	w/o Generator	5.91	12.05	32.12	72.54	19.43
	w/o Booster	16.26	15.96	31.58	77.79	20.04
ConvAI2 → SPC	MCL	**16.98**	**23.10**	**29.52**	**67.93**	18.42
	w/o Generator	10.22	16.71	29.19	65.57	**18.48**
	w/o Booster	14.23	20.03	27.49	66.21	17.93
Neighbor Identifier						
ConvAI2 → ConvAI2	BM25	16.55	16.10	34.02	79.22	21.57
	Splade-Index	**17.78**	**17.63**	34.17	78.69	21.22
	GTE	17.13	16.98	**34.71**	79.01	**21.80**
ConvAI2 → SPC	BM25	18.02	23.50	29.67	66.67	18.30
	Splade-Index	**18.75**	**24.68**	**29.88**	67.88	18.28
	GTE	16.98	23.10	29.52	**67.93**	**18.42**
Reward Function						
ConvAI2 → ConvAI2	Coherence-only	16.34	16.76	32.75	77.18	19.76
	Diversity-only	16.59	16.13	33.25	76.77	19.52
	Persona-only	16.70	15.67	33.21	76.58	19.83
	Ours	**17.13**	**16.98**	**34.71**	**79.01**	**21.80**
ConvAI2 → SPC	Coherence-only	16.89	22.38	28.89	67.72	16.97
	Diversity-only	17.43	23.24	**30.37**	**69.08**	15.57
	Persona-only	**19.08**	**24.76**	28.12	66.99	17.45
	Ours	16.98	23.10	29.52	67.93	**18.42**

Table 2: Ablation and in-depth analysis of MCL. The table studies component ablation, retrieval backbones for the Neighbor Identifier, and reward designs.

MCL maintains stable gains over baselines, suggesting that the Booster can selectively distill useful personalization cues rather than relying on perfectly matched neighbors.

6 Conclusion

In this work, we explore and tackle the critical challenges of generalization in personalized dialogue generation, where models must generalize to novel users and domains without explicit supervision. We introduce **Meta-Collaborative Learning (MCL)**, a novel learning paradigm that integrates neighbor identification, collaborative context enrichment, and preference-aligned response generation. By decomposing the personalization process into these synergistic components and optimizing them through an iterative co-training scheme, MCL enables dialogue models to generalize effectively under compound conditions involving unseen users and domain shifts. Comprehensive experiments demonstrate that MCL consistently enhances personalization fidelity and robustness, outperforming competitive baselines across all metrics.

Limitations

While the findings highlight the promise of our approach, several directions remain open for future research: (1) As the first study to extend DG rationale to personalized dialogue, MCL prioritizes robustness and generalization for realistic challenges over minimal training complexity: although inference involves only a single forward pass with latency comparable to standard PEFT-based deployments, the iterative optimization and trial-based procedures introduce offline training cost. This design reflects a common trade-off in personalization systems, where increased upfront training is indispensable for efficient and stable deployment. Importantly, this limitation does not detract from the empirical advantages demonstrated by MCL, but instead highlights a natural avenue for future exploration, such as distilling the modules for scalable deployment. (2) While our study focuses on the generalizability of PDG systems, broader concerns such as biases embedded in public datasets remain outside the scope of the current study. Extending our work to explicitly mitigate such biases in PDG represents a potential step toward more fair, reliable, and trustworthy NLP applications.

References

Anonymous. 2026. Diper: Disentangled personal direction discovery for personalized generation steering. *Under review*.

Junbum Cha, Sanghyuk Chun, Kyungjae Lee, Han-Cheol Cho, Seunghyun Park, Yunsung Lee, and Sungrae Park. 2021. Swad: Domain generalization by seeking flat minima. *Preprint*, arXiv:2102.08604.

Sadia Sultana Chowa, Riasad Alvi, Subhey Sadi Rahman, Md Abdur Rahman, Mohaimenul Azam Khan Raiaan, Md Rafiqul Islam, Mukhtar Hussain, and Sami Azam. 2026. From language to action: a review of large language models as autonomous agents and tool users. *Artificial Intelligence Review*, 59(2).

Emily Dinan, Varvara Logacheva, Valentin Malykh, Alexander Miller, Kurt Shuster, Jack Urbanek, Douwe Kiela, Arthur Szlam, Iulian Serban, Ryan Lowe, Shrimai Prabhumoye, Alan W Black, Alexander Rudnicky, Jason Williams, Joelle Pineau, Mikhail Burtsev, and Jason Weston. 2019. The second conversational intelligence challenge (convai2). *Preprint*, arXiv:1902.00098.

Kawin Ethayarajh, Winnie Xu, Niklas Muennighoff, Dan Jurafsky, and Douwe Kiela. 2024. Kto: Model alignment as prospect theoretic optimization. *Preprint*, arXiv:2402.01306.

Nicklas Hansen and Xiaolong Wang. 2021. Generalization in reinforcement learning by soft data augmentation. *Preprint*, arXiv:2011.13389.

Qiushi Huang, Shuai Fu, Xubo Liu, Wenwu Wang, Tom Ko, Yu Zhang, and Lilian Tang. 2023. Learning retrieval augmentation for personalized dialogue generation. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 2523–2540, Singapore. Association for Computational Linguistics.

Qiushi Huang, Xubo Liu, Tom Ko, Bo Wu, Wenwu Wang, Yu Zhang, and Lilian Tang. 2024. Selective prompting tuning for personalized conversations with LLMs. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 16212–16226, Bangkok, Thailand. Association for Computational Linguistics.

Maximilian Ilse, Jakub M. Tomczak, Christos Louizos, and Max Welling. 2019. Diva: Domain invariant variational autoencoders. *Preprint*, arXiv:1905.10427.

Pegah Jandaghi, XiangHai Sheng, Xinyi Bai, Jay Pujara, and Hakim Sidahmed. 2023. Faithful persona-based conversational dataset generation with large language models. *Preprint*, arXiv:2312.10007.

Hannah Rose Kirk, Andrew M. Bean, Bertie Vidgen, Paul Röttger, and Scott A. Hale. 2023. The past, present and better future of feedback learning in large language models for subjective human preferences and values. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 2409–2430, Singapore. Association for Computational Linguistics.

Kimin Lee, Kibok Lee, Jinwoo Shin, and Honglak Lee. 2020. Network randomization: A simple technique for generalization in deep reinforcement learning. *Preprint*, arXiv:1910.05396.

Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2016. 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.

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.

Jianxin Lin, Yongqiang Tang, Junping Wang, and Wensheng Zhang. 2022. Constrained maximum cross-domain likelihood for domain generalization. *Preprint*, arXiv:2210.04155.

Jiahong Liu, Zexuan Qiu, Zhongyang Li, Quanyu Dai, Wenhao Yu, Jieming Zhu, Minda Hu, Menglin Yang, Tat-Seng Chua, and Irwin King. 2025. A survey of personalized large language models: Progress and future directions. *Preprint*, arXiv:2502.11528.

Yifan Liu, Wei Wei, Jiayi Liu, Xianling Mao, Rui Fang, and Dangyang Chen. 2022. Improving personality consistency in conversation by persona extending. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, CIKM '22, page 1350–1359. ACM.

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.

Vihari Piratla, Praneeth Netrapalli, and Sunita Sarawagi. 2020. Efficient domain generalization via common-specific low-rank decomposition. *Preprint*, arXiv:2003.12815.

Alireza Salemi, Sheshera Mysore, Michael Bendersky, and Hamed Zamani. 2024. LaMP: When large language models meet personalization. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7370–7392, Bangkok, Thailand. Association for Computational Linguistics.

Alireza Salemi and Hamed Zamani. 2025. Comparing retrieval-augmentation and parameter-efficient fine-tuning for privacy-preserving personalization of large language models. *Preprint*, arXiv:2409.09510.

Shiv Shankar, Vihari Piratla, Soumen Chakrabarti, Siddhartha Chaudhuri, Preethi Jyothi, and Sunita Sarawagi. 2018. Generalizing across domains via cross-gradient training. *Preprint*, arXiv:1804.10745.

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, Wei-Nan Zhang, Yiming Cui, Dong Wang, and Ting Liu. 2019. Exploiting persona information for diverse generation of conversational responses. *Preprint*, arXiv:1905.12188.

Yihong Tang, Bo Wang, Miao Fang, Dongming Zhao, Kun Huang, Ruifang He, and Yuexian Hou. 2023. Enhancing personalized dialogue generation with contrastive latent variables: Combining sparse and dense persona. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5456–5468, Toronto, Canada. Association for Computational Linguistics.

Yihong Tang, Bo Wang, Dongming Zhao, Xiaojia Jin, Jijun Zhang, Ruifang He, and Yuexian Hou. 2024. Morpheus: Modeling role from personalized dialogue history by exploring and utilizing latent space. *ArXiv*, abs/2407.02345.

Jindong Wang, Cuiling Lan, Chang Liu, Yidong Ouyang, and Tao Qin. 2021. Generalizing to unseen domains: A survey on domain generalization. *IEEE Transactions on Knowledge and Data Engineering*, 35:8052–8072.

Stanisław Woźniak, Bartłomiej Koptyra, Arkadiusz Janz, Przemysław Kazienko, and Jan Kocoń. 2024. Personalized large language models. *Preprint*, arXiv:2402.09269.

Kai Zhang, Yejin Kim, and Xiaozhong Liu. 2025. Personalized llm response generation with parameterized memory injection. *Preprint*, arXiv:2404.03565.

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? *Preprint*, arXiv:1801.07243.

You Zhang, Jin Wang, Liang-Chih Yu, Dan Xu, and Xuejie Zhang. 2024a. Personalized lora for human-centered text understanding. *Preprint*, arXiv:2403.06208.

Zhehao Zhang, Ryan A. Rossi, Branislav Kveton, Yijia Shao, Diyi Yang, Hamed Zamani, Franck Dernoncourt, Joe Barrow, Tong Yu, Sungchul Kim, Ruiyi Zhang, Jiuxiang Gu, Tyler Derr, Hongjie Chen, Ju-Ying Wu, Xiang Chen, Zichao Wang, Subrata Mitra, Nedim Lipka, and 2 others. 2024b. Personalization of large language models: A survey. *ArXiv*, abs/2411.00027.

Wanjun Zhong, Duyu Tang, Jiahai Wang, Jian Yin, and Nan Duan. 2021. UserAdapter: Few-shot user learning in sentiment analysis. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1484–1488, Online. Association for Computational Linguistics.

Kaiyang Zhou, Ziwei Liu, Yu Qiao, Tao Xiang, and Chen Change Loy. 2022. Domain generalization: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, page 1–20.

Kaiyang Zhou, Chen Change Loy, and Ziwei Liu. 2021. Semi-supervised domain generalization with stochastic stylematch. *Preprint*, arXiv:2106.00592.

Yuchen Zhuang, Haotian Sun, Yue Yu, Rushi Qiang, Qifan Wang, Chao Zhang, and Bo Dai. 2024. Hydra: Model factorization framework for black-box llm personalization. *Preprint*, arXiv:2406.02888.

A The Pseudo-code of MCL

As shown in Algorithm 1, the Booster and Generator are trained alternately under an iterative optimization scheme.

B Additional Experiment Setup

Implementation Details. Since the official ConvAI2 test set is not publicly available, we construct a held-out test set from the original training corpus under a strict persona-level partition. Specifically, we first group all dialogues by persona/user identity. Then, users are partitioned into train/validation/test splits with no shared persona identity across splits. All dialogues belonging to the same user are assigned to only one split. The Neighbor Identifier is built only over the source-side training users, and no target test dialogue or target test response is indexed during training or retrieval. We repeat the main experiments over five random seeds and observe negligible variance, with maximum standard deviation below 0.5 across all reported metrics.

General Text Embeddings (GTE) is employed as the backbone of Neighbor Identifier to figure out semantically similar users, while Booster and Generator are instantiated from Llama-3.1-8B-Instruct[3]. We set topk=2 neighbors per user and J=3 in all experiments, which is the minimum number of

[3] https://huggingface.co/meta-llama/Llama-3.1-8B-Instruct

Algorithm 1 Iterative Co-training Optimization in MCL

Require: User set $\mathcal{U}$, Neighbor Identifier Θ_N, initial Booster $\Theta_B^{(0)}$, initial Generator $\Theta_G^{(0)}$, reward function $\mathcal{R}(\cdot)$, number of candidates J

for each iteration k **do**
  Sample a user mini-batch $\mathcal{U}_k \subseteq \mathcal{U}$
  (1) Booster Update
  for each user $u \in \mathcal{U}_k$ **do**
    Retrieve neighbor set $\mathcal{N}_u = \Theta_N(H_u, P_u; \mathcal{U})$
    for $j = 1$ to J **do**
      Generate enhanced context C_u^j using $\Theta_B^{(k)}$ ▷ Eq. 1
      Generate response y^j using $\Theta_G^{(k)}$ ▷ Eq. 2
      Compute reward r^j ▷ Eq. 3
    end for
    Select C_u^+, C_u^- with highest and lowest r^j
  end for
  Compute $\mathcal{L}_B$ ▷ Eq. 4
  Update Booster parameters $\Theta_B^{(k+1)}$ ▷ Eq. 5
  (2) Generator Update
  for each user $u \in \mathcal{U}_k$ **do**
    Obtain the enhanced context C_u^1 using $\Theta_B^{(k+1)}$ ▷ Eq. 1
    for $j = 1$ to J **do**
      Generate response y^j using $\Theta_G^{(k)}$ ▷ Eq. 2
      Compute reward r^j ▷ Eq. 3
    end for
    Select y^+, y^- with highest and lowest r^j
  end for
  Compute $\mathcal{L}_G$ ▷ Eq. 4
  Update Generator parameters $\Theta_G^{(k+1)}$ ▷ Eq. 6
  $k \leftarrow k + 1$
end for
return optimized Booster $\Theta_B^{(*)}$ and Generator $\Theta_G^{(*)}$

stochastic trials required to obtain a reliable best-worst reward signal while keeping computation efficient. This choice balances training cost and signal quality, providing stable supervision for modules without excessive forward passes. LaMP PEFT-RAG is implemented by first performing PEFT on source data to learn a global personalized LLM, then retrieving relevant information from each target user's own dialogue history as context for generation. The reward function $\mathcal{R}(\cdot)$ is defined as the normalized average across all five automatic metrics introduced above. The best–worst pair of outcomes is treated as a holistic reward signal to guide the components towards well-rounded improvements.

The training of the proposed method is conducted using 2 NVIDIA A40 GPUs, which takes approximately 8.5 hours. We set the temperature to 0.1 for the final prediction to ensure stability, and 0.5 for candidate generation to promote output diversity during exploratory sampling. Consequently, only *a single forward pass* is involved during inference, with negligible inference latency (up to 210ms).

MCL retrieves neighboring users only to provide collaborative personalization signals, rather than to copy user-specific content. All experiments are conducted on publicly available crowdsourced datasets with anonymized users. In practical deployments, the retrieval index can be built with standard sanitization, anonymization, and access control, and the Booster can operate on aggregated or k-anonymized neighbor representations.

Evaluation Metrics. We adopt a suite of established metrics for persona-based personalized dialogue generation, which are categorized into the following three dimensions:

- **Coherence**: BLEU-1 (Papineni et al., 2002) and ROUGE-L (Lin, 2004) are employed to measure lexical overlap between the generated response and the reference response.
- **Diversity**: We compute the proportion of unique unigrams and bigrams to assess the richness of generated utterances, denoted as DIST-1 and DIST-2, respectively (Li et al., 2016).
- **Persona Consistency**: The IDF-weighted word-overlap metric Persona Cosine Comparison (P-Co) (Song et al., 2019) evaluates the semantic similarity between the model's output and user personas.

C Human Evaluation

We further conduct human evaluation to validate whether the learned preferences translate into perceived improvements. We recruit three qualified annotators and randomly sample a total of 300 dialogue instances from the four formulated generalized PDG tasks. Each annotator evaluates the responses from our MCL and the strongest baseline (PLoRA) across two key dimensions using a 0-2 Likert scale: persona-consistency (Per-Con: how well the response aligns with the given user profile) and coherence (how naturally it follows the dialogue history).

The outcomes are presented in Table 3. The Fleiss' Kappa score 0.63, which indicates substantial inter-annotator agreement, confirms that our automatic metric gains correlate reliably with human judgment, and strengthens the validity of our main conclusions.

Source → Target	Method	Per-Con	Coherence
ConvAI2 → ConvAI2	MCL	1.74	1.86
	PLoRA	1.70	1.62
ConvAI2 → SPC	MCL	1.60	1.75
	PLoRA	1.46	1.53

Table 3: Results of human evaluation.

Dataset	BLEU-1	ROUGE-L	DIST-1	DIST-2	P-Co
ConvAI2	+1.9%	+3.6%	+0.3%	+1.8%	+2.5%
SPC	+2.1%	+1.8%	+0.7%	+2.2%	+3.3%

Table 4: Relative improvements of MCL over the strongest competing baseline PLoRA on seen users.

D Additional Robustness Analysis

D.1 Performance on Seen Users

Although our main experiments focus on generalized PDG scenarios involving unseen users and cross-domain generalization, practical systems must also serve users whose data has been observed during training. Table 4 shows that MCL remains consistently effective in this setting. The improvements are observed on both ConvAI2 and SPC, covering persona consistency, lexical overlap, semantic overlap, and response diversity metrics. These results suggest that the gains of MCL are not limited to cold-start or transfer-oriented scenarios. Rather, the collaborative optimization among the Neighbor Identifier, Booster, and Generator also preserves in-distribution personalization ability, making MCL applicable to deployment settings involving both existing and new users.

D.2 Robustness Under Information Flaws

In practical PDG deployment, user profiles may be incomplete, outdated, or expressed in different surface forms, especially when they are derived from static self-reports (Liu et al., 2025; Jandaghi et al., 2023). Such imperfect profile signals make generalized personalization more challenging, as models must infer stable user preferences from noisy or partially reliable evidence.

To examine this issue, we construct perturbed profile settings by paraphrasing persona descriptions[4]. This introduces lexical and syntactic variations while preserving the underlying user attributes, allowing us to test whether models can remain robust when explicit persona signals become less reliable. We evaluate both intra-domain and cross-domain generalization under these information flaws.

Table 5 reports results when either the source-side profiles are perturbed. Compared with baselines, MCL consistently maintains stronger coherence, diversity, and persona consistency across both intra-domain and cross-domain settings. This suggests that MCL does not rely on exact surface-form matching of persona descriptions. Instead, the Booster helps distill transferable personalization cues from imperfect contextual evidence, making the framework more robust to moderate information flaws.

D.3 Dual-Side Perturbation

We further consider a more challenging setting where both seen users and unseen users contain perturbed persona information. This setting tests whether MCL remains effective when collaborative references and target-user evidence are simultaneously imperfect, which more closely reflects noisy deployment conditions.

As shown in Table 6, MCL remains consistently superior even when both collaborative references and target-user profiles are perturbed. Although some baselines perform well on isolated metrics, they fail to maintain balanced performance across coherence, diversity, and persona consistency. In contrast, MCL achieves stable gains across all dimensions, with average improvements ranging from **47.84%** to **155.68%** over baselines. These results further support that the Booster can filter imperfect neighbor information and extract useful personalization signals without relying on perfectly matched or noise-free profiles.

[4]Specifically, persona descriptions are rewritten via rephrasing, introducing lexical and syntactic variations without altering underlying user attributes.

Source → Target	Method	Coherence		Diversity		Persona-Consistency	AVG$_\uparrow$
		BLEU-1	ROUGE-L	DIST-1	DIST-2	P-Co	
ConvAI2* → ConvAI2	CLV	10.28	9.24	30.29	68.53	11.64	52.42%
	MORPHEUS	4.82	9.85	20.28	61.34	12.06	100.43%
	SPT	4.14	6.09	20.90	49.72	10.18	144.07%
	LaMP PEFT-RAG	9.28	12.93	26.40	70.67	17.08	36.31%
	PLoRA	8.07	14.32	29.34	70.78	17.46	36.21%
	MCL	**17.09**	**16.02**	**34.33**	**77.93**	**22.75**	-
SPC* → SPC	CLV	11.66	11.07	21.75	58.17	12.17	54.83%
	MORPHEUS	10.94	14.13	20.91	56.49	13.08	46.40%
	SPT	14.28	23.37	22.34	50.87	9.22	37.09%
	LaMP PEFT-RAG	13.37	23.16	22.58	53.78	14.35	23.28%
	PLoRA	15.34	26.56	20.66	54.61	14.38	19.24%
	MCL	**16.82**	**25.68**	**29.82**	**65.87**	**17.96**	-
ConvAI2* → SPC	CLV	8.91	9.69	28.29	**68.01**	10.22	54.60%
	MORPHEUS	7.03	8.81	17.09	48.01	11.83	85.32%
	SPT	11.97	**22.47**	21.38	48.84	9.23	38.95%
	LaMP PEFT-RAG	10.48	10.09	25.72	64.52	15.89	38.07%
	PLoRA	10.31	18.05	26.31	61.61	15.41	21.33%
	MCL	**15.83**	21.07	**29.07**	67.06	**18.04**	-

Table 5: Robustness under information flaws. ConvAI2* and SPC* respectively denote datasets with perturbed profiles.

Source → Target	Method	Coherence		Diversity		Persona-Consistency	AVG$_\uparrow$
		BLEU-1	ROUGE-L	DIST-1	DIST-2	P-Co	
ConvAI2* → ConvAI2*	CLV	9.28	9.03	20.59	58.53	10.25	79.52%
	MORPHEUS	4.34	7.64	16.20	49.67	10.69	139.41%
	SPT	4.09	6.12	20.67	47.57	8.93	155.68%
	LaMP PEFT-RAG	12.33	12.48	20.74	54.09	15.00	47.84%
	PLoRA	7.22	12.74	28.75	61.50	16.86	50.38%
	MCL	**16.91**	**15.95**	**34.38**	**79.87**	**24.12**	-

Table 6: **Performance comparison when both seen-user and unseen-user personas are perturbed.** ConvAI2* denotes ConvAI2 with information flaws. Best results are in **bold**, sub-optimal ones are underlined. **AVG**$_\uparrow$ indicates the average improvement of our method over the corresponding baselines.