DR GENRÉ: Reinforcement Learning from Decoupled LLM Feedback for Generic Text Rewriting

Anonymous ACL submission

Abstract

Generic text rewriting is a prevalent large language model (LLM) application that covers diverse real-world tasks, such as style transfer, fact correction, and email editing. These tasks vary in rewriting objectives (e.g., factual consistency vs. semantic preservation), making it challenging to develop a unified model that excels across all dimensions. Existing methods often specialize in either a single task or a specific objective, limiting their generalizability. In this work, we introduce a generic model proficient in *factuality*, *stylistic*, and *conversational* rewriting tasks. To simulate real-world user rewrite requests, we construct a conversational rewrite dataset, CHATREWRITE, that presents "natural"-sounding instructions, from raw emails using LLMs. Combined with other popular rewrite datasets, including LONGFACT for the factuality rewrite task and REWRITELM for the stylistic rewrite task, this forms a broad benchmark for training and evaluating generic rewrite models. To align with task-specific objectives, we propose DR GENRÉ, a **D**ecoupled-**r**eward learning framework for **Gen**eric **re**writing, that utilizes objective-oriented reward models with a task-specific weighting. Evaluation shows that DR GENRÉ delivers higher-quality rewrites across all targeted tasks, improving objectives including instruction following (*agreement*), internal consistency (*coherence*), and minimal unnecessary edits (*conciseness*).

1 Introduction

Text rewriting is a fundamental NLP task with applications spanning style transfer (Shu et al., 2024), summarization, and fact correction (Wei et al., 2024). Existing models often specialize in a particular transformation type, such as paraphrasing (Siddique et al., 2020), sentence fusion (Mallinson et al., 2022), or focus on optimizing a specific objective, such as syntactic overlap with reference texts, during post-training (Shen et al., 2017; Hu et al., 2017). This specialization limits their applicability in real-world scenarios where diverse rewriting capabilities are required. In this work, we address the challenges of building a generic rewrite model capable of handling multiple rewriting tasks, including fact correction, style transfer, and conversational rewriting.

Versatile rewriting tasks. *Factuality rewrite* involves correcting content that contains factual errors in initial responses generated by a large language model (LLM) responding to fact-seeking prompts on open-ended topics (Hu et al., 2023). The rewrite model takes the prompt, initial response, and critique outputs (e.g., span-level checks from autoraters), and generates revised responses that correct non-factual claims, maintain global coherence (a challenge for standard post-training methods, as shown in Table 3), and minimally edit the factual claims. *Stylistic rewrite* focuses on transforming a source text into another style without introducing new information (Shu et al., 2024), including formalization, paraphrasing, and summarization (Li et al., 2018; Zhang et al., 2020). While this task has been studied extensively in the in-context learning (ICL) setting (Brown et al., 2020; Raffel et al., 2020), few-shot methods struggle to follow user-specified instructions accurately (as shown in Table 4). Both factuality and stylistic rewrite expose limited user applicability. To mimic real-world user requests, we introduce a new rewriting task—*conversational rewrite*—which addresses the need for modifying specific parts of a text based on user instructions that may lack detailed context. For instance, a user might say, "This email is a bit dry; let's celebrate our success! Add some enthusiastic phrases like 'We nailed it!'". We construct a benchmark dataset, CHATREWRITE, through multi-turn instruction prompting (see Table 1) and revised conversation generation using LLMs.

Multi-objective intrinsic. From analyzing these

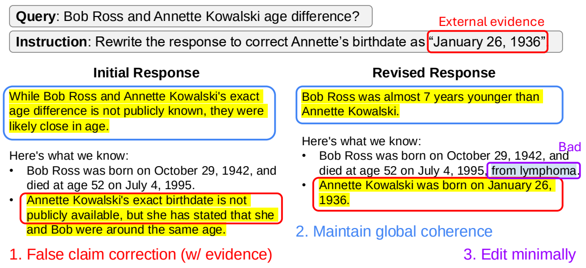


Figure 1: An example illustrating the three objectives for generic text rewriting: Agreement: Follow the rewrite instruction (e.g., false claim correction). Coherence: Maintain global coherence across revisions (e.g., updating the age difference to maintain logical flow). Conciseness: Avoid unnecessary edits (e.g., specifying Bob's death place "lymphoma" is irrelevant).

tasks, we decouple three major objectives that are preferred for a high-quality rewrite, as shown in Figure 1. *Agreement*: the revised response should strictly follow the rewrite instruction, such as accurately correcting false claims (e.g., revising "Annette Kowalski's exact birthdate is not publicly available..." to "Annette Kowalski was born on January 26, 1936"). *Coherence*: the revised response should maintain global coherence after the desired corrections are made at a local level, such as modifying contradictory contents (e.g., changing "Bob Ross and Annette Kowalski's exact age difference is not publicly known..." to "Bob Ross was almost 7 years younger than Annette Kowalski"). *Conciseness*: the revision should avoid unnecessary edits and make minimal changes, such as updating the age difference calculation without altering unrelated details (e.g., "from lymphoma").

A generic framework. To build a versatile rewrite model, we propose DR GENRÉ, a **D**ecoupled-**r**eward learning framework for **Gen**eric **re**writing. At the starting supervised fine-tuning (SFT) stage, we train a student model on a mixture of the three task-specific datasets. During reward modeling, due to limited human preferences, we distill both the agreement and coherence preferences from a teacher LLM to derive task-agnostic, objective-oriented reward models (RMs) (Stiennon et al., 2020; Lee et al., 2023), and employ rule-based edit ratio as the conciseness reward. At the reinforcement learning (RL) stage, we compute the final preference for a input rewrite prompt and its corresponding rewritten response by performing a task-specific weighting of objective-oriented reward models. DR GENRÉ offers a fine-grained control over the alignment direction by adjusting the weights according to specific task requirements. Our contributions are threefold:

- We introduce conversational rewrite—a new rewriting task that is more challenging yet user-applicable—and create a dataset CHATREWRITE for benchmarking.
- We propose DR GENRÉ, a post-training framework for generic text rewriting, and establish robust baselines using few-shot LLMs, SFT, and single-reward RL across the three tasks.
- Experiments show that weighted decoupled rewards offer enhanced control over the alignment direction, leading to improved performance across multiple rewriting objectives.

2 Related Work

Text rewriting. Existing research on rewriting often focus on a particular set of rewriting tasks, including factuality correction and style transfer, aiming to improve accuracy, tone, or coherence.

Factuality rewrite addresses factual inaccuracies in generated responses where merely prompting LLMs such as GPT-4 (Han et al., 2024) cannot promise fact-related (e.g., dates, locations, statistics) accuracy in generation (Li et al., 2023), especially when dealing with open-ended or fact-seeking prompts. Existing methods often leverage external knowledge bases (Shen et al., 2017), fact-checking modules (Hu et al., 2017), or post-editing strategies (Hu et al., 2023), though balancing correctness with minimal edits remains challenging.

Style transfer, covering tasks like paraphrasing (May, 2021), formalization (Rao and Tetreault, 2018; Li et al., 2024b), and elaboration (Iv et al., 2022), adapts tone without altering meaning. Recent methods like RewriteLM (Shu et al., 2024) extend rewriting to broader domains but are limited by single reward granularity. Our work unifies these objectives, leveraging decoupled rewards to handle diverse rewriting tasks while ensuring instruction adherence, coherence, and minimal edits.

Data augmentation with LLMs. Leveraging LLMs for data augmentation has emerged as a widely adopted approach to enhance model performance by generating synthetic data for training. Early works (He et al., 2020; Huang et al., 2023) focused on augmenting data distributions to improve performance, while more recent advancements, such as PEER (Schick et al., 2023), demonstrated the effect of infilling missing data with synthetic samples. Methods like Self-Instruct (Wang

Raw generated email	**Rewrite instruction** (Raw→Specific→Natural)
Dear [Customer Name], We hope this email finds you well. We're writing to you today to let you know about a new way we're improving our social media presence. We've created a new account on [social media platform], and we'd love for you to follow us! We'll be using this account to share news about our company, our products, and our industry. We'll also be posting photos and videos, and we'll be running contests and giveaways. We hope you'll join us on [social media platform]! We're looking forward to connecting with you there. Sincerely, [Your Name]	**Raw**: Specify the social media platform and highlight the specific benefits for customers. **Specific**: Specify the social media platform as TikTok and highlight the specific benefits for customers such as exclusive behind-the-scenes content, early access to product launches, and the chance to win prizes in contests and giveaways. **Natural**: Instead of just saying [social media platform], say we're now on TikTok! Also, let's tell them about the cool stuff they'll find there, like exclusive behind-the-scenes content, early access to new products, and even the chance to win prizes!

Table 1: An example of multi-turn instruction generation from our CHATREWRITE dataset. The instruction is first specified with more details, and then reorganized and expressed in a more human-like speaking style.

et al., 2023) bootstrap instructions and model outputs to boost task-specific accuracy. Further research (Li et al., 2024a; Han et al., 2024) highlights the capability of LLMs to produce high-quality augmented data for diverse downstream tasks. Building upon these principles, our work uses LLMs to synthesize CHATREWRITE—a more challenging conversational rewrite dataset that aligns with real-world user-LLM interaction scenarios.

LLM-as-a-judge (AutoRater). LLMs, with their superiority in language understanding and knowledge integration, have been widely used as AutoRaters to judge the quality of generated responses due to high costs of human evaluation (Vu et al., 2024). In some benchmarks, LLM agents even exhibit better annotation capabilities than humans (Wei et al., 2024). We use Gemini-1.5-Ultra (Team et al., 2023) to auto-evaluate agreement, coherence and compare pairwise responses in their instruction satisfaction granularity.

Reinforcement Learning from AI feedback (RLAIF). RL from human feedback (RLHF) is a technique that combines RL with human evaluations to fine-tune LLMs (Christiano et al., 2017). In RLHF, a LLM generates outputs that are assessed by human evaluators, who provide feedback indicating preferences or ratings based on certain criteria (e.g., helpfulness, correctness, style). This feedback is used to train a RM that predicts the human-provided scores. The LLM is then fine-tuned using RL algorithms like Proximal Policy Optimization (PPO) (Schulman et al., 2017), optimizing the rewards to improve alignment with human preferences (Raffel et al., 2020). However, collecting large-scale human feedback is resource-intensive, motivating the exploration of RLAIF (Lee et al., 2024) such as leveraging LLMs to simulate human evaluations and derive reward preferences.

Dataset	Size	IR len	RR len	ER
LongFact	21,294	316	296	0.281
RewriteLM	29,985	131	127	0.729
ChatRewrite	82,290	108	111	0.650

Table 2: Statistics of the three datasets. "IR len" and "RR len" denote the lengths of initial and revised responses. "ER" denotes the edit ratio.

3 Dataset Generation

We describe the construction of our dataset, covering the three distinct tasks. Its statistics is shown in Table 2. Each dataset is generated through structured prompting that ensures high-quality rewrites tailored to specific objectives.

3.1 Factuality Rewrite

We follow LONGFACT (Wei et al., 2024) to generate factually rich, multi-faceted responses. First, we prompt LLMs to generate fact-seeking queries, such as "What happened in the first modern Olympics?". For each query, we obtain an initial response (IR) from LLMs, which is then passed to a critic model[1] that provides span-level factuality critiques and suggested revisions (see Table 9 in Appendix B). To construct the dataset, we integrate the query, IR, and critique outputs to prompt LLMs (see Table 11 in Appendix C) and generate the revised responses. Each example (see Table 8 in Appendix B) in the dataset is a quadruplet (<query>, <initial response>, <critique outputs>, <revised response>).

[1]We employ SAFE (Wei et al., 2024) which enables external fact-checking calls from LLMs.

3.2 Stylistic Rewrite

We follow RewriteLM (Shu et al., 2024) and apply chain-of-thought (CoT) prompting (Wei et al., 2022) to generate stylistic rewrite instructions for source texts from the C4 corpus (Raffel et al., 2020). The dataset is constructed by integrating source text and generated rewrite instructions using a structured template (see Table 12 in Appendix C). We then prompt LLMs to produce the revised text. Each example consists of a triplet (`<source>`, `<instruction>`, `<revised text>`), as exemplified by Table 10 in Appendix B.

3.3 Conversational Rewrite

We begin with large-scale natural prompts paired with raw generated emails from those prompts (see Table 7). This dataset serves as the foundation for improving clarity, tone, and personalization in conversation-based text generation.

Multi-turn instruction generation. To generate human-like rewrite instructions that focus on modifying specific details of the original conversation, we introduce a multi-turn refinement process (see Table 1) that iteratively enhances instruction specificity and naturalness. We first integrate natural prompts and raw emails, using few-shot demonstrations (see Tables 14, 15 in Appendix C) to generate raw instructions. These raw instructions, however, often lack and fail to guide nuanced rewrites (e.g., adding or removing details). For example, they might simply request specifying placeholders and benefits for customers in the raw email without providing concrete details.

We refine *raw* instructions into a more *specific* version via few-shot prompting (see Tables 16, 17 in Appendix C). For example, instantiate a concrete social media example (“TikTok”) and specific benefits (“chance to win prizes in contests and giveaways”). Then, we further refine specific instructions into a *natural*, human-like style (see Tables 18, 19 in Appendix C). For example, use oral expressions (“say we’re now on...”, “let’s tell them about the cool stuff...”) to improve engagement and conversational fluency.

Rewrite generation. Finally, we generate revised emails by combining the natural prompt, raw email, and final rewrite instruction via structured template (see Table 13 in Appendix C) to prompt LLMs. Each example in CHATREWRITE consists of a quadruplet (`<natural prompt>`, `<raw email>`, `<instruction>`, `<revised email>`).

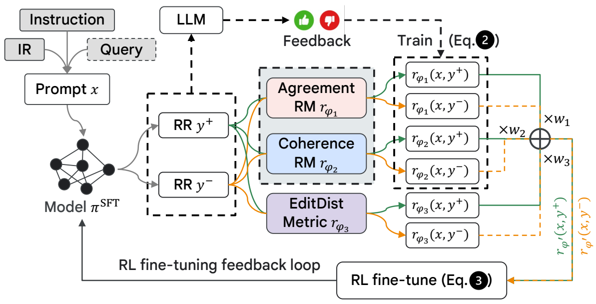


Figure 2: DR GENRÉ: RL fine-tuning with weighted decoupled rewards. Dashed lines represent workflows of reward modeling (for agreement and coherence). IR, RR denote initial and revised responses.

4 Framework of DR GENRÉ

4.1 Supervised Fine-Tuning (SFT)

SFT involves training a LLM on our synthetic (prompt, revisions) datasets. The prompt specify the task to be performed as well as the source text, and the responses are the generated revised outputs. This process distills the basic knowledge of reasonable text rewriting patterns from LLMs to a unified student model, and teaches the model to learn to follow diverse instructions and perform a variety of tasks. In our approach, we fine-tune a pretrained model on the combined datasets of factuality, style, and conversational rewrites. By providing explicit instructions and corresponding rewrites, the model learns to generalize across different rewriting tasks under supervised learning, acting as a reference policy π^{SFT} for the later RL stage.

4.2 Reward Modeling with LLM Preference

Preference data annotation. After SFT, the model has a basic knowledge of what high-quality revisions be like but not well-aligned with implicit preferences. For each input, we sample 10 SFT responses and compute their agreement and coherence scores by few-shot prompting LLMs. We design an agreement judging prompt for each of the three tasks (see Tables 20∼22 in Appendix D). For coherence, as it only depends on the revised response, we use consistent prompting (see Table 23 in Appendix D) across all tasks. We select a pair of revised responses with highest and lowest scores (denoted as y^+ and y^-) for each prompt x to formulate the RM training set.

RM training. We then train a generic agreement RM r_{φ_1}, and coherence RM r_{φ_2}[2], using a

[2] Conciseness reward is a rule-based edit distance metric.

mixture of three preference datasets. We adopt the Bradley-Terry (BT) (Bradley and Terry, 1952) model where for each pair of revised responses $y^+, y^- \sim \pi^{\text{SFT}}(y|x)$ given a prompt x from training data $\mathcal{D}$, the preference probability is defined as

$$P(y^+ \succ y^- | x) = \frac{e^{r^*(y^+,x)}}{e^{r^*(y^+,x)} + e^{r^*(y^-,x)}}, \quad (1)$$

where r^* represents a RM (e.g., $r_{\varphi_1}, r_{\varphi_2}$). The BT loss function is then formulated as

$$\mathcal{L}_R = -\mathbb{E}_{(x,y^+,y^-)\sim\mathcal{D}}[\log \sigma(r^*(y^+, x) - r^*(y^-, x))], \quad (2)$$

where σ is the sigmoid function. Using this framework, we independently train the agreement and coherence reward models.

4.3 Reinforcement Learning with Decoupled Rewards

During the reward modeling phase, the SFT model takes prompts x and generates response pairs $y^+, y^- \sim \pi^{\text{SFT}(y|x)}$, which are then evaluated by each reward function to generate objective-oriented preference signals. In the RL phase, we integrate all reward functions to guide policy optimization. Specifically, we incorporate the decoupled rewards into the PPO[3] (Schulman et al., 2017) objective

$$\max_{\pi_\theta} \mathbb{E}_{x\sim\mathcal{D}, y\sim\pi_\theta(y|x)}[r_{\varphi'}(x, y)] \\ - \beta \cdot \mathbb{D}_{\text{KL}}[\pi_\theta(y|x) || \pi_{\text{ref}}(y|x)], \quad (3)$$

where β is a parameter that controls the deviation from the reference policy π_{ref} (the initial SFT model π^{SFT}), and $r_{\varphi'}$ represents the aggregated decoupled reward function

$$r_{\varphi'}(x, y) = \sum_{o=1}^{O} w_o^t \cdot r_{\varphi_o}(x, y), \text{where } (x, y) \sim \mathcal{D}^t. \quad (4)$$

Here, $O = 3$ is the number of objectives, $\mathcal{D}^t$ denotes the dataset for task t (e.g., LONGFACT), and $w_o^t \in [0, 1]$ is a task-specific weight for objective o-oriented reward r_{φ_o}. For example, conversational rewriting samples will be assigned a higher agreement weight w_1 than stylistic rewriting, as they involve more intricate detailed editing. An illustration of DR GENRÉ is shown in Figure 2.

[3]We use PPO instead of Direct Policy Optimization (DPO) (Rafailov et al., 2024) since our approach relies on explicit reward modeling rather than implicit preference scores. Additionally, PPO allows for exploration beyond the initial SFT policy, potentially discovering better rewriting strategies.

5 Evaluation

5.1 Setup

We use instruction-tuned PaLM 2-L (denoted PaLM 2-L-IT) (Anil et al., 2023) as the teacher model for generating both training data (§3) and reward modeling data (§4.2). The base (student) model is a PaLM 2-S[4], which serves as the foundation for few-shot experiments and the initial checkpoint for SFT. Both agreement and coherence RMs are PaLM 2-M, fine-tuned to capture the teacher model's preferences and guide the student model during RL. For evaluation, we employ Gemini-1.5-Ultra (Team et al., 2023) as the AutoRater, which are used together with rule-based metrics to rate the quality of generated rewrites. We detail the other experimental setups in Appendix A.

Metrics. We evaluate performance across three objective-oriented metrics using AutoRaters:

- **Agreement**: Measures how well the revised text adheres to the instruction in terms of atomic requirements (e.g., correcting non-factual statements). We design task-specific agreement prompts to capture these requirements.
- **Coherence**: Judge whether the revised response is internally consistent. We use few-shot LLM prompting to assess coherence.
- **Edit Ratio** (Ristad and Yianilos, 1998): Quantifies the word-level textural difference between the original and revised texts. This is computed as the relative edit distance normalized by the length of the original text, reflecting conciseness—the proportion of text modified.

For LONGFACT, we follow Wei et al. (2024) to evaluate fact correction accuracy **F1@K**, where K is the expected number of facts per response. We consider two K values: the medium and maximum number of facts averaged in LONGFACT.

For REWRITELM, we follow Shu et al. (2024) to measure **NLI** (Bowman et al., 2015) and **Reverse NLI** score over the source-revision pairs. These scores estimate how well the rewrite retains the original information. Higher NLI and lower edit ratio are desirable, as excessive edits can introduce hallucinations if the NLI scores are low.

For CHATREWRITE, we use auto Side-by-Side (**AutoSxS**) (Zheng et al., 2023) for pairwise agreement evaluation. AutoSxS compares responses generated by different models for the same prompt (see Tables 24, 25 in Appendix D). It is particularly

[4]We choose PaLM 2 models to be consistent with that of the prior rewrite work, i.e., RewriteLM (Shu et al., 2024).

Method	Length	F1@13↑	F1@35↑	Agreement↑	Coherence↑	Edit Ratio↓
Few-shot	287	0.7232	0.4579	0.7607	0.6367	0.0420
Surgical	369	0.8255	0.5480	**0.9238**	0.4120	**0.0108**
SFT-LongFact	348	0.7967	0.5192	0.8011	0.5141	0.0214
SFT	348	0.8209	0.5204	0.7950	0.5400	0.0210
DR GENRÉ-static	389	**0.8261**	**0.5531**	0.7926	**0.6520**	0.0583
DR GENRÉ	365	0.8091	0.5409	0.7878	0.6400	0.0270

Table 3: Performance on LONGFACT. Methods are grouped into ICL-based, SFT-based, and RL-based.

Method	Length	NLI↑	Reverse NLI↑	Agreement↑	Coherence↑	Edit Ratio↓
Few-shot	108	0.8790	0.8418	0.8235	**0.6960**	0.1168
SFT-RewriteLM	106	0.8806	0.8563	0.9524	0.6720	0.1242
SFT	102	0.8914	0.8718	0.9163	0.6840	**0.1078**
RL-CoComposer	181	**0.9256**	**0.8830**	0.9042	0.6480	0.2499
DR GENRÉ-static	107	0.9173	0.8591	0.9433	0.6800	0.1200
DR GENRÉ	101	0.8937	0.8684	**0.9641**	0.6834	0.1541

Table 4: Performance on OPENREWRITEEVAL (Shu et al., 2024). Length is the averaged output length.

useful for capturing nuanced differences between responses, as pointwise agreement checks often neglect implicit differences.

Baselines. We compare DR GENRÉ with three rewrite generation baselines[5] across all tasks:

- **Few-shot (ICL)**: Direct generate rewritten texts by few-shot prompting PaLM 2-S.
- **SFT**: PaLM 2-S fine-tuned on a mixture of all datasets, serving as a supervised baseline.
- **DR GENRÉ-static**: PaLM 2-S fine-tuned with RL using static weights, i.e., w_o^t becomes task-agnostic w_o, for the three reward objectives.

5.2 Results on LONGFACT

For the factuality rewriting task, we include two additional baselines:

- **Surgical**: Directly substituting non-factual contents identified in the critique outputs with the factual revisions, without further refinement.
- **SFT-LongFact**: Training the PaLM 2-S exclusively on LONGFACT. This baseline allows isolating the effect of task-specific training compared to the generic training used in SFT.

Table 3 presents the results on the sampled 250 query subset (Wei et al., 2024) of LONGFACT. DR GENRÉ-static achieves the highest factual correction performance (F1@13: 0.82, F1@35: 0.55), along with the highest coherence, while maintaining competitive edit ratio. DR GENRÉ, with dynamically adjusted reward weights, balances factuality and edit preservation, yielding a lower edit ratio and strong coherence while maintaining factual accuracy (F1@13: 0.81, F1@35: 0.54).

SFT outperforms SFT-LongFact, confirming that multi-task learning (Vilalta and Drissi, 2002) benefits factuality rewriting. However, all SFT-based methods experience a significant drop in coherence (around 10%) compared to few-shot prompting, highlighting *the challenge of maintaining internal consistency while improving instruction adherence in factuality rewriting*. Such limitation is further exemplified by the surgical baseline, which, despite achieving the highest agreement score, exhibits poor coherence due to its lack of refinement.

DR GENRÉ mitigates this trade-off by dynamically adjusting reward weights based on task properties. For example, by assigning higher edit distance weights to tasks requiring more substantial revisions (e.g., factuality vs. stylistic), it ensures robustness and no severe deviation (e.g., lower edit ratio compared to DR GENRÉ-static) throughout the post-training process.

5.3 Results on OPENREWRITEEVAL

For the stylistic rewriting task, we investigate two additional baselines:

- **SFT-RewriteLM**: Trains the PaLM 2-S exclusively on REWRITELM, without leveraging the patterns from other tasks.
- **RL-CoComposer**: An existing RL-based method (Shu et al., 2024) that optimizes a single reward function, where outputs receive a binary score (0 or 1) based on edit ratio constraints, NLI scores, and length-based transformations, effectively acting as a binary filter.

Table 4 summarizes the results on OPEN-REWRITEEVAL (the evaluation set of RewriteLM). DR GENRÉ achieves the highest agreement, demon-

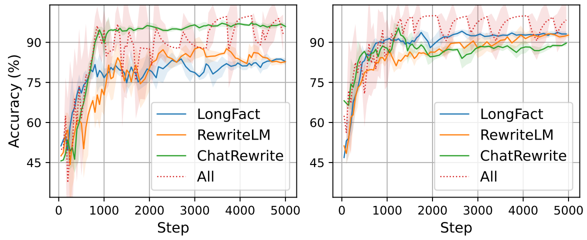
[5] We discuss more selection rationale in Appendix A.

R1 \ R2	Few-shot	SFT	SFT-CR	Dr Genré-s.	Dr Genré
Few-shot	–	0.044 \ **0.661**	0.036 \ **0.685**	0.012 \ **0.806**	0.020 \ **0.754**
SFT	**0.552** \ 0.133	–	0.145 \ **0.319**	0.068 \ **0.512**	0.097 \ **0.435**
SFT-CR	**0.565** \ 0.105	0.161 \ **0.254**	–	0.065 \ **0.484**	0.093 \ **0.444**
Dr Genré-s.	**0.681** \ 0.089	**0.371** \ **0.173**	**0.298** \ 0.185	–	0.206 \ **0.262**
Dr Genré	**0.649** \ 0.113	**0.258** \ 0.206	**0.298** \ 0.234	**0.239** \ 0.225	–

Table 5: AutoSxS results comparing different models on ChatRewrite dataset. Each value denotes the average confidence in one side of the pairwise responses, with **bold** texts highlighting the preferred side. SFT-CR represents SFT-ChatRewrite. Dr Genré-s. denotes Dr Genré-static.

Method	Length	Agreement↑	Coherence↑	Edit Ratio↓
Few-shot	118	0.7786	0.8347	0.1277
SFT-CR	114	0.9200	0.8347	**0.1003**
SFT	114	0.9308	0.8468	0.1036
Dr Genré-s.	123	0.9584	0.8548	0.1278
Dr Genré	119	**0.9648**	**0.8669**	0.1243

Table 6: Model performance on our ChatRewrite.



(a) Agreement (b) Coherence

Figure 3: Pairwise learning curve of (a) agreement and (b) coherence reward modeling on both mixture of tasks ("All") and each individual one.

strating superior adherence to stylistic rewrite instructions. It also maintains high coherence while balancing semantic preservation. Its edit ratio (0.1541) is significantly lower than RL-CoComposer (0.2499), highlighting the advantage of *fine-grained, decoupled rewards* over binary filtering in balancing multiple objectives.

Few-shot prompting achieves competitive coherence (0.6960) and semantic preservation (NLI: 0.8790, Reverse NLI: 0.8418), but its limited edits result in low agreement (0.8235), indicating insufficient transformation. SFT-RewriteLM performs well in agreement (0.9524) but lags in coherence compared to generic SFT, reinforcing the benefit of multi-task training.

While stylistic rewriting is a relatively simpler task that primarily involves transformation without introducing new information, Dr Genré effectively balances instruction adherence, semantic consistency, and stylistic flexibility, making it a robust and adaptable solution.

5.4 Results on ChatRewrite

Conversational rewriting requires balancing instruction adherence, coherence, and conciseness while preserving the intended tone and nuances. We introduce an additional baseline, **SFT-ChatRewrite**, which trains PaLM 2-S exclusively on the ChatRewrite dataset to isolate the effect of task-specific training.

Overall comparison. Table 6 presents the results on ChatRewrite. Post-training methods yield larger improvements in agreement over few-shot prompting compared to other tasks, suggesting that *conversational rewrite is a more complex task that cannot be adequately solved by in-context learning*. Dr Genré achieves the highest agreement (0.9648) and coherence (0.8669), proving its ability in adhering to conversational rewrite instructions while maintaining fluency. Dr Genré-static performs competitively in agreement but lags slightly in coherence, highlighting the limitations of fixed reward weighting in handling diverse conversational styles. While SFT-ChatRewrite and SFT perform well in agreement (0.9200 and 0.9308), they fall behind in coherence, reinforcing the benefits of RL in refining response consistency. Notably, edit ratios across models remain similar, indicating that improvements are not due to excessive modifications but rather more effective rewrites.

Side-by-side comparison. Table 5 provides AutoSxS results, offering a more fine-grained perspective on pairwise preferences. Dr Genré consistently outperforms SFT and SFT-ChatRewrite, confirming that it refines responses beyond what supervised fine-tuning can achieve. While Dr Genré-static competes closely, its performance declines against Dr Genré, underscoring the benefits of dynamical reward weighting in handling diverse conversational styles. Overall, the strong AutoSxS results validate that Dr Genré not only improves instruction adherence but also generates more preferred in direct comparisons.

5.5 Case Study

Preference alignment. As shown in Figure 3, both the agreement and coherence RMs demonstrate

Natural prompt: Congratulate myself and the team on a successful project. **Raw email:** Dear team, I'm writing to congratulate everyone on a successful project launch. Your hard work and dedication were essential to our success. I'm especially proud of the way we all came together to support each other and solve problems. We faced some challenges along the way, but we never gave up. I'm grateful for the opportunity to work with such a talented and dedicated team. Thank you for all your hard work. **Instruction:** This email is a bit dry, let's celebrate our success! Add some enthusiastic phrases like "We nailed it!" or "What a triumph!". Also highlight specific achievements like exceeding our target goal by 15% or landing that big client contract.	
Revised email (SFT): Dear team, We nailed it! What a triumph! We successfully launched our project and exceeded our target goal by 15%. We also landed that big client contract! Your hard work, dedication, and teamwork were essential to our success. I'm so proud of the way we all came together to support each other and solve problems. We faced some challenges along the way, but we never gave up. Thank you for all your hard work and dedication. You are a talented and dedicated team, and I'm grateful for the opportunity to work with you.	**Revised email (Dr Genré):** Dear team, We nailed it! Our successful project launch was a triumph for all of us. Your hard work and dedication were essential to our success. I'm especially proud of the way we all came together to support each other and solve problems. We faced some challenges along the way, but we never gave up. Not only did we exceed our target goal by 15%, but we also landed that big client contract! This is a huge win for our team and our company. We are truly unstoppable. I'm so grateful for the opportunity to work with such a talented and dedicated team. Thank you for all your hard work and dedication. We're just getting started! Celebrate with me!
Score (SFT): 0.7	**Score (Dr Genré):** 0.9
Reasoning: Response (Dr Genré) does a better job of infusing enthusiasm and highlighting specific achievements in a natural way. It effectively uses phrases like "We nailed it!" and "This is a huge win," and it also adds a celebratory call to action, "Celebrate with me!", which further enhances the celebratory tone of the email.	

Table 7: An example of AutoSxS result comparing SFT and RL (Dr Genré) responses on ChatRewrite.

progressive learning of preference knowledge from the teacher LLM AutoRaters. The pairwise accuracy steadily improves across training steps, with the mixture-trained ("All") reward model achieving the highest accuracy, indicating better generalization across different rewriting objectives. However, mixed rewards are also more fluctuated with larger shaded regions compared to individual task-specific RMs (e.g., LongFact, RewriteLM), reflecting the difficulty of aligning preferences (e.g., factual vs. stylistic rewrites). This variance across different evaluation runs highlights the importance of learning dynamics due to task complexity.

RL fine-tuning. Table 7 shows an example from ChatRewrite, comparing responses generated by the SFT and Dr Genré for a celebratory email revision task. The instruction emphasizes enhancing enthusiasm and highlighting specific achievements. While the SFT response incorporates phrases from the instruction, it mainly mirrors the given prompt rather than naturally enhancing the overall celebratory tone. In contrast, the Dr Genré response exhibits greater creativity and engagement by integrating expressive phrases. These additions not only follow the instruction but also improve emotional resonance in a more natural way. We show examples for LongFact (Table 8) and RewriteLM (Table 10) in Appendix B.

6 Conclusion

Generic text rewriting is inherently a *multi-task*, *multi-objective* problem, requiring models to adapt to diverse rewriting needs, such as factual correction, stylistic transformation, and conversational enhancement. Existing approaches, which either consider a single task or apply static reward functions, struggle to generalize across these objectives. To address this, we introduce a more comprehensive evaluation setup, including a newly constructed conversational rewrite dataset, ChatRewrite, which emphasizes detailed instructions and personalized editing in interactive scenarios.

To tackle the complexities of generic rewriting, we propose Dr Genré, an RL framework that decouples reward signals and dynamically adjusts their contributions based on task-specific requirements. Our method outperforms existing SFT and RL baselines across LongFact, OpenRewriteEval, and ChatRewrite, demonstrating its ability to balance instruction adherence, coherence, and edit efficiency. Notably, AutoSxS results validate its superiority in nuanced rewrites, where traditional metrics fall short. Future work will explore human-in-the-loop optimization and context-aware reward modeling to further refine its performance in complex rewriting scenarios.

Limitation

While DR GENRÉ demonstrates strong performance in generic text rewriting, several limitations should be acknowledged along with potential considerations for improvement.

First, our dataset generation process relies on LLMs for critique and rewriting, which may introduce biases, inconsistencies, or hallucinations inherited from the teacher model. To mitigate this, we employ reward models trained on multiple datasets and enable external fact-checking calls from LLMs to refine the generated outputs. We also ensure the quality of CHATREWRITE by randomly sampling rewrite pairs and checking both instructions and revised responses. However, future work could explore incorporating human-annotated critique to enhance reliability.

Second, while AutoRaters provide a scalable evaluation mechanism, they may not fully capture nuanced human preferences, especially in conversational rewrites. We mitigate this by also considering rule-based metrics (e.g., edit ratio, NLI scores) as basic judgements for rewriting quality, but further improvements could involve human-in-the-loop evaluation pipelines, or hybrid scoring systems that integrate both automatic and human judgements.

Last, our approach relies on proprietary LLMs for training all components, which may pose challenges for reproducibility. To facilitate practitioners, we provide detailed prompts with demonstrations in Appendix C, along with comprehensive methodology description 4.

Acknowledgment

We acknowledge the use of LLMs for assisting with writing refinement and prompt design throughout this work. Specifically, GPT-4 was used to polish textual clarity and Gemini-1.5-Ultra was used to enhance prompt engineering strategies. While the core methodology and analyses were conducted independently, LLM-based assistance helped streamline certain aspects of content presentation.

References

Rohan Anil, Andrew M Dai, Orhan Firat, Melvin Johnson, Dmitry Lepikhin, Alexandre Passos, Siamak Shakeri, Emanuel Taropa, Paige Bailey, Zhifeng Chen, et al. 2023. Palm 2 technical report. *arXiv preprint arXiv:2305.10403*.

Samuel Bowman, Gabor Angeli, Christopher Potts, and Christopher D Manning. 2015. A large annotated corpus for learning natural language inference. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 632–642.

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.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. 2017. Deep reinforcement learning from human preferences. *Advances in neural information processing systems*, 30.

Pengrui Han, Rafal Kocielnik, Adhithya Saravanan, Roy Jiang, Or Sharir, and Anima Anandkumar. 2024. ChatGPT based data augmentation for improved parameter-efficient debiasing of LLMs. In *Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity, Inclusion*, pages 73–105, St. Julian's, Malta. Association for Computational Linguistics.

Junxian He, Jiatao Gu, Jiajun Shen, and Marc'Aurelio Ranzato. 2020. Revisiting self-training for neural sequence generation. In *International Conference on Learning Representations*.

Linmei Hu, Zeyi Liu, Ziwang Zhao, Lei Hou, Liqiang Nie, and Juanzi Li. 2023. A survey of knowledge enhanced pre-trained language models. *IEEE Transactions on Knowledge and Data Engineering*.

Zhiting Hu, Zichao Yang, Xiaodan Liang, Ruslan Salakhutdinov, and Eric P Xing. 2017. Toward controlled generation of text. In *International conference on machine learning*, pages 1587–1596. PMLR.

Jiaxin Huang, Shixiang Gu, Le Hou, Yuexin Wu, Xuezhi Wang, Hongkun Yu, and Jiawei Han. 2023. Large language models can self-improve. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 1051–1068.

Robert Iv, Alexandre Passos, Sameer Singh, and Ming-Wei Chang. 2022. Fruit: Faithfully reflecting updated information in text. In *Proceedings of the 2022 Conference of the North American Chapter of the*

Association for Computational Linguistics: Human Language Technologies, pages 3670–3686.

Harrison Lee, Samrat Phatale, Hassan Mansoor, Thomas Mesnard, Johan Ferret, Kellie Lu, Colton Bishop, Ethan Hall, Victor Carbune, Abhinav Rastogi, et al. 2023. Rlaif: Scaling reinforcement learning from human feedback with ai feedback. *arXiv preprint arXiv:2309.00267*.

Harrison Lee, Samrat Phatale, Hassan Mansoor, Thomas Mesnard, Johan Ferret, Kellie Ren Lu, Colton Bishop, Ethan Hall, Victor Carbune, Abhinav Rastogi, et al. 2024. Rlaif vs. rlhf: Scaling reinforcement learning from human feedback with ai feedback. In *Forty-first International Conference on Machine Learning*.

Juncen Li, Robin Jia, He He, and Percy Liang. 2018. Delete, retrieve, generate: a simple approach to sentiment and style transfer. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1865–1874.

Junyi Li, Xiaoxue Cheng, Wayne Xin Zhao, Jian-Yun Nie, and Ji-Rong Wen. 2023. Halueval: A large-scale hallucination evaluation benchmark for large language models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 6449–6464.

Yichuan Li, Kaize Ding, Jianling Wang, and Kyumin Lee. 2024a. Empowering large language models for textual data augmentation. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 12734–12751, Bangkok, Thailand. Association for Computational Linguistics.

Zelong Li, Wenyue Hua, Hao Wang, He Zhu, and Yongfeng Zhang. 2024b. Formal-llm: Integrating formal language and natural language for controllable llm-based agents. *arXiv preprint arXiv:2402.00798*.

Jonathan Mallinson, Jakub Adamek, Eric Malmi, and Aliaksei Severyn. 2022. EdiT5: Semi-autoregressive text editing with t5 warm-start. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 2126–2138, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Philip May. 2021. Machine translated multilingual sts benchmark dataset.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2024. Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems*, 36.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140):1–67.

Sudha Rao and Joel Tetreault. 2018. Dear sir or madam, may i introduce the gyafc dataset: Corpus, benchmarks and metrics for formality style transfer. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 129–140.

Eric Sven Ristad and Peter N Yianilos. 1998. Learning string-edit distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(5):522–532.

Timo Schick, A Yu Jane, Zhengbao Jiang, Fabio Petroni, Patrick Lewis, Gautier Izacard, Qingfei You, Christoforos Nalmpantis, Edouard Grave, and Sebastian Riedel. 2023. Peer: A collaborative language model. In *The Eleventh International Conference on Learning Representations*.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Noam Shazeer and Mitchell Stern. 2018. Adafactor: Adaptive learning rates with sublinear memory cost. In *International Conference on Machine Learning*, pages 4596–4604. PMLR.

Tianxiao Shen, Tao Lei, Regina Barzilay, and Tommi Jaakkola. 2017. Style transfer from non-parallel text by cross-alignment. *Advances in neural information processing systems*, 30.

Lei Shu, Liangchen Luo, Jayakumar Hoskere, Yun Zhu, Yinxiao Liu, Simon Tong, Jindong Chen, and Lei Meng. 2024. Rewritelm: An instruction-tuned large language model for text rewriting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 18970–18980.

AB Siddique, Samet Oymak, and Vagelis Hristidis. 2020. Unsupervised paraphrasing via deep reinforcement learning. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1800–1809.

Hrituraj Singh, Gaurav Verma, Aparna Garimella, and Balaji Vasan Srinivasan. 2021. Drag: Director-generator language modelling framework for non-parallel author stylized rewriting. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 863–873.

Nisan Stiennon, Long Ouyang, Jeffrey Wu, Daniel Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul F Christiano. 2020. Learning to summarize with human feedback. *Advances in Neural Information Processing Systems*, 33:3008–3021.

Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, et al. 2023. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.

Katherine Tian, Eric Mitchell, Huaxiu Yao, Christopher D Manning, and Chelsea Finn. 2024. Fine-tuning language models for factuality. In *The Twelfth International Conference on Learning Representations*.

Ricardo Vilalta and Youssef Drissi. 2002. A perspective view and survey of meta-learning. *Artificial intelligence review*, 18:77–95.

Tu Vu, Kalpesh Krishna, Salaheddin Alzubi, Chris Tar, Manaal Faruqui, and Yun-Hsuan Sung. 2024. Foundational autoraters: Taming large language models for better automatic evaluation. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 17086–17105, Miami, Florida, USA. Association for Computational Linguistics.

Yizhong Wang, Yeganeh Kordi, Swaroop Mishra, Alisa Liu, Noah A Smith, Daniel Khashabi, and Hannaneh Hajishirzi. 2023. Self-instruct: Aligning language models with self-generated instructions. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13484–13508.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

Jerry Wei, Chengrun Yang, Xinying Song, Yifeng Lu, Nathan Hu, Jie Huang, Dustin Tran, Daiyi Peng, Ruibo Liu, Da Huang, Cosmo Du, and Quoc V Le. 2024. Long-form factuality in large language models. In *Advances in Neural Information Processing Systems*, volume 37, pages 80756–80827. Curran Associates, Inc.

Akhila Yerukola, Xuhui Zhou, Elizabeth Clark, and Maarten Sap. 2023. Don't take this out of context!: On the need for contextual models and evaluations for stylistic rewriting. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 11419–11444.

Yi Zhang, Tao Ge, and Xu Sun. 2020. Parallel data augmentation for formality style transfer. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3221–3228.

Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing, et al. 2023. Judging llm-as-a-judge with mt-bench and chatbot arena. *Advances in Neural Information Processing Systems*, 36:46595–46623.

A Experimental Details

We describe the hardware setup, training configurations, and dataset weighting strategies used in different phases of our experiments, covering Supervised Fine-Tuning (SFT), Reward Modeling, and Reinforcement Learning (RL).

Hardware setup. All training experiments were conducted on 64 Tensor Processing Units (TPU) chips per phase:

- SFT & Reward modeling: TPU V3.
- RL fine-tuning: TPU V4.

For inference, we use a temperature of 1.0 with top-K sampling (K=40).

Supervised fine-tuning (SFT). We fine-tune the base PaLM 2-S model on our dataset mixture using Adafactor (Shazeer and Stern, 2018) with the following configuration:

- Batch size: 64.
- Max training steps: 1000.
- Learning rate: 1e-5.
- Dropout: 0.1.
- Max context length: 2048.
- Max decoding length: 1024.

Reward modeling. We train reward models on preference data collected from LLM comparisons using the following setup:

- Batch size: 64.
- Max training steps: 5000.
- Learning rate: 3e-3.
- Dropout: 0.05.
- Max context length: 1280.
- Max decoding length: 1024.
- Optional Z_loss: 1e-2.

Reinforcement learning (RL) fine-tuning. For policy optimization, we employ PPO with dynamically weighted multi-objective rewards. The policy and value functions are optimized separately:

- Batch size: 64.
- Max training steps: 3000 (with a warm-up phase of the first 100 steps where we train only value functions and freeze policy).
- Learning rate: 1e-7 for policy and 1e-5 for value.
- Dropout: None.
- Max context length: 2048.
- Max decoding length: 1024.

Dataset weighting strategy. We assign different dataset weights based on task-specific objectives to balance training across agreement, coherence, and edit conciseness.

For DR GENRÉ-static (task-agnostic), we use a fixed weighting of $w_1 = 9/16$, $w_2 = 2/16$, $w_3 =$

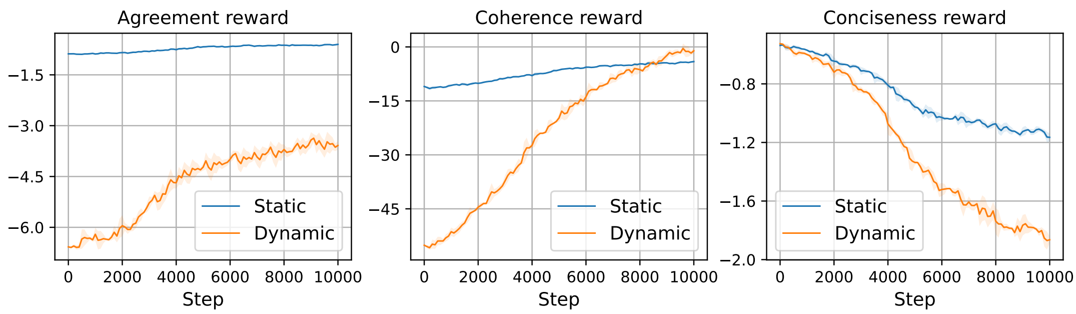


Figure 4: Reward learning curves during RL fine-tuning under static and dynamic weighting.

$5/16$ for agreement, coherence, and conciseness.

For Dr Genré (task-specific), we empirically set:

- LongFact: $w_1 = 8/16$, $w_2 = 6/16$, $w_3 = 2/16$.
- RewriteLM: $w_1 = 3/9$, $w_2 = 4/9$, $w_3 = 2/9$.
- ChatRewrite: $w_1 = 9/16$, $w_2 = 5/16$, $w_3 = 2/16$.

The dynamic weighting scheme ensures that different datasets prioritize their most relevant rewrite objectives, allowing for more effective RL fine-tuning.

Baseline selection. In our experiments, we focus on three major categories of baselines: ICL-based, SFT-based, and RL-based methods. There are some existing works in factual or stylistic rewriting focus on either direct editing heuristics or single-objective models that do not fit well with our multi-objective formulation.

For factual rewriting, works like knowledge-grounded editing rely on retrieval-based fact verification or human annotations (Tian et al., 2024), whereas our approach optimizes for factuality without requiring explicit retrieval.

For stylistic rewriting, previous works often rely on large supervised datasets for a single transformation (e.g., formality change, politeness adjustment, style matching (Singh et al., 2021)) or context integration (Yerukola et al., 2023), whereas our model generalizes across multiple stylistic transformations.

Even if the above task-specific approaches perform well in their domain, they do not necessarily generalize across diverse rewriting tasks, making them less suitable as baselines in our setting.

Static and dynamic weights. Figure 4 represents the reward curves of the RL fine-tuning phase (Dr Genré-static and Dr Genré). Dynamic RL (Dr Genré) adapts objectives over time, focusing more on agreement and coherence, and less on the conciseness. In contrast, static RL (Dr Genré-static) is more stable (balanced) but less optimized learning across objectives. Dynamic RL exhibits stronger overall improvement rates across all three objectives, confirming its ability to adjust to task needs more effectively.

B Generated Examples

This section presents qualitative examples of factuality and stylistic rewrite cases generated by Dr Genré, illustrating its ability to correct errors while preserving coherence and adhering to task-specific instructions.

Factuality rewrite. The first example in Table 8 showcases a factuality rewrite on the topic of the United States' involvement in the East Asia Summit (EAS). The initial response contains several factual inaccuracies, such as "Incorrectly stating that the U.S. has been involved in the EAS since its inception in 2005 (corrected to 2011)". The critique outputs (highlighted in red for incorrect and blue for revised content) pinpoint these errors, allowing Dr Genré to generate a factually accurate response while maintaining internal coherence. We also show an example of critique outputs from SAFE (Wei et al., 2024) in Table 9, where for each span, the outputs contain a revision (from external fact-checking calls) and a reason. Compared to the initial response, the revised version (i) corrects all factual errors without introducing unnecessary modifications, (ii) preserves the original structure and key ideas, and (iii) Improves clarity by streamlining redundant information (e.g., simplifying the explanation of U.S. participation). This demonstrates Dr Genré's capability to integrate factual verification and controlled editing, ensuring reliability in long-form factual rewriting tasks.

Stylistic rewrite. The second example in Table 10 illustrates a style transfer task, where the instruc-

Query	How is the United States related to the East Asia Summit (EAS)? Provide as many specific details and examples as possible (such as names of people, numbers, events, locations, dates, times, etc.)
Initial response	The United States is not a member of the East Asia Summit (EAS), but participates in the EAS as a guest. The EAS is a forum for leaders of 18 countries in the Asia-Pacific region to discuss issues of common interest, such as economic cooperation, security, and sustainability. The United States has been invited to participate in the EAS since its inception in 2005, and has attended every summit since then. The United States plays an important role in the EAS, even though it is not a member. The United States is a major economic and military power in the Asia-Pacific region, and its participation in the EAS allows it to engage with other regional leaders on issues of mutual concern. The United States also uses the EAS to promote its own interests in the region, such as expanding trade and investment and countering the rise of China. Some specific examples of the United States' involvement in the EAS include: • In 2015, President Obama announced the launch of the US-ASEAN Connect initiative, which aimed to deepen ties between the United States and the Association of Southeast Asian Nations (ASEAN), a key regional organization. • In 2017, President Trump attended the EAS and met with several regional leaders, including Chinese President Xi Jinping. • In 2018, the United States announced that it would provide $113 million in funding for regional infrastructure projects, as part of its commitment to the EAS. The United States' participation in the EAS is a reflection of its commitment to the Asia-Pacific region. The EAS provides a valuable forum for the United States to engage with other regional leaders and promote its own interests.
Critique outputs	- The United States is ~~not a member of the East Asia Summit (EAS), but participates in the EAS as a guest.~~ *a member of the East Asia Summit (EAS).* - The United States has been invited to participate in the EAS since ~~its inception in 2005~~ *2011*, and has attended every summit since then. - In 2017, President Trump ~~attended the EAS and met with several regional leaders, including Chinese President Xi Jinping.~~ *did not attend the EAS, but sent his Secretary of State, Rex Tillerson, instead.*
Revised response (Dr Genré)	The United States is a member of the East Asia Summit (EAS). The EAS is a forum for leaders of 18 countries in the Asia-Pacific region to discuss issues of common interest, such as economic cooperation, security, and sustainability. The United States has been invited to participate in the EAS since 2011, and has attended every summit since then. The United States plays an important role in the EAS. As a major economic and military power in the Asia-Pacific region, its participation allows it to engage with other regional leaders on issues of mutual concern. The United States also uses the EAS to promote its own interests in the region, such as expanding trade and investment and countering the rise of China. Specific examples of U.S. involvement in the EAS include: • In 2015, President Obama announced the US-ASEAN Connect initiative to deepen ties with ASEAN. • In 2017, President Trump did not attend the EAS but sent Secretary of State Rex Tillerson. • In 2018, the United States provided $113 million in funding for regional infrastructure projects. The United States' participation in the EAS reflects its commitment to the Asia-Pacific region, offering a valuable forum for engaging with regional leaders and advancing its interests.

Table 8: An example of factuality rewrite involving the East Asia Summit (EAS) on LongFact. The critique identifies inaccuracies (~~strike through~~ and *italic* denote non-factual and corrected statements) in the initial response, and the revised response (Dr Genré) corrects these errors while maintaining consistency and relevance.

Query	What is the game Undertale? Provide as many specific details and examples as possible (such as names of people, numbers, events, locations, dates, times, etc.)
Initial response	Undertale is a role-playing video game created by Toby Fox. It was released in 2015 for Microsoft Windows and macOS, and in 2016 for Linux. The game has been ported to PlayStation 4, PlayStation Vita, Nintendo Switch, Xbox One, and iOS. Undertale is a story about a human child who falls into the Underground, a world inhabited by monsters. The player must navigate through the Underground, solving puzzles and battling monsters, in order to return to the surface. The game is notable for its unique combat system, which allows the player to choose whether to kill or spare the monsters they encounter. Undertale also features a branching narrative, with multiple endings depending on the player's choices. The game has been praised for its writing, music, and characters. It has been nominated for and won numerous awards, including Game of the Year from several publications. Here are some specific details and examples about Undertale: * **Characters:** The game features a cast of memorable characters, including the protagonist, a human child named Frisk; Flowey, a sentient flower; Toriel, a motherly goat monster; Sans and Papyrus, two skeleton brothers; and Undyne, a fish-like warrior. * **Combat:** The combat system in Undertale is unique in that it allows the player to choose whether to kill or spare the monsters they encounter. Killing monsters will earn the player EXP and gold, but it will also make the game more difficult. Sparing monsters will reward the player with less EXP and gold, but it will also allow them to progress through the game more easily. * **Story:** The story of Undertale is complex and multifaceted, with multiple endings depending on the player's choices. The game explores themes of morality, choice, and consequence. * **Music:** The music in Undertale is composed by Toby Fox and is widely considered to be one of the game's best features. The soundtrack is diverse, with songs ranging from upbeat and catchy to somber and atmospheric. * **Awards:** Undertale has been nominated for and won numerous awards, including Game of the Year from several publications. Here are some additional details about Undertale: * The game was developed by Toby Fox over the course of three years. * It was originally released for Microsoft Windows and macOS in September 2015. * The game was ported to Linux in July 2016. * Undertale has been ported to PlayStation 4, PlayStation Vita, Nintendo Switch, Xbox One, and iOS. * The game has sold over 1 million copies.
Critique outputs	**Span (& Revision) 1**:The game has been ported to PlayStation 4, PlayStation Vita, Nintendo Switch, and Xbox One~~, and iOS~~. - Reason: The evidence suggests that Undertale has been ported to PlayStation 4, PlayStation Vita, Nintendo Switch, and Xbox One. However, there is no evidence that it has been officially ported to iOS. The search results mention discussions about a possible port and unofficial ports, but none of them are official. **Span (& Revision) 2**: It has been nominated for ~~and won numerous~~ several awards, including Game of the Year from several publications. - Reason: The Wikipedia snippet states that the game was nominated for Game of the Year awards, but there is no information on whether it won any. **Span (& Revision) 3**: Killing monsters will earn the player EXP and gold, but it will ~~also~~ not make the game more difficult. In fact, it can make the game easier by providing the player with more resources. - Reason: The sentence contains two factual claims: 1. Killing monsters will earn the player EXP and gold. 2. Killing monsters will make the game more difficult. The first claim is accurate and can be verified by playing the game or by consulting online sources. The second claim is inaccurate. Killing monsters does not make the game more difficult; in fact, it can make the game easier by providing the player with more resources. The search result provides further evidence for the inaccuracy of the second claim. The Reddit post describes a player who killed most of the monsters in their first playthrough and found the game to be easy. **Span (& Revision) 4**: Sparing monsters will reward the player with less EXP and gold, but it will also ~~allow them to progress through the game more easily.~~ make the game more difficult in some ways, as the player will have less resources to use. - Reason: The sentence contains two factual claims: 1. Sparing monsters will reward the player with less EXP and gold. 2. Sparing monsters will allow the player to progress through the game more easily. The first claim is accurate, as evidenced by evidence 3. However, the second claim is inaccurate. Evidence 1 and 2 suggest that sparing monsters can make the game more difficult, as the player will have less EXP and gold to buy items and power ups. Therefore, the sentence is partially inaccurate. **Span (& Revision) 5**:* The game was developed by Toby Fox over the course of approximately three years. - Reason: The sentence claims that Toby Fox developed Undertale over the course of three years. The Evidence [1] states that Toby Fox started developing Undertale in December 2012 and the game was released in September 2015. This means that the development time is roughly 2 years and 9 months, which is close to three years but not exactly three years. **Span (& Revision) 6**: * Undertale has been ported to PlayStation 4, PlayStation Vita, Nintendo Switch~~, Xbox One~~, and iOS. - Reason: The evidence shows that Undertale has been ported to PlayStation 4, PlayStation Vita, Nintendo Switch, and iOS. However, there is no evidence that it has been ported to Xbox One.

Table 9: An example of critique outputs from SAFE (Wei et al., 2024).

Source	Modern jet combat aircraft and helicopters have a variety of complex electronic and weapon systems that are specific to a particular aircraft. This led to a genre called the "study sim" that focuses on modelling an aircraft's systems as accurately as possible. Advancements in computer technology led to the development of highly detailed models that improved upon the fidelity of avionics, weapons systems, physics, flight models, graphics, etc. and allowed for fully interactive cockpits where virtually every control was mapped and functional.
Instruction	Make sure to put the definition at the beginning of the paragraph.
Type (optional)	Wiki
Revised text (DR GENRÉ)	A "study sim" is a genre of flight simulator that focuses on modelling an aircraft's systems as accurately as possible. Modern jet combat aircraft and helicopters have a variety of complex electronic and weapon systems that are specific to a particular aircraft. Advancements in computer technology led to the development of highly detailed models that improved upon the fidelity of avionics, weapons systems, physics, flight models, graphics, etc. and allowed for fully interactive cockpits where virtually every control was mapped and functional.

Table 10: An example of style transfer involving a "study sim" genre, where the instruction emphasizes putting the definition at the beginning of the paragraph. The revised response follows this structure while maintaining the original meaning.

tion specifies reordering the content to place the definition at the beginning of the paragraph. The original text describes modern jet combat aircraft systems before introducing the term "study sim." The revised response (i) restructures the content by moving the definition ("A study sim is a genre of flight simulator...") to the start, (ii) retains all relevant information, ensuring semantic consistency, and (iii) maintains fluency and coherence, keeping the technical details intact. This highlights DR GENRÉ's ability to follow explicit structural modifications while preserving meaning and style, a critical requirement for controlled text rewriting.

C Prompt Templates

Factuality rewrite prompt. The factuality rewrite prompt (Table 11) is structured to ensure precise factual corrections while preserving the overall coherence of the response. Key considerations include: (i) Explicit span-level corrections: The prompt provides non-factual spans and their corresponding factual replacements, ensuring that the model understands exactly what needs to be corrected. (ii) Minimal edit constraints: The instruction explicitly asks the model to modify only necessary parts to correct inaccuracies while preventing unnecessary alterations that could introduce inconsistencies. (iii) Instructional clarity: The step-by-step instructions help the model integrate factual updates without disrupting the original flow, addressing a common challenge in factuality-driven text revision. This structure ensures that factual inaccuracies are corrected while preserving fluency and minimizing unintended modifications.

Stylistic rewrite prompt. The stylistic rewrite prompt (Table 12) differs in its focus on open-ended transformations rather than strict factual accuracy. Key aspects include: (i) Flexible rewrite instructions: The model is provided with a high-level comment (e.g., "formalize," "elaborate") rather than explicit edits, requiring it to understand and interpret the transformation intent. (ii) Preserving meaning: The instructions emphasize maintaining the original meaning and context, which is crucial in tasks such as paraphrasing and formalization where semantic drift is a risk. (iii) Encouraging coherence: Stylistic changes often require restructuring the sentence flow, so the model is explicitly directed to ensure internal consistency. This design enables adaptability across various rewriting styles while ensuring that the output remains natural and aligned with the intent.

Conversational rewrite prompt. Table 13 presents the prompt template for conversational rewrites, which is specifically designed to improve clarity, tone, and personalization in email responses. This prompt structure ensures that the model follows a controlled yet flexible rewriting process, preserving the intent of the original message while refining it according to a given instruction. There are three key features of the prompt design: (i) Explicit input segmentation: The prompt clearly separates the natural prompt, the raw generated email, and the rewrite instruction, ensuring that the model understands the original context before making modifications. (ii) Focus on instruction adherence: Unlike factuality and stylistic rewrites, which

Task: Rewrite your initial response to a user query according to the factuality corrections. Your rewritten response should keep internal consistency while making minimal edits. You will be given: 1. Query: The user query. 2. Initial Response: The original text generated by an LLM to answer the user query. 3. Non-Factual Spans and Replacements: A list of text segments from the initial response that were potentially incorrect, along with their intended factual replacements. Instructions: 1. Carefully review the query and the initial response. 2. Examine each non-factual span and its corresponding factual replacement. 3. Rewrite the initial response to incorporate the factual replacements while ensuring the response remains coherent and consistent. 4. Make minimal edits to the original text, altering what is necessary to correct factual inaccuracies as well as any incoherent or inconsistent content. 5. Output the rewritten response. Test example: 1. Query: `[QUERY]` 2. Initial Response: `[INITIAL RESPONSE]` 3. Non-Factual Spans and Replacements: Span 1: - Span: `[SPAN #1]` - Revision: `[REVISION #1]` ... Span N: - Span: `[SPAN #N]` - Revision: `[REVISION #N]` 4. Your Output (Rewritten Response):

Table 11: Prompt template used to generate factuality rewrites. The `[QUERY]` and `[INITIAL RESPONSE]` placeholders will be replaced by a given user query and its initial response, and the `[SPAN #1` $\rightarrow$ `N]` and `[REVISION #1` $\rightarrow$ `N]` placeholders will be replaced by each of the N spans and revisions from the SAFE (Wei et al., 2024) outputs.

Task: Rewrite a source text according to the comment. You will be given: 1. Source: The original text from a public corpus. 2. Comment: An open-ended rewrite requirement, such as formalize, paraphrase, shorten, elaborate, etc. Instructions: 1. Carefully review the source text and the comment. 2. Understand the intent of the comment and how it should influence the rewrite. 3. Rewrite the source text to align with the comment's requirement, ensuring that the text remains internally consistent and coherent. 4. Ensure the rewritten text maintains the original meaning and context as much as possible. 5. Output the rewritten text. Test example: 1. Source: `[SOURCE]` 2. Comment: `[REWRITE INSTRUCTION]` 3. Your Output (Rewritten Response):

Table 12: Prompt template used to generate stylistic rewrites. The `[SOURCE]` and `[REWRITE INSTRUCTION]` placeholders will be replaced by a given source text and rewrite instruction.

Task: Rewrite an email based on the given instruction. The rewritten email should maintain internal consistency and align with the provided instruction. You will be given: 1. Natural Prompt: The original prompt used to generate the initial email. 2. Raw Generated Email: The initial email generated by the LLM based on the natural prompt. 3. Rewrite Instruction: The instruction for how to improve or modify the raw generated email. Instructions: 1. Carefully read the 'Natural Prompt' and 'Raw Generated Email'. 2. Analyze the 'Rewrite Instruction' to understand the required modifications. 3. Rewrite the 'Raw Generated Email' according to the 'Rewrite Instruction'. Test example: 1. Natural Prompt: `[NATURAL PROMPT]` 2. Raw Generated Email: `[EMAIL]` 3. Rewrite Instruction: `[REWRITE INSTRUCTION]` 4. Your Output (Rewritten Email):

Table 13: Prompt template used to generate conversation rewrites. The `[NATURAL PROMPT]`, `[EMAIL]`, and `[REWRITE INSTRUCTION]` placeholders will be replaced by a given natural prompt, its raw generated email, and rewrite instruction.

Task: Generate a rewrite instruction for a raw generated email based on the natural prompt. The instruction should guide the rewriting process to improve clarity, tone, structure, or other aspects. You will be given: 1. Natural Prompt: The original prompt given to the LLM. 2. Raw Generated Email: The initial email generated by the LLM based on the natural prompt. 3. Demonstrations: Examples of existing emails and their rewrite instructions. Instructions: 1. Read the 'Natural Prompt' and 'Raw Generated Email' carefully. 2. Analyze the 'Raw Generated Email' and identify areas for improvement based on the 'Natural Prompt'. 3. Refer to the provided 'Demonstrations' to understand different types of rewrite instructions and their contexts. 4. Generate a concise and clear rewrite instruction for the 'Raw Generated Email'. Demonstrations: Example 1: - Email: `[EMAIL #1]` - Instruction: `[REWRITE INSTRUCTION #1]` ... Example N: - Email: `[EMAIL #N]` - Instruction: `[REWRITE INSTRUCTION #N]` Test example: 1. Natural Prompt: `[NATURAL PROMPT]` 2. Raw Generated Email: `[EMAIL]` 3. Your Output (Rewrite Instruction):

Table 14: Prompt template used to generate raw conversation rewrite instructions. The `[NATURAL PROMPT]` and `[EMAIL]` placeholders will be replaced by a natural prompt and its generated raw email, and the `[EMAIL #1` $\rightarrow$ `N]` and `[REWRITE INSTRUCTION #1` $\rightarrow$ `N]` placeholders will be replaced by each of the N (email, instruction) pairs.

prioritize correctness and rewording, this task emphasizes cohesively integrating the instruction into the existing email while maintaining internal consistency. (iii) Preserving conversational flow: The instructions explicitly require the model to analyze the provided email and apply the requested changes while ensuring the rewritten email remains natural and engaging.

By structuring the prompt in a way that guides but does not over-restrict the model, this template ensures that email rewrites maintain fluency, correctness, and engagement while following specific improvement instructions. The step-by-step breakdown aids the model in handling more challenging contextual refinements, such as enhancing tone, incorporating enthusiasm, or making the message more concise.

Conversational rewrite instruction generation. Table 14 outlines the prompt template for generating rewrite instructions for conversational emails. This step is crucial in structuring the dataset, as it determines the type and specificity of modifications the model will learn to perform. Unlike factuality or stylistic rewrites, conversational rewrites often involve subtle refinements in tone, structure, and content balance.

Table 15 illustrates the diversity of generated instructions, ranging from simple grammatical refinements (e.g., "use complete sentences") to substantial structural modifications (e.g., "add structure and boilerplate to make the email more professional"). The ability to generate rich, context-aware instructions ensures that conversational rewrites cover various real-world use cases, enhancing the model's generalization ability.

Conversational rewrite instruction refinement. Table 16 presents the prompt template for refining generic rewrite instructions by making them more specific and actionable. This step is crucial in ensuring that rewrite instructions provide clear, detailed guidance, reducing ambiguity for the rewriting model. As seen in Table 17, refining instructions significantly improves task clarity and execution. Instead of broad instructions (e.g., "Make it more persuasive"), the specified versions provide concrete actionable changes (e.g., "Highlight increased brand visibility, direct customer engagement, and networking opportunities"). This ensures that the rewriting model receives precise, contextually relevant directives, ultimately leading to higher-quality, more controlled rewrites.

Table 18 presents the prompt template for making rewrite instructions more natural and linguistically diverse. This step is essential in ensuring that rewrite instructions feel human-like, conversational, and engaging, while still preserving clarity and specificity. Table 19 showcases how structured rewrite instructions are transformed into more intuitive, engaging requests. Instead of formal, mechanical directives (e.g., "Make the email more specific by mentioning the product name and key features."), the modified instructions use more natural phrasing (e.g., "This email is too generic—specify the product name and key features, and highlight why these updates matter for media professionals.").

D AutoRater Prompting

The LLM-as-a-judge framework systematically evaluates rewrites across agreement, coherence, and pairwise comparisons (AutoSxS) to ensure fine-grained, objective assessments. Below, we describes its design rationales.

Agreement evaluation across tasks. Agreement evaluation measures how well the rewritten response adheres to the given instruction across different tasks:

- **Factuality rewrite**: Table 20 illustrates a span-based approach, checking whether all identified factual inaccuracies in the initial response have been corrected. Each non-factual span is individually assessed, ensuring precise, granular feedback.
- **Stylistic rewrite**: Table 21 deconstructs the instruction into multiple transformation requirements (e.g., formalization + conciseness) and evaluate whether each aspect is incorporated while preserving meaning.
- **Conversational rewrite**: Table 22 introduces a context-aware evaluation, ensuring that modifications align with the natural prompt while accurately implementing tone, structure, or engagement-related changes.

Coherence evaluation. Coherence evaluation in AutoRater is designed to assess whether a rewritten response maintains internal logical consistency while ensuring fluency and readability. Unlike agreement evaluation, which focuses on compliance with rewrite instructions, coherence evaluation directly measures whether a revised response contains contradictions, logical inconsistencies, or abrupt structural breaks. The AutoRater employs a structured prompt (Table 23) that systematically

Raw email	Raw instruction
Dear James, I'd like to invite your to join ponding on [DATE]. Weather cold lately, recommend the wetsuit, gloves, woolly hat. Hope you can make it. Charles Campbell, the lifeguard, will also be joining us. John	Rewrite using complete sentences.
Sorry Alexander Thompson, my son, Frank Williams, can't attend, too young. Cancel registration & refund. Sorry for any inconvenience.	Write in a formal tone.
Benjamin, check attached delivery notice for item. Delivery person is Derek Johnson, arriving on December 14th between 10 am - 2 pm. Need signature from someone over 18 to accept delivery. Any questions, contact John Wilson.	Make it longer and more polite.
Dear Anna Barret, I am writing to request a change of class because I am currently in Kevin Smith's class and I think the teaching quality is very poor. Sincerely, Charles White	Mention that I don't learn well with Kevin's teaching style.
Dear Connor, I am writing today to request that you transfer my phone number from AT&T to T-Mobile. I have been an AT&T customer for over 10 years and have been very happy with the service. However, I recently switched to T-Mobile because they have a better data plan for my needs. I would like to keep my AT&T phone number so that my friends and family can still reach me. Please let me know if you need any additional information from me. Thank you for your help. Sincerely, Katherine Walker	Makes it shorter.
Hi Avery, I'm writing to request a refund for my valet reservation. Unfortunately, I've come down with a bad case of the flu and won't be able to travel. I'm so sorry for the inconvenience. I would really appreciate it if you could process the refund as soon as possible. Thank you for your understanding. Sincerely, Jerry Jones	This is too formal, makes it more casual.
Dear Liam, I hope this email finds you well. I'm writing to see if you're available to cat-sit for James again. My family and I will be away on vacation from December 20th to January 2nd, and we would love it if you could take care of him while we're gone. As you know, James is a very sweet and affectionate cat. He loves to be petted and cuddled, and he's always up for a game of fetch. He's also very good at entertaining himself, so you won't have to worry about him getting bored. We would be happy to pay you the same rate as last time. Please let me know if you're available and if you have any questions. Thank you, Fred	Don't mention that we will pay Liam.
Hi Hailey, I hope you're doing well. My name is Jerry Walker, and I'm currently in the process of purchasing a home in San Francisco. I was referred to you by George Wilson, who's a close friend of mine. I recently had the home inspected, and the inspector identified a few electrical issues that I'd like to get fixed before moving in. Would you be able to provide me with a quote for the repairs? I'm planning to use the quote in my negotiations with the home seller. Thanks in advance for your help. Best, Jerry Walker	Make the second paragraph shorter.
Hi Ava, Just following up on the insurance enrollment. I've been trying to reach Justin but no luck. Can u send me a written confirmation of my enrollment? Thx. Larry	Soften the tone to make this request more appealing.
Hi Ian, I will donate some SAS books, lmk if u want them. Best, Christopher	Add that the SAS books are "SAS Survival Handbook" and "The Little SAS Book".
Hi Andrew, My gmail chat stopped working!!! I tried to log out/in, nothing works. Help!!! Thx, Ariana	Add structure and boilerplate sentences to this email to make it more professional.

Table 15: Examples (demonstrations) of raw conversational rewrite instructions used in Table 14.

Task: Modify the existing rewrite instruction to make it more specific. The modified instruction should provide clear and detailed guidance by specifying necessary entities, nouns, or details.

You will be given:
1. Natural Prompt: The original prompt given to the LLM.
2. Raw Generated Email: The initial email generated by the LLM based on the natural prompt.
3. Existing Instruction: The general rewrite instruction that needs to be refined.
4. Demonstrations: Examples showing how to modify general instructions to be more specific.

Instructions:
1. Read the 'Natural Prompt', 'Raw Generated Email', and 'Existing Instruction' carefully.
2. Refer to the provided 'Demonstrations' to understand how to modify an instruction to be more specific.
3. Modify the 'Existing Instruction' to make it more specific by including necessary entities, nouns, or details.

Demonstrations:
Example 1:
- Email: `[EMAIL #1]`
- Original Instruction: `[ORIGINAL INSTRUCTION #1]`
- Modified Instruction: `[MODIFIED INSTRUCTION #1]`
...
Example N:
- Email: `[EMAIL #N]`
- Original Instruction: `[ORIGINAL INSTRUCTION #N]`
- Modified Instruction: `[MODIFIED INSTRUCTION #N]`

Test example:
1. Natural Prompt: `[NATURAL PROMPT]`
2. Raw Generated Email: `[EMAIL]`
3. Existing Instruction: `[ORIGINAL INSTRUCTION]`
4. Your Output (Modified Instruction):

Table 16: Prompt template used to specify conversational rewrite instructions with more details. The `[NATURAL PROMPT]`, `[EMAIL]`, and `[ORIGINAL INSTRUCTION]` placeholders will be replaced by a given natural prompt, its raw generated email, and original instruction. The `[EMAIL #1` $\rightarrow$ `N]`, `[ORIGINAL INSTRUCTION #1` $\rightarrow$ `N]`, and `[MODIFIED INSTRUCTION #1` $\rightarrow$ `N]` placeholders will be replaced by each of the N instruction refinement pairs.

Raw email	Original instruction	Modified instruction
Dear [Recipient's name], I hope you have a great holiday season and enjoy your time with your family.	Make it more specific and personal.	Make it more specific and personal by mentioning the holiday plans you know about, such as their planned visit to Hawaii.
Dear [Boss's name], Happy Mother's Day to you! I hope you have a wonderful day filled with love, laughter, and relaxation. I know that you are a great mother to your children, and I appreciate all that you do for them. You are always there for them, no matter what. You are a role model for them, and they are so lucky to have you. I also want to thank you for being a great boss. You are always supportive and understanding, and I appreciate your guidance and advice. I am lucky to have you as a mentor. I hope you have a very special Mother's Day! Sincerely, [Your name]"	Make it more personal and specific to my boss by mentioning her children's names or something special about her.	Make it more personal and specific to my boss by mentioning Eva, her daughter, and how she likes piggy-back.
Dear [Name], I hope this email finds you well. My name is [Your Name] and I am the event coordinator for [Event Name]. I am writing to you today to request a sponsorship from [Company Name]. [Event Name] is an annual event that raises money for [Charity Name]. Last year, we were able to raise over [Amount] for the charity. This year, we are hoping to raise even more money. We believe that [Company Name] would be a great fit for our event. Your company's products and services would be a great addition to our event and we believe that your customers would be interested in learning more about your company. In return for your sponsorship, we would be happy to provide you with a number of benefits, including: * Your logo would be prominently displayed on all promotional materials for the event. * You would have a booth at the event to promote your company. * You would be given the opportunity to speak at the event. * You would be invited to a VIP reception before the event. We believe that a sponsorship from [Company Name] would be a mutually beneficial relationship. We hope that you will consider our request. Thank you for your time and consideration. Sincerely, [Your Name]	Make the email more persuasive by highlighting the benefits of sponsorship for the company.	Make the email more persuasive by highlighting increased brand visibility, direct customer engagement at the event booth, and networking opportunities at the VIP reception.
Dear [Patient Advocate Name], I am writing to you today to express my concern about the lack of informed consent in my recent medical treatment. On [date], I was admitted to [hospital name] for [procedure]. Prior to the procedure, I was asked to sign a consent form. However, I was not given any explanation of the procedure, the risks involved, or the alternatives to treatment. I was simply told that I needed to sign the form in order to receive treatment. I feel that I was not given the opportunity to make an informed decision about my treatment. I had no idea what I was consenting to, and I feel that I was taken advantage of. I am requesting that you investigate this matter and take steps to ensure that all patients are given the opportunity to make informed decisions about their treatment. Thank you for your time and consideration. Sincerely, [Your Name]	Add details about the procedure and the risks involved.	Add specific details about a knee replacement surgery and the potential risks of infection, blood clots, and damage to surrounding tissues.
Dear Marketing Team, I hope this email finds you well. I'm writing to confirm that the account update has been completed. All of the necessary changes have been made to the account, and everything should be working properly. If you have any questions, please don't hesitate to contact me. Thank you for your cooperation. Sincerely, [Your Name]	Specify what account was updated and what changes were made.	Specify the account details such as client database, and describe the changes, like updated contact information, fixed email delivery issues, and added new segmentation tags.

Table 17: Examples (demonstrations) of specified conversational rewrite instructions used in Table 16.

Task: Modify the given rewrite instruction to make it more linguistically diverse and natural. The modified instruction should retain specificity and clarity while sounding more like human-asked natural language.

You will be given:
1. Natural Prompt: The original prompt given to the LLM.
2. Raw Generated Email: The initial email generated by the LLM based on the natural prompt.
3. Previous Instruction: The specific instruction generated previously that needs to be made more linguistically diverse and natural.
4. Demonstrations: Examples showing how to modify instructions to make them more natural and conversational.

Instructions:
1. Read the 'Natural Prompt', 'Raw Generated Email', and 'Existing Instruction' carefully.
2. Refer to the provided 'Demonstrations' to understand how to modify an instruction to be more specific.
3. Modify the 'Existing Instruction' to make it more specific by including necessary entities, nouns, or details.

Demonstrations:
Example 1:
- Email: `[EMAIL #1]`
- Previous Instruction: `[ORIGINAL INSTRUCTION #1]`
- Enhanced Instruction: `[MODIFIED INSTRUCTION #1]`
...
Example N:
- Email: `[EMAIL #N]`
- Previous Instruction: `[ORIGINAL INSTRUCTION #N]`
- Enhanced Instruction: `[MODIFIED INSTRUCTION #N]`

Test example:
1. Natural Prompt: `[NATURAL PROMPT]`
2. Raw Generated Email: `[EMAIL]`
3. Previous Instruction: `[ORIGINAL INSTRUCTION]`
4. Your Output (Enhanced Instruction):

Table 18: Prompt template used to make conversational rewrite instructions more natural. The `[NATURAL PROMPT]`, `[EMAIL]`, and `[ORIGINAL INSTRUCTION]` placeholders will be replaced by a given natural prompt, its raw generated email, and previous instruction. The `[EMAIL #1` $\rightarrow$ `N]`, `[ORIGINAL INSTRUCTION #1` $\rightarrow$ `N]`, and `[MODIFIED INSTRUCTION #1` $\rightarrow$ `N]` placeholders will be replaced by each of the N instruction refinement pairs.

Raw email	Original instruction	Modified instruction
Dear [Media Name], I hope this email finds you well. I'm writing to you today to invite you to a product update webinar on [date] at [time]. During this webinar, we'll be sharing the latest updates on our product, including new features, functionality, and improvements. We'll also be answering any questions you may have. We'd love for you to be able to attend. Please RSVP to [email address] by [date]. Thanks, [Your Name]	Make the email more specific by mentioning the product name and highlighting the key features of the update, such as the new AI-powered analytics tool and the redesigned user interface. Additionally, emphasize the benefits of these updates for media professionals, such as improved workflow efficiency and more insightful data analysis.	This email is too generic, specify the product name and highlight the key features of the update, such as the new AI-powered analytics tool and the redesigned user interface. Also emphasize the benefits of these updates for media professionals, such as improved workflow efficiency and more insightful data analysis.
Dear [Teacher's name], I hope this email finds you well. I'm writing to let you know that I recently got my dream job! I'm so excited to be starting this new chapter in my career. I wanted to thank you for all of your support and guidance over the years. You've helped me to become the confident and capable professional that I am today. I'll be sure to keep you updated on my progress. In the meantime, I wish you all the best. Sincerely, [Your name]	Make it more specific by mentioning the dream job as a [position name] at [company name], and how the teacher helped you achieve it by providing specific examples like [encouragement, career advice, mock interview practice, etc.].	I just got a job as a graphic designer at disney, include a story about the time when Mr. Miller went through a bunch of van gogh paintings and how that really inspired me.
Dear Mom, I hope this email finds you well. I just wanted to take a moment to thank you for coming to my birthday party last night. I had so much fun celebrating with you and all of my friends and family. It meant the world to me to have you there. I know you were busy, but I'm so glad you made the time. I hope you had a good time as well. I know I did! Love, [Your name]	Make it more personal and heartfelt by mentioning specific moments you enjoyed with your mom during the party, like dancing together or her funny stories. You can also express your love and appreciation for her presence and support.	Make the email more heartfelt, including reference to the times we were dancing together and her funny stories.

Table 19: Examples (demonstrations) of naturalized conversational rewrite instructions used in Table 18.

Task: Analyze a revised text passage for factual accuracy by calculating the percentage of replacements correctly incorporated into the revised version.

You will be given:
1. Query: The user query.
2. Initial Response: The original text generated by an LLM to answer the user query.
3. Non-Factual Spans and Replacements: A list of text segments from the initial response that were potentially incorrect, along with their intended factual replacements.
4. Rewritten Response: A revised text to answer the user query where the goal was to replace the non-factual spans with their factual counterparts while making minimal edits.

Instructions:
1. Carefully compare the ‘Non-Factual Spans and Replacements’ and the ‘Rewritten Response’. Note whether sentence structure changes were necessary to integrate the replacements while striving for accuracy.
2. Check Accuracy: Meticulously examine the ‘Rewritten Response’. Your focus is to determine if ALL intended factual replacements are correctly incorporated.
3. Calculate Percentage: Calculate the following:
Total Number of Replacements: The number of items in the ‘Non-Factual Spans and Replacements’ list.
Correct Replacements: The number of replacements accurately integrated into the ‘Rewritten Response’.
Percentage: (Correct Replacements / Total Number of Replacements) * 100
4. Output Percentage.

Test example:
1. Query: `[QUERY]`
2. Initial Response: `[INITIAL RESPONSE]`
3. Non-Factual Spans and Replacements:
Span 1:
- Span: `[SPAN #1]`
- Revision: `[REVISION #1]`
...
Span N:
- Span: `[SPAN #N]`
- Revision: `[REVISION #N]`
4. Rewritten Response: `[REWRITTEN RESPONSE]`
5. Your Output:

Table 20: Prompt template used to generate factuality rewrite agreement score. The `[QUERY]`, `[INITIAL RESPONSE]`, and `[REWRITTEN RESPONSE]` placeholders will be replaced by a given user query, its initial response, and the rewritten response, and the `[SPAN #1` $\rightarrow$ `N]` and `[REVISION #1` $\rightarrow$ `N]` placeholders will be replaced by each of the N spans and revisions from the SAFE (Wei et al., 2024) outputs.

Task: Analyze a revised text passage for agreement with the given rewrite instruction.

You will be given:
1. Source: The original text from a public corpus.
2. Instruction: Open-ended rewrite requirements, such as formalize, paraphrase, shorten, elaborate, etc.
3. Rewritten text: The revised text that has been altered according to the instruction’s requirements.

Instructions:
1. Carefully compare the ’source’ and the ’rewritten text’ based on the ’instruction’. Note whether the structure, tone, and content changes were necessary to integrate the instruction’s requirements while maintaining coherence and consistency.
2. Check Agreement: Meticulously examine the ’rewritten text’. Your focus is to determine if ALL requirements mentioned in the ’instruction’ are correctly and effectively incorporated.
3. Calculate Percentage: Calculate the following:
Total Number of Requirements: The number of distinct requirements mentioned in the ’instruction’.
Correct Requirements: The number of requirements accurately integrated into the ’rewritten text’.
Percentage: (Correct Requirements / Total Number of Requirements) * 100
4. Output Percentage.

Test example:
1. Source: `[SOURCE]`
2. Instruction: `[INSTRUCTION]`
3. Rewritten text: `[REWRITTEN TEXT]`
4. Your Output:

Table 21: Prompt template used to generate stylistic rewrite agreement score. The `[SOURCE]`, `[INSTRUCTION]`, and `[REWRITTEN TEXT]` placeholders will be replaced by a given source text, rewrite instruction, and the rewritten text.

Task: Analyze a revised email for agreement with the rewrite instruction.

You will be given this XML input:
<natural_prompt>The original prompt given to the LLM to generate the initial email.</natural_prompt>
<raw_generated_email>The initial email generated by the LLM based on the natural prompt.</raw_generated_email>
<rewrite_instruction>The instruction provided for rewriting the initial email.</rewrite_instruction>
<rewritten_email>The revised email that has been altered according to the rewrite instruction.</rewritten_email>

Instructions:
1. Carefully compare the 'Raw Generated Email' and the 'Rewritten Email' based on the 'Rewrite Instruction'. Note whether the structure, tone, and content changes were necessary to integrate the instruction's requirements while maintaining coherence and consistency.
2. Check Agreement: Meticulously examine the 'Rewritten Email'. Your focus is to determine if ALL requirements mentioned in the 'Rewrite Instruction' are correctly and effectively incorporated.
3. Calculate Percentage: Calculate the following:
Total Number of Requirements: The number of distinct requirements mentioned in the 'Rewrite Instruction'.
Correct Requirements: The number of requirements accurately integrated into the 'Rewritten Email'.
Percentage: (Correct Requirements / Total Number of Requirements) * 100
4. Output Percentage.

You will output in XML form:
<output_explanation>Explain your reasoning for each requirement, indicating whether it's satisfied, and why.</output_explanation>
<output_percentage>Provide the final accuracy percentage. Do not include anything besides a percentage.</output_percentage>

Begin!

<natural_prompt>`[NATURAL PROMPT]`</natural_prompt>
<raw_generated_email>`[EMAIL]`</raw_generated_email>
<rewrite_instruction>`[REWRITE INSTRUCTION]`</rewrite_instruction>
<rewritten_email>`REWRITTEN EMAIL`</rewritten_email>

Your XML output with explanation and percentage:
<output_explanation>

Table 22: Prompt template used to generate conversational rewrite agreement score. The `[NATURAL PROMPT]`, `[EMAIL]`, `[REWRITE INSTRUCTION]`, and `REWRITTEN EMAIL` placeholders will be replaced by a given natural prompt, its raw generated email, rewrite instruction, and rewritten email.

evaluates coherence by requiring explicit yes/no judgments alongside explanatory reasoning. This approach enhances interpretability and ensures reliable scoring. (i) Binary Coherence Decision: Each rewritten response is judged as either internally consistent ("YES") or inconsistent ("NO"), ensuring clarity in evaluation. (ii) Justification for Each Judgment: The prompt mandates a reason for the decision, encouraging the model to articulate why a response is or is not coherent. (iii) Few-Shot Learning with Examples: The prompt provides multiple illustrative cases, demonstrating correct coherence judgments across different scenarios, including numerical inconsistencies, logical contradictions, and unsupported claims.

AutoSxS evaluation. AutoSxS is designed to directly compare rewritten responses by evaluating how well they satisfy a given rewrite instruction. Unlike individual scoring metrics (e.g., agreement, coherence), AutoSxS provides pairwise judgments, making it a more fine-grained and human-aligned evaluation method. As shown in Table 24, the AutoSxS framework follows a structured decision-making process: (i) The model is presented with two rewritten responses (A, B) for the same raw email and rewrite instruction. (ii) It must analyze both responses and select the one that better satisfies the rewrite instruction (or declare them as equal). (iii) It provides explicit reasoning for its choice and assigns numerical scores (0–1) to each response, ensuring that the relative ranking is interpretable. To enhance reliability, the prompt includes: (i) Clear task instructions emphasizing rewrite alignment. (ii) Demonstrations with diverse examples covering different types of modifications (removal, addition, tone change, etc.). (iii) Structured output, ensuring consistency in comparisons.

Table 25 presents two examples illustrating AutoSxS decisions. In Example 1, Response A is rated 0.91, significantly higher than Response B (0.58), as it strictly adheres to the instruction of removing unnecessary details. In Example 2, Response B is preferred (0.83 vs. 0.76) for expanding on the religious holiday's significance in a more formal and comprehensive way. These results demonstrate AutoSxS's ability to: (i) Identify minor instruction violations (e.g., residual details in Response B of Example 1). (ii) Capture nuanced preferences in tone and detail level (e.g., preference for a more formal explanation in Example 2). (iii) Provide interpretable ranking beyond binary correctness. By leveraging structured evaluation, explicit reasoning, and numerical scoring, AutoSxS enhances automated rewrite assessment, ensuring more human-aligned and context-aware comparisons across rewriting tasks.

Your task is to evaluate whether or not a piece of text is internally consistent.

Please provide your answer in the following format:
<text>[text to evaluate for internal consistency]</text>
<answer>[YES or NO]</answer>
<reason>[reason for giving the answer above]</reason>

Example #1:
<text>Paul has 5 daughters named Ava, Brittney, and Claire.</text>
<answer>NO</answer>
<reason>The response states that Paul has 5 daughters but only names 3 of them.</reason>

Example #2:
<text>IPhones are better than Samsung. The reason is because Samsung has a shorter battery life.</text>
<answer>YES</answer>
<reason>Everything is internally consistent.</reason>

Example #3:
<text>While Bob Ross and Annettte Kowalski's exact age difference is not publicly known, they likely had several years age difference.

Here's what we found online:
* Bob Ross was born on October 29, 1942 and died at age 52 on July 4, 1995.
* Annette Kowalski was born on January 26, 1936.
</text>
<answer>NO</answer>
<reason>The beginning says that the age difference is not publicly known, but later the text says that both birth dates were found online.</reason>

Question:
<text>[REWRITTEN RESPONSE]</text>
<answer>

Table 23: Prompt template used to generate rewrite coherence score. The [REWRITTEN RESPONSE] placeholder will be replaced by a rewritten response.

Task: Judge which response better satisfies the rewrite instruction, choose among two choices: (A), (B) or same, and rate each response a score between 0 and 1.
Instructions:
1. Carefully analyze the ‘Response A’ and ‘Response B’ based on the ‘Instruction’.
2. Provide an explanation comparing the two responses based on the instruction.
3. Make your choice between ‘A’, ‘B’, or ‘same’.
4. Provide your score for each response (A, B) based on how well they satisfy the rewrite instruction.

Demonstrations:
Example #1
<original_email>[EMAIL #1]</original_email>
<instruction>[REWRITE INSTRUCTION #1]</instruction>
<response_A>[RESPONSE A #1]</response_A>
<response_B>[RESPONSE B #1]</response_B>
<choice>[CHOICE #1]</choice>
<explanation>[REASON #1]</explanation>
<score>[SCORE TUPLE #1]</score>
...
Example #N
<original_email>[EMAIL #N]</original_email>
<instruction>[REWRITE INSTRUCTION #N]</instruction>
<response_A>[RESPONSE A #N]</response_A>
<response_B>[RESPONSE B #N]</response_B>
<choice>[CHOICE #N]</choice>
<explanation>[REASON #N]</explanation>
<score>[SCORE TUPLE #N]</score>

You will output in XML format:
<choice>Please choose among three choices: (A), (B) or same. Do not include anything besides a choice.</choice>
<explanation>Explain the reason of your choice.</explanation>
<score>A tuple of two scores between 0 and 1, rating how well each response satisfies the rewrite instruction.</score>

Begin!
<original_email>[EMAIL] </original_email>
<instruction>[REWRITE INSTRUCTION]</instruction>
<response_A>[RESPONSE A]</response_A>
<response_B>[RESPONSE B]</response_B>

Your XML output with choice, explanation, and score:
<choice>

Table 24: Prompt template used to generate AutoSxS scores of two rewritten responses [RESPONSE A], [RESPONSE B] regarding the same raw generated email [EMAIL] and instruction [REWRITE INSTRUCTION].

Raw email: Dear [Name],
I am writing to cancel the interview with the shareholders that was scheduled for [date]. I am very sorry for any inconvenience this may cause. I have a personal emergency that I need to attend to, and I will not be able to make it to the interview.
I have already contacted [Name], and they have agreed to reschedule the interview for [date].
I apologize again for any inconvenience this may cause. Thank you for your understanding.
Sincerely, [Your Name]

Instruction: Remove the sentences mentioning the personal emergency and contacting [Name]. Instead, state that the interview is being canceled due to unforeseen circumstances.

Response (A): Dear [Name],
I am writing to inform you that, due to unforeseen circumstances, I must cancel the interview with the shareholders scheduled for [date]. Please accept my sincerest apologies for any inconvenience this may cause.
I understand this may be disappointing news, and I appreciate your understanding during this time. If necessary, we can discuss rescheduling at a later date. Thank you for your continued support.
Sincerely, [Your Name]

Response (B): Dear [Name],
I am writing to cancel the interview with the shareholders scheduled for [date] due to unforeseen circumstances. I apologize for any inconvenience this may cause. Unfortunately, I won't be able to attend as planned, and I will follow up with [Name] regarding rescheduling.
Thank you for your understanding and patience.
Best regards, [Your Name]

Score (A): 0.91

Score (B): 0.58

Choice: (A)

Reasoning: Response (A) directly follows the instruction by focusing entirely on the unforeseen circumstances and does not mention contacting [Name] or personal emergencies. It also offers to discuss rescheduling in a more formal tone. Response (B) still includes a mention of following up with [Name], which was not in the instruction, and has a more casual tone, which might make it slightly less aligned with the rewrite instruction.

Raw email: Dear [Recipient Name],
I am writing to request leave for [number of days] days on [date] and [date] for [religious holiday].
I understand that this is a busy time for the team, but I would be grateful if you could approve my request. I have already completed all of my tasks for the week, and I have made arrangements for [coworker's name] to cover my duties while I am away. Thank you for your understanding.
Sincerely, [Your Name]

Instruction: Please add a sentence or two about the significance of the religious holiday and how you plan to observe it.

Response (A): Dear [Recipient Name],
I am writing to request leave for [number of days] days on [date] and [date] for [religious holiday]. This is a special time for me as it allows me to connect with my family and our traditions. We plan to celebrate with a gathering and special prayers.
I have ensured that all my tasks are completed and that [coworker's name] will handle any urgent matters in my absence. I hope for your approval of my leave request. Thank you for your understanding.
Best regards, [Your Name]

Response (B): Dear [Recipient Name],
I am writing to request leave for [number of days] days on [date] and [date] for [religious holiday]. This holiday holds significant spiritual importance for me and my family. We plan to observe it by participating in religious ceremonies and spending quality time together.
I realize that my absence may impact the team's workload, and I have made arrangements for [coworker's name] to cover my duties during this period. I would greatly appreciate your approval for this leave. Thank you for your consideration.
Sincerely, [Your Name]

Score (A): 0.76

Score (B): 0.83

Choice: (B)

Reasoning: Both responses add information about the significance of the religious holiday, but they do so in different ways. Response (A) emphasizes the family connection and traditional practices, while Response (B) focuses on the spiritual importance and includes participation in ceremonies. Both are valid, but Response (B) might be seen as slightly more formal and aligned with the instruction to expand on the holiday's significance.

Table 25: Two examples used for prompting AutoSxS evaluations in Table 24.