Persona-Aware Alignment of LLMs Using Synthetic Dialogue Data

Annick Grob[0009−0008−7002−6587], Hans Friedrich Witschel[0000−0002−8608−9039] and Andreas Martin[0000−0002−7909−7663]

FHNW University of Applied Sciences and Arts Northwestern Switzerland, School of Business, Riggenbachstrasse 16, 4600, Olten, Switzerland
annick.grob@bluewin.ch, {hansfriedrich.witschel, andreas.martin}@fhnw.ch

Abstract. Large Language Models (LLMs) are increasingly deployed in interactive public information systems, yet often fail to adapt their outputs to diverse user expectations. This paper presents a persona-aware alignment pipeline that fine-tunes LLMs using synthetic dialogue data to improve the stylistic alignment and communicative relevance of generated responses. The approach is applied in the context of RepoChat, a dialogue system developed with Nagra, the Swiss agency for the disposal of radioactive waste. RepoChat provides public access to complex technical and regulatory information and must respond appropriately to diverse audiences, including citizens, journalists, politicians, and subject matter experts.

We introduce a modular training process combining retrieval-augmented answer generation with persona-specific rewriting, supervised fine-tuning, and preference-based optimization. In total, the pipeline produced approximately 6000 persona-specific prompt–response pairs for supervised fine-tuning and 2500 contrastive triplets for Direct Preference Optimization, ensuring sufficient coverage across four personas.

To evaluate the system, we employ both automated assessments using LLM-as-a-Judge methods (with GPT-4o scoring prompts explicitly defined for factuality and stylistic alignment) and qualitative user feedback through interviews. Findings show that the fine-tuned model demonstrates moderate and inconsistent improvements in tone, clarity, and user-perceived alignment—particularly for non-expert audiences. However, limitations remain in handling emotional nuance and maintaining consistency across multi-turn dialogue.

This work contributes a reproducible alignment pipeline for persona-sensitive LLM deployment and highlights the value of synthetic training data in human-centred, high-stakes communication domains.

Keywords: Large Language Models · Persona-Aware AI · Synthetic Training Data · Human-Centred AI · Dialogue Systems.

1 Introduction

LLMs have become a central component in public-facing dialogue systems. Their ability to deliver fluent, informative, and context-aware responses makes them valuable for interactive platforms where users seek knowledge from large document corpora. However, current systems often struggle to adapt their linguistic output to the needs of diverse audiences. Especially in high-stakes public communication, tailoring the tone, complexity, and rhetorical framing of LLM outputs to different user personas is essential for accessibility, trust, and engagement.

This study addresses this challenge in the context of *RepoChat*, a document-based conversational assistant developed in collaboration with Nagra, the Swiss National Cooperative for the Disposal of Radioactive Waste. The system provides access to technical and regulatory information surrounding deep geological repositories. Its users include laypersons, journalists, politicians, and subject matter experts—each expecting stylistically distinct, yet factually accurate responses.

Existing LLMs are typically fine-tuned for general purpose outputs but fall short in persona-sensitive adaptation. Despite advancements in retrieval-augmented generation and model alignment, there remains a methodological gap in systematically aligning LLM outputs to stylistic preferences without hard-coded persona labels.

Research Question: How can persona-sensitive synthetic training data be used to align the stylistic behaviour of LLMs for audience-specific yet accurate response generation in public dialogue systems?

In this work, we generated a synthetic dataset of around 6000 supervised prompt–response pairs and 2500 DPO training triplets to enable systematic persona alignment. The evaluation combined quantitative LLM-as-a-Judge scoring, with explicit prompting templates to ensure reproducibility, and qualitative user interviews capturing persona-specific feedback

This paper presents a modular training and evaluation pipeline that uses synthetic prompt–response pairs, stylistic rewriting, and preference-based fine-tuning to enable persona-aware alignment of LLMs.

The remainder of this paper is structured as follows: Section 2 reviews related work on LLM alignment and persona modelling. Section 3 outlines the research design. Section 4 describes the synthetic data generation pipeline. Section 5 details the fine-tuning and alignment process. Section 6 presents the evaluation setup and results. Section 7 discusses implications and limitations, and Section 8 concludes the paper.

2 Related Work

Recent advances in large language models have enabled their application across a wide range of dialogue-based use cases, including personalized information retrieval, citizen interaction platforms, and decision support systems [2]. However, aligning these models with the needs, expectations, and communication styles of different user groups remains a central research challenge [10]. This section reviews existing research in four core areas relevant to this work: persona modeling in dialogue systems, synthetic data generation as a strategy for customizing LLM behaviour, supervised fine-tuning and preference-based optimization approaches such as Direct Preference Optimization, and human-centred as well as socio-technical perspectives on language model alignment.

2.1 Persona Modeling in Dialogue Systems and LLMs

User modeling and persona adaptation have become key research areas in the development of dialogue systems [3]. Prior work shows that tailoring outputs to user characteristics increases trust, engagement, and communicative effectiveness [3]. Personas are typically defined by traits such as expertise, goals, and preferred communication styles, and are either inferred dynamically or specified through rule-based or template-driven methods [15]. Cheng et al. [3] introduced a model for persona-based response generation using predefined personality profiles to steer output in multi-turn conversations. Similarly, Schuller et al. [15] emphasized that grounding dialogues in user profiles leads to more coherent and relevant interactions, particularly in personalized assistants.

Building on these earlier works, recent studies have explored more flexible approaches to persona conditioning. For example, [16] examined the use of implicit persona signals in LLM dialogue generation, while [11] demonstrated that persona prompts can be combined with fine-tuning to improve coherence and consistency across multi-turn dialogues. Other work has proposed evaluation frameworks for systematically assessing persona realism and alignment in generated responses (e.g., persona consistency benchmarks; [15]).

However, most persona conditioning approaches rely on hard-coded attributes or metadata, which limits their flexibility in dynamic, open-domain scenarios [8]. LLM-based systems, in contrast, offer the potential to model personas implicitly via prompt engineering and data-driven fine-tuning [1]. Yet, there is limited empirical evidence on how to train such models to distinguish between fine-grained persona expectations without explicit labels.

In particular, while existing methods have shown the benefits of template-driven or prompt-based persona steering, little work has addressed how synthetic data and alignment techniques (such as SFT and DPO) can achieve more systematic, reproducible persona adaptation [4]. This gap motivates the contribution of the present study.

2.2 Synthetic Data Generation for Alignment and Fine-Tuning

Synthetic training data has become a central strategy for aligning LLMs to specific behaviours or domains [9]. Ouyang et al. [12] note the high cost and complexity of collecting human preference data, but show the value of applying RLHF with larger quantities of preference data. In response, several studies have proposed using LLMs themselves to generate alignment data, such as preference comparisons or stylized responses, thereby reducing the reliance on manual labeling [4].

This work follows that trend by automatically generating both chosen (persona-aligned) and rejected (misaligned) examples for Direct Preference Optimization (DPO) [13]. The combination of guided prompt generation, retrieval-augmented grounding, and stylistic rewriting provides a lightweight alternative to RLHF for behaviour alignment [13].

2.3 Supervised Fine-Tuning vs. Preference-Based Optimization

Supervised Fine-Tuning (SFT) is the most common method for adapting LLMs to downstream tasks [1]. It requires a large number of high-quality input–output pairs and allows for controlled learning of task-specific behaviour [2]. However, SFT does not always reflect nuanced user preferences or subtle aspects of communication, such as tone or rhetorical structure [2].

To address this, Direct Preference Optimization has recently emerged as a promising method that trains models on contrastive examples rather than idealized targets [14]. DPO eliminates the need for separate reward models and optimizes directly on pairwise preferences, which can be particularly effective for stylistic alignment tasks [14]. In this study, both approaches were used sequentially—SFT to establish stylistic baselines, and DPO to refine outputs based on persona-specific preferences.

2.4 Human-Centred and Socio-Technical AI Approaches

The growing deployment of LLMs in public information systems raises ethical and communicative challenges [17]. Human-centred design emphasizes the need for inclusive, transparent, and context-sensitive AI systems [2]. In high-stakes environments such as public policy or scientific communication, dialogue systems must not only be factually correct but also linguistically appropriate, emotionally aware, and sensitive to audience expectations, and adapted to the users' varying levels of domain knowledge, including the ability to simplify complex content where necessary [17].

Socio-technical frameworks suggest that LLM alignment cannot be addressed through model optimization alone but must consider user goals, institutional norms, and communicative settings [2]. This study contributes to that discourse by integrating qualitative feedback from real users and by contextualizing persona alignment in the domain of radioactive waste communication, where credibility and trust are paramount.

3 Research Design and Methodology

This study follows an iterative, artifact-centred design process inspired by the Design Science Research (DSR) methodology [5]. The goal was to develop and evaluate a training pipeline that enables LLMs to generate persona-aligned responses through stylistic fine-tuning on synthetic dialogue data. The design process was driven by a real-world application context, incorporated both theoretical and empirical inputs, and applied multi-level evaluation criteria.

3.1 Iterative Artifact Development Process

The artifact—a modular training pipeline for persona-aware alignment—was developed using the five-phase DSR cycle [6]: (1) problem awareness, (2) solution suggestion, (3) artifact development, (4) evaluation, and (5) reflection. This iterative process allowed for progressive refinement based on both domain needs and observed user behaviour.

During the *problem awareness phase*, theoretical and empirical challenges in generating persona-appropriate LLM outputs were identified through a literature review and exploratory user interviews. Based on these insights, the *suggestion phase* defined key requirements and sketched a conceptual design for the pipeline. This included a persona-specific data generation workflow and a two-stage training approach.

In the *development phase*, the proposed design was implemented using Jupyter notebooks and open-source tools. The pipeline was executed locally using the Mistral-7B language model—an open-source LLM released under the Apache 2.0 license—and fine-tuned via the Unsloth library, which supports efficient LoRA-based supervised and preference-aligned training.

The *evaluation phase* combined two complementary approaches: (a) automated scoring through LLM-as-a-Judge methods, using explicit prompting templates to assess factuality and stylistic fit, and (b) qualitative interviews with domain users to capture persona-specific perspectives.

Finally, the *reflection phase* synthesised these results, identifying both strengths and limitations of the pipeline. While this project was limited to a single macro-cycle, micro-iterations occurred within phases: for example, the persona-specific question generation strategy was refined across multiple runs, and the evaluation approach was expanded from automated scoring to include qualitative feedback. In line with Hevner's principles [5], the reflection stage outlined clear opportunities for refinement—such as expanding dataset diversity, strengthening evaluation design, and running additional training cycles—that could guide subsequent macro-cycles in future research.

3.2 Application Context: RepoChat and Nagra

The research was conducted in the context of *RepoChat*, a document-grounded chatbot system developed in collaboration with *Nagra*, the Swiss National Cooperative for the Disposal of Radioactive Waste. RepoChat is designed to provide trustworthy, accessible answers to questions about deep geological repositories—a domain that requires both factual accuracy and communication sensitivity.

The system must serve diverse user groups, including laypersons concerned about environmental risks, journalists seeking precise information, politicians engaged in regulatory debates, and technical experts evaluating site suitability. This diversity places high demands on linguistic style, tone, and explanatory depth—motivating the need for persona-aware training data and alignment techniques.

3.3 Justification and Design of User Personas

To operationalise persona alignment, four representative user personas were defined:

- **Citizen**: concerned, non-expert, seeks clarity and reassurance
- **Journalist**: fact-focused, requires structured and source-based answers
- **Politician**: argumentative, expects persuasive and concise reasoning
- **Subject Matter Expert (SME)**: technically proficient, values precision and completeness

These personas were derived from a requirement analysis that combined insights from literature on persona-sensitive communication with findings from think-aloud interviews involving representative users from the domains of journalism, politics, and citizen. The interviews were conducted as part of the awareness phase and encouraged participants to interact with a non-aligned version of the chatbot, verbalising their observations and concerns regarding its responses. They revealed recurrent issues such as inappropriate tone, inconsistent terminology, and factual distortion through oversimplification.

Based on the interview insights, we described each persona along three dimensions: (1) knowledge level, (2) communicative expectations, and (3) stylistic preferences. These informed the design of prompts, rewriting instructions, and evaluation criteria used throughout the training pipeline.

3.4 Design Objectives and Evaluation Criteria

The core design objective was to create a reproducible training process that enables LLMs to stylistically adapt to different user personas without compromising factual accuracy. To achieve this, the system was designed with the following objectives:

Design Objectives

- **Synthetic training data** generation that reflects persona-specific communication needs.
- **Persona-adapted fine-tuning** through Supervised Fine-Tuning and Direct Preference Optimization.
- **Retrieval-augmented grounding** to ensure factual accuracy based on source documentation.
- **Automated and qualitative evaluation** mechanisms to assess stylistic alignment and user satisfaction.

Evaluation Criteria To measure the effectiveness of the artifact, three key evaluation dimensions were defined:

- **Factual Accuracy**: Consistency with retrieved documents and domain knowledge.

– **Stylistic Alignment**: Clarity, tone, and rhetorical appropriateness with respect to the intended persona.
– **User Satisfaction**: Perceived helpfulness and persona-fit, derived from qualitative feedback.

4 Synthetic Data Generation Pipeline

To enable stylistic adaptation in LLM outputs, a modular pipeline for synthetic dialogue data generation was developed. The pipeline was designed to produce high-quality, persona-specific training data that preserves factual accuracy while tailoring tone and complexity to different user groups. It consists of four main components: (1) persona-based prompt design, (2) guided response generation with retrieval grounding, (3) quality assurance and stylistic refinement, and (4) data structuring and storage for supervised fine-tuning and Direct Preference Optimization.

Figure 1 illustrates the overall structure of the synthetic data generation pipeline. It visualizes this four-step process, showing how prompts are created, grounded in retrieved knowledge, refined for stylistic alignment, and finally stored in structured formats for use in model training. Each stage corresponds directly to one element in the diagram, ensuring consistency between the textual description and the visual representation.

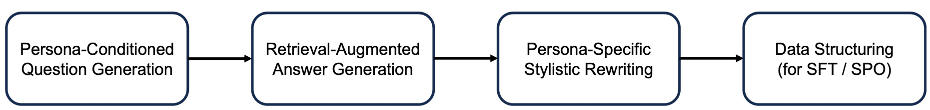

Fig. 1. Overview of the synthetic data generation pipeline. The process includes four stages: persona-specific data preparation, persona-specific question generation, retrieval-augmented response creation and persona-specific rewriting.

4.1 Persona-Conditioned Question Generation

The process begins with a small set of real, thematically relevant user questions derived from domain documentation and previous user interactions. For each of these original prompts, additional persona-specific questions were generated using GPT-4o. The objective was not merely to rephrase existing questions, but to produce thematically related and stylistically varied prompts that reflect the communication style, interests, and knowledge level of each predefined persona (citizen, journalist, politician, SME).

Based on our interview findings, the generation process for each persona was guided by a short instruction describing their expectations. For example, questions tailored to the *citizen* persona focused on clarity and accessibility, while those for the *politician* emphasized argumentative framing and regulatory relevance.

4.2 Retrieval-Augmented Answer Generation

Each generated question was answered using a retrieval-augmented generation (RAG) setup based on vector search and source documents from the Nagra domain [7]. Relevant document chunks were retrieved using a vector index, and GPT-4o was used via API to generate factually grounded responses based on the retrieved context. This step ensured domain coherence and factual accuracy.

At this stage, all responses were formulated in a neutral tone, independent of persona-specific stylistic characteristics. The generated outputs served as a factual base for further stylistic adaptation.

4.3 Persona-Specific Stylistic Rewriting

To adapt the answers to different communication styles without changing their content, a second rewriting step was applied. Using GPT-4o, the previously generated answers were rewritten according to persona-specific style guidelines. Each persona followed a dedicated instruction prompt specifying tone, structure, and linguistic conventions:

- **Citizen**: Clear, concise, easy to understand, emotionally attuned. Avoid jargon and technical complexity. Keep the message short and reassuring.
- **Journalist**: Structured, precise, objective. Use neutral and professional tone. Ensure traceability and factual integrity.
- **Politician**: Strategically phrased, rhetorically effective, aimed at audience impact. Moderate complexity, persuasive but grounded in facts.
- **Subject Matter Expert (SME)**: Technically accurate, analytical, using correct terminology. Style should be formal, efficient, and expert-level.

The resulting prompt–response pairs formed the dataset for supervised fine-tuning. For DPO, an additional rejected version was generated by deliberately violating one or more stylistic expectations—e.g., using vague language, excessive simplification, or inconsistent tone.

4.4 Data Structuring

The final stage of the pipeline organizes the generated data into structured formats suitable for supervised fine-tuning and preference-based optimization. Data was stored in CSV files, with each row containing the original persona-specific question, the generated response, and relevant metadata. For DPO training, paired responses were included, distinguishing between the "chosen" (persona-aligned) and "rejected" (misaligned) outputs.

In total, the data generation process produced approximately 6,000 prompt–response pairs for supervised fine-tuning across four personas (citizen, journalist, politician, and subject matter expert). Additionally, around 2,500 contrastive triplets were created for Direct Preference Optimization, each consisting of a prompt, a chosen response, and a rejected response. This ensured sufficient coverage of persona-specific stylistic variation while maintaining manageable dataset sizes for local fine-tuning experiments.

To ensure reusability and transparency, all datasets were versioned and annotated with metadata fields describing persona, question type, and generation method (e.g., retrieval-based, rewritten, or augmented). This structure facilitated both model training and subsequent evaluation of persona alignment.

5 Model Training and Alignment

To align the language model's responses with persona-specific communication styles while preserving factual accuracy, two training approaches were applied: Supervised Fine-Tuning and Direct Preference Optimization. Both were performed on a local instance of the Mistral-7B model using the Unsloth framework for parameter-efficient fine-tuning.

5.1 Supervised Fine-Tuning

The SFT phase used approximately 6,000 synthetic prompt–response pairs generated in the earlier pipeline steps. Each training sample consisted of a persona-specific question and its corresponding stylistically adapted answer. The model was trained to predict the aligned answer given the original prompt, without injecting explicit persona instructions or system prompts. The persona-specific style was learned implicitly through the structure and wording of both the questions and the responses.

The model used for fine-tuning was Mistral-7B-Instruct, an open-source LLM. Fine-tuning was performed using LoRA adapters and Unsloth's training routines, executed locally on a CUDA-enabled GPU, which allowed resource-efficient training without external infrastructure.

Training setup:

- LoRA config: rank r = 8; trained `q_proj` and `v_proj`; no dropout, no bias
- Epochs: 5
- Batch size: 8
- Learning rate: 2e-5
- Device: CUDA-enabled GPU (RTX 4070, 24GB VRAM) with `device_map={"":0}`

5.2 Direct Preference Optimization

The DPO phase refined stylistic alignment using approximately 2,500 contrastive triplets, each consisting of a prompt, a chosen (persona-aligned) response, and a rejected (misaligned) response. As in the SFT phase, persona alignment was not enforced through explicit instructions but learned implicitly by contrasting well-formed answers with less appropriate ones.

Training was initialized from the previously fine-tuned model, ensuring that DPO optimization refined outputs on top of a stylistically aware baseline.

Training setup:

- LoRA config: rank r = 8; trained `q_proj` and `v_proj`; no dropout, no bias
- Epochs: 3
- Batch size: 8
- Learning rate: 2e-5
- DPO-specific: $\beta = 0.1$ (preference margin strength); fine-tuned baseline model frozen as reference
- Device: CUDA-enabled GPU (RTX 4070, 24GB VRAM)

5.3 Tools and Model Architecture

The overall training architecture combined open-source components and commercial API access to ensure reproducibility and efficiency. This included data generation pipelines, fine-tuning frameworks, and evaluation tools, integrated into a modular workflow.

Components:

- **Base model:** Mistral-7B-Instruct, an open-source instruction-tuned LLM
- **Training framework:** Unsloth with LoRA adapters for parameter-efficient fine-tuning
- **Hardware environment:** CUDA-enabled GPU (RTX 4070, 24 GB VRAM)
- **Data management:** Structured CSV/JSON datasets with ∼6,000 SFT pairs and ∼2,500 DPO triplets
- **Evaluation tools:** GPT-4o used as LLM-as-a-Judge with explicit scoring prompts for factuality and stylistic alignment; complemented by qualitative user interviews

This combination of open-source tooling, lightweight parameter-efficient methods, and transparent dataset structuring ensured that the alignment process can be reproduced in other high-stakes application domains.

6 Evaluation

To evaluate the effect of fine-tuning on the stylistic quality and factual accuracy of model responses, an automated scoring procedure using GPT-4o was conducted. The evaluation compared the base model to two fine-tuned variants—one trained with SFT and the other with DPO. Each test was run twice to minimize random variation and ensure more robust results.

6.1 Quantitative Evaluation Using GPT-4o Scoring

For each model variant (SFT and DPO), two independent evaluation runs were carried out. Each output was rated on a 1–5 Likert scale, with 1 indicating very poor alignment and 5 indicating strong alignment. This ensured comparability across personas and runs. In each case, GPT-4o was shown a prompt and two responses: one from the base model and one from the fine-tuned variant. The model was instructed to evaluate the answers along four dimensions:

1. **Factual accuracy** - Assessed whether the trained model preserved the factual content of the base model's response. This ensured that stylistic adaptation did not degrade correctness.
2. **Stylistic quality of the base model response** - Evaluated how well the original output matched the communicative style and language expectations of the respective persona. Served as a baseline for comparison.
3. **Stylistic quality of the trained model response** - Measured the degree to which the fine-tuned output improved linguistic alignment with the persona, focusing on tone, clarity, formality, and rhetorical appropriateness.
4. **Relative stylistic preference between the two outputs** - Captured which response, base (scale 1) or trained (scale 5), was stylistically preferred.

Prompt template (simplified for readability)

```
You are evaluating the output of a dialogue system. Assess the
following response along two dimensions:
1. factual accuracy (1 = factually incorrect, 5 = fully correct) 2.
stylistic alignment with the given persona (1 = not fitting, 5 =
perfectly fitting).
Question: [insert question] Persona: [insert persona] Answer: [insert
model output]
Return your evaluation in the format: [Score factual accuracy]
- [Short justification in German] [Score stylistic alignment] -
[Short justification in German]
```

Table 1 summarizes the averaged results across both evaluation runs for each training method.

Table 1. Average GPT-4o evaluation scores for factual and stylistic performance

Model	Factual Score	Base Style Score	Trained Style Score	Base-Trained Preference
SFT 1	4.10	2.75	2.70	2.80
SFT 2	3.85	2.75	2.75	2.90
DPO 1	4.20	2.90	2.95	2.90
DPO 2	4.00	2.80	2.80	3.00

Scores in the table were produced by GPT-4o using the standardized prompt shown above, applied consistently across all personas.

Overall, the DPO-trained model achieved the highest factual accuracy (4.20), while both fine-tuned models showed modest but consistent stylistic improvements over the base model.

We acknowledge that relying on GPT-4o as the sole evaluator introduces potential bias, since an external LLM is used as the evaluator. While this approach provides scalability and consistency, future work should triangulate these results with human raters to strengthen validity.

6.2 Qualitative Evaluation through User Interviews

To complement the automated scoring, three qualitative interviews were conducted with participants representing the defined personas: citizen, journalist, and politician. Each participant was asked to review both baseline and fine-tuned responses and to provide feedback regarding clarity, tone, and relevance from their respective perspectives.

The key observations were as follows:

- **Citizen:** Participants noticed some improvements in clarity and language simplification. However, the tone often remained neutral or overly formal rather than empathetic and reassuring, as expected for a general public audience. One participant commented on the following response to the question *"Ist das Grundwasser durch das Tiefenlager gefährdet?"*:

  > *"Der geologische Tiefenlager ist aufgrund der geringen Gefährdungspotenzialität für die Sicherheit des Grundwassers von geringer Bedeutung. Die Einbauten im Tiefenlager, die den Durchfluss des Grundwassers beeinflussen, sind auf die Sicherheit des Tiefenlagers fokussiert und haben keine signifikanten Auswirkungen auf das Grundwasser oberhalb der Lagerstätte. [...] Die Sicherheit des Grundwassers ist daher nicht durch das Tiefenlager gefährdet."*

  While factually correct, the participant described the answer as "technically accurate but too abstract and bureaucratic in tone", and wished for a more direct and emotionally reassuring explanation.
- **Journalist:** Some answers showed clearer paragraph structure and more precise vocabulary. Still, the participants felt that the trained responses often lacked contextual framing and sourcing that would be typical in journalistic communication. One participant reviewed the following response to the question *"Wie wird der Standort für ein Tiefenlager ausgewählt?"*:

  > *"Der Standort für ein geologisches Tiefenlager wird auf Grundlage sicherheitstechnischer Kriterien ausgewählt, insbesondere hinsichtlich der geologischen Barriereeigenschaften."*

  The participant noted that this statement, while accurate, lacked critical context and evidence: "I would expect a reference to the selection process, stakeholders involved, and perhaps a link to published reports or data. As a journalist, I need more than a textbook summary."

– **Politician:** A few responses included more consensus-oriented phrasing or highlighted societal relevance. Nevertheless, rhetorical structuring and political framing were largely absent, limiting the perceived adaptation to this persona. One participant reviewed the following answer to the question *"Gibt es eine Volksabstimmung über das Tiefenlager?"*:

  > *"Es ist vorgesehen, dass die Bevölkerung in der Diskussion über die Rahmenbewilligung für geologische Tiefenlager einen fakultativen Referendumstermin wählen kann. Das Stimmvolk hat das letzte Wort, wenn der Entscheid vom Parlament angenommen und ein Referendum ergriffen wird. Die Auflage käme voraussichtlich um das Jahr 2031 stattfinden. [...] Der vorläufige Schutzbereich für das Tiefenlager wird auf einer Karte übertragen und schützt den Bereich tief im Untergrund, wo das spätere Tiefenlager steht."*

  The participant welcomed the mention of a concrete date ("2031 is memorable"), but criticised the overall lack of focus: "The answer goes far beyond what was asked. It floods the reader with technical details instead of strategically framing the referendum as a democratic milestone. I want to know what this means for public trust, not for borehole permits."

6.3 Summary of Evaluation Findings

The evaluation results suggest that the fine-tuning approach preserved factual accuracy across all variants and led to minor improvements in stylistic adaptation. The DPO-tuned model variants showed slightly stronger stylistic alignment compared to the SFT-tuned models, particularly in preference scores assigned by GPT-4o.

However, the stylistic differences were often subtle and inconsistent. In the qualitative feedback, some persona-specific adjustments were perceived, but they did not consistently reflect the distinct rhetorical expectations of the target user groups. Rather than indicating a failure, these limitations point to concrete opportunities for future work. Our findings suggest that broader training variation, more distinctive persona prompting, and the use of base models with greater stylistic capacity may further strengthen the effectiveness of persona-aware alignment.

In addition, future work should adopt more structured evaluation instruments, such as Likert-scale questionnaires or standardized surveys, to increase the comparability and reliability of user feedback.

7 Discussion

This section reflects on the evaluation findings and discusses methodological, technical, and socio-technical implications of persona-aware LLM alignment. Four aspects are considered: the observed stylistic improvements, trade-offs in human-centred system design, challenges in nuanced and context-aware generation, and the broader applicability of the approach in other sensitive domains.

7.1 Interpretation of Alignment Improvements

Particularly in the qualitative feedback, participants noted some persona-specific refinements (e.g., clearer structure for journalists or simpler language for citizens), but also pointed out that the adaptations did not fully match their expectations in tone or rhetorical framing. This indicates that the current approach—while effective at maintaining content quality—was limited in achieving clear, audience-specific communication styles.

Furthermore, the relatively limited dataset size (around 6,000 SFT pairs and 2,500 DPO triplets) restricted the depth of stylistic learning. While this was sufficient to demonstrate feasibility, larger and more varied datasets would be required to achieve more robust and generalisable stylistic differentiation.

7.2 Socio-Technical Design Trade-offs and Limitations

The system was developed with a strong emphasis on factuality, traceability, and fairness, particularly due to the high-stakes nature of the application domain. While these priorities supported trust and reliability, they may have constrained stylistic creativity and expressiveness. The lack of significant persona-specific divergence could be a result of cautious prompt design and the deliberate exclusion of emotionally charged or polarising language. Moreover, using only synthetic data—without incorporating real conversational feedback—limited the variability and naturalness of stylistic adaptation. These trade-offs illustrate the tension between ethical, communicative, and technical requirements in socio-technical AI system design.

Another limitation concerns the evaluation procedure. Automated scoring with GPT-4o enabled consistent large-scale comparisons, but it also introduces potential bias, since one LLM was used to judge the performance of another. Although partially mitigated by qualitative interviews, future work should triangulate automated judgments with structured human evaluation methods.

7.3 Challenges in Emotional Nuance and Multi-Turn Context

The results also reveal limitations in handling emotional nuance and maintaining consistent stylistic tone across multi-turn interactions. While single-turn responses occasionally reflected persona traits, longer dialogues may require deeper context tracking and emotion modelling—capabilities that go beyond current prompt-based stylistic control. In particular, empathetic or persuasive responses require sensitivity to both content and conversational history, which was not explicitly encoded in the training pipeline. This presents a future challenge for extending persona-aware alignment to more complex interaction formats.

7.4 Applicability to Other High-Stakes Communication Domains

Despite its limitations, the proposed alignment pipeline offers a reproducible and modular approach that could be adapted to other domains with diverse user needs. The combination of retrieval grounding, synthetic data based on persona, and contrastive fine-tuning allows for domain-specific tailoring without requiring human feedback at scale. Potential application areas include healthcare communication, public safety systems, or legal information platforms—contexts where factual precision and audience-sensitive delivery are equally critical. However, successful transfer would depend on domain-specific persona design, careful prompt engineering, and validation with target users.

8 Conclusion and Future Work

This paper introduced a modular and persona-aware pipeline for the synthetic generation of dialogue data to align the outputs of LLMs with the communicative expectations of different user groups. By combining prompt-based question expansion, retrieval-augmented answer generation, and persona-specific stylistic rewriting, the pipeline enabled the creation of targeted training datasets for both SFT and DPO. Evaluation results showed that factual consistency was preserved across all model variants, while improvements in stylistic adaptation were present but limited in scope.

The main contribution of this work lies in demonstrating a reproducible and low-cost approach for persona-driven alignment without requiring human-labelled datasets. The proposed pipeline offers practical value for domains where audience-sensitive communication is essential, such as public policy, healthcare, and scientific communication. By ensuring transparency in the design of both training data and model behaviour, the approach supports responsible use of LLMs in high-stakes, trust-dependent environments.

Nonetheless, the findings also revealed several limitations. Stylistic differences between base and fine-tuned models were often marginal and inconsistently realised. Emotional nuance and persona-specific tone were only partially achieved, particularly in cases where affective expression or rhetorical framing was expected. Moreover, the single-turn interaction format limited the ability to evaluate sustained dialogue coherence or adaptive behaviour in multi-turn conversations.

Future research should explore methods for enhancing the emotional and rhetorical expressiveness of LLMs, for example through sentiment conditioning or affective persona attributes. The development of dynamic persona modelling—where system responses adapt to user behaviour over time—may also lead to more natural and engaging interactions. In addition, hybrid evaluation strategies that combine automated scoring with qualitative user feedback could offer a more comprehensive assessment of alignment quality. Finally, the pipeline should be tested with larger foundation models beyond Mistral-7B, in order to examine whether model capacity plays a critical role in achieving more refined and consistent persona alignment.

Acknowledgements

This work was conducted as part of a master's thesis at FHNW and supported by Nagra Switzerland. We thank the interview participants and domain experts who contributed feedback during the evaluation phase.

References

1. Balavadhani Parthasarathy, V., Zafar, A., Khan, A., Shahid, A.: The Ultimate Guide to Fine-Tuning LLMs from Basics to Breakthroughs: An Exhaustive Review of Technologies, Research, Best Practices, Applied Research Challenges and Opportunities. arXiv e-prints pp. arXiv–2408 (2024), https://ui.adsabs.harvard.edu/abs/2024arXiv240813296B/abstract
2. Chen, J., Zhang, Y., Wang, B., Zhao, X., Wen, J.R., Chen, W.: Unveiling the Flaws: Exploring Imperfections in Synthetic Data and Mitigation Strategies for Large Language Models. In: Al-Onaizan, Y., Bansal, M., Chen, Y.N. (eds.) Findings of the Association for Computational Linguistics: EMNLP 2024. pp. 14855–14865. Association for Computational Linguistics, Miami, Florida, USA (Nov 2024). https://doi.org/10.18653/v1/2024.findings-emnlp.873, https://aclanthology.org/2024.findings-emnlp.873/
3. Cheng, Y., Liu, W., Xu, K., Hou, W., Ouyang, Y., Leong, C.T., Wu, X., Zheng, Y.: AutoPal: Autonomous Adaptation to Users for Personal AI Companionship (Oct 2024). https://doi.org/10.48550/arXiv.2406.13960, http://arxiv.org/abs/2406.13960, arXiv:2406.13960 [cs]
4. Gallego, V.: Refined Direct Preference Optimization with Synthetic Data for Behavioral Alignment of LLMs (Feb 2024), http://arxiv.org/abs/2402.08005, arXiv:2402.08005
5. Hevner, A., Chatterjee, S.: Design research in information systems: theory and practice, vol. 22. Springer Science & Business Media (2010)
6. Kuechler, W., Vaishnavi, V., Kuechler Sr, W.L.: Design [science] research in IS: a work in progress. In: Proceedings of the second international conference on design science research in information systems and technology (DESRIST 2007). pp. 1–17 (2004)
7. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.t., Rocktäschel, T., et al.: Retrieval-augmented generation for knowledge-intensive nlp tasks. Advances in neural information processing systems **33**, 9459–9474 (2020)
8. Li, H., Dong, Q., Chen, J., Su, H., Zhou, Y., Ai, Q., Ye, Z., Liu, Y.: Llms-as-judges: a comprehensive survey on llm-based evaluation methods. arXiv preprint arXiv:2412.05579 (2024)
9. Liu, R., Wei, J., Liu, F., Si, C., Zhang, Y., Rao, J., Zheng, S., Peng, D., Yang, D., Zhou, D.: Best Practices and Lessons Learned on Synthetic Data. In: First Conference on Language Modeling (2024), https://openreview.net/forum?id=OJaWBhh61C
10. Liu, Y., Yao, Y., Ton, J.F., Zhang, X., Guo, R., Cheng, H., Klochkov, Y., Taufiq, M.F., Li, H.: Trustworthy LLMs: a Survey and Guideline for Evaluating Large Language Models' Alignment (Mar 2024). https://doi.org/10.48550/arXiv.2308.05374, http://arxiv.org/abs/2308.05374, arXiv:2308.05374

11. Mullick, A., Bose, S., Saha, R., Bhowmick, A.K., Goyal, P., Ganguly, N., Dey, P., Kokku, R.: On the persona-based summarization of domain-specific documents. arXiv preprint arXiv:2406.03986 (2024)
12. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A.: Training language models to follow instructions with human feedback. Advances in neural information processing systems **35**, 27730–27744 (2022), https://proceedings.neurips.cc/paper_files/paper/2022/hash/b1efde53be364a73914f58805a001731-Abstract-Conference.html
13. Pan, J., Shen, W., Huang, S., Zhou, Q., Zhang, Y.: Pre-DPO: Improving Data Utilization in Direct Preference Optimization Using a Guiding Reference Model. arXiv preprint arXiv:2504.15843 (2025)
14. Rafailov, R., Sharma, A., Mitchell, E., Manning, C.D., Ermon, S., Finn, C.: Direct preference optimization: Your language model is secretly a reward model. Advances in Neural Information Processing Systems **36** (2024), https://proceedings.neurips.cc/paper_files/paper/2023/hash/a85b405ed65c6477a4fe8302b5e06ce7-Abstract-Conference.html
15. Schuller, A., Janssen, D., Blumenröther, J., Probst, T.M., Schmidt, M., Kumar, C.: Generating personas using LLMs and assessing their viability. In: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems. pp. 1–7. ACM, Honolulu HI USA (May 2024). https://doi.org/10.1145/3613905.3650860, https://dl.acm.org/doi/10.1145/3613905.3650860
16. Tseng, Y.M., Huang, Y.C., Hsiao, T.Y., Chen, W.L., Huang, C.W., Meng, Y., Chen, Y.N.: Two tales of persona in llms: A survey of role-playing and personalization. arXiv preprint arXiv:2406.01171 (2024)
17. Yun, L., Yun, S., Xue, H.: Improving citizen-government interactions with generative artificial intelligence: Novel human-computer interaction strategies for policy understanding through large language models. PloS one **19**(12), e0311410 (2024), https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0311410, publisher: Public Library of Science San Francisco, CA USA