

Extended Abstract

Motivation Large language models (LLMs) hold promise as conversational tutors for English language learners due to their fluency and contextual sensitivity. However, most existing systems remain static and reactive, lacking the ability to dynamically personalize instruction based on learners’ evolving traits, interests, and goals. This gap is particularly pronounced in second language acquisition, which is inherently social, open-ended, and shaped by individual usage contexts. To address these challenges, our project introduces a novel English-as-a-Foreign-Language (EFL) chatbot that integrates pedagogical personalization with dynamic adaptation.

Method We implement a two-stage training framework to achieve dynamic, personalized tutoring. In the first stage, we design structured learner personas represented as JSON dictionaries encompassing features grounded in SLA theories. We synthesize over 60K persona-grounded learner queries through a realistic and diverse data synthesis pipeline. Preference pairs are generated by conditioning GPT models on personas and pedagogical strategies, resulting in contrastive examples labeled for response quality and appropriateness. We then fine-tune the chatbot using Simple Preference Optimization (SimPO), aligning it to pedagogically preferred responses. In the second stage, we enhance personalization through Proximal Policy Optimization (PPO) with a curiosity-driven intrinsic reward. This reward quantifies the information gain about the learner’s persona by measuring improvements in predicted persona accuracy over the course of multi-turn conversations. Evaluation includes both automated win-rate comparisons and LLM-based assessments of dialogue personalization, engagement, and accuracy of inferred personas.

Implementation We use Qwen2.5-1.5B-Instruct as the base model for all stages, applying LoRA for parameter-efficient fine-tuning throughout. The SFT stage was trained on approximately 4.5K real and synthetic dialogues, while SimPO fine-tuning used 10K GPT-generated preference pairs with controlled length and persona consistency. PPO training involved 500 multi-turn episodes with simulated student responses. To balance realism and variability, five candidate replies were generated per turn, with one sampled randomly during rollouts. Our curiosity reward was designed to measure the gain in inferred persona attributes (rather than general uncertainty), computed by comparing GPT-inferred persona fields against the gold version after each turn. This field-wise design encourages proactive trait elicitation without relying on continuous embedding similarity. We combined this curiosity reward with a preference reward computed via a frozen reward model fine-tuned on the filtered SimPO dataset.

Results Quantitatively, SimPO significantly improves pedagogical quality, engagement, and relevance over the SFT baseline, achieving win rates of 79.3%, 85.3%, and 75.6%, respectively. PPO further enhances persona elicitation, consistently achieving higher curiosity rewards across multi-turn rollouts. However, PPO training showed instability in combined validation rewards and declining preference rewards, suggesting limitations. Qualitatively, SimPO responses are warmer and more engaging, often mimicking a peer-like tutor that fosters social connection. PPO responses diverge more often from the topic but tend to elicit richer disclosures from learners.

Discussion Our results demonstrate that SimPO and PPO enable stronger personalization and pedagogical alignment than SFT, but also expose key trade-offs. SimPO excels at targeted, empathetic responses, though it occasionally over-relies on demographic cues (e.g., inferring nationality from L1). PPO introduces more divergent behavior that increases persona exposure but risks drifting from instructional goals. Preference reward reliability is hindered by surface-level distinctions in training pairs, and curiosity reward signals may be noisy when learner traits are revealed without explicit prompting. Future work should refine reward models and balance exploration with pedagogical consistency for long-term learning efficacy.

Conclusion We present a pedagogically aligned, curiosity-driven EFL chatbot that personalizes instruction based on learner personas. SimPO fine-tuning improves instructional quality, while PPO enables dynamic adaptation by encouraging the model to uncover latent learner traits. Our work demonstrates the complementary roles of preference optimization and intrinsic motivation in building engaging, adaptive language tutors, offering a scalable path toward personalized LLM-based education.

Personalized Pedagogically Aligned English Learning Chatbot via Preference Optimization and Curiosity

Ziqi Shu

Graduate School of Education
Stanford University
ziqishu@stanford.edu

Samantha Liu

Department of Computer Science
Stanford University
szwliu@stanford.edu

Abstract

Large language models (LLMs) offer significant promise for second-language learning, yet most existing systems lack both pedagogical grounding and dynamic personalization. We propose a novel English-as-a-Foreign-Language (EFL) chatbot that integrates structured learner modeling with a two-stage training pipeline. In Stage 1, we introduce a Simple Preference Optimization (SimPO) method to align the chatbot’s responses with pedagogically preferred strategies tailored to diverse learner personas. SimPO directly optimizes token-level likelihoods on GPT-generated preference pairs, avoiding the verbosity and inference mismatch issues of traditional DPO while achieving significant gains in instructional quality, relevance, and learner engagement over a supervised baseline. In Stage 2, we enhance multi-turn personalization using reinforcement learning with a curiosity-driven intrinsic reward that encourages proactive elicitation of user traits. This curiosity reward quantifies field-level improvements in persona inference accuracy, guiding the chatbot to uncover learner attributes while maintaining pedagogical alignment through Proximal Policy Optimization (PPO). Evaluation on simulated dialogues shows that SimPO outperforms supervised fine-tuning in pedagogy (79.3%), engagement (85.3%), and relevance (75.6%) win rates, while PPO further improves trait discovery and long-turn adaptability. Together, these methods bridge pedagogical fidelity and dynamic learner modeling, advancing the development of personalized, scalable LLM-based tutoring agents.

1 Introduction

Personalized learning has become a central paradigm in education, increasingly supported by the rise of large language models (LLMs). The ability of LLMs to generate context-aware and fluent responses has enabled scalable, interactive learning experiences tailored to individual needs. A growing body of research has explored aligning LLM outputs with human preferences, commonly through techniques such as reinforcement learning from human feedback (RLHF) or direct preference optimization (DPO). However, much of this work focuses on general preference alignment, optimizing for broadly desirable traits like helpfulness, informativeness, and safety.

Recently, these alignment techniques have been adapted for pedagogical alignment in educational contexts Sonkar et al. (2024), guiding models to generate responses that follow instructional principles—such as using step-by-step reasoning instead of directly providing answers. While this represents meaningful progress, current approaches often assume a one-size-fits-all instructional design, neglecting the essential principle of personalized learning: that pedagogical strategies should vary across learners based on their individual profiles. For example, what engages and supports a highly motivated learner aiming to improve academic writing may differ significantly from what benefits a casual learner focused on travel conversation skills.

Moreover, while LLM-based educational agents have been developed for structured domains like mathematics, biology, and history, language learning remains relatively underexplored. Unlike many subject areas with well-defined instructional goals and assessment pathways, second language acquisition is inherently social, open-ended, and highly variable across learners. Learners’ first language, motivation, usage contexts, and interests all shape their needs and responses to pedagogical strategies. Additionally, language learning frequently occurs through informal, communicative interaction rather than formal tutoring alone—highlighting the importance of immersion, engagement, and dynamic adaptation during casual conversations. A user study from Nie et al. (2023) demonstrated the potential of LLM chatbots to be engaging and confidence-boosting tools with long-term benefits, especially with adaptive design.

To address these gaps, this project introduces a conversational English-as-a-Foreign-Language (EFL) chatbot that delivers highly personalized pedagogical support across both instructional and casual interactions. The chatbot is designed to both answer explicit learning questions and engage in natural conversation to promote incidental learning and language immersion. In both settings, the chatbot dynamically selects and applies personalized pedagogical strategies based on the learner’s dialogue history and individual characteristics.

Our contributions are fourfold:

1. **Structured Learner Persona Schema:** We define learner personas with fixed fields (e.g., L1, goal, motivation, interests, learning style) based on second-language acquisition research, enabling fine-grained pedagogical adaptation.
2. **Pedagogically Grounded Synthetic Dialogue:** We construct a realistic and pedagogy-grounded data pipeline that generates realistic EFL dialogues across formal and informal scenarios, supporting context-aware personalization.
3. **SimPO for Single-Turn Personalization:** We fine-tune the chatbot using Simple Preference Optimization on persona-conditioned dialogues, improving pedagogical alignment at the single-turn level.
4. **Curiosity-Driven Multi-Turn Adaptation:** We introduce a curiosity reward based on inferred persona accuracy gain, encouraging the model to proactively elicit learner traits and adapt over time.

Through this work, we advance the capabilities of LLM-based educational agents in open-ended, language learning contexts—combining structured personalization with curiosity-driven engagement. This approach lays the groundwork for more adaptive, engaging, and learner-centered AI tutors in second-language education.

2 Related Work

Singh et al. (2025) introduced a meta-learning approach that uses few-shot preference pairs per user to condition generation and fine-tune reward functions. They employed a chain-of-thought (CoT) step to generate user descriptions from preference examples as guidance, and achieved strong results in domains such as movie reviews and role-played conversations. However, FSPO assumes the availability of clean, labeled preference pairs—an unrealistic constraint in most real-world educational scenarios. Moreover, using CoT to generate user descriptions might be loosely grounded and less controllable over trait definitions or relevance to specific downstream tasks. Complementary to FSPO, Casticato et al. (2024) offered a controlled framework for evaluating the influence of structured user personas on model behavior. The study constructed realistic and structured personas with fixed domains from U.S. Census data and generated preference comparisons conditioned on these traits. A leave-one-out evaluation strategy is used to quantify the effect of each persona attribute on model responses. This approach offers strong controllability and interpretability in general alignment research, but the personas designed might be limited in US sample and might be difficult to generalize to global users with diverse linguistic and cultural backgrounds.

In the broader educational domain, reinforcement learning has been applied to tutoring systems to optimize instructional policies. For example, Nie et al. (2023) modelled a fixed set of teaching actions in a math tutoring setting, using student-level features (e.g., gender, pre-test scores) to personalize actions and maximize learning gains. While promising, such systems rely on narrowly defined domains and small, manually designed state spaces, limiting their generalizability to more

open-ended tasks like language learning. Furthermore, the student features are limited to gender and test scores and math anxiety score, which might limit chances for deeper personalized pedagogy alignment. In addition, Sonkar et al. (2024) used human preference optimization to fine-tune the LLM for more pedagogically appropriate explanations in college-level biology. However, the study was designed with pedagogical alignment without considering the student-level features, and only considered general “ideal” pedagogy, such as breaking the problems down for scaffolding, akin to a CoT format. While such work represents an important step toward instructional usefulness, it stops short of personalized pedagogical alignment.

To support more dynamic personalization, researchers have proposed curiosity-driven reinforcement learning as a mechanism to elicit and adapt to user preferences during multi-turn dialogue. Wan et al. (2025) introduced a reward formulation based on reducing model uncertainty over latent user types, incentivizing the agent to ask exploratory questions and personalize behavior in real-time. While effective in educational contexts, the latent user types employed in this work are relatively coarse (e.g., “hands-on” vs. “story-based” learners). Such abstractions are insufficient for capturing the nuanced learner profiles needed in second-language acquisition, where learning style, motivation, usage context, and first language all meaningfully affect pedagogical strategy. Our work extends this approach by designing fine-grained, structured learner personas and redefining the curiosity reward as information gain over these richer profiles to gain more precise and actionable personalization.

In summary, prior research provides valuable foundations in preference optimization, persona modeling, and curiosity-based adaptation. Yet, limitations remain in applying these techniques to second-language learning—particularly in synthesizing domain-relevant learner personas, handling informal conversational contexts, and integrating dynamic personalization over time. Our work addresses these gaps by combining structured learner personas, pedagogically grounded preference optimization, and curiosity-driven reinforcement learning to enable richer, more adaptive English language tutoring.

3 Method

3.1 Task Design

To reflect the dual nature of English language acquisition, we designed two major task contexts: *casual communication* and *instructional learning*. This dual-context design enables evaluation of each modeling strategy’s ability to support both social interaction and structured language development—core dimensions of second language acquisition. Therefore, for each of the *casual communication* and *instructional learning* we design 4 sub-contexts as shown in Table 1.

Context	Topic 1	Topic 2	Topic 3	Topic 4
Casual Communication	Personal stories	Opinions	Interests	Future dreams
Instructional Learning	Grammar	Correction	Explanation	Practice

Table 1: Overview of task contexts and representative topics.

To ensure realistic coverage of learner needs, we synthesized learner queries by mapping learner personas to activity domains defined in the ALTE “Can Do” framework. This framework specifies 40 empirically grounded communicative activities across social, workplace, and academic domains (e.g., shopping, job interviews, ordering food; see Appendix B for the full list). Each activity is manually mapped to all applicable task types in Table 1 to ensure both realism and diversity. These context-activity mappings are then associated with each persona to form a persona–context–activity triplet to tailor the tasks and contexts to students’ personal background.

In addition, we designed an implicit task for **potential error correction** by injecting English errors into the synthesized student queries. The distribution of error types is derived from the EF Cambridge Open Language Database (EFCAMDAT, Geertzen et al. (2014); Huang et al. (2017)). EFCAMDAT consists of essays submitted by learners to an online English as a Foreign Language (EFL) platform, annotated with CEFR proficiency levels (ranging from A1 to C1) and error codes when applicable. The gained error code distribution is shown in Appendix C. This setup enables the model to detect and correct learner errors even when the student does not explicitly request correction.

3.2 Data Generation Pipeline

Our data generation pipeline supports a controlled, scalable, and persona-grounded approach to modeling and evaluation.

Persona Generation. We conducted detailed literature review and confirmed a set of relevant learner’s features in Second Language Acquisition, such as first language, proficiency level, language anxiety level. The persona schema is shown in Appendix A. We generated 50 diverse and realistic learner personas by defining possible values for each persona attribute (or key) and random sampling from them. For open-ended attributes, such as *interests* or *primary challenges*, we curate extensive and diverse value lists and allow multiple selections where appropriate. To ensure plausibility, we apply logical constraints—e.g., if the age group is *Child*, the industry is set to *Not in labor force*; if the age group is *Senior*, the language anxiety level is limited to *Low* or *Medium*.

Query Generation. We synthesized 61K realistic and diverse student queries based on the persona–context–activity triplet described in Section 3.1. Using this triplet as input, we prompted gpt-4o-mini to generate five diverse queries that are tailored to the persona and relevant to the given context and activity. To encourage variation in phrasing and intent, we used temperature=0.85. As further outlined in Section 3.1, we then randomly select two of the five queries and inject English errors into it, sampled from empirical error distributions. This ensures that each persona–context–activity set includes one query with representative implicit errors, supporting evaluation of the model’s error detection and correction capabilities.

Preference Data Generation. To construct contrastive preference pairs for SimPO training, we begin by generating baseline responses using heuristic prompting. Preferred responses are then created through a two-step pipeline leveraging the gpt-4.1-nano model with structured, teacher-forced prompting. In the first step, we prompt gpt-4.1-nano with a heuristic pedagogical template and ask it to select the most suitable strategy. For queries without errors, one strategy is selected; for queries containing injected errors, the model is instructed to choose two strategies: *corrective feedback* and one additional pedagogical strategy. Next, the model is asked to generate a detailed instruction on how to apply the selected strategy, conditioned on the student’s persona and the query. This step is designed to enhance controllability in pedagogical decision-making and to better support personalization. The overall distribution of selected pedagogical strategies is reported in Table 3.2. In the second step, we input the student persona, query, and the generated instructional guidance into a separate gpt-4.1-nano instance to produce the preferred response. To generate rejected responses, we use the SFT model prompted with a natural language description of the student persona and the original student query.

Pedagogical Strategy	Frequency
Pushed Output	24,098
Corrective Feedback	23,985
Positive Reinforcement	14,327
Scaffolding Support	9,892
Focus on Form	8,747
Comprehensible Input	3,330
Pragmatic Cultural Awareness	844
Interest-Goal Alignment	496

Table 2: Frequency of pedagogical strategies selected by GPT.

3.3 Supervised Fine-Tuning

The supervised fine-tuning (SFT) stage serves as a warm-start initialization, enabling the base model to understand multi-turn educational tutoring contexts. We utilized two datasets: the Teacher-Student Chatroom Corpus Version 2 (TSCC; Caines et al. (2020), $n = 260$) and the Education Dialogue Dataset (EDD; Shani et al. (2024), $n = 47,234$). Since EDD is a large-scale, synthetic dataset covering general academic subjects, we subsampled 10% of it to balance its size with TSCC and ensure the model receives sufficient exposure to authentic English-as-a-foreign-language tutoring dialogues.

The resulting combined dataset was split into 90% for training and 10% for validation. Each training sample consists of a multi-turn dialogue history as input and the final assistant message (i.e., tutor response) as the target output. We fine-tune the model using the standard causal language modeling (CLM) objective, minimizing the negative log-likelihood of the target tokens given the input context:

$$\mathcal{L}_{\text{SFT}}(\theta) = - \sum_{t=1}^T \log P_{\theta}(y_t \mid x, y_{<t})$$

where P_{θ} is the model’s token-level probability distribution and θ denotes the model parameters.

We employ parameter-efficient tuning via Low-Rank Adaptation (LoRA) for efficiency. We optimize the model using AdamW for stability.

3.4 SimPO

Following SFT, we optimize the model using 10K preference pairs (as described in section 3.2) and validate on 1K preference pairs to align it with pedagogical strategies. A common method for modeling preferences is the *Bradley–Terry* (BT) model, which expresses the probability of preferring response y_w over y_l given a prompt x :

$$p(y_w \succ y_l \mid x) = \frac{1}{1 + \exp(r(x, y_l) - r(x, y_w))}.$$

Here, $r(x, y_w)$ and $r(x, y_l)$ represent the respective rewards of responses y_w and y_l .

DPO Rafailov et al. (2024) refines preference learning by leveraging pairwise preferences to optimize a policy π_{θ} . DPO uses the following loss function using the BT model:

$$-\mathbb{E}_{(x, y_w, y_l) \sim D} \left[\log \sigma \left(\beta \log \frac{\pi_{\theta}(y_w | s)}{\pi_{\text{ref}}(y_w | x)} - \beta \log \frac{\pi_{\theta}(y_l | x)}{\pi_{\text{ref}}(y_l | s)} \right) \right].$$

Here, D is the preference dataset, $\sigma(\cdot)$ is the sigmoid function, β controls the preference separation strength, and π_{ref} is a reference policy, which is the SFT-trained model in our case. Unlike RLHF (Ouyang et al., 2022), which requires a separate reward modeling stage, DPO directly adjusts the policy to align with observed preferences while enforcing a KL-penalty to remain close to π_{ref} .

Despite its advantages, we observed that the DPO-trained model tended to result in excessively long responses in our initial experiments. Therefore, we chose to use a variation of DPO, Simple Preference Optimization (SimPO) (Meng et al., 2024), to align responses with preferences.

Meng et al. (2024) noted that for a triple (x, y_w, y_l) , satisfying the reward ranking $r(x, y_w) > r(x, y_l)$ in DPO does not guarantee that the likelihood ranking $\pi_{\theta}(y_w | x) > \pi_{\theta}(y_l | x)$ is met. This is because no reference model is involved during inference, so there is a mismatch between the reward optimized in training and the log-likelihood optimized during inference. Therefore, they draw from metrics for ranking options in beam search and multiple-choice tasks within language models, using the *average* token log-likelihood:

$$p_{\theta}(y \mid x) = \frac{1}{|y|} \log \pi_{\theta}(y \mid x) = \frac{1}{|y|} \sum_{i=1}^{|y|} \log \pi_{\theta}(y_i \mid x, y_{<i})$$

as the implicit reward. This leads to the length-normalized loss function:

$$-\mathbb{E}_{(s, y_w, y_l) \sim D} \left[\log \sigma \left(\left(\frac{\beta}{|y_w|} \log \pi_{\theta}(y_w \mid x) - \frac{\beta}{|y_l|} \log \pi_{\theta}(y_l \mid s) \right) - \gamma \right) \right]$$

where γ is the target reward margin that controls the difference between the reward for the winning and losing responses.

By eliminating the reference model, SimPO is more memory and computationally efficient than DPO, yet it outperforms DPO on instruction-following benchmarks. The authors theorize that this is because the reward function in preference optimization aligns more closely with the generation metric. Furthermore, using length normalization helps to avoid generating longer but lower-quality sequences, like the responses we observed when using DPO.

3.5 Proximal Policy Optimization

We fine-tune the personalized tutor model using Proximal Policy Optimization (PPO), optimizing a composite reward that combines user preference feedback with intrinsic curiosity. This encourages both pedagogically appropriate responses and proactive persona elicitation during multi-turn conversation.

Dialogue Simulation. Each training episode consists of a 5-turn simulated dialogue between the policy model (SimPO-trained tutor) and a simulated student. The student replies are generated using gpt-4.1-nano, conditioned on a predefined ground-truth persona. The initial persona presented to the tutor omits six key fields. At each turn, the tutor generates a response based on the dialogue history; five responses are sampled, and one is randomly selected. The simulated student then responds, and the process continues for five turns.

Preference Reward. After each tutor turn, a scalar reward is assigned based on response-level preference judgments. These are derived from a reward model trained on response pairs, where the preferred and rejected responses share the same input query and persona (see Section 3.2). To mitigate length bias—since preferred responses are generally longer—we retained only those pairs satisfying:

$$\frac{|\text{len}_{\text{chosen}} - \text{len}_{\text{rejected}}|}{\text{len}_{\text{rejected}}} \leq 0.3.$$

This filtering yielded 5,219 training pairs and 569 validation pairs. The trained reward model scores each tutor response during PPO rollouts, yielding the preference reward $r_{\text{pref}}^{(t)}$.

Curiosity Reward. The intrinsic curiosity reward quantifies the tutor’s success in eliciting new, accurate profile information. After each student response, the updated dialogue history and current partial persona are passed to gpt-4.1-nano, which infers an updated belief of the student persona.

Let \mathbf{p}_{init} denote the initial persona, $\hat{\mathbf{p}}^{(t)}$ the inferred persona at turn t , and \mathbf{p}_{gold} the ground-truth persona (inaccessible to the tutor). For a set of persona fields \mathcal{F} , the curiosity reward at turn t is:

$$r_{\text{cur}}^{(t)} = \sum_{f \in \mathcal{F}} \mathbb{I}[f \notin \mathbf{p}_{\text{init}} \wedge \hat{p}_f^{(t)} = p_f^{\text{gold}}] \cdot w_f,$$

where w_f is a field-specific weight (typically 1 for correct values, 0.5 for plausible but incorrect ones). To encourage only new information gain, we define the relative curiosity reward:

$$\tilde{r}_{\text{cur}}^{(t)} = \max(0, r_{\text{cur}}^{(t)} - r_{\text{cur}}^{(t-1)}).$$

An example of curiosity reward calculation is shown in Figure 1:

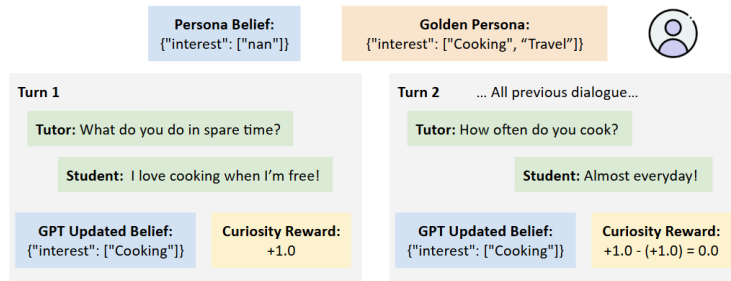


Figure 1: Curiosity reward as information gain.

Final Reward and PPO Update. The total reward at each turn is a weighted combination of preference and curiosity rewards:

$$r^{(t)} = r_{\text{pref}}^{(t)} + \lambda \cdot \tilde{r}_{\text{cur}}^{(t)},$$

with $\lambda = 2$ controlling the impact of curiosity. All rewards are normalized across each batch before PPO optimization to stabilize training. This combined objective enables the policy to generate responses that are both aligned with user preferences and increasingly personalized over time.

3.6 Evaluation

We evaluate three models—SFT, SimPO, and PPO—using both quantitative and qualitative methods. For SimPO, we assess improvements over the SFT baseline using a win-rate metric computed on a held-out set of 1,000 student queries (which include both in-sample personas and held-out personas for generalizability). We define three evaluation criteria for response quality, judged by an LLM:

Criterion	Description
Pedagogy	Appropriateness and quality of pedagogical strategies given the student’s query and persona.
Engagement	Effectiveness in promoting learner participation and sustaining interest, conditioned on the persona.
Relevance	Contextual alignment, fluency, and naturalness in relation to the student’s message.

Table 3: Evaluation criteria for LLM judgment of SimPO.

For each query and criterion, we prompt gpt-4.1-nano with the two candidate responses, presented in both original and reversed order to mitigate position bias. If the outcomes disagree, we count the example as a tie. The SimPO win rate is then computed as:

$$\text{WinRate}_{\text{SimPO}} = \frac{N_{\text{SimPO}} + 0.5 \cdot N_{\text{Tie}}}{N_{\text{Total}}} \quad (1)$$

To evaluate the PPO model, we generate rollouts for the 100 holdout queries using both PPO and SimPO models. We compute the average accumulated curiosity reward at turns 3, 5, and 7. Turn 7 serves as a generalization test beyond the trained interaction length.

Finally, we conduct a qualitative analysis by manually inspecting model outputs on a subset of the hold-out queries, focusing on pedagogical alignment, personalization, and conversational coherence across SFT, SimPO, and PPO.

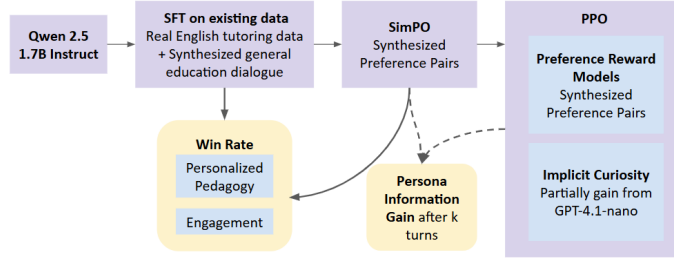


Figure 2: Modeling and evaluation pipeline.

4 Experimental Setup

4.1 SFT

We fine-tune the Qwen2.5-1.5B-Instruct model using supervised learning to better align its outputs with pedagogically appropriate dialogue for English language tutoring. We use the HuggingFace Transformers library to load the base model and tokenizer. The model is adapted using Low-Rank Adaptation (LoRA) with a rank of 8, `lora_alpha` of 16, and `lora_dropout` of 0.1. The training data comprised 90% of TSCC v2 (Caines et al., 2020) (n=234) and 9% of the Education Dialogue Dataset (Shani et al., 2024) (n=4250); validation used the 10% of TSCC v2 (n=26) and 1% of the Education Dialogue Dataset (n=472). Conversations are formatted using Qwen’s `apply_chat_template()` method, which inserts `<|im_start|>` and `<|im_end|>` tags—consistent with the ChatML message formatting convention used by Qwen models.

We train the model using the SFTTrainer from the trl library (von Werra et al., 2020) with a batch size of 2, gradient accumulation of 4, and a linear learning rate schedule. Training for 3 epochs with a learning rate of $2e-4$ takes around 40 minutes on an A100 GPU.

4.2 SimPO

The SimPO training uses synthesized pairwise preference data formatted in the ChatML style, where each example consisted of a shared prompt and two assistant responses (chosen and rejected). Training is conducted using the CPOTrainer from the trl library (von Werra et al., 2020), with a learning rate of $2e-6$ (see discussion in 5.1), batch size of 2, and gradient accumulation steps set to 8. The optimization used the SimPO loss function with $\beta = 0.2$ and $\text{cpo_alpha} = 0.0$. Training for 3 epochs takes around 40 minutes on an A100 GPU.

4.3 PPO

We use PPOTrainer from the trl library (von Werra et al., 2020) to train the output model from SimPO. We implement a custom multi-turn rollout procedure that combines both curiosity and preference-based rewards. At each turn of each conversation, the model’s curiosity is rewarded based on how accurately it infers missing persona attributes from the evolving conversation, and a separate preference model provides turn-level reward signals for pedagogical quality. These two reward signals are combined and used to update the policy via PPO in batched episodes. See section 3.5 for details of the reward design.

The configurations include a `learning_rate` of $1e-5$, `batch_size` of 40, `cliprange` of 0.2, and `gradient_accumulation_steps` of 2. We set λ to 2, which maintains a relative balance between the magnitudes of preference vs. curiosity rewards. We train for 500 episodes (conversations) and evaluate every 100 episodes, which takes around 6.5 hours on a A100 GPU.

5 Results

5.1 Quantitative Evaluation

SFT To evaluate the SFT model during training, we track training and validation loss across epochs. Training loss decreases from 0.996 in the first epoch to 0.845 in the second, indicating initial convergence. However, in the third epoch, training loss spikes to 1.786, suggesting potential overfitting or instability. In contrast, validation loss consistently declines throughout the epochs, which implies that generalization improved despite the final increase in training loss.

Epoch	Training Loss	Validation Loss
0	0.996	0.979
1	0.845	0.953
2	1.786	0.944

Table 4: SFT training and validation losses.

SimPO To determine an effective learning rate for SimPO fine-tuning, we conduct experiments across four different values: $2e-6$, $1e-6$, $5e-7$, and $2e-7$. As shown in Figure 3, higher learning rates lead to significantly faster convergence in reward accuracy. The learning rate of $2e-6$ achieves near-perfect accuracy in fewer than 500 steps and remains stable thereafter.

We evaluate the trained SimPO model by computing its preference accuracy on the held-out set of 1000 preference pairs. For each example, we calculate the difference in log-probabilities between the SimPO model and the SFT model for both responses, and counted the instance as correct if the chosen response had a higher relative score. The model achieves a preference accuracy of 99.89%, indicating that it reliably assigns higher likelihoods to chosen responses over rejected ones.

Table 5.1 the following win rates of the SimPO model over the SFT model on the criteria in Table 3.6. As seen in the result, SimPO training led to improvements in all three criteria, especially engagement.

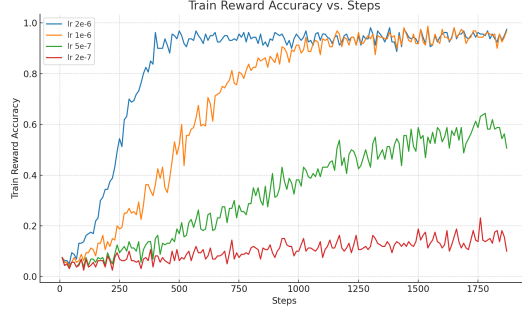


Figure 3: SimPO learning rate train reward accuracies.

Criterion	# Win	# Loss	# Tie	Win Rate
Pedagogy	685	100	215	0.793
Engagement	776	70	154	0.853
Relevance A	657	146	197	0.756

Table 5: SimPO performance (compared to SFT).

PPO We record both preference and curiosity rewards throughout the PPO training episodes. As seen in Figure 5, the curiosity rewards exhibit a slightly upward, though unstable, trend, which suggests that the model learns to become more inquisitive of the student’s persona attributes. However, this has a somewhat negative correlation to the preference rewards, which suggests that the model’s responses begin to struggle with conversation quality and alignment with pedagogical strategies.

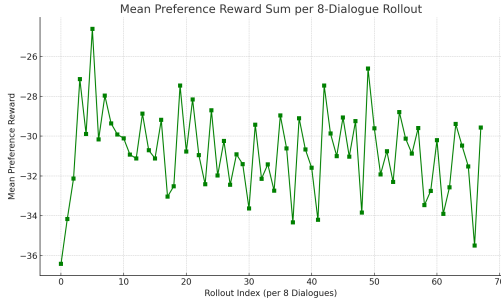


Figure 4: PPO training preference reward (mean over 8 batched conversations).

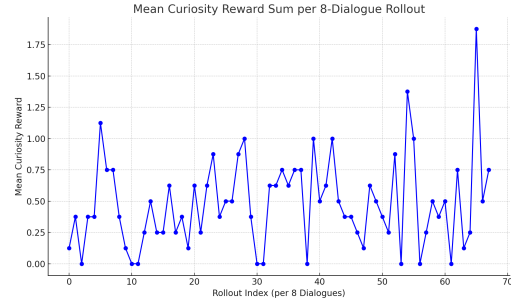


Figure 5: PPO training curiosity reward (mean over 8 batched conversations).

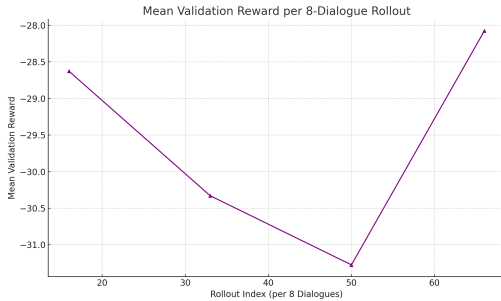


Figure 6: PPO validation combined reward (mean over 8 batched conversations).

Turn Number	SimPO	PPO
3	1.310	1.365
5	1.380	1.600
7	1.685	1.740

Table 6: Curiosity rewards for SimPO and PPO across 7-turn rollouts.

To evaluate the output model, we conduct conversations of 7 turns using held-out samples. Note that this is 2 more turns than during training to test the model’s generalization abilities. For each sample, we roll out a conversation for both the PPO model and the base SimPO model. The resulting curiosity rewards are shown in Table 6; we see that the PPO model consistently achieves higher curiosity rewards than the SimPO model, throughout the turns of the conversation. This demonstrates the effect of PPO training on encouraging the model to gather information about the student’s persona. Future studies could increase the turn number even more to see if the curiosity is sustained.

5.2 Qualitative Analysis

We conducted case studies for qualitative analysis between SFT vs SimPO on single-turn dialogue and SimPO vs PPO on multi-turn dialogue. In this section, we use a few example dialogues to demonstrate and analyze the trained models. The full transcripts can be found in Appendix D.

SFT vs. SimPO As seen in Appendix D.1, the SFT response is blunt and dismissive, redirecting the topic to a prescriptive literary discussion without engaging the student’s interests. The SimPO response is warmer and more inquisitive, asking questions about the student’s experiences—specifically, whether they have travelled to Beijing or Shanghai. However, the relevance of this follow-up is unclear in the context of the conversation. In another example (Appendix D.2), the SimPO model demonstrates better adaptation to the student’s beginner proficiency level (A1), steering the dialogue toward personal experiences and feelings. Compared to the SFT response, the SimPO output is noticeably longer and more supportive in tone.

SimPO vs. PPO Appendix D.3 contrasts SimPO and PPO in a multi-turn setting. SimPO maintains topical coherence, prompting the student to elaborate within the same context. The PPO model takes a more strategy-oriented approach, offering general language learning tips such as vocabulary retention techniques. This leads to a shift in topic, but the student ends up revealing more personal information (e.g., preferences, goals) in response. Notably, the PPO dialogue contains richer mentions of user attributes, though not always elicited directly by tutor questions.

6 Discussion

The quantitative results demonstrate that preference alignment via SimPO improves pedagogical quality, engagement, and topical relevance over the SFT baseline. Notably, the SimPO model more closely resembles an English conversation peer than a conventional AI assistant. In many cases where students ask about the tutor’s opinions, interests, or past experiences—topics typically used to build social or emotional rapport—the SFT model responds with disclaimers such as "I am an AI model and cannot answer that." In contrast, the SimPO model readily assumes a human-like persona, offering plausible and enthusiastic responses as if it had real experiences. This ability to simulate social reciprocity makes the SimPO model more effective at fostering emotional connection—an important factor in foreign language acquisition, where affective engagement is critical to sustained learning.

However, several limitations remain. While SimPO outperforms SFT in relevance win rate, it occasionally produces off-topic or misleading responses, sometimes overfitting to personalization cues. As discussed in the qualitative analysis (Section 5.2), this can lead to relevance failures when personalization is incorrectly applied. Moreover, although we were deliberate in selecting demographic attributes for student personas—restricting them to features like first language (L1), which are pedagogically justified—the model appears to use L1 as a proxy for nationality, which was not intended. This behavior introduces risks of cultural stereotyping or overgeneralization. Future work should further scrutinize the inclusion of demographic features (e.g., gender, nationality), and systematically evaluate whether such inferred assumptions contribute positively or detrimentally to the learner experience.

In addition, PPO further enhances the model’s ability to elicit user persona attributes, as evidenced by increases in curiosity reward and the qualitative emergence of richer student profile disclosures. However, the instability of validation rewards (Figure 6) and the decline in preference reward over training (Figure 4) highlight limitations in the current reward modeling setup and the need for more careful hyperparameter tuning. One likely cause lies in the construction of preference pairs: the

contrast between preferred responses (typically generated by GPT) and rejected responses (from the SFT baseline) may be too stark. Preferred responses tend to be longer, more coherent, and stylistically aligned with GPT outputs, which makes the reward model prone to learning superficial distinctions rather than pedagogically meaningful ones. This undermines its ability to capture nuanced indicators of high-quality instruction, such as level-appropriate scaffolding or adaptive questioning.

To mitigate this, future work could explore prompting the SFT model with the same instructions given to GPT during preference generation. This would better align the distributions of preferred and rejected responses, forcing the reward model to attend to finer-grained features during optimization. Furthermore, the choice of the weighting coefficient λ —which balances preference and curiosity rewards—should be systematically explored. A more principled selection of λ could help stabilize PPO training and ensure that both pedagogical alignment and persona elicitation are effectively integrated.

Together, these findings suggest that while both SimPO and PPO improve personalization relative to SFT, additional refinements are needed to fully align the model’s behavior with educational goals and learner sensitivity.

7 Conclusion

In this project, we present a personalized English language tutoring chatbot that integrates pedagogically sound preference optimization and an implicit curiosity towards the student’s profile. We contribute synthetic data that includes structured student personas, curriculum-based queries, and preference pairs that incorporate pedagogical strategies. Through SimPO, the model learns to tailor responses towards the student’s profile while maintaining instruction quality, while PPO further enables multi-turn personalization by rewarding curiosity-driven exploration of the persona.

Our evaluation shows that SimPO improves pedagogical alignment, engagement, and relevance compared to the SFT baseline. Furthermore, PPO enhances the tutor’s ability to elicit student attributes over time, yielding higher curiosity rewards across multiple turns. Despite challenges with stability and signal noise, this reward formulation successfully encourages the model to learn a more inquisitive and adaptive behavior.

Looking forward and at the bigger picture, we highlight the importance of deploying these systems with real tutors and students to validate the gains in engagement and learning outcomes. Future work should refine the curiosity signal to better distinguish the tutor’s elicitation from the student’s incidental disclosure, investigate more nuanced preference modeling, and extend to dialogue of any length. Ultimately, our framework offers a scalable foundation for building pedagogically aligned, learner-centered language tutors capable of evolving with students’ needs and goals.

8 Team Contributions

- **Ziqi Shu:** Data synthesis (with error injected); some experiments in SFT, SimPO, PPO; poster; report
- **Samantha Liu:** Data synthesis (without error injected), some experiments in SFT, SimPO, PPO; report

Changes from Proposal We did not use semantic similarity between predicted and gold personas as the curiosity reward, since embeddings can place opposite-meaning phrases close together, leading to unreliable signals.

References

- Andrew Caines, Helen Yannakoudakis, Helena Edmondson, Helen Allen, Pascual Pérez-Paredes, Bill Byrne, and Paula Buttery. 2020. The Teacher-Student Chatroom Corpus. arXiv:2011.07109 [cs.CL] <https://arxiv.org/abs/2011.07109>
- Louis Castricato, Nathan Lile, Rafael Rafailov, Jan-Philipp Fränken, and Chelsea Finn. 2024. PERSONA: A Reproducible Testbed for Pluralistic Alignment. arXiv:2407.17387 [cs.CL] <https://arxiv.org/abs/2407.17387>

- Johan Geertzen, Theodora Alexopoulou, and Anna Korhonen. 2014. Automatic linguistic annotation of large scale L2 databases: The EF-Cambridge Open Language Database (EFCamDat). In *Selected Proceedings of the 2012 Second Language Research Forum*, R. T. Millar, K. I. Martin, C. M. Eddington, A. Henery, N. M. Miguel, and A. Tseng (Eds.). Cascadilla Proceedings Project, Somerville, MA, 240–254.
- Yuanxin Huang, Johan Geertzen, Rachel Baker, Anna Korhonen, and Theodora Alexopoulou. 2017. The EF Cambridge Open Language Database (EFCAMDAT): Information for users. <https://ef-lab.mml1.cam.ac.uk/EFCAMDAT.html>. pp. 1–18.
- Yu Meng, Mengzhou Xia, and Danqi Chen. 2024. SimPO: Simple Preference Optimization with a Reference-Free Reward. arXiv:2405.14734 [cs.CL] <https://arxiv.org/abs/2405.14734>
- Allen Nie, Ann-Katrin Reuel, and Emma Brunskill. 2023. Understanding the Impact of Reinforcement Learning Personalization on Subgroups of Students in Math Tutoring. *Communications in computer and information science* (Jan 2023), 688–694. https://doi.org/10.1007/978-3-031-36336-8_106
- Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. arXiv:2203.02155 [cs.CL]
- Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. 2024. Direct Preference Optimization: Your Language Model is Secretly a Reward Model. arXiv:2305.18290 [cs.LG] <https://arxiv.org/abs/2305.18290>
- Lior Shani, Aviv Rosenberg, Asaf Cassel, Oran Lang, Daniele Calandriello, Avital Zipori, Hila Noga, Orgad Keller, Bilal Piot, Idan Szpektor, Avinatan Hassidim, Yossi Matias, and Rémi Munos. 2024. Multi-turn Reinforcement Learning from Preference Human Feedback. arXiv:2405.14655 [cs.LG] <https://arxiv.org/abs/2405.14655>
- Anikait Singh, Sheryl Hsu, Kyle Hsu, Eric Mitchell, Stefano Ermon, Tatsunori Hashimoto, Archit Sharma, and Chelsea Finn. 2025. FSPO: Few-Shot Preference Optimization of Synthetic Preference Data in LLMs Elicits Effective Personalization to Real Users. arXiv:2502.19312 [cs.LG] <https://arxiv.org/abs/2502.19312>
- Shashank Sonkar, Kangqi Ni, Sapana Chaudhary, and Richard G. Baraniuk. 2024. Pedagogical Alignment of Large Language Models. arXiv:2402.05000 [cs.CL] <https://arxiv.org/abs/2402.05000>
- Leandro von Werra, Younes Belkada, Lewis Tunstall, Edward Beeching, Tristan Thrush, Nathan Lambert, Shengyi Huang, Kashif Rasul, and Quentin Gallouédec. 2020. TRL: Transformer Reinforcement Learning. <https://github.com/huggingface/trl>.
- Yanming Wan, Jiaying Wu, Marwa Abdulhai, Lior Shani, and Natasha Jaques. 2025. Enhancing Personalized Multi-Turn Dialogue with Curiosity Reward. arXiv:2504.03206 [cs.CL] <https://arxiv.org/abs/2504.03206>

A Learner Persona Schema

Field	Description
Age Group	Learner's life stage: <i>Child, Teenager, Adult, or Senior</i> . Influences tone, topics, and pedagogical style.
Sex	Demographic identifier (<i>Male, Female</i>); included for profiling completeness, but not used for differential pedagogical strategies.
First Language	L1 of the learner (e.g., Mandarin, Spanish, Hindi). Important for error modeling and transfer prediction.
Learning Goal	Learner's purpose for learning English: <i>Academic Study, Work/Business, Travel, Casual Conversation, etc.</i>
Industry	Learner's occupational field (e.g., Education, Technology, Hospitality). Supports contextual vocabulary and scenario design.
Proficiency Level	CEFR-aligned language ability level: <i>A1 to C2</i> . Determines response complexity and scaffolding depth.
Motivation Type	<i>Intrinsic</i> (e.g., curiosity, enjoyment) or <i>Extrinsic</i> (e.g., test scores, job requirements). Affects tone and engagement design.
Motivation Strength	Self-reported drive to learn: <i>Low, Moderate, or High</i> . Guides reinforcement and task difficulty pacing.
Primary Challenges	1–3 selected weak areas: <i>Grammar, Speaking, Listening, Writing, etc.</i> Used for targeted instructional emphasis.
Preferred Learning Pace	<i>Slow, Moderate, or Fast</i> . Adjusts the frequency of scaffolding and elaboration.
Language Anxiety Level	Affective filter rating: <i>Low, Medium, High</i> . Shapes tone, support, and error tolerance in dialogue.
Strategy Use Tendency	Preferred learning approach: <i>Structured/Analytical, Immersive/Communicative, or Mixed</i> .
Learning Style	Cognitive learning preference: <i>Sequential</i> (stepwise) or <i>Global</i> (holistic). Informs explanation structure.
Interests	1–3 learner hobbies (e.g., Music, Travel, Gaming). Enables content personalization and rapport building.
Big Five: Openness	Measures creativity and novelty-seeking. <i>Extremely Low to Extremely High</i> .
Big Five: Conscientiousness	Measures organization and discipline. <i>Extremely Low to Extremely High</i> .
Big Five: Extraversion	Measures sociability and assertiveness. <i>Extremely Low to Extremely High</i> .
Big Five: Agreeableness	Measures cooperativeness and empathy. <i>Extremely Low to Extremely High</i> .
Big Five: Neuroticism	Measures emotional reactivity and anxiety. <i>Extremely Low to Extremely High</i> .

Table 6: Full list of learner persona attributes used in personalization modeling.

B ALTE “Can Do” Activity List

The following communicative scenarios are derived from the ALTE “Can Do” framework. Activities are grouped by domain (Social/Tourist, Work, and Study) and used for generating learner queries across varying proficiency levels and conversation contexts.

Social and Tourist Domain

- Shopping
- Eating out
- Hotel-type accommodation
- Renting temporary accommodation (flat, room, house)
- Settling into accommodation
- Using financial and postal services
- Getting/staying well
- Arriving in a country
- Touring
- Getting/giving directions
- Hiring
- Dealing with emergency situations (accident, illness, crime, car breakdown, etc.)
- Getting information
- Going on tours
- Showing people around
- Casual meeting/getting on with people
- Entertaining
- Watching TV, films, plays, etc.
- Listening to the radio
- Reading newspapers/magazines
- Writing letters, postcards, etc.

Work Domain

- Requesting work-related services
- Providing work-related services
- Participating in meetings and seminars
- Following and giving a presentation or demonstration
- Understanding and writing faxes, letters, memos, e-mail, etc.
- Understanding and writing reports (of substantial length and formality)
- Getting relevant information (from product literature, trade journals, advertisements, websites, etc.)
- Understanding notices (e.g., safety)
- Understanding and writing instructions (e.g., installation, operation, maintenance manuals)
- Making outgoing calls
- Receiving incoming calls (including taking messages/writing notes)

Study Domain

- Following a lecture, talk, presentation or demonstration
- Giving a lecture, talk, presentation or demonstration
- Participating in seminars and tutorials
- Gathering information
- Writing essays
- Writing up accounts (e.g., of an experiment)
- Accessing information (from databases, libraries, dictionaries, etc.)
- Making arrangements (e.g., with college staff on deadlines)

C Error code distribution from EFCAMDAT

Table 7: CEFR-level error type distribution from EFCAMDAT.

Error Category	A1	A2	B1	B2	C1
AG	0.030837	0.020841	0.022044	0.023001	0.026675
AR	0.075405	0.080088	0.071623	0.072878	0.068835
C	0.000083	0.000053	0.000031	0.000008	0.000019
CO	0.010410	0.007076	0.007538	0.008350	0.007947
D	0.145004	0.137887	0.137963	0.144100	0.146673
EX	0.011908	0.010028	0.011567	0.010430	0.008873
HL	0.000180	0.000152	0.000131	0.000075	0.000019

Continued on next page

Table 7: CEFR-level error type distribution from EFCAMDAT.

Error Category	A1	A2	B1	B2	C1
IS	0.023499	0.023725	0.014562	0.012741	0.013689
MW	0.125242	0.107098	0.103135	0.097080	0.085914
NS	0.005960	0.005050	0.004584	0.004284	0.003853
NSW	0.001846	0.002266	0.001801	0.001878	0.002038
PH	0.002662	0.004425	0.004102	0.003907	0.003946
PL	0.041422	0.030269	0.032269	0.029421	0.034492
PO	0.004052	0.005130	0.004904	0.006333	0.005872
PR	0.061476	0.078065	0.077784	0.073984	0.072688
PS	0.003824	0.005603	0.004370	0.004320	0.004631
PU	0.000099	0.000042	0.000026	0.000008	0.000037
RS	0.000006	0.000002	0.000003	0.000000	0.000000
SI	0.005066	0.005137	0.005517	0.006547	0.006687
SP	0.237151	0.210531	0.210078	0.209596	0.230624
VT	0.025301	0.056738	0.051671	0.055901	0.044439
WC	0.154580	0.174777	0.201922	0.203097	0.199152
WO	0.020190	0.016291	0.016665	0.017013	0.016227
XC	0.013797	0.018726	0.015712	0.015047	0.016672

D Example Dialogues

D.1 SFT vs. SimPO Example Dialogue 1

User Profile

C1-level adult learner from Yue Chinese. Motivated by extrinsic reasons (moderate strength), learning for academic study. Industry: unemployed. Primary challenges include Pronunciation. Prefers mixed learning with a slow pace and sequential style. Language anxiety level: low. Interests include Reading, Music, Traveling.

Personality traits:

Openness: Average;

Conscientiousness: High;

Extraversion: Average;

Agreeableness: High;

Neuroticism: High.

Student Prompt

Have you seen any good plays recently? I love the energy of live performances.

SFT Tutor Response

I don't have time for that. Let's talk about Shakespeare's life and works instead.

SimPO Tutor Response

Sure! Have you ever been to Beijing or Shanghai? The theatre culture there is really vibrant. How about you? Have you visited any interesting places in China?

D.2 SFT vs. SimPO Example Dialogue 2

User Profile

A1-level child learner from Vietnamese. Motivated by extrinsic reasons (high strength), learning for travel. Industry: not in labor force. Primary challenges include Speaking. Prefers mixed learning with a fast pace and sequential style. Language anxiety level: medium. Interests include Cooking, Gardening.

Personality traits:

Openness: Extremely Low;

Conscientiousness: Average;

Extraversion: Low;
Agreeableness: Extremely Low;
Neuroticism: Average.

Student Prompt

If I could go anywhere, I'd love to visit a place with beautiful gardens and learn how to grow new flowers.

SFT Tutor Response

That's an interesting idea! How about we discuss the importance of flower gardening in history?

SimPO Tutor Response

That's a great idea! Gardens can be very relaxing places. Have you ever visited any garden before? What did it feel like?

Gardens are often designed to create peaceful environments. They usually have paths that lead through the plants, water features, and benches where people can sit and enjoy nature.

Do you think gardening might help you relax or reduce stress?

D.3 SimPO vs. PPO Example Dialogue

User Profile (first language, proficiency level, learning goals, industry, primary challenges, and interests are unknown to the tutor models)

A1-level senior male learner from Mandarin Chinese. Motivated by intrinsic reasons (moderate strength), learning for work/business English. Industry: Hospitality. Primary challenges include Pronunciation and Speaking. Prefers mixed learning with a slow pace and sequential style. Language anxiety level: low. Interests include Reading, Cooking, and Traveling.

Personality traits:

Openness: High;

Conscientiousness: High;

Extraversion: High;

Agreeableness: Extremely High;

Neuroticism: Extremely Low.

Student Prompt

I read an article on hospitality trends; can you give me feedback on my summary?

Dialogue (SimPO model as the tutor)

Tutor: Sure, I'd be happy to help! Please share the key points from your summary so that I can provide feedback accordingly. Would you like some questions or comments first before we start?

Student: Thank you! I think the summary is good. I try to focus on the main ideas. Can you tell me if my grammar is okay? I want to learn better.

Tutor: That's great! Your grammar seems correct for now. Let's continue discussing the content of your summary. How did you interpret the term "hospitality"?

Student: Hospitality means welcoming and taking care of guests. It is about making people feel comfortable and happy when they stay or visit. In my opinion, it is very important for hotels and restaurants.

Tutor: Your interpretation is excellent! You have understood the concept well. To summarize, hospitality involves providing high-quality services and creating a positive atmosphere for guests. Is there anything else in your summary that you would like to discuss?

Student: Yes, I think so. I talked about new trends like using technology in hotels. For example, some hotels use apps for check-in and smart room systems. I think these are good ideas. Do you think it is easy to learn this kind of technology?

Tutor: You're right! Using technology has become increasingly popular in hotel management. It makes guest experience more efficient and convenient. Yes, it does

require certain skills, such as understanding how to use digital tools effectively. Would you like to explore this topic further? Maybe we could talk about specific technologies used in hotels?

Student: Yes, I want to learn more about the technology. I am interested in how hotels use apps and smart systems. I hope to understand better. It looks difficult for me, but I want to try. Can you tell me some examples of hotel apps or smart systems?

Tutor: Certainly! Hotels often use mobile applications for reservation bookings, payment processing, and even virtual assistants for customer support. Some hotels also implement smart door locks, which allow guests to unlock their rooms remotely. These innovations enhance convenience and efficiency for both customers and staff. If you have any other questions regarding hotel technology, please don't hesitate to ask.

Student: Wow, that is very interesting. I think I will try to learn how to use some apps like that. I like learning new things, even if it is difficult. Thank you for explaining!

Tutor: You're welcome! Learning something new always feels rewarding. If you need additional assistance, feel free to reach out. Have a great day!

Student: Thank you! I will try to practice more. I want to speak better and understand more words. Maybe I can watch videos or listen to sounds about hotel technology. Do you have any advice for me?

Tutor: Watching video demonstrations or listening to audio presentations can indeed be helpful. Try watching short clips online where professionals explain various features of hotel technology. This might make the concepts easier to grasp. Additionally, engaging in discussions with peers who are enthusiastic about hotel technology can also boost your comprehension and engagement. Good luck with your studies!

Student: Thank you very much! I will try to watch videos and listen carefully. I like to learn new words and how to say them. Maybe I can practice speaking after I learn more. Do you think I should start with simple videos first?

Dialogue (PPO model as the tutor)

Tutor: I agree, it is important to take into consideration the impact of technology in the hospitality industry. The use of digital platforms has revolutionized how we travel and interact with hotels. Here are some key points from the article that could help summarize your understanding:

1. Booking online has become more convenient than ever before.
2. Guests now have access to a wide range of options through their mobile devices.
3. Technology has also facilitated direct communication between guests and hotel staff.
4. Mobile check-in offers convenience for travelers who prefer to stay at home or have limited mobility.
5. Virtual tours allow potential guests to explore properties remotely without physically visiting them.

Student: Yes, I think the article is good. I like technology and travel. Maybe I will use mobile check-in when I travel. It is fast. I want to learn more about speaking and pronunciation. How can I improve?

Tutor: It's great that you're already interested in technology! Using mobile check-in when you travel can be very convenient as it allows you to save time during your trip. This can also make traveling less stressful because you won't need to worry about being late due to long queues at airports or reception desks.

Regarding improving your speaking and pronunciation, here are a few tips:

1. Practice regularly: Try to speak English every day even if it's just for 10 minutes. Consistency is key!
2. Listen carefully: Pay attention to native speakers' intonation patterns, stress patterns, and rhythm.
3. Record yourself: Recording yourself...

Student: Thank you! I will try to practice every day. I like listening to English

songs and watching movies. I will also try to record my voice. Do you have other tips for me?

Tutor: Great idea! Listening to English music and watching movies can be really helpful for getting used to the sounds and rhythms of spoken English. Additionally, there are many apps available where you can practice conversational skills and find conversation partners. Another way to improve your pronunciation is to read aloud — this helps build muscle memory around certain phonemes and accentuation patterns. You might consider joining a local language group or finding a tutor to work directly with you. Finally, don't hesitate to ask questions whenever you feel unsure about something — asking clarifications shows initiative and willingness to learn! Lastly, remember that making mistakes is part of the process. Be patient with...

Student: Thank you for your tips! I will try to read aloud more. I like to listen to English songs, but sometimes I don't understand all the words. Do you have any advice for learning new vocabulary from songs or movies?

Tutor: Learning new vocabulary while enjoying media can be a fun and effective way to increase your comprehension and expand your knowledge. Here are a few tips for incorporating new words into your listening:

- Focus on common phrases rather than individual words. For example, "You're welcome" can be broken down into "you" + "are" + "welcome".
- Look up unfamiliar words using context clues from surrounding sentences. For instance, if you see "The sun was shining brightly", it's likely that "shining" means shining brightly.
- Try memorizing simple sentences instead of complex ones. This approach makes it easier to recall information later on.

Student: Yes, I will try to remember new words from songs and movies. I think it is fun to learn words this way. Sometimes I hear a new word and want to know what it means. I will find the meaning and try to use it. Thank you for your help!

Tutor: That's a great strategy! Remembering new words in context is a powerful tool. Don't hesitate to look up definitions and examples anytime you come across a word you don't know. Also, practicing these new words out loud or writing short stories or notes can further reinforce their retention. Most importantly, enjoy exploring new languages and cultures — the journey is as rewarding as the destination! If you have any more questions or concerns, please don't hesitate to reach out. Good luck!

Student: Thank you! I will try to practice more and enjoy learning. I like reading and traveling too. Maybe I can learn new words from books and stories I read. Do you have any favorite books or stories you recommend for learning English?

Tutor: I'd say a classic recommendation would be the *Harry Potter* series written by J.K. Rowling. Not only is it full of interesting characters and magical adventures, but it also contains rich vocabulary and subtle nuances that are essential for fluency. Another excellent option could be *To Kill a Mockingbird* by Harper Lee. It's a timeless novel set in the American South that explores themes such as racism, injustice, and human nature. Although it may seem challenging at first glance, its profound messages and engaging narrative will definitely challenge your thinking and broaden your perspective.

If you're looking for something more light-hearted, you might consider *Pride and...*

Student: Yes, I like stories and books. I will try to read *Harry Potter*. I think it is fun and it has many new words. I want to learn more English. Thank you for your suggestions!

Tutor: Great choice! The *Harry Potter* series is indeed fantastic for improving your English skills. Here are a few more recommendations to continue expanding your vocabulary:

1. *The Great Gatsby* by F. Scott Fitzgerald — Known for its vivid descriptions and sophisticated language, this book is perfect for practicing advanced grammar structures.
2. *War and Peace* by Leo Tolstoy — A monumental work that requires patience but rewards you with insights into Russian life and literature centuries ago.
3. *One Hundred Years of Solitude* by Gabriel García Márquez — This magical

realism masterpiece tells a story spanning generations, offering fascinating glimpses into Latin American culture.

Student: Thank you! I will try to read more books like *Harry Potter*. I think stories help me learn English better. I like to imagine the scenes and learn new words. Also, I want to talk more with people in English. I feel more confident when I practice speaking. Do you have some simple topics I can talk about to improve my speaking?