# ReadFact: A Workflow Framework for Readability and Factual Consistency in Medical Text Simplification

**Anonymous ACL submission**

## Abstract

Biomedical literature often remains inaccessible to lay readers due to technical complexity. Medical text simplification (MTS) aims to improve readability while preserving factual accuracy. We propose *ReadFact*, a workflow that integrates three complementary components: (i) a simplifier trained with Direct Preference Optimization (DPO), (ii) a readability reward model trained with Proximal Policy Optimization (PPO), and (iii) a PICO-based fact checker for structured factual alignment.

Our system is trained on the Cochrane Database of Systematic Reviews. Intermediate simplifications are first generated using DeepSeek-V3, and each (source, mid, target) triple is expanded into preference pairs to supervise both DPO and PPO training. Factual consistency is evaluated using SciBERT-based PICO similarity, while readability is optimized through preference-driven learning.

Experiments show that ReadFact improves factual consistency by more than 23% over the DPO baseline and increases readability by over 5 points. On the NapSS benchmark, ReadFact–DPO achieves the highest BERTScore, demonstrating closer alignment with human references.

## 1 Introduction

### 1.1 Background and Motivation

A large amount of medical information is embedded in complex biomedical texts, research papers, and clinical reports. For lay readers, these texts are difficult to understand due to the use of technical terminology and long, intricate sentence structures. This complexity makes it challenging for patients to interpret their own health conditions, treatment options, or even their physician's advice. Automatic Text Simplification (ATS) offers a promising solution by transforming technical biomedical language into more accessible text, while retaining essential content. Simplified medical texts benefit both patients and practitioners: patients gain clearer access to knowledge, and clinicians can communicate research evidence more efficiently.

With the rapid progress of natural language processing (NLP) and large language models (LLMs), ATS systems have become increasingly sophisticated. Modern systems can adapt simplifications to different reader groups while maintaining fidelity to the underlying content (Sun et al., 2022; Pergola et al., 2021).

### 1.2 Gaps

Despite recent progress, Medical Text Simplification (MTS) methods still struggle to jointly optimize readability and factual accuracy. Many existing studies prioritize readability but neglect the preservation of specialized medical content. For instance, reinforcement learning (RL) methods such as TESLEA (Phatak et al., 2022) reduce textual complexity by optimizing the Flesch-Kincaid Grade Level (FKGL) (Kincaid, 1975), but often at the cost of omitting or distorting key clinical details.

Multi-Agent Systems (MAS) have been applied to some language tasks, yet their use in medical text simplification remains limited. Most existing MAS approaches rely on static prompt engineering for role assignment, lacking adaptive control mechanisms that would allow agents to dynamically coordinate based on text complexity (Lyu and Pergola, 2024b). Furthermore, such frameworks are rarely integrated with supervised fine-tuning or RL, leaving open the challenge of end-to-end optimization across multiple objectives such as readability, factual alignment, and terminology preservation.

A further critical limitation is the lack of robust factual consistency control. Current approaches often rely on embedding-based cosine similarity to evaluate semantic fidelity. While useful, such methods cannot reliably detect fine-grained factual

errors, such as incorrect substitutions, omissions, or distortions of medical entities (Jianping et al., 2024). This highlights the need for more structured approaches that explicitly verify factual consistency.

Finally, traditional simplification metrics such as BLEU capture only surface-level lexical overlap and fail to adequately measure the readability or factual accuracy of medical simplifications.

### 1.3 ReadFact: Proposed Pipeline

The objective of this study is to improve medical text simplification by simultaneously optimizing readability and factual consistency. To achieve this, we introduce ReadFact, an end-to-end workflow that integrates three complementary components.

First, the simplifier is a Qwen3-based model fine-tuned with Direct Preference Optimization (DPO). By leveraging human preference pairs consisting of chosen and rejected simplifications, the model directly learns to optimize for readability in line with human judgments.

Second, a Learned Readability Model (LRM) is trained following the pairwise ranking strategy commonly used in PPO-style reward modeling, on pairwise preference data from the Cochrane simplification corpus. This model outputs scalar scores that distinguish between low-, medium-, and high-quality simplifications, providing a continuous and reliable optimization signal.

Third, we incorporate a PICO-based fact checker, which extracts structured trial descriptors (Participants, Interventions, Comparisons, and Outcomes) from both the source text and the generated simplification. By aligning these fields, the module detects factual inconsistencies and revises the simplification when necessary, similar in spirit to structured QA-based consistency methods (Zha et al., 2023; Li et al., 2022).

### 1.4 Summary of Contributions

In summary, this work presents ReadFact, a workflow that combines DPO-based preference optimization, a Learned Readability Model (LRM) for readability, and a PICO-guided fact checker for structured correction. On the Cochrane simplification corpus, ReadFact achieves significant improvements over strong baselines: factual consistency improves by approximately 23% compared to the DPO baseline (0.8601 vs. 0.6306 average PICO similarity), and readability increases by more than 5 points compared to a non-finetuned 8B workflow (9.06 vs. 3.53).

These results demonstrate that ReadFact advances the state of medical text simplification by effectively balancing readability and factual fidelity. The integration of preference optimization, Learned Readability Model (LRM), and structured fact-checking establishes a robust and extensible framework for biomedical simplification tasks.

## 2 Related Work

### 2.1 Current Works

Transformer-based models such as BART (Lewis, 2019) and T5 (Raffel et al., 2020) have achieved strong results in text simplification, including biomedical domains (Devaraj et al., 2021b). However, they often compromise factual accuracy (Lu et al., 2023a). Reinforcement learning (RL) methods like TESLEA (Phatak et al., 2022) combine RL with maximum-likelihood training to balance readability and fidelity, but still rely on coarse similarity metrics unable to detect fine-grained medical errors (e.g., “hypertension” vs. “hyperglycemia”). More stable preference-based methods such as DPO (Rafailov et al., 2023) offer a promising alternative.

Traditional readability metrics (FKGL, ARI, BLEU) mainly capture surface-level properties and are limited in domain-specific contexts. Recent metrics such as SciGisPy (Lyu and Pergola, 2024a) or learned models (Rashid et al., 2025) better assess scientific readability by modeling semantics directly.

Multi-agent systems (e.g., SoMS (Lyu and Pergola, 2024b)) and instruction-tuned LLM frameworks provide modular roles for simplification and verification but lack end-to-end optimization and structured factual control.

### 2.2 Key Insights

From past studies, a few important things have been learned to help make a better system for simplifying medical text. Having different agents specialize in specific jobs, like the Society of Medical Simplifiers (SoMS) does, works well. They split up the work among experts, which helps keep the text easy to read and still full of accurate facts.

A thing called reinforcement learning (RL) can be used to optimize multiple goals at once. There’s this example called TESLEA where using RL to directly improve how easy text is to read really

made a difference. This could also be a good way to make sure facts stay accurate in simplified medical texts. More recently, Direct Preference Optimization (DPO) (Rafailov et al., 2023) has been proposed as a more stable alternative to PPO-style methods, helping to address instability during training and accelerating readability alignment.

Another key insight is that giving clear instructions is really important for helping agents work together better. This lets big language models do specialized jobs, like explaining medical terms or picking out complex words, more accurately and clearly. Recent work on Chain-of-Thought (CoT) prompting shows that carefully designed workflows and instructions can significantly enhance LLM reasoning (Chernodub et al., 2025), supporting the idea that GPT-style workflows and prompt engineering can boost performance in text simplification.

In addition, rankwise training strategies such as the reward model within PPO (Schulman et al., 2017) demonstrate that reward models can be trained to distinguish better from worse simplifications, making them directly applicable for learned readability scoring in medical text simplification.

Finally, being exact with medical terms and keeping facts straight is crucial. Simple embedding similarity can miss omissions or substitutions of critical medical terms (e.g., Participants or Outcomes). Therefore, structured information like PICO is essential for rigorous consistency evaluation. Using tools that extract terms and check medical facts makes the simplified text more trustworthy. This helps solve big problems in making sure the simplified text is both clear and correct (Wan and Bansal, 2022).

# 3 Methodology

## 3.1 Dataset Preparation

We use the Cochrane Database of Systematic Reviews (GEM benchmark), which provides three official splits: 3568 training examples, 411 validation examples, and 480 test examples. Each example consists of a source abstract (expert-written) and a simplified reference (layperson-oriented).

We do not always use the full 3568 training pairs. For SFT and LRM training, we sample a subset of the official training set. For evaluation, we follow the official validation and test splits without modification.

To construct preference data for Direct Preference Optimization (DPO), we augment the dataset with automatically generated intermediate simplifications. DeepSeek-V3 produces $\frac{4185}{3} \approx 1395$ mid-level outputs, which form (source, mid, target) triples with the original source and target.

Each triple is then expanded into three preference pairs:

$$(y^+, y^-) \in \{(\text{target}, \text{mid}), (\text{target}, \text{source}), (\text{mid}, \text{source})\}. \quad (1)$$

We obtain 4185 pairs, split 80%/20% for training and validation. The same pairs are reused for training the LRM, although in this case only half of the data is employed to avoid overfitting.

## 3.2 ReadFact Overview

ReadFact is a LangGraph-based workflow composed of sequential nodes for simplification, factual verification, and correction. The main nodes are:

- Simplifier: generates layperson-oriented simplifications from Cochrane abstracts using the DPO-trained model.
- PICO Extractor: extracts structured PICO fields (Participants, Interventions, Comparisons, Outcomes) from both source and simplified texts.
- Similarity Computation: computes cosine similarity between source and simplified PICO elements.
- Corrector: revises the simplified text when inconsistencies are detected to restore factual alignment.
- Reward Scoring: evaluates readability of both simplified and corrected outputs via the LRM.
- Iterative Controller: selects the best candidate based on PICO similarity and readability, stopping after three iterations or convergence.

## 3.3 ReadFact Nodes and Associated Models

ReadFact integrates three distinct models: (i) a DPO-trained simplifier, (ii) a general-purpose LLM as a PICO extractor, and (iii) a Learned Readability Model (LRM) for readability. Each node is instantiated with one of these models, guided by carefully designed system prompts. A full mapping between workflow nodes, associated models, and prompts is provided in Appendix A.

#### 3.3.1 Pipeline Execution Flow

Beyond the description of individual nodes, it is important to outline how the full ReadFact pipeline operates from input to output. Given a Cochrane source abstract, the workflow proceeds as follows:

1. The source text is passed to the Simplifier Node, which generates an initial layperson-oriented candidate. 2. Both the source and the candidate are fed into the PICO Extractor Node, which outputs structured JSON with four fields (Participants, Interventions, Comparisons, Outcomes). 3. The Similarity Computation Node then calculates cosine similarity between the extracted PICO fields of the source and candidate. 4. If the similarity falls below a threshold or inconsistencies are detected, the Corrector Node revises the candidate to restore factual alignment. 5. The Reward Scoring Node assigns a readability score to each candidate (simplified and corrected) using the learned reward model. 6. The Iterative Loop Controller selects the best candidate based on highest factual similarity, breaking ties by readability score. The process repeats for up to three iterations or until no further improvements are observed.

All nodes are implemented and orchestrated via LangGraph, which provides modular execution and explicit message passing between nodes. To ensure compatibility, every node communicates using lightweight JSON schemas (e.g., the Simplifier returns `{"simplified text": "..."}`). This design allows intermediate outputs to be stored and inspected, and enables controlled intervention when factual drift is detected.

### 3.4 Simplifier Training

#### 3.4.1 Supervised Fine-Tuning (SFT)

We fine-tune Qwen3-8B under a QLoRA configuration (rank $r = 32$, $\alpha = 32$, 4-bit quantization). Training is performed on two 4090 GPUs (48 GB RAM each), with per-device batch size $= 2$, gradient accumulation $= 4$, giving an effective batch size of 16. We train for 3 epochs with a learning rate of $1 \times 10^{-4}$ using AdamW optimizer (8-bit) with linear scheduling.

The training dataset contains approximately 2864 examples in the training split. Across 3 epochs, the observed 537 optimization steps correspond to roughly $537 \times 16 \approx 8592$ instances, or about 2864 per epoch.

The training objective minimizes the negative log-likelihood:

$$\mathcal{L}_{\text{SFT}}(\theta) = -\sum_{(x,y)} \log p_\theta(y|x), \tag{2}$$

where $x$ is the input paragraph and $y$ is the simplified reference.

#### 3.4.2 Direct Preference Optimization (DPO)

We fine-tune Qwen3-8B using DPO under a QLoRA regime. Training is performed with two 4090 GPUs, 4-bit loading with LoRA adapters (rank $r = 32$, $\alpha = 64$), mixed-precision (bf16), and implicit KL via a frozen reference model. Hyperparameters: 3 epochs, per-device batch $= 2$, gradient accumulation $= 8$, effective batch size $= 16$, learning rate $= 5 \times 10^{-6}$, DPO temperature $\beta = 0.1$, and maximum sequence length $= 4096$ (prompt cap $= 2048$).

The DPO objective directly contrasts preferred and dispreferred outputs. For each pair $(y^+, y^-)$, the loss is:

$$\mathcal{L}_{\text{DPO}}(\theta) = -\mathbb{E}_{(x,y^+,y^-)}\Big[\log \sigma\Big(\beta\Big(\log \frac{\pi_\theta(y^+|x)}{\pi_{\text{ref}}(y^+|x)} - \log \frac{\pi_\theta(y^-|x)}{\pi_{\text{ref}}(y^-|x)}\Big)\Big)\Big] \tag{3}$$

where $\pi_\theta$ is the trainable policy, $\pi_{\text{ref}}$ is the frozen reference, and $\sigma$ is the sigmoid.

### 3.5 Learned Readability Model

#### 3.5.1 Motivation

Traditional readability metrics such as Flesch-Kincaid Grade Level (FKGL) are unreliable in the medical domain. On the Cochrane validation set, the mean FKGL of sources is 11.63, while references average 13.77. Only 25.3% of references have lower FKGL than their source, whereas 74.7% are equal or higher. This reflects that simplifications often add clarifications or restructure content, increasing surface complexity while improving actual readability.

Therefore, we design a learned reward model that better aligns with human readability judgments.

#### 3.5.2 Training Data

The LRM uses the same Cochrane dataset as DPO, but only half the size of preference pairs, since its goal is to learn a stable scoring function rather than directly optimize generation. Each (good, medium, poor) triple is expanded into three preference pairs.

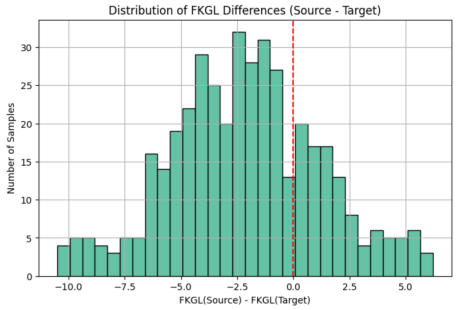


Figure 1: Distribution of FKGL differences between sources and human references. Most simplified texts have higher FKGL, indicating FKGL is unreliable for medical simplification.

#### 3.5.3 Model and Objective

We use DeepSeek-R1-0528-Qwen3-8B with a scalar regression head (`num_labels=1`). The model is trained with LoRA adapters ($r = 64$, $\alpha = 128$, dropout $= 0.05$) in 4-bit mode. Training uses bf16 precision, cosine learning rate scheduling with warmup ratio 0.05, learning rate $2 \times 10^{-5}$, batch size 16 (via gradient accumulation), for 3 epochs.

The objective is a pairwise margin-ranking loss: given scores $r(y^+)$ and $r(y^-)$,

$$\mathcal{L}_{\text{reward}}(\theta) = -\log \sigma\big(r_\theta(y^+) - r_\theta(y^-)\big). \quad (4)$$

### 3.6 PICO Fact Checker Module

To quantitatively assess factual consistency, we integrate a PICO-based fact checker. The module extracts structured PICO elements (Participants, Interventions, Comparisons, Outcomes) from both the source text and the generated simplifications. We then compute cosine similarity between corresponding fields using SciBERT embeddings, which are pretrained on scientific text and provide better semantic coverage of biomedical terminology compared to general-purpose sentence encoders. The averaged similarity serves as an indicator of factual alignment.

## 4 Experiments

In this section, we conduct a comprehensive evaluation of our proposed system ReadFact on the Cochrane simplification corpus. We benchmark against both existing systems (e.g., NapSS (Lu et al., 2023b)) under widely used simplification metrics, and against SFT/DPO baselines using our proposed evaluation setup (PICO similarity and LRM). This dual evaluation allows us to (i) situate ReadFact in the context of prior medical text simplification methods, and (ii) validate the effectiveness of our PICO correction and reward-based optimization modules. Finally, we analyze trade-offs between readability and factual consistency, and provide case studies to highlight strengths and limitations.

### 4.1 System Overview Results

#### 4.1.1 Comparison with Existing Metrics

Following the experimental setup of NapSS (Lu et al., 2023b), we evaluate our systems on the Cochrane test split used in their work. NapSS constructed a paragraph-level simplification benchmark from Cochrane systematic review abstracts, where human-written plain language summaries (PLS) serve as references. Table 1 summarizes performances, and all methods are evaluated on standard automatic metrics widely adopted in text simplification, including Flesch-Kincaid Grade Level (FKGL, lower is better), Automated Readability Index (ARI, lower is better), BLEU and BERTScore. These results allow for direct comparison with previously published systems.

Here, Baseline–SFT and Baseline–DPO denote single-stage fine-tuning with supervised learning or direct preference optimization only, while ReadFact–SFT and ReadFact–DPO denote the full pipeline with PICO-based correction and reward scoring on top of the corresponding simplifier.

Our ReadFact models demonstrate clear advantages on semantic similarity metrics. In particular, ReadFact–DPO achieves the highest BERTScore (29.59), exceeding the best NapSS variant (25.73) by a large margin. This result highlights that our simplifications are semantically closer to human-written references, despite relatively low BLEU scores, which are less reliable in medical simplification due to extensive lexical and syntactic rewritings.

Meanwhile, the DPO baseline achieves the lowest FKGL (9.17) and ARI (11.04) across all systems, indicating that preference-based optimization effectively reduces surface-level complexity and improves readability when measured by traditional readability formulas. This complements the semantic strength of ReadFact–DPO: while DPO alone produces highly readable outputs, integrating PICO correction within ReadFact–DPO further ensures factual consistency and preserves semantic fidelity, pushing the system towards a more balanced trade-off between readability and accuracy.

Table 1: Cochrane paragraph-level simplification results on the NapSS test split. Lower is better for FKGL/ARI; higher is better for BLEU/SARI/BERTScore. NapSS rows are reproduced from (Lu et al., 2023a).

| Model | FK ↓ | ARI ↓ | BLEU ↑ | SARI | BERTScore ↑ |
|---|---|---|---|---|---|
| Vanilla BART | 10.89 | 14.32 | 11.50 | 38.72 | 23.94 |
| UL-BART (Devaraj et al., 2021a) | 9.30 | 12.40 | 7.90 | 40.08 | 24.64 |
| NapSS (BART) | 10.97 | 14.27 | 12.30 | 40.37 | 25.73 |
| NapSS (BioBART) | 10.98 | 14.24 | 11.90 | 40.21 | 25.61 |
| NapSS (+UL) | 8.67 | 11.80 | 9.10 | 41.12 | 23.13 |
| ReadFact – DPO | 11.38 | 12.78 | 3.00 | 37.53 | 29.59 |
| ReadFact – SFT | 15.16 | 16.92 | 13.90 | 40.05 | 28.17 |
| Baseline – DPO | 9.17 | 11.04 | 3.04 | 36.82 | 28.24 |
| Baseline – SFT | 17.87 | 20.00 | 14.17 | 40.51 | 27.85 |

#### 4.1.2 Comparisons on proposed metric

Building on the above comparisons with existing readability metrics (FKGL, ARI, BLEU, SARI, BERTScore), we further introduce our proposed evaluation setup, which combines factual consistency (PICO similarity) and readability (reward model score). We benchmark ReadFact against several baselines, including non-finetuned Qwen3 models and SFT/DPO-trained simplifiers without factual correction. Table 2 reports factual consistency (PICO cosine similarity), readability (reward model score), and BLEU against human references.

Here, Baseline–SFT and Baseline–DPO denote single-stage supervised fine-tuning or preference optimization without correction, while ReadFact–SFT and ReadFact–DPO apply our full pipeline, adding the PICO-based correction and reward scoring modules. Variants with 8B and 32B indicate different backbone model sizes.

Readability (pairwise): Baseline–DPO (8.92) > Baseline–SFT (8.45) and ReadFact–DPO+PICO (9.06) > ReadFact–SFT+PICO (8.01). This shows that DPO, by directly contrasting chosen vs. rejected samples, learns preference distinctions more effectively. When combined with the PPO-trained readability reward model (which is trained on the same chosen/reject pairs), the optimization converges faster and yields larger gains in readability.

Factual consistency: ReadFact–SFT+PICO achieves the highest PICO average similarity (0.9041). Meanwhile, PICO correction substantially boosts DPO's factual consistency (0.6306 → 0.8601, +0.2295), demonstrating that the correction module largely mitigates DPO's weakness in factual alignment. SFT also benefits strongly from correction (0.7120 → 0.9041, +0.1921).

Limitations of BLEU: Although the SFT baseline shows the highest BLEU (0.1417), BLEU does not correlate with factual consistency or readability. After applying PICO correction, BLEU slightly decreases (e.g., SFT: 0.1417 → 0.1386; DPO: 0.0304 → 0.0300), reflecting structural adjustments made to ensure factual alignment rather than surface $n$-gram overlap. More importantly, BLEU is not suitable for medical text simplification tasks, since simplified outputs typically involve large-scale lexical and syntactic rephrasing; thus, higher or lower BLEU values provide little meaningful reference in this domain.

Readability–factuality trade-off: ReadFact demonstrates a controllable trade-off between readability and factual consistency: DPO-based paths maximize readability, while SFT-based paths maximize factual consistency. With PICO correction, both methods are pushed closer to the Pareto frontier, with factual alignment improved substantially while DPO still maintains superior readability (8.92 → 9.06).

These results validate the effectiveness of ReadFact: the structured correction step substantially boosts factual consistency, while DPO-driven simplification maintains superior readability. All reported improvements were tested for statistical significance using a paired t-test ($p < 0.01$).

### 4.2 Training Dynamics and Validation Analysis

Figure in Appendix B.1 reports the supervised fine-tuning (SFT) training and validation loss. Both curves exhibit rapid early convergence, followed by a plateau with a small and stable generalization gap, indicating that the SFT model learns quickly and does not suffer from severe overfitting.

Similarly, Figure in Appendix B.2 shows the training dynamics for Direct Preference Optimization (DPO). Here, the exponentially smoothed reward margin $r(y^+) - r(y^-)$ rises quickly and saturates at a large positive value. At the same time, the

Table 2: Comparison of factual consistency (per PICO element and average), readability, and BLEU across models. Best values are underlined.

| Model / Method | P | I | C | O | Avg | Readability | BLEU |
|---|---|---|---|---|---|---|---|
| Qwen3-8B (no finetune) | 0.6608 | 0.7907 | 0.5153 | 0.7568 | 0.6809 | 7.82 | 0.0429 |
| Qwen3-32B (no finetune) | 0.6368 | 0.7408 | 0.6791 | 0.7470 | 0.7009 | 8.27 | 0.0260 |
| Baseline – SFT | 0.7349 | 0.7976 | 0.5872 | 0.7283 | 0.7120 | 8.45 | 0.1417 |
| Baseline – DPO | 0.5916 | 0.7124 | 0.5412 | 0.6771 | 0.6306 | 8.92 | 0.0304 |
| ReadFact – 8B + PICO | 0.8711 | 0.9208 | 0.7419 | 0.8525 | 0.8466 | 3.53 | 0.0452 |
| ReadFact – 32B + PICO | 0.7269 | 0.7587 | 0.6907 | 0.7657 | 0.7355 | 3.70 | 0.0279 |
| ReadFact – SFT + PICO | 0.9300 | 0.9159 | 0.8302 | 0.9401 | 0.9041 | 8.01 | 0.1386 |
| ReadFact – DPO + PICO | 0.8782 | 0.9207 | 0.7802 | 0.8614 | 0.8601 | 9.06 | 0.0300 |

chosen reward steadily increases while the rejected reward decreases, demonstrating clear preference separation. The DPO loss itself decays smoothly toward zero, confirming stable convergence under reference-policy regularization.

#### 4.2.1 Readability Comparison: DPO vs SFT

The readability comparison between DPO and SFT is shown in Figure in Appendix B.3, which compares the readability LRM reward score across training epochs. DPO directly contrasts chosen versus rejected outputs, forcing the model to enlarge the reward margin between them. This pairwise preference signal provides a sharper gradient than standard likelihood training, allowing the model to learn the distinction between better and worse simplifications more efficiently. As a result, DPO converges rapidly, with its loss dropping close to 0.1 by the end of training, while SFT plateaus earlier and at a lower readability score. This explains why DPO-trained simplifications are not only more readable but also more robust to preference alignment.

#### 4.2.2 Validation Outcome

On the held-out set, the reward scores clearly separate the three categories (Figure in Appendix B.4). Mean scores are source $-3.43$, mid $7.40$, and target $17.81$. Thresholding at $(t_1 = 1.98, t_2 = 12.60)$ yields 84.7% accuracy in distinguishing low-, medium-, and high-quality simplifications, with a brute-force grid search improving to 84.9%.

This confirms that the reward model provides a reliable optimization signal for PPO/DPO training and downstream readability evaluation in our workflow.

#### 4.2.3 Upper Bound of Factual Consistency

Table 3 compares PICO similarity between the source text and two different references: the human-written plain language summary (PLS, ground truth) and the outputs of an automatic text simplifier (ATS). The ATS baseline is obtained by training a BART model with maximum likelihood estimation (MLE), using the source abstracts as input and the PLS references as supervision. This represents a standard supervised learning setup for automatic text simplification in the medical domain.

We observe that the source vs PLS similarity reaches 0.9301, which can be considered the empirical upper bound of achievable factual consistency, since even the expert-authored PLS does not perfectly align with the source at the extractor level. In contrast, the ATS simplifications achieve an average similarity of only 0.6676, highlighting the factual drift problem in automatic simplification.

This establishes that the fact checker not only provides a numeric measure of alignment but also sets an interpretable benchmark: simplification systems cannot realistically exceed the source–PLS similarity of about 0.93 due to inherent extractor noise and paraphrasing differences.

### 4.3 Case Analysis: Why Average PICO Similarity Sometimes Drops

Although our DPO-based workflow generally improves factual consistency, in a few cases the average PICO similarity decreases. This is not due to the simplification or correction modules, but rather due to errors from the PICO extractor. A detailed example is provided in Appendix B.5, where the reference PICO, the extractor output on the DPO simplification, and the extractor output on the corrected simplification are compared side by side.

Original text: We identified 11 RCTs with a total of 821 participants... Seven trials examined green tea and four examined black tea... No studies reported cardiovascular events... Black tea reduced LDL cholesterol (-0.43 mmol/L) and blood pressure slightly... Green tea reduced total cholesterol, LDL cholesterol, and blood pressure... Combining

Table 3: PICO cosine similarity between source text, ATS simplifications (BART MLE baseline), and ground-truth PLS.

| Comparison | Average cosine similarity | Per-element cosine similarity (P / I / C / O) |
|---|---|---|
| Source vs PLS | 0.9301 | 0.9463 / 0.8992 / 0.9124 / 0.9626 |
| Source vs ATS simplification | 0.6676 | 0.8782 / 0.6154 / 0.4582 / 0.7187 |

both teas showed favourable effects on LDL cholesterol and blood pressure... Adverse events were measured in five trials and included prostate cancer, influenza hospitalisation, appendicitis and retinal detachment, but these are unlikely attributable to...

The workflow outputs (both DPO simplification and correction) faithfully preserve the original trial facts (number of participants, interventions, outcomes). However, the PICO extractor hallucinated extra details not present in the text, such as:

- Adults with cardiovascular risk factors (added to the Participants field)
- placebo or no intervention and individual use (added to the Comparisons field)

These hallucinations reduce the computed similarity score, even though the workflow itself did not introduce factual errors. Thus, the observed drop in average PICO similarity is attributable to extractor noise, not to the simplification or correction pipeline.

## 5 Limitations

While ReadFact improves factual alignment and readability, it still relies on automatic PICO extraction, which can introduce noise and limit fine-grained factual verification. Future work could extend this framework to broader biomedical corpora, explore multi-agent collaboration for further robustness, investigate domain-specific reward functions for finer-grained control, and further enhance the controller mechanism for adaptive iteration management. Another promising direction is to refine the fact-checking module to operate at the sentence level, allowing the system to explicitly verify whether each simplified sentence faithfully preserves or includes the key information from the source.

## 6 Conclusion

In this dissertation, we presented *ReadFact*, a workflow-based framework for medical text simplification that jointly optimizes readability and factual consistency. The system integrates three key components: a DPO-trained simplifier, a PPO-trained reward model for readability scoring, and a PICO-based fact checker for structured factual alignment. Through the use of preference pairs derived from the Cochrane simplification corpus, the simplifier is able to distinguish between higher- and lower-quality simplifications, while the reward model provides a stable optimization signal beyond traditional readability metrics such as FKGL. The fact checker explicitly compares extracted PICO elements between source and simplified texts, ensuring that clinical information is faithfully preserved.

Our experiments demonstrate that ReadFact substantially improves over strong baselines. Compared to the DPO baseline, factual consistency improves by approximately 23% (average PICO similarity 0.8601 vs. 0.6306), while readability increases by more than 5 points relative to a non-finetuned simplifier (9.06 vs. 3.53). Moreover, ReadFact–DPO achieves the highest BERTScore (29.59) on the NapSS test split, outperforming all previously reported systems, and thereby confirming that our simplifications are semantically closer to human-written references. These results also highlight a trade-off: DPO-based paths achieve higher readability, while SFT-based paths maximize factual consistency, both being further improved with PICO correction.

Beyond empirical performance, this work highlights the limitations of traditional simplification metrics such as BLEU or FKGL, which fail to capture the true readability and factuality of medical simplifications. By combining preference optimization, learned reward modeling, and structured factual verification, ReadFact provides a principled and effective approach to medical text simplification.

## References


Artem Chernodub, Aman Saini, Yejin Huh, Vivek Kulkarni, and Vipul Raheja. 2025. Apio: Automatic prompt induction and optimization for grammatical error correction and text simplification. *arXiv preprint arXiv:2508.09378*.

Ashwin Devaraj, Iain Marshall, Byron Wallace, and Junyi Jessy Li. 2021a. Paragraph-level simplification of medical texts. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4972–4984, Online. Association for Computational Linguistics.

Ashwin Devaraj, Byron C Wallace, Iain J Marshall, and Junyi Jessy Li. 2021b. Paragraph-level simplification of medical texts. In *Proceedings of the conference. Association for Computational Linguistics. North American Chapter. Meeting*, volume 2021, page 4972. NIH Public Access.

LIU Jianping, CHU Xintao, WANG Jian, GU Xunxun, WANG Meng, and WANG Yingfei. 2024. Semantic matching model for chinese scientific datasets. *Journal of Zhengzhou University: Engineering Science*, 45(6).

JP Kincaid. 1975. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. *Chief of Naval Technical Training*.

Mike Lewis. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*.

Yiyang Li, Lei Li, Qing Yang, Marina Litvak, Natalia Vanetik, Dingxin Hu, Yuze Li, Yanquan Zhou, Dongliang Xu, and Xuanyu Zhang. 2022. Just cloze! a fast and simple method for evaluating the factual consistency in abstractive summarization. *arXiv preprint arXiv:2210.02804*.

Junru Lu, Jiazheng Li, Byron C Wallace, Yulan He, and Gabriele Pergola. 2023a. Napss: Paragraph-level medical text simplification via narrative prompting and sentence-matching summarization. *arXiv preprint arXiv:2302.05574*.

Junru Lu, Jiazheng Li, Byron C. Wallace, Yulan He, and Gabriele Pergola. 2023b. Napss: Paragraph-level medical text simplification via narrative prompting and sentence-matching summarization. *Preprint*, arXiv:2302.05574.

Chen Lyu and Gabriele Pergola. 2024a. Scigispy: a novel metric for biomedical text simplification via gist inference score. *arXiv preprint arXiv:2410.09632*.

Chen Lyu and Gabriele Pergola. 2024b. Society of medical simplifiers. *arXiv preprint arXiv:2410.09631*.

Gabriele Pergola, Elena Kochkina, Lin Gui, Maria Liakata, and Yulan He. 2021. Boosting low-resource biomedical qa via entity-aware masking strategies. *arXiv preprint arXiv:2102.08366*.

Atharva Phatak, David W Savage, Robert Ohle, Jonathan Smith, and Vijay Mago. 2022. Medical text simplification using reinforcement learning (teslea): Deep learning–based text simplification approach. *JMIR Medical Informatics*, 10(11):e38095.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 36:53728–53741.

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.

Ahmad Rashid, Ruotian Wu, Rongqi Fan, Hongliang Li, Agustinus Kristiadi, and Pascal Poupart. 2025. Towards cost-effective reward guided text generation. In *Forty-second International Conference on Machine Learning*.

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

Zhaoyue Sun, Jiazheng Li, Gabriele Pergola, Byron C Wallace, Bino John, Nigel Greene, Joseph Kim, and Yulan He. 2022. Phee: A dataset for pharmacovigilance event extraction from text. *arXiv preprint arXiv:2210.12560*.

David Wan and Mohit Bansal. 2022. Factpegasus: Factuality-aware pre-training and fine-tuning for abstractive summarization. *arXiv preprint arXiv:2205.07830*.

Yuheng Zha, Yichi Yang, Ruichen Li, and Zhiting Hu. 2023. Alignscore: Evaluating factual consistency with a unified alignment function. *arXiv preprint arXiv:2305.16739*.


## A Workflow Node Definitions and Prompts

Table 4 summarizes the ReadFact pipeline nodes, their associated models, and the system prompts used for each task.

# B Additional Training Figures

## B.1 SFT Training Curves

Figure 2: SFT training and evaluation loss across global steps. The loss decreases rapidly at first and then stabilizes, with no indication of severe overfitting.

## B.2 DPO Training Curves

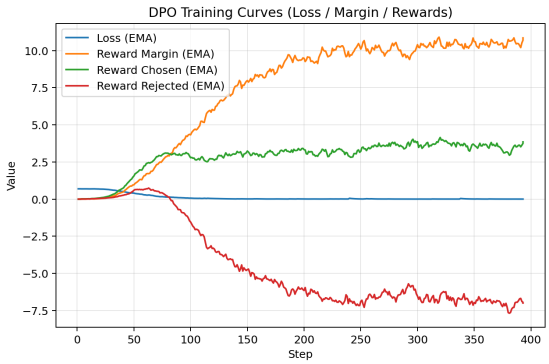


Figure 3: DPO training curves (exponentially smoothed): loss, reward margin, chosen reward, and rejected reward. The loss converges toward zero, while reward margins widen over training.

## B.3 DPO vs SFT Readability Curves

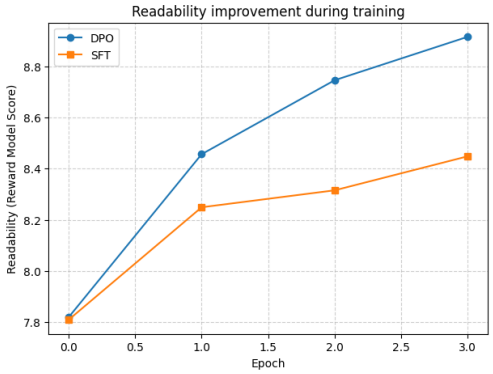


Figure 4: Readability improvement during training. DPO shows faster convergence and higher final readability than SFT.

## B.4 Reward Score Distribution on Validation Set

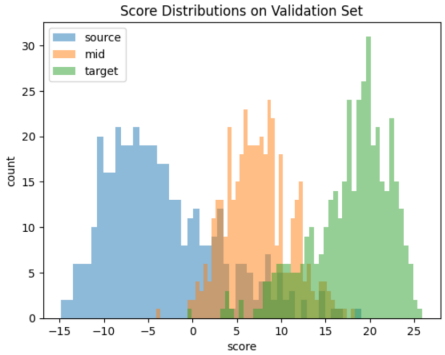


Figure 5: Reward score distributions on validation data. The reward model successfully separates low, medium, and high-quality simplifications.

## B.5 Detailed Case Example of PICO Extraction Errors

Table 5 presents the detailed PICO fields from the reference, the extractor output on the DPO simplification, and the extractor output on the corrected output for the same Cochrane example.

Table 4: Overview of workflow nodes, associated models, and system prompts.

| Node | Model Used | System Prompt (excerpt) |
|---|---|---|
| Simplifier | DPO/SFT fine-tuned Qwen3-8B | "You are a medical language simplification assistant. Rewrite any given Cochrane review paragraph into simpler language for laypeople, while preserving clinical facts, numbers and uncertainties. Return ONLY this exact JSON: {"simplified text": "<text>"}." |
| PICO Extractor | Zero-shot Qwen3 (base) | "You are a PICO information extraction assistant. Please extract the PICO components from the following medical text paragraph and strictly output them in JSON format: {"Participants": "...", "Interventions": "...", "Comparisons": "...", "Outcomes": "..." }. Only output valid JSON with double quotes." |
| Similarity Computation | SciBERT | No prompt (cosine similarity computed over embeddings of extracted PICO fields). |
| Corrector | Zero-shot Qwen3 (base) | "You are a medical language reviewer. Compare the original PICO and simplified PICO. If inconsistencies are found, revise the simplified abstract to restore alignment. Return ONLY this JSON: {"corrected text": "<revised simplified abstract>"}." |
| Reward Scoring | LRM (Qwen3 backbone + LoRA adapter) | No prompt (scalar score produced directly by the trained reward head). |
| Iterative Loop Controller | – | Implements stopping rule: select the best candidate by highest PICO similarity, tie-broken by reward score, until convergence or maximum of 3 iterations. |

Table 5: PICO fields extracted from the reference, DPO output, and corrected output.

| Field | Reference | From DPO output | From corrected output |
|---|---|---|---|
| Participants | 821 participants from 11 RCTs | 821 people across 11 studies | Adults with cardiovascular risk factors (e.g., elevated cholesterol or blood pressure) |
| Interventions | Green tea (including green tea powder capsules) and black tea interventions with varying dosages and forms | green tea (including capsules) and black tea consumption | Green tea (including capsules) and black tea consumption with varying dosages and forms |
| Comparisons | Comparison between green tea and black tea effects on cardiovascular risk factors | placebo or no intervention, and combination of green and black tea | Green tea versus black tea; combination of both teas versus individual use |
| Outcomes | Reductions in LDL cholesterol, total cholesterol, systolic/diastolic blood pressure; no cardiovascular events reported; adverse events unlikely attributable | reduction in LDL cholesterol, slight blood pressure lowering, total cholesterol reduction, and potential side effects; no heart-related events reported | Reduction in LDL cholesterol, total cholesterol, and blood pressure; absence of reported heart-related events; potential side effects not definitively linked |