Evaluating Direct Preference Optimization for Personalizing German Automatic Text Simplifications for Persons with Intellectual Disabilities

Yingqiang Gao[†UZH] **Kaede Johnson**[EPFL]
David Fröhlich[capito] **Luisa Carrer**[zh aw] **Sarah Ebling**[†UZH]

[UZH]Department of Computational Linguistics, University of Zurich, Switzerland
[EPFL]School of Computer and Communication Sciences, EPFL, Switzerland
[zh aw]School of Applied Linguistics, Zurich University of Applied Sciences, Switzerland
[capito]capito.ai, Graz, Austria

{yingqiang.gao, ebling}@cl.uzh.ch

Abstract

Automatic text simplification (ATS) aims to enhance language accessibility for various target groups, particularly persons with intellectual disabilities. Recent advancements in large language models (LLMs) have substantially improved the quality of machine-generated text simplifications, however, existing LLM-based ATS systems do not incorporate preference feedback during post-training, resulting in a lack of personalization tailored to the specific needs of target group persons. In this work, we propose an ATS personalization framework using direct preference optimization (DPO). Specifically, we post-trained LLM-based ATS models using human feedback collected from persons with intellectual disabilities, reflecting their preferences of paired text simplifications generated by mainstream LLMs. Our pipeline for developing personalized LLM-based ATS systems encompasses data collection, model selection, supervised fine-tuning (SFT) and DPO post-training, and result evaluation. Our findings underscore the necessity of active participation of target group persons in designing personalized inclusive AI solutions aligned with human preferences.

Dataset Code

1 Introduction

Automatic text simplification (ATS) is a natural language processing (NLP) task that converts a standard-language text into an easier-to-understand version by improving text readability, increasing lexical and syntactic simplicity, and optimizing content complexity (Hansen-Schirra et al., 2020; Al-Thanyyan and Azmi, 2021). Nowadays most often being tackled through AI approaches, ATS is oriented at diverse target groups, such as non-native language learners (Higasa et al., 2023, 2024), persons with low literacy (Fu et al., 2024), persons with hearing difficulties (Alonzo et al., 2020, 2024), and persons with intellectual disabilities (Säuberli et al., 2024). Among these target groups, persons with intellectual disabilities may encounter fundamental challenges in comprehending complex sentence structures, domain-specific jargon, implicit metaphors, and high information density, all of which can pose significant barriers to access of daily-life information flows (Säuberli et al., 2024).

Research in ATS and accessibility technology has focused on enhancing the diversity of machine-generated simplifications, incorporating techniques such as text splitting, semantic paraphrasing, lexical substitution, and information deletion (Alva-Manchego et al., 2020; Maddela et al., 2021; Yamaguchi et al., 2023; Vendeville et al., 2025), as well as developing more robust evaluation metrics that better match with human judgment and perception (Maddela et al., 2023; Cripwell et al., 2023; Heineman et al., 2023; Souayed et al., 2025; Korobeynikova et al., 2026). However, in practice, the preferences of persons with intellectual disabilities are often overlooked in the development of inclusive AI technologies, largely because they are rarely consulted to provide feedback on AI-generated text simplifications tailored to their individual preferences (Birhane et al., 2022). Moreover, communication barriers are often bidirectional for them, affecting both comprehension and expression (Cashin et al., 2024); as a result, their involvement in ATS research is usually confined to late-stage human evaluation, with feedback that rarely informs further system refinement.

In this work, we aim to personalize LLM-based ATS models for persons with intellectual disabilities (henceforth referred to as the target group) within a lightweight and inclusive framework, in-

[†]Corresponding authors.

Proceedings of the Sixth Workshop on Language Technology for Equality, Diversity, Inclusion, pages 43–62

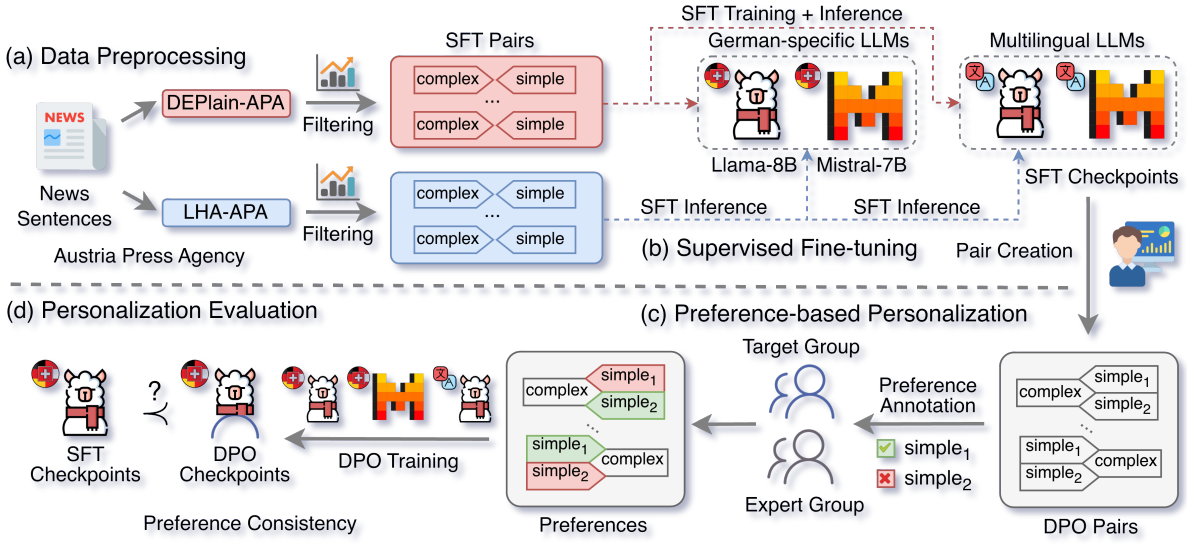


Figure 1: **Our personalization pipeline for LLM-based ATS models**. (a) Data filtering of sentence-level complex–simple pairs from two datasets; (b) supervised fine-tuning of German-specific and multilingual LLMs; (c) preference-based personalization via DPO post-training using preference data from target and expert group annotators; (d) evaluation comparing DPO checkpoints against their SFT precursors for personalization success.

corporating human-in-the-loop (HITL; Wu et al. (2022); Mosqueira-Rey et al. (2023)) participation. As our primary personalization methodology, we propose investigating direct preference optimization (DPO; Rafailov et al. (2023)), an LLM alignment algorithm that does not require explicit reward modeling. By integrating target group participants throughout all phases and adhering to the validate-annotate-evaluate HITL principle, we aim not only to develop LLM-based ATS models that are post-trained on the preferences of the target group persons but also to establish an inclusive workflow for personalizing LLM-based ATS models.

The **main contributions** of this work are: (1) an inclusive workflow for collecting human preference data from target group persons and text simplification experts; (2) HF4ATS, the largest German human preference dataset for ATS pairs generated by mainstream LLMs; (3) open-source ATS models post-trained with DPO on HF4ATS; (4) extensive experiments analyzing model- and data-level factors in group-level ATS personalization.

2 Related Works

While most ATS models were trained on English data (Scarton and Specia, 2018; Sheang and Saggion, 2021; Agrawal and Carpuat, 2024), German ATS research has gained increasing attention in recent years, driven by active political and legal initiatives in German-speaking countries (Ebling et al., 2022). These efforts have significantly advanced German ATS research, particularly in areas such as dataset construction (Klaper et al., 2013; Battisti et al., 2020; Säuberli et al., 2020; Gonzales et al., 2021; Aumiller and Gertz, 2022; Seiffe et al., 2022; Toborek et al., 2023; Stodden et al., 2023; Klöser et al., 2024), alignment of texts (Spring et al., 2022, 2023), and training of models (Spring et al., 2021; Anschütz et al., 2023; Hewett et al., 2024).

Preference learning algorithm such as DPO is a key approach to personalizing ATS models, alongside methods such as personalized prompting and personalized adaptation (Liu et al., 2025). It is a method that directly incorporates subjective human feedback into the personalization process without the need of identifiable user profile data (Zhao et al., 2025), thus is ideal for inclusive AI research as the construction of user profiles of target group persons is considered unethical and, in many jurisdictions, legally prohibited.

In the context of ATS, given a preference dataset $\mathcal{D}$ consisting of triples of $(x, y_w, y_l) \sim \mathcal{D}$, where x is the complex text, y_w is the preferred LLM-generated text simplification and y_l is the dispreferred counterpart, DPO aims at learning a policy model π_θ that assigns a higher preference score to y_w than to y_l. The human preference can be modeled as probabilistic ranking with the Bradley-Terry

model (Bradley and Terry, 1952)

$$P(y_w \succ y_l \mid x) = \sigma\big(R_\psi(x, y_w) - R_\psi(x, y_l)\big) = \frac{\exp(R_\psi(x, y_w))}{\exp(R_\psi(x, y_w)) + \exp(R_\psi(x, y_l))},$$

where $\sigma(\cdot)$ is the sigmoid function. With some reparametrization trick that essentially gets rid of the explicit reward modeling $R_\psi(x, y)$ and estimates the implicit reward directly from the training samples, the DPO training objective becomes

$$\mathcal{L}_{\text{DPO}}(\pi_\theta; \pi_{\text{ref}}) = -\mathbb{E}\left[\log\sigma\left(\beta\log\frac{\pi_\theta(y_w \mid x)}{\pi_{\text{ref}}(y_w \mid x)} - \beta\log\frac{\pi_\theta(y_l \mid x)}{\pi_{\text{ref}}(y_l \mid x)}\right)\right],$$

where the parameter β actively regulates the deviation of the policy model π_θ from the reference model π_{ref}, ensuring that the log-odd differences remain within a controlled range. This log-odd difference is the so-called implicit reward margin

$$\hat{r}(x, y_w, y_l) = \beta\left(\log\frac{\pi_\theta(y_w|x)}{\pi_{\text{ref}}(y_w|x)} - \log\frac{\pi_\theta(y_l|x)}{\pi_{\text{ref}}(y_l|x)}\right).$$

A common practice when post-training with DPO is to initialize both π_{ref} and π_θ with the SFT model checkpoint and freeze π_{ref} during post-training, so that the gradient $\nabla_\theta\mathcal{L}_{\text{DPO}}(\pi_\theta; \pi_{\text{ref}})$ will only be back-propagated to the policy model π_θ.

In this work, we focus on user-agnostic, group-level LLM personalization using DPO, aiming to develop LLM-based ATS systems that cater to the needs of target group persons as a whole. Our approach relies solely on preference data over LLM-generated text simplifications collected from target group persons, ensuring group-level personalization without any user profiling. We propose the following research questions (RQs):

RQ1. Can DPO post-training with pairwise human preferences further improve the quality of ATS, as measured by automatic evaluation metrics?

RQ2. To what extent do factors such as preference source, information equality, and generalization of LLMs influence the effectiveness of DPO post-training?

RQ3. Can DPO post-training enable successful group-level personalization of ATS models?

Next, in Section 3, we introduce our research pipeline, including 1) the introduction of HF4ATS, our curated human preference dataset designed for post-training German LLM-based ATS models; 2) the LLM models we used for training; and 3) the training and hyper-parameter tuning of the LLM-based ATS models.

3 Data, Model, and Method

We introduce Human Feedback for Automatic Text Simplification (HF4ATS), a dataset designed to enhance German ATS through learning with human preferences. To the best of our knowledge, HF4ATS is the first and largest German-language preference dataset collected directly from the target group for this purpose. HF4ATS consists of two key datasets: (1) HF4ATS-SFT ($\mathcal{D}_{\text{SFT}}$), a dataset of complex-simple sentence pairs for supervised fine-tuning (SFT) of German LLM-based ATS models, and (2) HF4ATS-DPO ($\mathcal{D}_{\text{DPO}}$), an ATS preference pair dataset annotated by native German speakers. $\mathcal{D}_{\text{DPO}}$ can be adapted for use in several preference alignment frameworks. In this work, we use it to post-train ATS models with DPO.

3.1 SFT Phase

SFT Model Selection. To develop robust ATS models, we started with four open-source LLMs as backbones, prioritizing models that (1) were either multilingual or specifically tuned for the German language; and (2) had been instruction-tuned to effectively follow text simplification guidelines. Based on these criteria, we chose LLMs around 8 billion parameters, including Llama-3.1-8B-Instruct, DiscoLeo-Llama-3-8B-Instruct, Mistral-7B-Instruct, and LeoLM-Mistral-7B-Chat.

Data Filtering. We curated HF4ATS-SFT from the DEPLAIN (Stodden et al., 2023) dataset containing parallel complex-simple pairs professionally written and manually aligned by human. To ensure the inclusion of high-quality pairs during SFT, we incorporated the following data filtering steps: First, we excluded pairs with many-to-many or many-to-one mappings, retaining only those with one-to-one or one-to-many mappings. This selection ensured a focus on pairs that did not introduce overly dense information. Second, we sought to remove pairs in which the simplified text was not entailed by the corresponding complex text.

We employed a semantics-based approach by computing the cosine similarity between complex and simplified texts. Specifically, we utilized a

Table 1: **Annotator agreement scores** measured for target and expert group participants.

(a) **Intra-annotator agreement (Intra-AA)** for target and expert groups, measured using Cohen's Kappa (Cohen, 1960). NA indicates unavailable data due to missing Intra-AA pairs or exclusion from DPO post-training.

Target						Expert	
id	κ	**id**	κ	**id**	κ	**id**	κ
ta01	-0.037	ta06	NA	ta11	0.063	ea01	0.420
ta02	0.040	ta07	-0.045	ta12	0.155	ea02	0.755
ta03	-0.026	ta08	NA	ta13	NA	ea03	0.745
ta04	0.168	ta09	NA	ta14	0.008	ea04	0.376
ta05	0.300	ta10	0.065	ta15	NA		

(b) **Inter-annotator agreement (Inter-AA)** for target and expert groups, measured using Krippendorff's Alpha (Krippendorff, 2004). We report Inter-AA scores for pairs annotated by at least four annotators, stratified by the generating SFT checkpoint.

SFT Checkpoint	α	
	Target	**Expert**
DiscoLeo-Llama-SFT-2800	0.019	0.324
Llama-SFT-2400	0.003	0.248
LeoLM-Mistral-SFT-1600	-0.016	0.536

Table 2: **Overall statistics of the HF4ATS dataset**. We report SFT and DPO splits derived from target and expert preference annotations; Pref. % 1st denotes the rate at which the left-displayed simplification was preferred.

Dataset	**# Instances**			**# words**			**Pref. % of 1st. ATS**	
	Train	**Dev**	**Test**	**Train**	**Dev**	**Test**	**Target**	**Expert**
HF4ATS-SFT ($\mathcal{D}_{\text{SFT}}$)	3,600	800	800	252,285	55,208	55,852	-	-
HF4ATS-DPO ($\mathcal{D}_{\text{DPO}}$)	4,814	602	602	372,687	45,857	45,992	36.65	47.44

Sentence-BERT model (Reimers and Gurevych, 2019) fine-tuned for German language (cross-en-de-roberta-sentence-transformer; May (2020)). Based on empirical analysis, we filtered out 591 text pairs with cosine similarity scores below the threshold of 0.5, which is a deliberately relaxed filter given the involvement of human input when creating preference pairs at a later stage.

We then further removed pairs in which the simplified texts were overly similar to the complex texts, as indicated by a high degree of n-gram overlap. To address this issue, we applied an additional heuristic filter, removing 2,322 pairs whose F1 score across ROUGE-1, ROUGE-2, and ROUGE-L (Lin, 2004) exceeded a threshold of 0.8. Lastly, to ensure task completion would only introduce a moderate amount of information, we removed 116 pairs in which the simplification exceeded 30 words. In Appendix A we show examples of these two classes of low-quality pairs.

The remaining set amounted to 9,359 pairs following our four-step data filtering process. To create the training, development, and test sets for SFT, we applied a stratified approach with a 70%-15%-15% split. Specifically, we randomly allocated 3,600 train, 800 development, and 800 test pairs from a subset of 5,200 pairs purposefully sampled to balance the sentence length distribution. To achieve this, we employed Gaussian sampling based on the word count of the complex texts. Formally, for a given complex text x, the sampling weight w_x is defined as $w_x = \exp\big(-(|x| - 13)^2/(2\sigma^2)\big)$ where $|x|$ denotes the word count of the complex text and the standard deviation σ is set to 3. This sampling formula yielded a less skewed word count distribution for complex texts in our final subset of 5,200 pairs.

Input Prompts. In collaboration with a native German text simplification expert, we developed ten prompts for SFT. Eight of these prompts were later re-used for DPO post-training (see Appendix A and B). To ensure a consistent response format and facilitate post-processing, all prompts include the instruction: "*Bitte gib nur die Vereinfachung an, ohne Einleitung, Alternativen und Kommentare*". (English: "*Please provide only the simplification, without introduction, alternatives, or comments*".)

SFT Training. We performed SFT using the 3,600 training pairs from HF4ATS-SFT. Following the findings of (Zhou et al., 2023), which suggest that the optimal SFT checkpoint may emerge after a few thousand training instances, we periodically evaluated model performance on the 800-pair development set. Specifically, evaluations were conducted after every 400 training instances during cross-model comparisons and every 448 training instances during hyper-parameter tuning.

SFT Checkpoint Evaluation. We employed an offline evaluation strategy to assess 36 SFT checkpoints saved at regular training intervals. The eval-

uation focused on performance across three key dimensions:

- **Simplification Quality**: We computed BERTScore F1 measure (Zhang et al., 2019), BLEU (Papineni et al., 2002), and SARI (Xu et al., 2016) on the development set. SARI was selected as the most salient metric for evaluation.
- **Simplification Readability**: We assessed average word count, Flesh Reading Ease (Kincaid et al., 1975), and Wiener Sachtextformel Variant 4 ($WSTF_4$; Bamberger and Vanacek (1984)) (English: Vienna Formula) on the development set. A $WSTF_4$ score of 4 indicates a very simple text, while a score of 15 indicates a very complex text. We selected $WSTF_4$ as the most salient readability metric over Flesch Reading Ease because it was specifically designed for non-fiction, German-language text.
- **SFT Implementation Quality**: We assessed the mirror rate, i.e., the ratio of generated simplifications being identical to the complex input.

Selected final SFT checkpoints are listed in Appendix A in Table 7.

3.2 DPO Phase

Preference Pairs Creation. After training the SFT models on HF4ATS-SFT, we created the ATS pairs for HF4ATS-DPO by first performing inferences with the selected SFT checkpoints. We generated 20 text simplifications per SFT checkpoint, with one of eight prompts was assigned to each complex sentence at random (see Appendix B), and we varied temperature and the top-p sampling parameter to achieve inference variety with four decoding configurations.

Once inference was completed, 13 proficient German-speaking human pair creators (CEFR-level C1 and above) with strong backgrounds in computational linguistics research reviewed the automatic simplifications to construct ATS pairs for preference annotation. We have programmed a Python script to facilitate the annotation process.

Preference Pair Annotation. To reduce the cognitive burden of complex crowd-sourcing interfaces, we developed a minimal web application for collecting preferences from target and expert participants (see Figure 5 in Appendix A). The tool was deployed on the university cloud, and participants were compensated at 10 EUR/hour (target group) and 120 EUR/hour (expert group).

To measure preference consistency, we injected repeated pairs (within-annotator) and shared pairs (within-group) into the annotation workflow, targeting 40–45 of each per participant, corresponding to about 10% of annotations. These pairs were randomly interleaved and presented with randomized order, and were used to compute intra- and inter-annotator agreement (Intra-AA and Inter-AA) separately for the target and expert groups (see Table 1). Table 2 summarizes the HF4ATS dataset: 70% of the 5,200 HF4ATS-SFT pairs were used for SFT, with development data for checkpoint selection and test data for evaluating DPO against SFT models. HF4ATS-DPO consists of 6,018 preference pairs (3,009 unique), each annotated by both groups; 80% were used for DPO training, 15% for model selection, and 15% for held-out evaluation. We also report left-side preference rates in Table 1, expected to be near 50% due to randomization.

DPO Post-training. Starting from pre-trained LLMs, we first conducted SFT on $\mathcal{D}_{\text{SFT}}^{\text{train}}$ and selected the best-performing SFT checkpoints based on offline evaluation of $\mathcal{D}_{\text{SFT}}^{\text{dev}}$. These SFT checkpoints served as the initialization for DPO policy models and were used as frozen reference models during the DPO phase. The DPO policy models were trained on $\mathcal{D}_{\text{DPO}}^{\text{train}}$ using either target or expert annotations, with the best DPO checkpoints selected by offline evaluation of $\mathcal{D}_{\text{DPO}}^{\text{dev}}$ (with win rates, see the coming Section 4). Apart from changing the training batch size to 8 from 16, we made no changes to the SFT phase training parameters described in Appendix C. Our β was 0.1.

To study the impact of various factors on DPO post-training, we took different subsets of preference pairs from $\mathcal{D}_{\text{DPO}}^{\text{train}}$ and trained DPO policy models on each. These subsets are denoted as follows:

- **all**: all annotated preference pairs;
- **all**$_{=}$: all preference pairs with equal information, as indicated by pair creators during annotation;
- **LLM**$_{=}$: all preference pairs generated by the SFT checkpoint being post-trained with DPO;
- **max. Intra-AA**: all preference pairs annotated

Table 3: **Automatic ATS quality evaluation** for SFT and DPO checkpoints on $\mathcal{D}^{\text{test}}_{\text{SFT}}$ (cols. 1–3) and $\mathcal{D}^{\text{test}}_{\text{DPO}}$ (col. 4). We report mean scores with standard deviations. DPO gains and drops over SFT are highlighted in blue and red, respectively, with the same scheme for win rates (threshold 0.50). Best scores per metric are bolded.

Checkpoint	Reference-based Metrics		Reference-free Metrics	
	SARI	BERTScore	WSTF_4	Win Rate
SFT Baselines				
DiscoLeo-Llama-SFT-2800	**46.22** ± 13.47	**0.9049** ± 0.054	6.515 ± 3.24	-
Llama-SFT-2400	45.94 ± 13.52	**0.8865** ± 0.054	5.852 ± 2.90	-
LeoLM-Mistral-SFT-1600	44.55 ± 13.95	**0.9054** ± 0.056	6.207 ± 3.55	-
DPO Target Checkpoints				
DiscoLeo-Llama-DPO-2160	44.41 ± 11.60	0.7854 ± 0.081	6.194 ± 2.17	0.5211
Llama-DPO-1440	46.11 ± 11.60	0.8756 ± 0.055	5.796 ± 2.56	0.5145
LeoLM-Mistral-DPO-1560	43.73 ± 13.36	0.7781 ± 0.113	5.683 ± 2.80	0.4382
DPO Expert Checkpoints				
DiscoLeo-Llama-DPO-1080	42.50 ± 13.28	0.7814 ± 0.059	**4.031** ± 2.68	**0.6118**
Llama-DPO-1320	**46.45** ± 12.25	0.8441 ± 0.052	**4.676** ± 2.59	**0.6099**
LeoLM-Mistral-DPO-2280	**44.92** ± 14.03	0.8340 ± 0.082	**4.802** ± 2.94	**0.6118**

by the four target group or two expert group participants exhibiting the highest Intra-AA scores;

- **max. Inter-AA**: all preference pairs annotated by the four target group or two expert group participants exhibiting the highest Inter-AA scores.

Table 7 in Appendix A lists all winning model checkpoints involved in our overall training pipeline. The numbers following the SFT and DPO checkpoints indicate the number of training instances associated with each checkpoint. Our final evaluation is concentrated on the "all" subset (which is essentially responsible for the winning DPO checkpoints referenced in Table 7).

4 Evaluation

4.1 Automatic Evaluation

Each subset of preferences listed above resulted in six DPO post-trainings, one for each combination of SFT checkpoint and annotator group. To evaluate the six winning DPO checkpoints trained on all group-appropriate preference pairs, we used greedy decoding to generate inferences for all 800 complex sentences in $\mathcal{D}^{\text{test}}_{\text{SFT}}$, calculated the reference-based metrics SARI (Xu et al., 2016) and BERTScore (Zhang et al., 2019) as well as the reference-free metric WSTF_4 (Bamberger and Vanacek, 1984), and compared these metrics to the same metrics calculated with our winning SFT checkpoints.

We then utilized $\mathcal{D}^{\text{test}}_{\text{DPO}}$ to calculate win rates (Rafailov et al., 2023) for all 30 winning DPO checkpoints from our trainings. Specifically, given $\mathcal{D}^{\text{test}}_{\text{DPO}} = \{(x^i, y^i_w, y^i_l)\}_{i=1}^N$, the win rate $W_{y_w \succ y_l}$ is defined as the proportion of preference pairs for which the DPO checkpoint assigns a higher implicit reward to y_w than y_l. That is,

$$W_{y_w \succ y_l} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}[\hat{r}(x^i, y^i_w, y^i_l) > 0],$$

where $\hat{r}(x, y_w, y_l)$ denotes the implicit reward margin computed with the policy model and reference model and $\mathbf{1}[\cdot]$ is the indicator function, which equals 1 if the condition holds and 0 otherwise. A win rate above 0.50 indicates that the DPO policy model more often assigns higher preference score to human-preferred simplifications than dispreferred simplifications, thereby achieving closer alignment with human judgments of ATS quality.

4.2 Human Evaluation

To test whether participants preferred models personalized with their own supervision, we computed the supremacy score $S_{\text{DPO} \succ \text{SFT}}$ for six winning DPO checkpoints trained on group-specific preferences. The score measures the proportion of cases in which a DPO model's simplification is preferred over its SFT precursor for the same input $x^i \in \mathcal{D}^{\text{test}}_{\text{DPO}}$, and is defined as

$$S_{\text{DPO} \succ \text{SFT}} = \frac{1}{N} \sum_{i=1}^{N} h(x^i), \text{ where}$$

$$\forall x^i \in \mathcal{D}^{\text{test}}_{\text{DPO}}, \quad h(x^i) = \begin{cases} 1, & y_{\text{DPO}}(x^i) \succ y_{\text{SFT}}(x^i) \\ 0, & \text{otherwise.} \end{cases}$$

In this context, the successful personalization of a DPO model would thus be indicated by a supremacy score greater than 50%.

Table 4: **Win rates) on $\mathcal{D}^{\text{test}}_{\text{DPO}}$ for winning DPO checkpoints trained on different $\mathcal{D}^{\text{train}}_{\text{DPO}}$ subsets**. Checkpoints were selected by highest dev-set win rate; bracketed values denote changes relative to the **all** setting.

DPO Checkpoint	all	all$_{=}$	LLM$_{=}$	max. Intra-AA	max. Inter-AA
Target	**Baseline**	**Subsets of HF4ATS-DPO training data**			
DiscoLeo-Llama-DPO	0.5211	0.4708 (9.65% ↓)	0.4861 (6.72% ↓)	0.5078 (2.55% ↓)	**0.5431** (4.22% ↑)
Llama-DPO	0.5145	0.4833 (6.06% ↓)	0.5385 (4.66% ↑)	**0.6094** (18.45% ↑)	0.5153 (0.16% ↑)
LeoLM-Mistral-DPO	0.4382	0.4625 (5.55% ↑)	0.4848 (10.63% ↑)	**0.5781** (31.93% ↑)	0.4917 (12.21% ↑)
Expert	**Baseline**	**Subsets of HF4ATS-DPO training data**			
DiscoLeo-Llama-DPO	0.6118	0.6333 (3.51% ↑)	0.5111 (16.46% ↓)	0.6118 (0.00% =)	**0.6438** (5.23% ↑)
Llama-DPO	0.6099	0.5833 (4.36% ↓)	**0.6538** (7.20% ↑)	0.6382 (4.64% ↑)	0.6313 (3.51% ↑)
LeoLM-Mistral-DPO	0.6118	0.6125 (0.11% ↑)	0.5871 (4.04% ↓)	**0.6776** (10.76% ↑)	0.5625 (8.06% ↓)

To compute the DPO supremacy score, for every complex sentence x in $\mathcal{D}^{\text{test}}_{\text{DPO}}$, we generated five text simplifications with the DPO checkpoint and five text simplifications with the corresponding SFT checkpoint using top-p sampling ($p = 0.9$). We then engaged one pair creator who had previously created pairs for HF4ATS-DPO to assemble from these inferences a final set of 300 ATS pairs, 50 for each of our six winning DPO checkpoints for the final calculation.

5 Results and Discussion

5.1 Quality Assessment of Generated ATS

To answer RQ1, we present an automatic quality assessment of ATS outputs generated by SFT and DPO checkpoints, evaluated using both reference-based and reference-free metrics (see Table 3). The best performance on SARI and WSTF_4 was achieved with the DPO checkpoints, whereas the highest BERTScore was obtained using the SFT checkpoints.

DPO post-training can improve readability (WSTF_4) across models and supervision sources, but leads to lower BERTScores, indicating semantic drift. Effects on faithfulness (SARI) are mixed: expert-supervised models largely preserve or recover baseline SARI, whereas target-supervised models sometimes show declines. Expert supervision also yields consistently higher win rates, suggesting greater preference consistency, which likely underlies the observed trade-off between simplification strength and meaning preservation.

Recent studies have revealed several core limitations of standard DPO post-training. These include a tendency to overfit to sparse or noisy preference signals (Fisch et al., 2024), catastrophic forgetting in continual learning settings (Qi et al., 2024), and the potential to undermine generalization and robustness in LLMs (Hu et al., 2024). In our study, these limitations may help explain the observed decline in DPO models' faithfulness with respect to the DEPLAIN data in $\mathcal{D}^{\text{test}}_{\text{SFT}}$. Nonetheless, our results highlight the critical role of preference consistency in the effectiveness of DPO for personalized ATS modeling.

The win rates in Table 3 as well as Inter- and Intra-AA scores in Table 1 indicate that target group preferences are more diverse or inconsistent than expert group preferences. It might be the case that offline LLM alignment methods such as DPO, which lack explicit reward modeling, are suboptimal for capturing nuanced preferences over text simplifications when trained with such data.

5.2 Factors on DPO Post-training

To answer RQ2, in Table 4 we show that DPO effectiveness is primarily driven by preference consistency rather than model- or perception-specific factors: training on high-consistency subsets (maximized Intra- or Inter-AA) yields consistent win-rate gains across models and supervision sources, including the largest observed improvement (+31.93%) for the target-group LeoLM-Mistral-DPO model. In contrast, subsets based on information equity or model matching often degrade performance. Expert-supervised DPO models consistently outperform target-supervised counterparts, never falling below the 0.50 win-rate threshold and exhibiting more stable training dynamics (see Figure 7), while target-supervised models show lower, noisier gains and smaller reward margins. Cross-group evaluation further reveals limited transfer from expert to target preferences, indicating that while preference consistency is crucial for DPO, offline alignment may be ill-suited for personalizing ATS systems.

Table 5 reports annotator-level DPO supremacy scores. Most target annotators prefer SFT over

Table 5: **DPO supremacy scores by LLM backbone**, measuring the proportion of cases where DPO outputs are preferred over SFT. Target-group results use the four most consistent annotators. Scores above and below 0.5 are highlighted in blue and red; asterisks denote significant DPO supremacy (binomial test, $p < 0.05$).

SFT Checkpoint	DPO Checkpoint	DPO Supremacy Score			
Baseline	**Target**	ta04	ta05	ta10	ta12
DiscoLeo-Llama-SFT-2800	DiscoLeo-Llama-DPO-2160	0.36	0.40	0.56	0.46
Llama-SFT-2400	Llama-DPO-1440	0.40	0.30	0.38	0.50
LeoLM-Mistral-SFT-1600	LeoLM-Mistral-DPO-1560	0.42	0.48	0.58	0.40
Baseline	**Expert**	ea01	ea02	ea03	ea04
DiscoLeo-Llama-SFT-2800	DiscoLeo-Llama-DPO-1080*,**	0.74	0.46	0.68	0.72
Llama-SFT-2400	Llama-DPO-1320	0.60	0.30	0.54	0.56
LeoLM-Mistral-SFT-1600	LeoLM-Mistral-DPO-2280*	0.68	0.44	0.52	0.56

DPO, with only one favoring DPO and mild model-specific variation, whereas most expert annotators consistently prefer DPO across backbones, with DiscoLeo-Llama-DPO-1080 performing best for all experts. Across analyses, DPO succeeds only when preference signals are consistent: expert-supervised models benefit reliably, while target-group supervision yields weaker, noisier gains and often favors SFT over DPO. These results indicate that preference consistency, rather than model choice or data heuristics, is the key determinant of DPO effectiveness.

5.3 Group-Level Personalization Success

To answer RQ3, and to verify group-level personalization success, we conducted one-sided binomial tests [*] for each model at the evaluation group level. Assuming each pair was evaluated independently, we defined the group-level preference for each test pair as the majority vote among the evaluators (tied pairs were assigned randomly). Our goal was to determine whether, across all 50 DPO supremacy test pairs, there was a statistically significant collective preference for ATS outputs generated by the DPO checkpoints. The asterisks in Table 5 indicate which models had a group-level DPO supremacy greater than 0.50 with a p-value less than 0.05. For the expert group, we indicate results for tests both including ($**$) and excluding ($*$) the outlier ea02.

Given the observed personalization failure in target group, DPO may be ill-suited for ATS alignment. While DPO can in principle learn from low-signal pairs at scale, such data collection is impractical for target groups, motivating exploration of lower-burden alternatives such as KTO (Ethayarajh et al., 2024), which replaces pairwise comparison with single-output judgments and may better capture human biases;restructuring HF4ATS-DPO to support other RLHF or even RLVR methods (Shao et al., 2024; Guo et al., 2025; Wen et al., 2026) is left for future work.

6 Conclusion and Future Work

In this work, we studied the effectiveness of DPO for personalizing LLM-based ATS to better reflect the preferences of persons with intellectual disabilities. To enable this, we developed a lightweight and accessible workflow for collecting pairwise human preferences from both target users and expert participants. We introduced HF4ATS, the first and largest German-language ATS dataset combining preference annotations from both target and expert group. We trained and analyzed models on various subsets of this dataset, systematically investigating how preference consistency, preference source, and LLM engagement impact personalization outcomes. Our findings expose a key limitation of preference-based LLM personalization: methods like DPO depend on consistent supervision, which is hard to obtain from target groups with diverse or uncertain preferences.

In future work, we will explore alternative LLM alignment techniques and personalization strategies that leverage small but high-quality human preference data. More broadly, we advocate for inclusive AI research that centers the voices of persons with disabilities, not merely as end-users or evaluators, but as active co-creators throughout the whole research process.

Limitations

Our study has several limitations. First, while HF4ATS is the largest German ATS preference dataset to date, the number of target group anno-

[*]We used `spacy` for the test, BSD-3-Clause license, available at https://github.com/scipy/scipy.

tators remains limited, and preference variability within this group may not be fully captured. Second, our experiments focus on group-level rather than individual-level personalization; stronger effects may emerge with more fine-grained or longitudinal preference modeling. Third, we evaluate DPO as a representative offline preference alignment method, but our findings may not generalize to alternative alignment frameworks or online learning settings. Finally, the cognitive demands of pairwise preference annotation constrained data scale, which may have limited DPO's effectiveness; future work should explore lower-burden feedback paradigms and alignment objectives better suited to accessibility contexts.

Ethical Considerations

This study involved the collection of human preference data from persons with intellectual disabilities. Ethical approval for all data collection procedures was obtained from the ethics committee of University of Zurich prior to the start of the study. Participation was voluntary, and all participants were informed about the study objectives and procedures in an accessible manner. Data were collected and stored in accordance with applicable data protection regulations. Particular care was taken to minimize cognitive burden during annotation and to ensure fair compensation for all participants.

Acknowledgment

This work was funded by the Swiss Innovation Agency (Innosuisse) Flagship Inclusive Information and Communication Technologies (IICT) under grant agreement PFFS-21-47. We sincerely thank all study participants, especially those from the target group in Austria.

References

Sweta Agrawal and Marine Carpuat. 2024. Do Text Simplification Systems Preserve Meaning? A Human Evaluation via Reading Comprehension. *Transactions of the Association for Computational Linguistics*, 12:432–448.

Suha S Al-Thanyyan and Aqil M Azmi. 2021. Automated Text Simplification: A Survey. *ACM Computing Surveys (CSUR)*, 54(2):1–36.

Oliver Alonzo, Sooyeon Lee, Akhter Al Amin, Mounica Maddela, Wei Xu, and Matt Huenerfauth. 2024. Design and Evaluation of an Automatic Text Simplification Prototype with Deaf and Hard-of-hearing Readers. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–18.

Oliver Alonzo, Matthew Seita, Abraham Glasser, and Matt Huenerfauth. 2020. Automatic Text Simplification Tools for Deaf and Hard of Hearing Adults: Benefits of Lexical Simplification and Providing Users with Autonomy. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13.

Fernando Alva-Manchego, Louis Martin, Antoine Bordes, Carolina Scarton, Benoît Sagot, and Lucia Specia. 2020. ASSET: A Dataset for Tuning and Evaluation of Sentence Simplification Models with Multiple Rewriting Transformations. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4668–4679.

Miriam Anschütz, Joshua Oehms, Thomas Wimmer, Bartłomiej Jezierski, and Georg Groh. 2023. Language Models for German Text Simplification: Overcoming Parallel Data Scarcity through Style-specific Pre-training. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 1147–1158.

Dennis Aumiller and Michael Gertz. 2022. Klexikon: A German Dataset for Joint Summarization and Simplification. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 2693–2701.

Richard Bamberger and Erich Vanacek. 1984. *Lesen-Verstehen-Lernen-Schreiben*. Diesterweg.

Alessia Battisti, Dominik Pfütze, Andreas Säuberli, Marek Kostrzewa, and Sarah Ebling. 2020. A Corpus for Automatic Readability Assessment and Text Simplification of German. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 3302–3311.

Abeba Birhane, William Isaac, Vinodkumar Prabhakaran, Mark Diaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed. 2022. Power to the people? Opportunities and challenges for participatory AI. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, pages 1–8.

Ralph Allan Bradley and Milton E Terry. 1952. Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika*, 39(3/4):324–345.

Andrew Cashin, Julia Morphet, Nathan J Wilson, and Amy Pracilio. 2024. Barriers to Communication with People with Developmental Disabilities: A Reflexive Thematic Analysis. *Nursing & health sciences*, 26(1):e13103.

Yanda Chen, Ruiqi Zhong, Sheng Zha, George Karypis, and He He. 2022. Meta-learning via Language Model In-context Tuning. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 719–730.

Jacob Cohen. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and psychological measurement*, 20(1):37–46.

Liam Cripwell, Joël Legrand, and Claire Gardent. 2023. Simplicity Level Estimate (SLE): A Learned Reference-Less Metric for Sentence Simplification. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 12053–12059.

Sarah Ebling, Alessia Battisti, Marek Kostrzewa, Dominik Pfütze, Annette Rios, Andreas Säuberli, and Nicolas Spring. 2022. Automatic Text Simplification for German. *Frontiers in Communication*, 7:706718.

Kawin Ethayarajh, Winnie Xu, Niklas Muennighoff, Dan Jurafsky, and Douwe Kiela. 2024. Model alignment as prospect theoretic optimization. In *Forty-first International Conference on Machine Learning*.

Adam Fisch, Jacob Eisenstein, Vicky Zayats, Alekh Agarwal, Ahmad Beirami, Chirag Nagpal, Pete Shaw, and Jonathan Berant. 2024. Robust Preference Optimization Through Reward Model Distillation. *arXiv preprint arXiv:2405.19316*.

Shihan Fu, Jianhao Chen, Emily Kuang, and Mingming Fan. 2024. Bridging the Literacy Gap for Adults: Streaming and Engaging in Adult Literacy Education through Livestreaming. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–15.

Annette Rios Gonzales, Nicolas Spring, Tannon Kew, Marek Kostrzewa, Andreas Säuberli, Mathias Müller, and Sarah Ebling. 2021. A New Dataset and Efficient Baselines for Document-level Text Simplification in German. In *Proceedings of the Third Workshop on New Frontiers in Summarization*, pages 152–161.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu Zhang, Shirong Ma, Xiao Bi, and 1 others. 2025. DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning. *arXiv preprint arXiv:2501.12948*.

Silvia Hansen-Schirra, Walter Bisang, Arne Nagels, Silke Gutermuth, Julia Fuchs, Liv Borghardt, Silvana Deilen, Anne-Kathrin Gros, Laura Schiffl, and Johanna Sommer. 2020. Intralingual Translation into Easy Language–or How to Reduce Cognitive Processing Costs. *Easy Language Research: Text and User Perspectives. Berlin: Frank & Timme*, pages 197–225.

Junxian He, Chunting Zhou, Xuezhe Ma, Taylor Berg-Kirkpatrick, and Graham Neubig. 2022. Towards a Unified View of Parameter-Efficient Transfer Learning. In *Proceedings of the Tenth International Conference on Learning Representations*.

David Heineman, Yao Dou, Mounica Maddela, and Wei Xu. 2023. Dancing Between Success and Failure: Edit-level Simplification Evaluation using SALSA. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 3466–3495.

Freya Hewett, Hadi Asghari, and Manfred Stede. 2024. Elaborative Simplification for German-language Texts. In *Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 29–39.

Taichi Higasa, Keitaro Tanaka, Qi Feng, and Shigeo Morishima. 2023. Gaze-Driven Sentence Simplification for Language Learners: Enhancing Comprehension and Readability. In *Companion Publication of the 25th International Conference on Multimodal Interaction*, pages 292–296.

Taichi Higasa, Keitaro Tanaka, Qi Feng, and Shigeo Morishima. 2024. Keep Eyes on the Sentence: An Interactive Sentence Simplification System for English Learners based on Eye Tracking and Large Language Models. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–7.

Edward J Hu, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, and 1 others. 2022. LoRA: Low-Rank Adaptation of Large Language Models. In *Proceedings of the Tenth International Conference on Learning Representations*.

Xiangkun Hu, Tong He, and David Wipf. 2024. New Desiderata for Direct Preference Optimization. In *ICML 2024 Workshop on Models of Human Feedback for AI Alignment*.

J Peter Kincaid, Robert P Fishburne Jr, Richard L Rogers, and Brad S Chissom. 1975. Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel.

David Klaper, Sarah Ebling, and Martin Volk. 2013. Building a German/Simple German Parallel Corpus for Automatic Text Simplification. In *Proceedings of the Second Workshop on Predicting and Improving Text Readability for Target Reader Populations*, pages 11–19.

Lars Klöser, Mika Beele, Jan-Niklas Schagen, and Bodo Kraft. 2024. German Text Simplification: Finetuning Large Language Models with Semi-Synthetic Data. In *Proceedings of the Fourth Workshop on Language*

Technology for Equality, Diversity, Inclusion, pages 63–72.

Maria Korobeynikova, Alessia Battisti, Lukas Fischer, and Yingqiang Gao. 2026. DETECT: Determining Ease and Textual Clarity of German Text Simplifications. In *Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2852–2882.

Klaus Krippendorff. 2004. Reliability in Content Analysis: Some Common Misconceptions and Recommendations. *Human communication research*, 30(3):411–433.

Chin-Yew Lin. 2004. ROUGE: A Package for Automatic Evaluation of Summaries. In *Text Summarization Branches Out*, pages 74–81. Association for Computational Linguistics.

Jiahong Liu, Zexuan Qiu, Zhongyang Li, Quanyu Dai, Jieming Zhu, Minda Hu, Menglin Yang, and Irwin King. 2025. A Survey of Personalized Large Language Models: Progress and Future Directions. *arXiv preprint arXiv:2502.11528*.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled Weight Decay Regularization. In *Proceedings of the Seventh International Conference on Learning Representations (ICLR 2019)*.

Ilya Loshchilov and Frank Hutter. 2022. SGDR: Stochastic Gradient Descent with Warm Restarts. In *Proceedings of the Tenth International Conference on Learning Representations (ICLR 2022)*.

Christiane Maaß. 2015. *Leichte Sprache. Das Regelbuch*. Deutsche Nationalbibliothek.

Mounica Maddela, Fernando Alva-Manchego, and Wei Xu. 2021. Controllable Text Simplification with Explicit Paraphrasing. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3536–3553.

Mounica Maddela, Yao Dou, David Heineman, and Wei Xu. 2023. LENS: A Learnable Evaluation Metric for Text Simplification. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 16383–16408.

Philip May. 2020. cross-en-de-roberta-sentence-transformer. Hugging Face Model Card.

Marius Mosbach, Tiago Pimentel, Shauli Ravfogel, Dietrich Klakow, and Yanai Elazar. 2023. Few-shot Fine-tuning vs. In-context Learning: A Fair Comparison and Evaluation. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 12284–12314.

Eduardo Mosqueira-Rey, Elena Hernández-Pereira, David Alonso-Ríos, José Bobes-Bascarán, and Ángel Fernández-Leal. 2023. Human-in-the-loop Machine Learning: A State of the Art. *Artificial Intelligence Review*, 56(4):3005–3054.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: A Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Biqing Qi, Pengfei Li, Fangyuan Li, Junqi Gao, Kaiyan Zhang, and Bowen Zhou. 2024. Online DPO: Online Direct Preference Optimization with Fast-slow Chasing. *arXiv preprint arXiv:2406.05534*.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D Manning, and Chelsea Finn. 2023. Direct Preference Optimization: Your Language Model is Secretly A Reward Model. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, pages 53728–53741.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992.

Andreas Säuberli, Sarah Ebling, and Martin Volk. 2020. Benchmarking Data-driven Automatic Text Simplification for German. In *Proceedings of the 1st workshop on tools and resources to empower people with reading dIfficulties (READI)*, pages 41–48.

Andreas Säuberli, Franz Holzknecht, Patrick Haller, Silvana Deilen, Laura Schiffl, Silvia Hansen-Schirra, and Sarah Ebling. 2024. Digital Comprehensibility Assessment of Simplified Texts among Persons with Intellectual Disabilities. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–11.

Carolina Scarton and Lucia Specia. 2018. Learning Simplifications for Specific Target Audiences. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 712–718.

Laura Seiffe, Fares Kallel, Sebastian Möller, Babak Naderi, and Roland Roller. 2022. Subjective text complexity assessment for German. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 707–714.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan

Zhang, YK Li, and 1 others. 2024. DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models. *arXiv preprint arXiv:2402.03300*.

Kim Cheng Sheang and Horacio Saggion. 2021. Controllable Sentence Simplification with a Unified Text-to-Text Transfer Transformer. In *Proceedings of the 14th International Conference on Natural Language Generation*, pages 341–352.

Zhengyan Shi, Adam X Yang, Bin Wu, Laurence Aitchison, Emine Yilmaz, and Aldo Lipani. 2024. Instruction Tuning with Loss Over Instructions. In *Proceedings of the 38th Conference on Neural Information Processing Systems (NeurIPS 2024)*.

Belkiss Souayed, Sarah Ebling, and Yingqiang Gao. 2025. Template-Based Text-to-Image Alignment for Language Accessibility A Study on Visualizing Text Simplifications. In *Proceedings of the Fourth Workshop on Text Simplification, Accessibility and Readability (TSAR 2025)*, pages 1–18.

Nicolas Spring, Annette Rios Gonzales, and Sarah Ebling. 2021. Exploring German Multi-Level Text Simplification. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 1339–1349.

Nicolas Spring, Marek Kostrzewa, David Fröhlich, Annette Rios, Dominik Pfütze, Alessia Battisti, and Sarah Ebling. 2023. Analyzing Sentence Alignment for Automatic Simplification of German Texts. In *Emerging Fields in Easy Language and Accessible Communication Research*, pages 339–369. Springer.

Nicolas Spring, Marek Kostrzewa, Annette Rios, and Sarah Ebling. 2022. Ensembling and Score-Based Filtering in Sentence Alignment for Automatic Simplification of German Texts. In *International Conference on Human-Computer Interaction*, pages 137–149.

Regina Stodden, Omar Momen, and Laura Kallmeyer. 2023. DEplain: A German Parallel Corpus with Intralingual Translations into Plain Language for Sentence and Document Simplification. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 16441–16463.

Vanessa Toborek, Moritz Busch, Malte Boßert, Christian Bauckhage, and Pascal Welke. 2023. A New Aligned Simple German Corpus. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11393–11412.

Benjamin Vendeville, Liana Ermakova, and Pierre De Loor. 2025. Resource for Error Analysis in Text Simplification: New Taxonomy and Test Collection. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 3723–3732.

Xumeng Wen, Zihan Liu, Shun Zheng, Shengyu Ye, Zhirong Wu, Yang Wang, Zhijian Xu, Xiao Liang, Junjie Li, Ziming Miao, Jiang Bian, and Mao Yang. 2026. Reinforcement Learning with Verifiable Rewards Implicitly Incentivizes Correct Reasoning in Base LLMs. In *The Fourteenth International Conference on Learning Representations*.

Xingjiao Wu, Luwei Xiao, Yixuan Sun, Junhang Zhang, Tianlong Ma, and Liang He. 2022. A Survey of Human-in-the-loop for Machine Learning. *Future Generation Computer Systems*, 135:364–381.

Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, and Chris Callison-Burch. 2016. Optimizing Statistical Machine Translation for Text Simplification. *Transactions of the Association for Computational Linguistics*, 4:401–415.

Daichi Yamaguchi, Rei Miyata, Sayuka Shimada, and Satoshi Sato. 2023. Gauging the Gap Between Human and Machine Text Simplification Through Analytical Evaluation of Simplification Strategies and Errors. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 359–375.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. BERTScore: Evaluating Text Generation with Bert. In *Proceedings of the Seventh International Conference on Learning Representations (ICLR 2019)*.

Siyan Zhao, Mingyi Hong, Yang Liu, Devamanyu Hazarika, and Kaixiang Lin. 2025. Do LLMs Recognize Your Preferences? Evaluating Personalized Preference Following in LLMs. In *Proceedings of the 15th International Conference on Learning Representations (ICLR 2025)*.

Chunting Zhou, Pengfei Liu, Puxin Xu, Srini Iyer, Jiao Sun, Yuning Mao, Xuezhe Ma, Avia Efrat, Ping Yu, Lili Yu, and 1 others. 2023. LIMA: Less is More for Alignment. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, pages 55006–55021.

A Data Curation and Model Training Details

A.1 Data Filtering

DEPLAIN (Stodden et al., 2023) consists of two sub-collections: DEPLAIN-APA and DEPLAIN-WEB. While the latter is derived from web-crawled documents consisting of non-news texts, the former comprises text from Austrian Press Agency (APA) news items published from May 2019 to April 2021 and covering a diverse range of topics. These topics include politics, crime, weather, economics, zoo births, and the coronavirus pandemic. Overall, DEPLAIN-APA contains 13,122 manually aligned sentence pairs from 483 documents pairs classified as A2 or B1 under the Common European Framework of Reference for Languages (CEFR). Given its diverse topic coverage, we selected DEPLAIN-APA as the base for HF4ATS-SFT. When discussing DEPLAIN-APA sentences, we refer to text from B1 articles as "complex" text and text from A2 articles as "simple" or "simplified" text.

In Table 6 we show examples in DEPLAIN we removed using the four-step data filtering approach.

Table 6: **Examples of invalid complex–simple pairs in DEPLAIN**. We identify (i) almost identical pairs that provide no simplification signal, and (ii) pairs that lack sufficient context, where the simplified sentence is not entailed by the complex sentence.

Issue Type	Complex Sentence	Simple Sentence
Almost identical	*Integration bedeutet, also dass jemand dazugehört.*	*Integration bedeutet also, dass jemand dazugehört.*
Lack of context	*Es gibt aber grosse Unterschiede.*	*Nicht in jedem Vanille-Eis ist gleich viel Luft drin.*

A.2 Input Prompt

The prompts used for SFT and DPO are assembled with one or more perspectives as follows:

- Description of the target audience (German-speaking persons with intellectual disabilities);
- Goal of easy language (German: *Leichte Sprache*);
- Suggestion of text simplification operations (including adding, removing, reordering, replacing, and splitting) according to the recommendations for German Easy Language (German: *Empfehlungen für Deutsche Leichte Sprache*[†] , and as in Maaß (2015));
- One-shot prompting with one concrete example;
- Two-shot prompting with two concrete examples.

All input prompts included at least one of the aforementioned perspectives. The inclusion of few-shot prompts leveraged in-context learning benefits for SFT (Chen et al., 2022; Mosbach et al., 2023).

A.3 SFT Training

To pad the input texts for LLMs, we set the padding token to `<finetune_right_pad_id>` for Llama-3.1-8B-Instruct, `<unk>` for the two Mistral models, and left it unchanged for DiscoLeo-Llama-3-8B-Instruct. Input padding was consistently applied to the right side of the prompts.

Research has shown that full-prompt tuning, i.e., tuning where instruction tokens are included in the training loss calculation, can enhance performance for open-ended tasks when the average ratio of prompt token count to completion token count exceeds five and the number of training instances amounts to a few thousand (Shi et al., 2024). These two conditions are met by HF4ATS-SFT. Therefore, to increase robustness, we adopted a mixed strategy and trained separate models with full-prompt tuning and completion-only tuning.

A.4 DPO Preference Pair Creation

Simplifications in HF4ATS-DPO were generated for 8,000 complex texts curated from two sources: (1) 3,200 complex texts sampled from all DEPLAIN pairs not included in HF4ATS-SFT, denoted $\mathcal{D}_{\text{DEPLAIN}} \setminus \mathcal{D}_{\text{SFT}}$, and (2) 4,800 complex texts sampled from the APA-LHA dataset (Spring et al., 2021), denoted $\mathcal{D}_{\text{LHA}}$. APA-LHA comprises automatically aligned sentence-level complex-simple text pairs pulled from APA news items classified as A2 and above. While we excluded this dataset from SFT because its automatic alignments could have induced hallucinations, we did involve its complex

[†] https://www.din.de/de/mitwirken/normenausschuesse/naerg/e-din-spec-33429-2023-04-empfehlungen-fuer-deutsche-leichte-sprache--901210

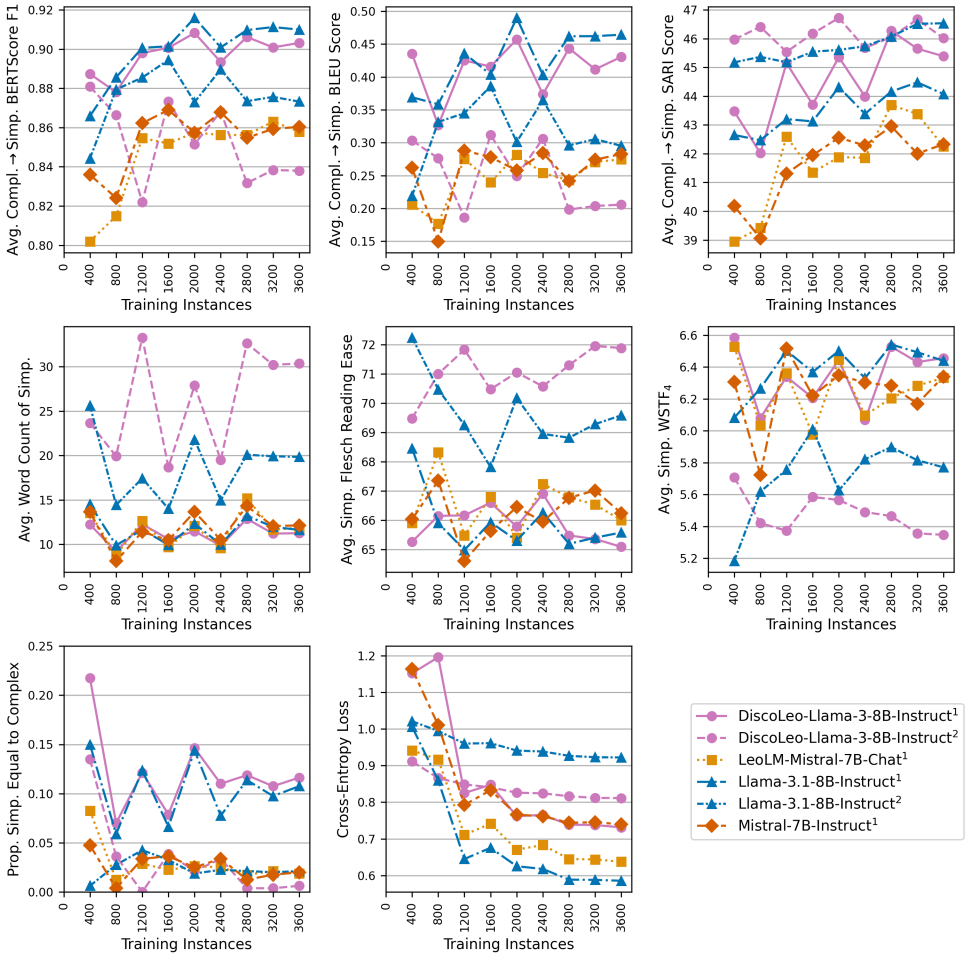


Figure 2: **Cross-model comparison for SFT checkpoint evaluation.** Full-prompt loss (denoted as 1 in the legend) includes both instruction and completion tokens in the loss calculation, while completion-only loss (denoted as 2) considers only the completion tokens.

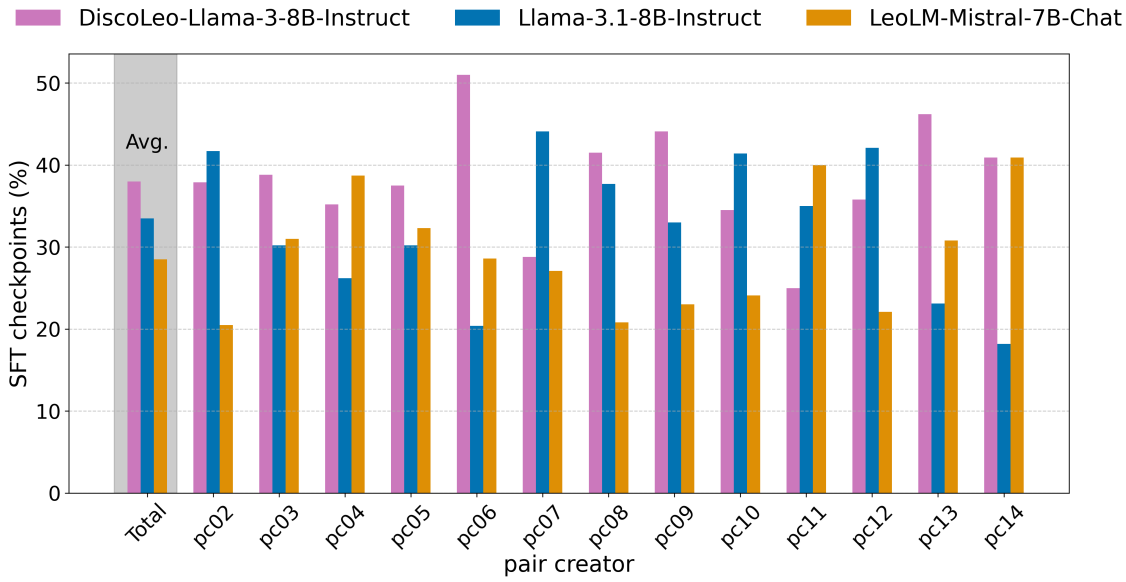


Figure 3: **Distribution of SFT model source for created ATS pairs.** This figure reflects the relative prevalence of different SFT backbone LLMs in the HF4ATS dataset and indicates that human preferences differ from the model perspective. The shaded left bar (Avg.) shows overall averages ranging from approximately 28% to 37%, with a plurality of pairs coming from the DiscoLeo-Llama model.

sentences during pair creation because its topic distribution is similar to that of DEPLAIN (in fact, the two datasets shared some complex sentences).

We applied Gaussian sampling with different weighting schemes to the two HF4ATS-DPO inference sources. From the DEPLAIN subset we sampled a complex text x with a Gaussian weight w_x defined as

$$w_{x \sim \mathcal{D}_{\text{DEPLAIN}} \setminus \mathcal{D}_{\text{SFT}}} = \exp\left(-\frac{\left(|x| - \mu_{\mathcal{D}_{\text{DEPLAIN}} \setminus \mathcal{D}_{\text{SFT}}}\right)^2}{2 \cdot \sigma^2}\right),$$

where the mean

$$\mu_{\mathcal{D}_{\text{DEPLAIN}} \setminus \mathcal{D}_{\text{SFT}}} = \frac{\sum_{x' \in \mathcal{D}_{\text{DEPLAIN}} \setminus \mathcal{D}_{\text{SFT}}} |x'|}{|\mathcal{D}_{\text{DEPLAIN}} \setminus \mathcal{D}_{\text{SFT}}|}$$

corresponds to the average word count of complex texts from the leftover DEPLAIN subset, with $|\mathcal{D}_{\text{DEPLAIN}} \setminus \mathcal{D}_{\text{SFT}}|$ denoting the subset's size and $|x|$ denoting a given complex text's word count. From APA-LHA, x was sampled with weight w_x defined as

$$w_{x \sim \mathcal{D}_{\text{LHA}}} = \exp\left(-\frac{\left(|x| - \left(\mu_{\mathcal{D}_{\text{LHA}}} + \eta \cdot \left(\mu_{\mathcal{D}_{\text{LHA}}} - \mu_{\mathcal{D}_{\text{DEPLAIN}} \setminus \mathcal{D}_{\text{SFT}}}\right)\right)\right)^2}{2 \cdot \sigma^2}\right),$$

where the mean

$$\mu_{\mathcal{D}_{\text{LHA}}} = \frac{\sum_{x' \in \mathcal{D}_{\text{LHA}}} |x'|}{|\mathcal{D}_{\text{LHA}}|}$$

represents the average word count of complex texts from APA-LHA and $\eta = 4,800/8,000$ is a scaling factor reflecting the share of LHA-APA (as opposed to leftover DEPLAIN) complex sentences present in the 8,000 instance inference set.

The 13 pair creators were trained with the following rubric to ensure high-quality ATS pairs:

- **Entailment**: Pair creators verified that the complex sentence entailed the simplification. Because adding information is a valid simplification strategy, pair creators were allowed to make an exception in an unambiguous situation (e.g., a simplification that identifies the "*Democratic candidate for the 2020 U.S. Presidential election*" as "*Joe Biden*").
- **Equal information**: Pair creators prioritized text simplifications that conveyed the same amount of information. Creators indicated via the creation tool whether each pair met this condition.
- **High simplification quality**: Pair creators prioritized simplifications that adhered to German language rules and were accessible for persons with intellectual disabilities. We also actively asked pair creators to avoid simplifications that were potentially non-ethical or non-faithful.
- **High simplification diversity**: Pair creators prioritized selecting two simplifications that differed in their applied simplification strategies (e.g., deletion, paraphrasing, or sentence splitting).

To facilitate the pair creation process, we developed an intuitive Python script that enabled human pair creators to review 20 inferences for a complex sentence, select two to pair together, and indicate whether their selected simplifications had equal levels of information. Pairs could only be created with two inferences from the same winning SFT checkpoint, and pair creators were able to skip inference sets if no suitable pairs could be identified. Along with the order of complex sentences, the order in which the three winning SFT checkpoint's inference sets appeared was randomized. Each complex sentence was only shown until one appropriate pair was created or all SFT checkpoint inference sets were skipped by the pair creator. Additionally, the SFT checkpoint responsible for each inference set remained masked during annotation.

A.5 DPO Preference Pair Annotation

As part of the inclusive workflow shown in Figure 5, we developed an easy-to-use web tool for preference data annotation (subfigure 5b). From a tablet browser, the preferred text simplification was highlighted with a light green background once the participant selected the corresponding button, "*Diesen Text verstehe ich besser*" (English: "*I understand this text better*"). After completing the current pair, participants could freely navigate to the previous or next pair using the "*Zurück*" (English: "*Back*") and "*Weiter*" (English: "*Next*") buttons. Annotations could be submitted at any time by clicking the "*Abschicken*" (English: "*Send*") button.

Prior to the start of each session, an educational caretaker (i.e., a proctor) supporting the target group participants read aloud the web tool instructions and consent form, both of which were written in simplified German language. The caretaker also

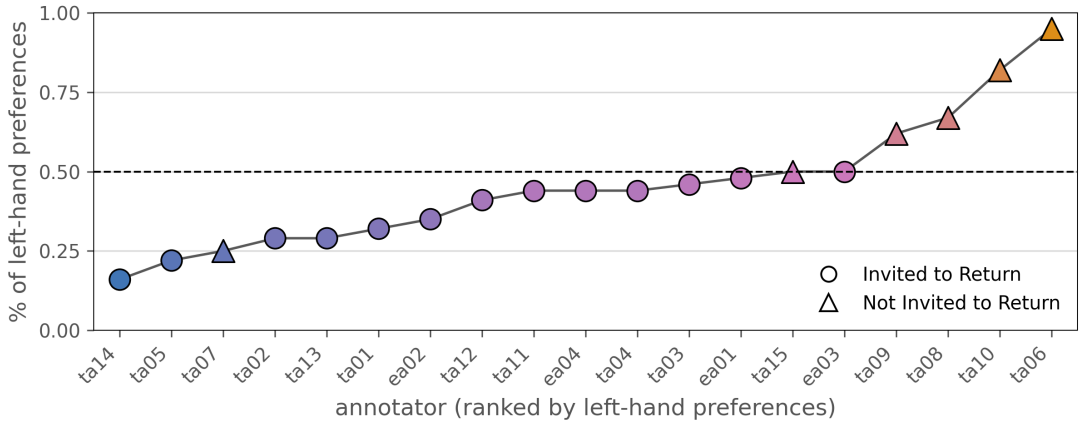
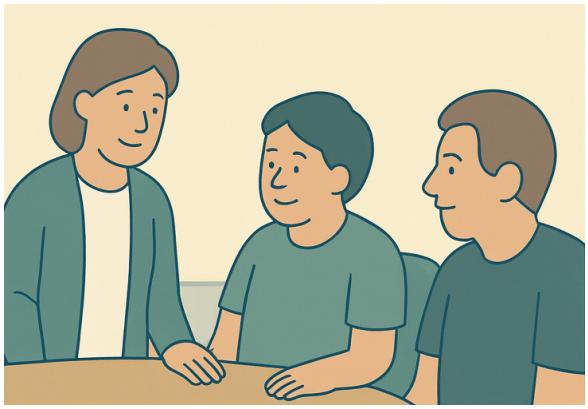
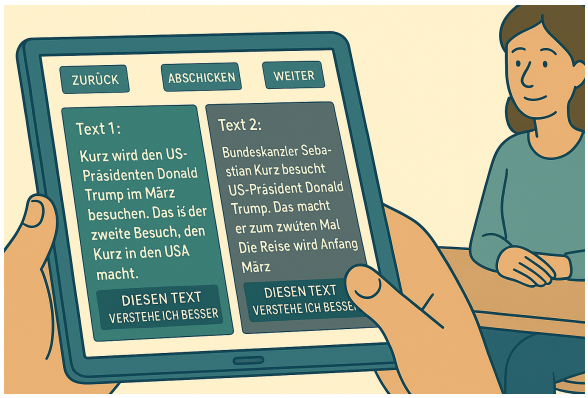


Figure 4: **Preference rate for the left-hand option by user**. Annotators such as ta06 exhibited an overwhelming preference for one side, suggesting they did not understand or adhere to task instructions. Apart from concentrating preferences on one side, annotators may not have been invited to return for other reasons (e.g. admitting they struggled to understand the task).



(a) The caretaker explains the task using simplified language.



(b) Participants indicate their text preferences on a tablet.

(c) Participants request additional clarification when needed.

(d) Caretaker and participants provide feedback on the user experience.

Figure 5: **An inclusive workflow to collect preference data from the target group participants**. We have actively involved a caretaker and a technical expert for all sessions with target group participants. English translations for feedback in (d): sentence too long; too many hard words; wrong use of hyphens.

demonstrated the annotation process through example tasks to familiarize the target group participants with the procedure. Each participant was provided with a tablet and unique log-in ID. Once they accessed the web tool, they annotated independently under the caretaker's supervision. Overall, we or-

ganized 15 annotation sessions, each attended by 1-10 of our 15 target group participants. The participants had an average age of 27.4, and each was previously assessed to have a mild to moderate intellectual disability.

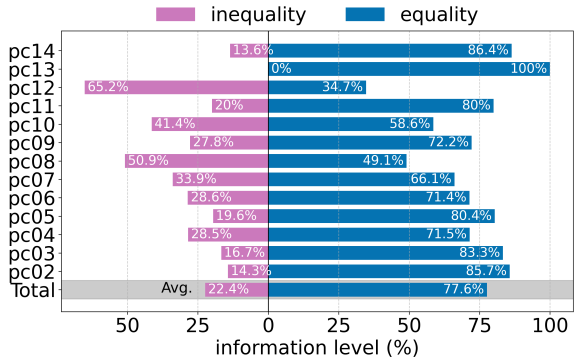


Figure 6: **Information-level annotation in the dataset HF4ATS-DPO.** This figure shows the percentage of ATS preference pairs labeled by each pair creator as containing either equal or unequal information, with almost 80% of all pairs possessing information equality.

In parallel, we recruited four native German-speaking human annotators with expertise in text simplification (the expert group) to perform the same preference annotation task on the HF4ATS-DPO dataset. To reduce inter-group bias, two of the expert group participants (denoted in our results as ea01 and ea03) were unable to see each pair's corresponding complex text, matching the target group's annotation conditions. To balance this inter-group bias reduction with a desire to leverage the expert group's professional training, we displayed corresponding complex texts to the two remaining expert group participants (ea02, ea04). No participant was made aware of the difference in annotation conditions.

Expert group participants were compensated for their annotations at an hourly rate. Ethical approval for the expert group annotations was not required by the university's ethics committee. A detailed task instruction sheet and a tutorial video were provided to all expert group annotators prior to the kick-off of the annotation task. Overall, the expert group participants completed the annotation tasks significantly faster than most target group participants; during the final evaluation annotations, expert annotators averaged 180 pairs per hour against the target group's 60 pairs per hour.

The pair creator was shown a complex sentence and a procession of possible ATS pairs in randomized order without being informed which checkpoints were responsible for each inference. The creator approved or rejected pairs based on the same criteria used during initial ATS pair creation. Only those complex sentences for which the pair creator could approve one pair for all six DPO checkpoints were included in the final evaluation round with human participants.

We invited the four target group participants with the highest Intra-AA scores (i.e., ta04, ta05, ta10, and ta12) and all four expert annotators to take part in the final human evaluation sessions. Apart from the fact that all pairs were shared within the annotation groups (albeit displayed in randomized order), annotation conditions were the same as before. Importantly, only the 150 pairs associated with the three target-group DPO checkpoints were shown to target group annotators, and only the 150 pairs associated with the three expert-group DPO checkpoints were shown to expert group annotators. Based on pairwise choices between SFT- and DPO-checkpoint-generated text simplifications, we computed DPO supremacy scores separately for each evaluator.

Table 7 shows the corresponding checkpoints of SFT and DPO phase.

B Prompt Templates Used for Constructing LLM Inputs

Table 8 lists the prompt templates in German used in SFT and DPO, optimized by an text simplification expert who was not involved in data annotation. We randomly sampled from this prompt bank to increase the diversity of ATS generation. We omit the English translations of the prompt templates for writing convenience.

C Details of SFT Evaluation

Based on SARI and WSTF_4 performance, we selected the hyper-parameter configuration with a gradient accumulation step size of 1 and a learning rate of $1e-4$ for cross-model comparison. We then trained all four models with this configuration, implementing full-prompt tuning generally and completion-only tuning for the two Llama-based models. We selected the following three SFT checkpoints: 1) DiscoLeo-Llama-3-8B-Instruct af-

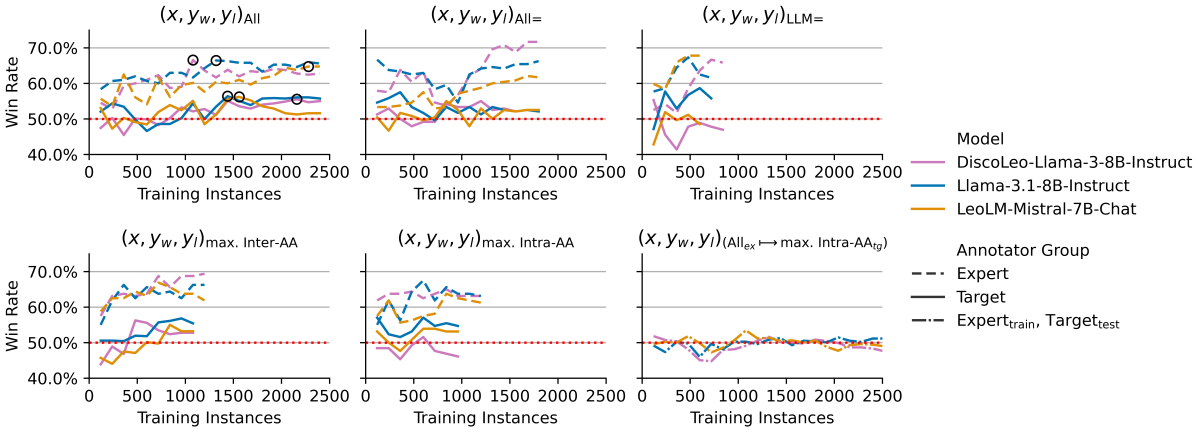


Figure 7: **Development-set win rates during DPO training across HF4ATS-DPO subsets**. Circles mark the best checkpoints; the bottom-right panel shows expert-trained models evaluated on target-group data (most consistent annotators).

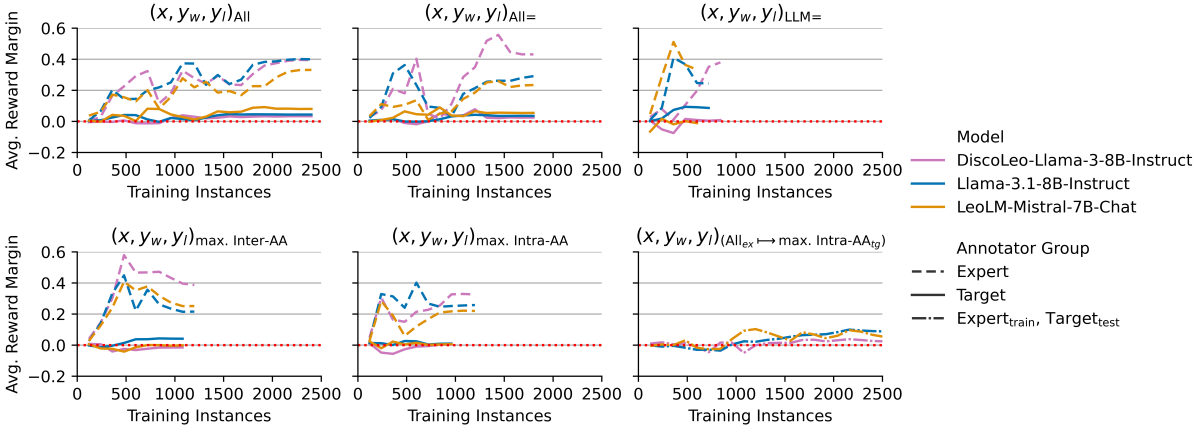


Figure 8: **Average reward margins with respect to the number of training instances** from different subsets of the HF4ATS-DPO training data, evaluated on the corresponding development sets.

Table 7: **Model sequences from our Pre-train → SFT → DPO pipeline** used to personalize LLM-based ATS. We reiterate that DPO checkpoints were trained separately using target and expert group annotations from HF4ATS-DPO, while SFT checkpoints were not group-specific.

Pre-trained LLMs	**SFT Checkpoint**	**DPO Checkpoint**	
		Target	**Expert**
DiscoLeo-Llama-3-8B-Instruct	DiscoLeo-Llama-SFT-2800	DiscoLeo-Llama-DPO-2160	DiscoLeo-Llama-DPO-1080
Llama-3.1-8B-Instruct	Llama-SFT-2400	Llama-DPO-1440	Llama-DPO-1320
LeoLM-Mistral-7B-Chat	LeoLM-Mistral-SFT-1600	LeoLM-Mistral-DPO-1560	LeoLM-Mistral-DPO-2280

ter 2,800 training steps of full-prompt tuning; 2) Llama-3.1-8B-Instruct after 2,400 training steps of completion-only tuning; 3) LeoLM-Mistral-7B-Chat after 1,600 training steps of full-prompt tuning.

Figure 2 illustrates the performance of SFT checkpoints during development across various metrics. For checkpoint selection, we prioritize models that generate high-quality, readable ATS outputs, as reflected by strong reference-based SARI and reference-free WSTF_4 scores.

To train all SFT models, we employed the AdamW optimizer (Loshchilov and Hutter, 2019) with a weight decay of 0.01, a cosine annealing learning rate scheduler (Loshchilov and Hutter, 2022), gradient norm clipping at 1, a maximum sequence length of 300 tokens, a batch size of 16, and FP16 mixed precision. We performed parameter-

Table 8: **Prompt templates used for SFT and DPO**. <complex_sentence> represents the complex text to be simplified, while <complex_sentence1> and <complex_sentence2> refer to example complex texts. Correspondingly, <simple_sentence1> and <simple_sentence2> serve as their respective simplifications.

No.	Prompt	Phase
1	Schreibe den folgenden Satz in Leichter Sprache um: <complex_sentence>. Bitte gib nur eine Vereinfachung an, ohne Einleitung, Alternativen oder Kommentare.	SFT + DPO
2	Vereinfache den folgenden Satz, sodass Menschen mit kognitiver Beeinträchtigung den vereinfachten Satz verstehen können: <complex_sentence>. Bitte gib nur die Vereinfachung an, ohne Einleitung, Alternativen oder Kommentare.	SFT + DPO
3	Schreibe den folgenden komplexen Satz um und verwende einfachere Wörter, kürzere Sätze und reduzierte grammatikalische Strukturen. Der Inhalt und die Bedeutung sollen nach dem Umschreiben unverändert bleiben. Bitte gib nur die Vereinfachung an, ohne Einleitung, Alternativen oder Kommentare. Komplex: <complex_sentence>. Leicht:	SFT + DPO
4	Formulieren Sie den komplexen Satz um, indem Sie mindestens einen neuen einfachen Satz bilden. Behalten Sie die gleiche Bedeutung des Ausgangssatzes bei. Geben Sie bitte nur die Vereinfachung an, ohne Einleitung, Alternativen oder Kommentare. Komplex: <complex_sentence>. Leich:	SFT + DPO
5	Schreibe den folgenden komplexen Satz in Leichter Sprache um. Die Vereinfachung soll kurz und von geringer Komplexität sein (durchschnittlich acht bis fünfzehn Wörter pro Satz) und eine geringe Anzahl von Aussagen pro Satz enthalten. Bitte gib nur die Vereinfachung an, ohne Einleitung, Alternativen oder Kommentare. Komplex: <complex_sentence>. Leicht:	SFT + DPO
6	Schreibe den folgenden komplexen Satz in Leichter Sprache um. Die Wörter in deiner Vereinfachung sollen kurz, beschreibend, und häufig verwendet von Menschen mit kognitiver Beeinträchtigung sein. Bitte gib nur die Vereinfachung an, ohne Einleitung, Alternativen oder Kommentare. Komplex: <complex_sentence>. Leicht:	SFT + DPO
7	Schreibe den folgenden komplexen Satz in Leichter Sprache um. Deine Vereinfachung soll für Menschen mit kognitiver Beeinträchtigung in Österreich verständlich sein. Bitte gib nur die Vereinfachung an, ohne Einleitung, Alternativen oder Kommentare. Komplex: <complex_sentence>. Leicht:	SFT + DPO
8	Schreiben Sie den folgenden komplexen Satz in Leichter Sprache um. Sie können 1) den Satz in mehrere Sätze aufteilen, 2) Die Wortstellung ändern, um die Grammatik zu vereinfachen, 3) Wörter hinzufügen, um schwierige Konzepte zu erklären, 4) Wörter, die sich mit unnötigen Informationen zusammenhängen, entfernen, und 5) schwierige Wörter durch einfache Vokabeln ersetzen. Achten Sie darauf, dass der Satz leichter verständlich bleibt, ohne die Bedeutung zu verändern. Bitte geben Sie nur die Vereinfachung an, ohne Einleitung, Alternativen oder Kommentare. Komplex: <complex_sentence>. Leicht:	SFT + DPO
9	Schreibe den folgenden Satz in Leichter Sprache um. Bitte gib nur die Vereinfachung an, ohne Einleitung, Alternativen oder Kommentare. Hier ist ein Beispiel. Komplex: <complex_sentence1>. Leicht: <simple_sentence1>. Schreibe deine Vereinfachung nach "Leicht:". Komplex: <complex_sentence>. Leicht:	SFT
10	Schreibe den folgenden komplexen Satz in Leichter Sprache um. Bitte gib nur die Vereinfachung an, ohne Einleitung, Alternativen oder Kommentare. Hier sind zwei Beispiele. Komplex: <complex_sentence1>. Leicht: <simple_sentence1>. Komplex: <complex_sentence2>. Leicht: <simple_sentence2>. Schreibe deine Vereinfachung nach "Leicht:". Komplex: <complex_sentence>. Leicht:	SFT

efficient fine-tuning (PEFT; He et al. (2022)) on a single NVIDIA A100 GPU using LoRA (Hu et al., 2022) with rank 16, a scaling factor of 32, and a dropout rate of 0.05. Our grid search for hyper-parameters scanned across gradient accumulation step size (1, 2, and 4) and learning rate ($1e-5$, $5e-5$, and $1e-4$). We selected DiscoLeo-Llama-3-8B-Instruct for hyper-parameter optimization due

to its abundance of German-language training data relative to Llama-3.1-8B-Instruct and Mistral-7B-Instruct as well as its extensive instruction tuning compared to LeoLM-Mistral-7B-Chat.

D Statistics of SFT/DPO Pair Creation

Figure 3 shows that pair creators exhibited individual preferences for specific SFT checkpoints despite the model-blind pair creation procedure.

Figure 4 suggests that some target group annotators consistently favored one side, indicating possible non-adherence to task instructions (we have randomly shuffled the sides of two text simplifications). Despite well-balanced preferences, ta15 decided to end study participation due to personal issues.

Figure 6 presents the percentage of pairs labeled as having equal or differing information, grouped by pair creator. The shaded bottom row (Avg.) shows the overall average, with nearly 80% of pairs labeled as having equal information.

E Implicit Reward Margin of During DPO Post-Training

Figure 8 visualizes the evolution of implicit reward margins during DPO post-training as a function of the number of training instances, across different HF4ATS-DPO data subsets and model backbones. Consistent with the win-rate analyses in the main paper, subsets with higher preference consistency (maximized Inter- or Intra-AA) exhibit faster and more stable increases in reward margins, particularly for expert-supervised models. In contrast, training on broader or mismatched subsets yields smaller or unstable margins, and cross-group evaluation (expert-trained, target-tested) shows margins fluctuating around zero, indicating limited transfer. These trends further support the conclusion that preference consistency is a key driver of effective DPO alignment.