

Title It Right: Faithful and Attractive Paper Title Generation via Pareto Dominance Alignment

Ke Xiang^{1,2} Yingchao Piao^{1,†} Zexin Wang¹ Yufei Hou^{1,2} Jianhui Li³
Jingjing Li¹ Zihan Liu^{1,2} Hang Cui^{1,2} Changhua Pei¹

¹Computer Network Information Center, Chinese Academy of Sciences

²University of Chinese Academy of Sciences ³Nanjing University

Abstract

Generating academic paper titles requires balancing two key goals: faithfulness to the content and attractiveness to readers. Existing methods often treat title generation as a simple summarization task, producing accurate but unengaging titles. While multi-objective Direct Preference Optimization (DPO) offers a potential solution, our empirical analysis shows that combining conflicting objectives into a single reward is associated with substantial gradient interference during training. To address this problem, drawing inspiration from Pareto dominance in multi-objective optimization, we propose **Multi-Objective Dominance Alignment (MODA)**, a data-centric training strategy tailored for academic title generation. MODA retains preference pairs with strict Pareto dominance, thereby reducing exposure to supervision with opposing reward directions. We implement MODA using dual reward models to independently score both dimensions. Across our experiments, this filtering strategy improves the observed balance between faithfulness and attractiveness, outperforms Human Author baselines, and approaches the performance of frontier models despite using an 8B backbone. Our code is publicly available at https://github.com/realyoume/title_it_right.

1 Introduction

Paper titles play a central role in scholarly communication, shaping how research is retrieved, indexed, recommended, and read. As scientific publications continue to grow, automatic title generation has attracted increasing attention (Mishra et al., 2021; Ravaut et al., 2022; Zhang et al., 2022; Chen and Eger, 2023; Bikku et al., 2025; Callegari et al., 2023). Unlike generic summaries, titles are highly condensed representations of papers and often form readers’ first impressions. As illustrated in Figure 1,

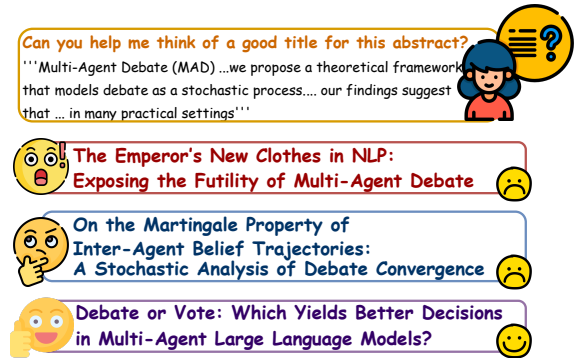


Figure 1: Comparison of title styles: faithful, attractive, and the optimal balance.

a good title should be both faithful to the paper content and attractive to readers, which can improve the visibility and impact of scientific work (Heard et al., 2023). Balancing these two goals, however, remains challenging.

Most existing methods formulate title generation as abstractive summarization and mainly optimize semantic consistency with the source document (Lewis et al., 2020; Raffel et al., 2020; Zhang et al., 2020a). While effective for preserving content, this formulation does not explicitly capture qualities that make titles distinctive and attractive. Conversely, overemphasizing attractiveness can produce vague, exaggerated, or unfaithful titles. Academic title generation is therefore naturally a multi-objective alignment problem involving two complementary objectives: *faithfulness* and *attractiveness*.

A straightforward solution is to combine multiple rewards into a scalar score and construct preference pairs for Direct Preference Optimization (DPO) (Rafailov et al., 2023). However, scalarization hides conflicts between individual objectives. Our gradient-level analysis shows that attractiveness-oriented (AO) and faithfulness-oriented (FO) samples induce strongly opposing gradients throughout training, and this conflict per-

[†]Corresponding author: Yingchao Piao (pyc@cnic.cn)

sists rather than disappearing during optimization. This suggests that the difficulty arises from the structure of the preference data itself.

To make these conflicts explicit, we formulate academic title generation with two independent rewards for faithfulness and attractiveness. Based on this dual-reward formulation, we propose **Multi-Objective Dominance Alignment (MODA)**, a minimal data-centric solution based on Pareto dominance. MODA retains only preference pairs in which one title is strictly better on both objectives and removes pairs with conflicting supervision. Importantly, MODA does not modify the standard DPO objective; it changes only the composition of the training data. Our controlled experiments further show that data composition strongly affects the learned faithfulness–attractiveness trade-off, supporting this simple filtering strategy.

Extensive automatic and human evaluations demonstrate that MODA achieves a strong balance between the two objectives. In blind human evaluation, our 8B-parameter model is preferred over human-authored titles on **73.5%** of the evaluated abstracts. Against GPT-5.1-High, MODA achieves a **46.0%** win rate. GPT-5.1-High is numerically preferred more often, but the difference is not statistically significant under our evaluation protocol. **Our contributions are summarized as follows:**

- (1) **A gradient-level explanation of scalarized multi-objective DPO.** We show that scalarized preference data can create persistent gradient conflicts between objective-specific samples, explaining why reward aggregation may fail to jointly optimize multiple objectives.
- (2) **A dual-reward formulation for faithful and attractive academic title generation.** We formulate title generation as a two-objective alignment problem and use separate reward models for faithfulness and attractiveness to provide more fine-grained signals for training and analysis.
- (3) **A minimal data-centric solution without modifying the DPO objective.** We propose MODA, which retains only strictly Pareto-dominant preference pairs and thereby reduces exposure to potentially conflicting supervision. This simple data filtering strategy improves the overall balance between faithfulness and attractiveness while keeping standard DPO unchanged.

The remainder of this paper is organized as follows: Section 2 reviews related work. Section 3 presents the dual-reward formulation and MODA. Section 4 reports the main experiments and analysis. Section 5 presents the human evaluation, and Section 6 concludes.

2 Related Work

2.1 Title Generation

Title generation is commonly formulated as abstractive summarization, from sequence-to-sequence models (Nallapati et al., 2016) to pre-trained BART, T5, and PEGASUS (Lewis et al., 2020; Raffel et al., 2020; Zhang et al., 2020a). Later work improves generation through selection, reranking, or refinement (Mishra et al., 2021; Ravaut et al., 2022; Bikku et al., 2025; Callegari et al., 2023). Such pipelines can improve the final candidate without explicitly representing the distinct qualities that determine whether a scientific title is both accurate and appealing. Most of these approaches optimize likelihood or a single aggregate score, so gains in fluency, salience, or stylistic appeal may obscure losses in factual specificity. This tension is particularly important for academic titles, where small shifts in scope, terminology, or emphasis can change both the perceived contribution and readers’ expectations of the underlying work. LLMs further increase title diversity (Chen and Eger, 2023; Rehman et al., 2024), yet lexical metrics do not fully reflect human judgments of title quality (Dristy et al., 2023). Existing methods therefore emphasize content fidelity, general quality, or narrow stylistic attributes such as humor, rather than jointly modeling faithfulness and broadly defined attractiveness. Our work studies this joint objective and the training conflicts it introduces.

2.2 Multi-Objective Preference Learning

DPO (Rafailov et al., 2023) directly learns from preference pairs without reinforcement learning. Data-centric variants show that pair construction matters: Curri-DPO (Pattnaik et al., 2024) organizes multiple pairs per prompt from easy to hard, whereas fDPO (Morimura et al., 2024) dynamically removes low-quality responses using a reward model. These methods address pair ordering or response quality, but not cases where different objectives reverse the preference direction for the same pair. A large preference margin or high-confidence label does not by itself resolve this issue, because

a pair can remain strongly preferred under one objective and strongly dispreferred under another.

Multi-objective methods instead aggregate rewards, modify DPO, or select data in reward space. MODPO (Zhou et al., 2024) introduces auxiliary-reward margins; SIPO (Li et al., 2025) combines sampling, response refinement, and filtering; and ParetoHqD (Gu et al., 2026) selects data along directional rays for SFT. MODA complements these approaches with an empirical gradient-level diagnosis in title generation: faithfulness-oriented and attractiveness-oriented pairs produce persistently opposing gradients. Unlike pipelines that regenerate responses or replace preference optimization with SFT, MODA filters a fixed candidate pool and retains the standard DPO objective, allowing us to study the effect of data composition under controlled conditions. This focus also differs from curriculum scheduling: MODA determines whether a pair supplies a preference direction consistent across objectives, rather than when that pair should be presented. Although the resulting dataset is smaller, each retained pair provides supervision that agrees on both faithfulness and attractiveness, directly reflecting the conflict identified in our gradient analysis. Its contribution is therefore not generic filtering, but an objective-consistency criterion for conflicting multi-objective supervision.

3 Methodology

As illustrated in Figure 2, our framework consists of two stages. First, we apply **Supervised Fine-Tuning (SFT)** to establish a strong initialization. Second, we propose **Multi-Objective Dominance Alignment (MODA)**, a preference learning method that filters training pairs using strict Pareto dominance. By leveraging dual reward models, MODA simultaneously optimizes the faithfulness and attractiveness of the generated titles.

3.1 Supervised Fine-Tuning

We use Qwen3-8B (Qwen Team, 2025) as our backbone model and apply LoRA (Hu et al., 2022) for parameter-efficient fine-tuning. The model is trained with the following instruction template: “generate a title based on the abstract. abstract: {abstract}. title: {title}”.

To ensure stable convergence, we train the model incrementally, adding 100,000 samples at each stage. The checkpoint from the eighth stage (800,000 samples in total) achieves the best val-

idation score and is saved as our base model, π_{sft} . Appendix A.1 provides the full hyperparameter details about our training.

3.2 Dual Reward Models

To avoid the high cost and subjectivity of human annotation, we follow the Reinforcement Learning from AI Feedback (RLAIF) paradigm. We divide title quality into two independent aspects: *faithfulness* and *attractiveness*. Each aspect is measured by a separate reward model to provide fine-grained optimization signals.

3.2.1 Faithfulness Reward

We define faithfulness as the semantic consistency between a generated **title** y and its source **abstract** x . Specifically, we use BERTScore F1 (Zhang et al., 2020b), which uses contextual embeddings to capture deep semantic similarities rather than just surface-level word overlap:

$$R_{\text{faith}}(y, x) = \text{BERTScore}_{\text{F1}}(y, x). \quad (1)$$

3.2.2 Attractiveness Reward

Attractiveness is subjective and varies by domain, making it hard to define with simple rules. Therefore, we learn it from AI-annotated preference data. Focusing on computer science papers from arXiv, we prompt DeepSeek V3.2 to rank five random candidate titles based on attractiveness and select the best and worst ones to form a pair. The clear quality gap between the best and worst titles helps stabilize the reward model training.

We fine-tune a RoBERTa-large (Liu et al., 2019) model with a scalar regression head, using the Bradley-Terry ranking loss (Bradley and Terry, 1952):

$$\mathcal{L}_{\text{BT}} = -\mathbb{E}_{(y^+, y^-)} [\log \sigma(s(y^+) - s(y^-))], \quad (2)$$

where $s(\cdot)$ is the predicted attractiveness score. We validate this reward model through human-correlation studies in Appendix E.

3.3 MODA

Jointly optimizing faithfulness and attractiveness is challenging. Combining rewards through a weighted sum forces a trade-off, while using preference pairs without a clear winner introduces conflicting gradients into DPO. We solve these issues with MODA, which only uses training pairs where one title is *strictly better* than the other.

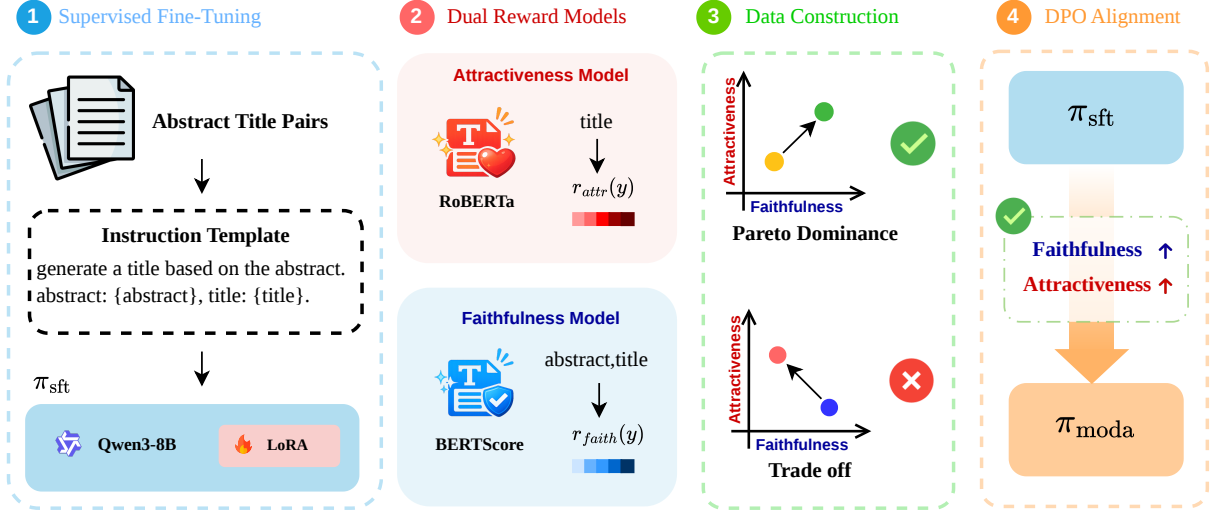


Figure 2: An overview of Title It Right.

Data Construction. For each abstract x , we pair the SFT-generated title y_{sft} with the original ground-truth title y_{origin} . We evaluate both titles using our dual reward models to obtain a two-dimensional score vector:

$$\mathbf{s}(y) = [R_{\text{faith}}(y, x), R_{\text{attr}}(y)] \in \mathbb{R}^2. \quad (3)$$

A candidate pair (y_A, y_B) is kept for DPO training if and only if y_A *strictly Pareto-dominates* y_B :

$$y_A \succ_{\text{Pareto}} y_B \iff \begin{cases} R_{\text{faith}}(y_A) > R_{\text{faith}}(y_B), \\ R_{\text{attr}}(y_A) > R_{\text{attr}}(y_B). \end{cases} \quad (4)$$

If $y_A \succ_{\text{Pareto}} y_B$, we set the chosen title $y_w = y_A$ and the rejected title $y_l = y_B$. Pairs without a strict dominance relationship are discarded to avoid confusing training signals. This allows the model to learn from its own high-quality outputs, serving as a form of self-improvement.

DPO Alignment. Out of 30,000 candidate pairs, our dominance filter selects about 12,448 high-quality pairs, forming the dataset $\mathcal{D}_{\text{pareto}}$. Starting from π_{sft} , we apply DPO on $\mathcal{D}_{\text{pareto}}$. By only learning from clear comparisons, MODA provides consistent training signals. This pushes the model beyond the standard trade-off curve, improving faithfulness and attractiveness at the same time.

4 Experiment

We conducted extensive experiments to evaluate our model. More implementation details are provided in the Appendix A.

4.1 Experiment Setup

4.1.1 Datasets

Train Data. We collected a large-scale training dataset consisting of 800,000 abstract-title pairs from arXiv. The data spans from 2000 to 2024 and encompasses multiple disciplines, including physics, mathematics, and computer science.

Test Data. For evaluation, we collected papers indexed on arXiv in 2025, focusing on the field of Computer Science. We sampled 1,000 papers in total from this dataset as our test set. This temporal train-test split reduces the risk of data leakage by ensuring that the test data postdates our supervised training corpus.

4.1.2 Training

We conduct all experiments using the Qwen3-8B model as our backbone, implemented via the LLaMA-Factory (Zheng et al., 2024) framework. For parameter-efficient fine-tuning, we apply LoRA to all linear layers. During the SFT stage, to balance scale and efficiency, we train incrementally on subsets of 100,000 samples per stage. We monitor semantic similarity using BERTScore F1 and apply early stopping if validation scores plateau for three consecutive stages. For the DPO stage, we initialize from the best SFT checkpoint and optimize the standard sigmoid preference loss. The complete hyperparameter settings for both stages are provided in Appendix A.1.

4.2 Experimental Results

Table 1 reports the evaluation results against various baselines, yielding two primary observations:

Table 1: Performance comparison under automatic quality evaluation and BT-Arena. **Faith./Attr.** denote LLM-judged faithfulness and attractiveness (1–5). BT-Arena scores are reported with 95% confidence intervals from 10,000 cluster-bootstrap resamples (Human Author = 1000). **Bold**: best result; underline: second-best.

Model	Faith.	Attr.	Gemma Judge	Qwen3.5 Judge
<i>Fine-tuned Language Models</i>				
T5-Large	4.772	2.890	900 [866, 933]	959 [926, 992]
BART-Large	4.768	2.915	879 [844, 913]	935 [901, 968]
Pegasus-Large	4.227	2.685	720 [679, 758]	752 [714, 788]
<i>Academic Title Generation</i>				
Mishra et al. (2021)	4.765	2.822	923 [889, 955]	1005 [971, 1039]
Callegari et al. (2023)	4.807	2.846	907 [871, 941]	964 [928, 998]
Chen and Eger (2023)	4.233	3.459	988 [951, 1023]	1008 [973, 1044]
<i>Large Language Models</i>				
DeepSeek-V3.2	4.973	3.069	<u>1179</u> [1146, 1213]	1225 [1189, 1262]
DeepSeek-V3.2-Thinking	<u>4.983</u>	3.075	1145 [1110, 1180]	<u>1242</u> [1208, 1277]
GPT-5.1-High	4.989	3.202	1288 [1251, 1327]	1336 [1299, 1377]
Human Author	4.650	3.043	1000	1000
<i>Preference Alignment</i>				
ParetoHqD	4.938	2.975	998 [965, 1031]	1093 [1059, 1127]
OURS (SFT)	4.901	2.990	1009 [975, 1043]	1075 [1041, 1109]
OURS (MODA)	4.913	<u>3.234</u>	1141 [1109, 1176]	1195 [1161, 1230]

MODA strikes a better balance between faithfulness and attractiveness while significantly improving preference-based evaluation. Compared to OURS (SFT), MODA improves attractiveness from 2.990 to 3.234 and marginally increases faithfulness from 4.901 to 4.913. More importantly, these gains consistently translate to preference-based evaluations: its BT-Arena score rises from 1009 to 1141 under the Gemma judge, and from 1075 to 1195 under the Qwen3.5 judge. Furthermore, MODA outperforms ParetoHqD, the closest controlled data-selection baseline. Although ParetoHqD yields a slightly higher faithfulness score (4.938 vs. 4.913), MODA delivers substantially higher attractiveness (3.234 vs. 2.975) and better BT-Arena scores under both judges (1141 vs. 998 and 1195 vs. 1093). In contrast, [Chen and Eger \(2023\)](#) achieves the highest attractiveness score (3.459) but suffers a severe drop in faithfulness (4.233), highlighting the challenge of optimizing one objective without compromising the other. Overall, these results demonstrate that using Pareto-dominant preference pairs effectively yields a more favorable trade-off between the two objectives.

MODA achieves competitive performance against frontier LLMs with an 8B backbone. Among all evaluated systems, MODA ranks second in attractiveness (3.234), surpassing DeepSeek-V3.2 (3.069), DeepSeek-V3.2-Thinking (3.075), and GPT-5.1-High (3.202), while maintaining a strong faithfulness score of 4.913. On BT-Arena, GPT-5.1-High remains the best-performing system, reaching 1288/1336 under the Gemma/Qwen3.5 judges compared to 1141/1195 for MODA. Nevertheless, MODA noticeably closes the performance gap: under the Gemma judge, its score is on par with DeepSeek-V3.2-Thinking (1141 vs. 1145), and under the Qwen3.5 judge, it closely follows DeepSeek-V3.2 and DeepSeek-V3.2-Thinking (1195 vs. 1225/1242). These findings suggest that an 8B model, when trained on carefully constructed multi-objective preference data, can achieve human preference quality competitive with frontier models. Detailed pairwise win rates are provided in Appendix F.

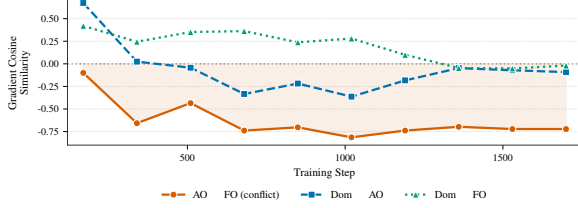


Figure 3: Gradient cosine similarity between different sample types throughout DPO training. AO and FO exhibit persistent negative similarity, while gradients involving Dom samples show weaker conflict.

4.3 RQ1: How Do Reward Conflicts Arise in DPO Training?

In multi-objective DPO, linearly aggregating reward scores into a single scalar obscures fine-grained trade-offs: the model only perceives an overall preference advantage without knowing which specific dimension contributes to it.

Sample Categorization. Based on this observation, we categorize the training data pairs into three types: (1) **Dom**: the chosen sample dominates on *both* objectives; (2) **FO**: the chosen sample excels in faithfulness but lacks attractiveness; (3) **AO**: the chosen sample is more attractive but less faithful. An analysis of 30,000 candidate pairs reveals that Dom, FO, and AO account for 41.68%, 23.35%, and 34.97%, respectively—meaning that, under our reward definitions, over half of the data exhibits opposing objective directions.

Gradient Conflict Analysis. Figure 3 reveals a clear and persistent opposition between **AO** and **FO** samples throughout DPO training. Their gradient cosine similarity decreases rapidly from around -0.1 in the early stage to below -0.6 , and remains strongly negative thereafter, fluctuating mostly around -0.7 .

Crucially, this opposition does *not* diminish as training proceeds. Instead, AO and FO continue to induce substantially opposing optimization directions until the final training steps, suggesting that scalar reward aggregation does not resolve the underlying cross-objective conflict.

In comparison, gradients involving **Dom** samples exhibit substantially weaker conflict. Their cosine similarities are positive or closer to zero during much of training and remain consistently higher than that between AO and FO. This contrast supports our hypothesis that Pareto-dominant pairs provide compatible supervision for optimization.

Reward-model robustness. Because our default faithfulness and attractiveness rewards use different encoders, we further examine whether the observed objective conflict could be an artifact of encoder mismatch. We construct three matched-backbone configurations using RoBERTa-base, RoBERTa-large, and SciBERT (Beltagy et al., 2019), where both reward scorers share the same backbone. Across 60,000 examples, reward rankings remain stable: Spearman correlations range from 0.723–0.820 for faithfulness and 0.875–0.901 for attractiveness, while the overlap between Pareto-selected subsets ranges from 0.634–0.689. These results suggest that MODA’s filtering behavior is not specific to the original encoder configuration and that the observed faithfulness–attractiveness conflict cannot be explained solely by encoder mismatch. Full results with 95% bootstrap confidence intervals are provided in Appendix B.

RQ1 Takeaway: The conflict between objective-specific samples persists throughout training rather than diminishing over time, suggesting that it originates from the structure of the preference data rather than transient optimization noise. This motivates a data-centric solution that removes conflicting supervision at the data level.

4.4 RQ2: Does Data Composition Determine Multi-Objective DPO Performance?

For RQ2 and RQ3, we evaluate all variants using the *same* proxy reward models as those used during data construction. This controlled evaluation measures how different training regimes optimize within a *fixed* objective space, so that performance differences can be attributed mainly to the training data rather than to changes in the evaluator. We therefore interpret the RQ2/RQ3 results as measurements of optimization behavior within this shared objective space. Table 1 and Section 5 provide complementary evaluations of absolute output quality.

To study how training data composition affects multi-objective DPO, we construct six controlled variants based on the sample categories identified in RQ1. All variants use the same training hyperparameters and differ only in sample composition or dataset size (see Appendix C). The results are summarized in Table 2 and Figure 4. We highlight two main findings.

Data composition has a stronger effect than data scale on the learned preference direction. Attr-Biased and Faith-Biased, which emphasize

Table 2: Performance of model variants on the test set. **Faith.** and **Attr.** are the internal reward-model scores. Adjacent Δ columns report relative gain over BASE as a percentage: $(\text{score} - \text{score}_{\text{base}}) / \text{score}_{\text{base}} \times 100\%$. **Bold** / underline mark the best / second-best *score* within each metric group; $\dagger$ denotes our proposed method.

Model	Faith $\uparrow$		Attr $\uparrow$		Avg $\uparrow$	
	Score	Δ	Score	Δ	Score	Δ
Base (SFT only)	0.406	—	0.323	—	0.364	—
Unfiltered-Full	0.618	(+52.2%)	<u>0.550</u>	(+70.3%)	0.584	(+60.4%)
Unfiltered-Sub	0.628	(+54.7%)	0.523	(+61.9%)	0.576	(+58.2%)
Attr-Biased	0.588	(+44.8%)	0.576	(+78.3%)	0.582	(+59.9%)
Faith-Biased	0.708	(+74.4%)	0.381	(+18.0%)	0.544	(+49.5%)
BalancedMix	0.678	(+67.0%)	0.492	(+52.3%)	<u>0.585</u>	(+60.7%)
ParetoSelect $\dagger$	<u>0.692</u>	(+70.4%)	0.507	(+57.0%)	0.599	(+64.6%)

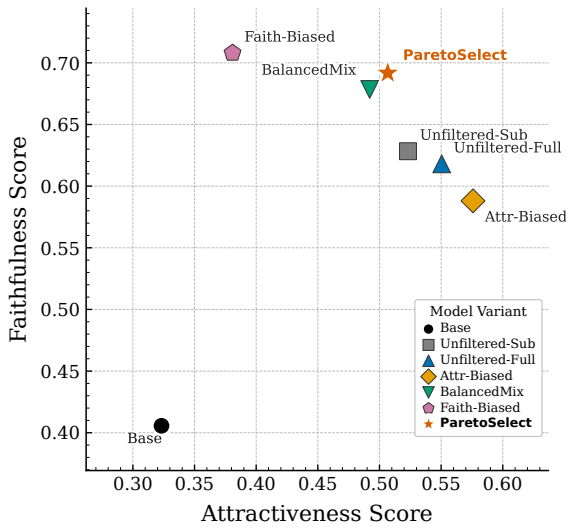


Figure 4: Faithfulness–attractiveness trade-off across variants, revealing an approximate Pareto frontier.

different objective-specific sample types, lead to clearly different outcomes. Attr-Biased achieves the highest attractiveness score (0.576) but relatively low faithfulness (0.588), whereas Faith-Biased achieves the highest faithfulness score (0.708) while reducing attractiveness to 0.381. These results show that changing the composition of preference data can strongly shift the model toward one objective at the expense of the other.

By comparison, increasing the amount of data alone provides much smaller gains. Unfiltered-Full and Unfiltered-Sub have nearly identical sample distributions but contain 30K and 12K pairs, respectively. Increasing the dataset size by $2.5\times$ improves the average score only from 0.576 to 0.584. In contrast, changing the sample composition at a fixed scale causes much larger changes. This sug-

gests that, under our controlled setting, the composition of preference data matters more than simply increasing the number of training pairs.

Pareto-dominant selection provides the best overall balance between the two objectives. ParetoSelect, trained only on Dom pairs, achieves a faithfulness score of 0.692 and an attractiveness score of 0.507, giving the highest average score (0.599) among all variants. Although objective-biased variants can achieve a higher score on one individual objective, they do so by substantially sacrificing the other. In contrast, PARETOSELECT maintains strong performance on both objectives.

A particularly informative comparison is with BalancedMix, which retains both AO and FO samples. At the same training scale, ParetoSelect improves both faithfulness (0.678 $\rightarrow$ 0.692) and attractiveness (0.492 $\rightarrow$ 0.507), increasing the average score from 0.585 to 0.599. This result is consistent with the gradient analysis in RQ1: balancing conflicting sample types can reduce directional bias, but it does not remove the conflicting supervision they introduce. In contrast, filtering for Pareto-dominant pairs provides more compatible supervision across the two objectives and leads to a better overall balance.

RQ2 Takeaway: Data composition strongly shapes the optimization behavior of multi-objective DPO. Increasing the amount of data while keeping a similar composition provides only limited improvement, whereas changing the proportions of objective-specific samples substantially shifts the learned trade-off. Selecting Pareto-dominant pairs achieves the highest average performance and the best balance between the two objectives among the evaluated variants.

4.5 RQ3: What Is the Essence of Reward Weight Tuning?

In multi-objective DPO, reward weighting is commonly used to control the relative importance of different objectives. We define

$$r_\alpha = \alpha r_{\text{attr}} + (1 - \alpha) r_{\text{faith}}, \quad (5)$$

where $\alpha \in [0, 1]$ denotes the weight assigned to attractiveness. We refer to the preference data constructed using r_α as **Weight- α** . Changing α affects not only the scalarized reward but also the resulting data composition, because some conflicting pairs change their chosen/rejected assignments.

Weight Tuning Induces Changes in Data Composition. As α varies, the proportion of **Dom** samples remains nearly unchanged, whereas the proportions of **AO** and **FO** samples shift systematically (see Appendix D). To separate this composition effect from reward weighting itself, we construct **Ratio- λ** . For each Weight- α setting, λ is defined as the proportion of **AO** samples in the preference dataset induced by that α :

$$\lambda_\alpha = \frac{N_{\text{AO}}(\mathcal{D}_\alpha)}{|\mathcal{D}_\alpha|}, \quad (6)$$

where $\mathcal{D}_\alpha$ denotes the dataset constructed by Weight- α . Ratio- λ fixes the reward weight at $\alpha = 0.5$, retains its preference assignments, and re-samples the data so that the AO proportion matches λ_α . For example, if Weight-0.8 induces an AO proportion of $\lambda_{0.8}$, the corresponding Ratio- $\lambda_{0.8}$ reproduces this proportion without changing the preference assignments defined at $\alpha = 0.5$.

Weight Tuning Largely Acts through Data Composition. As Figure 5 shows, Weight- α and its corresponding Ratio- λ produce similar faithfulness–attractiveness trade-offs, especially for moderate values of α . This suggests that a substantial part of the effect of reward-weight tuning is associated with the induced change in data composition. Larger differences appear at extreme weights, where Weight- α yields less balanced operating points. A possible reason is that extreme scalarization weights can reverse the preference assignments of low-margin pairs, whereas Ratio- λ changes only the sampling distribution.

Weight and Ratio Tuning Trace Similar Trade-offs. Both strategies mainly shift the model toward different regions of the faithfulness–attractiveness trade-off rather than consistently improving both objectives. In contrast, PARETOSELECT removes objective-specific conflicting pairs

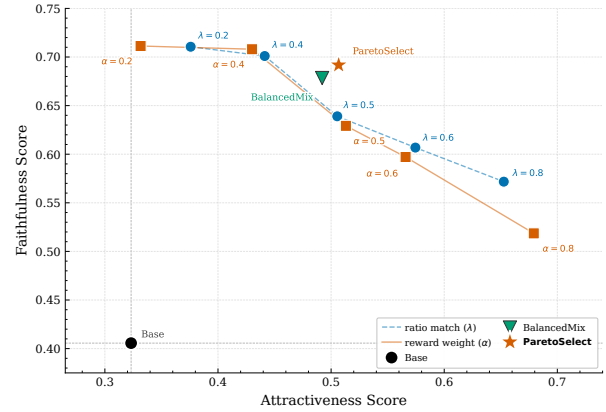


Figure 5: Faithfulness and attractiveness scores for Weight- α and Ratio- λ . α controls the attractiveness reward weight, while λ_α denotes the AO proportion induced by the corresponding Weight- α setting.

and provides a more balanced non-dominated operating point. Together with RQ2, these results indicate that data composition plays an important role in shaping the optimization behavior of multi-objective DPO.

RQ3 Takeaway: Reward-weight tuning changes both the scalarized objective and the composition of the induced preference data. Reproducing the AO proportion induced by Weight- α through Ratio- λ recovers much of the same faithfulness–attractiveness trade-off, suggesting that data composition mediates a substantial part of the effect of reward weighting.

5 Human Evaluation

To rigorously evaluate our DPO model, we conducted a blind human evaluation on 200 randomly sampled abstracts from the test set. For each abstract, we compared three candidate titles: the title generated by MODA, the title generated by GPT-5.1-High, and the original human-authored title. Three computer science researchers independently ranked the three titles by overall quality. Annotators were instructed to prioritize (1) accuracy in reflecting the core research content of the abstract, (2) attractiveness to readers, and (3) conciseness. The complete annotation instructions are provided in Appendix I.

5.1 Inter-Annotator Agreement

To assess annotation reliability, we calculated Kendall’s coefficient of concordance W (Kendall and Smith, 1939). The average agreement across

Table 3: Abstract-level human preference results after majority voting across three annotators. “Wins / Total” counts wins of the first system in each comparison. Confidence intervals are Wilson 95% confidence intervals, and p -values are from two-sided exact binomial tests against a 50% win rate.

Comparison	Wins / Total	Win Rate	95% CI	p -value
MODA vs. GPT-5.1-High	92 / 200	46.0%	[39.2%, 52.9%]	0.289
MODA vs. Human	147 / 200	73.5%	[67.0%, 79.1%]	2.10×10^{-11}
GPT-5.1-High vs. Human	143 / 200	71.5%	[64.9%, 77.3%]	1.00×10^{-9}

the 200 abstracts is $W = 0.44$ ($p < 0.001$), indicating moderate agreement. The distribution is bimodal: 61 samples fall near chance-level agreement ($W = 0.11$), whereas 56 samples exhibit strong agreement ($W = 0.78$). This pattern reflects the subjective nature of title-quality judgments, even when annotators follow the same evaluation criteria.

To avoid treating correlated judgments from the same abstract as independent observations, we use the abstract rather than the individual annotation as the statistical unit in our preference analysis. For each pairwise comparison, the three annotators’ judgments are aggregated by majority vote, yielding $n = 200$ independent outcomes. Additional agreement analyses are provided in Appendix G. We also note that the use of only three annotators remains a limitation of the evaluation.

5.2 Human Preference Results

Table 3 summarizes the abstract-level pairwise preference results. Evaluators preferred MODA over human-authored titles for 147 of the 200 abstracts, achieving a 73.5% win rate (95% CI: [67.0%, 79.1%]). This improvement is statistically significant based on a two-sided exact binomial test ($p = 2.10 \times 10^{-11}$). Similarly, GPT-5.1-High outperformed human titles with a 71.5% win rate (95% CI: [64.9%, 77.3%], $p = 1.00 \times 10^{-9}$).

In a direct comparison, MODA was preferred over GPT-5.1-High on 92 of 200 abstracts (46.0% win rate, 95% CI: [39.2%, 52.9%]). GPT-5.1-High was numerically preferred more often (54.0% vs. 46.0%), but the difference is not statistically significant ($p = 0.289$). Thus, we do not observe a statistically significant difference under our evaluation protocol; this result does not establish that the two systems are equivalent.

Overall, these results show that MODA is significantly preferred over human-authored titles under the adopted criteria. In the comparison with GPT-5.1-High, the observed difference does not reach

statistical significance, but the relatively broad confidence interval precludes a claim of equivalence or matched performance.

6 Conclusion

In this paper, we introduce a data-centric framework to generate academic titles that are both faithful and attractive. We empirically analyze the gradient interference associated with conflicting training pairs and find that simply tuning reward weights does not resolve it in our controlled setting. Our results further suggest that reward-weight tuning largely acts through changes in the induced data composition. To improve this composition, we propose MODA, which uses dual fine-grained reward models to retain strictly Pareto-dominant pairs and reduce exposure to ambiguous trade-off signals. Comprehensive automatic and human evaluations show that our fine-tuned 8B-parameter model improves the observed balance between faithfulness and attractiveness, outperforms standard baselines, and achieves competitive human preference scores against frontier LLMs.

Limitations

Domain coverage. Our attractiveness reward model is trained on computer science papers because rigorous evaluation requires accessible domain experts; the SFT corpus itself spans physics, mathematics, and computer science. This is a practical, not fundamental, restriction of MODA: the framework only requires a domain-appropriate preference dataset, which can be constructed via RLAIIF with any capable LLM. Quantifying cross-domain transfer is left to future work.

Practical scope. Frontier LLMs may be the most practical choice for occasional title generation. MODA instead targets sustained or institutional workflows that value local deployment, domain adaptation, reproducibility, data governance, or reduced reliance on proprietary APIs. Title gener-

ation also provides a concrete testbed for multi-objective alignment; future work should assess adoption, editing effort, and deployment trade-offs in real writing workflows.

Reward model reliability. The attractiveness reward model, trained on DeepSeek-V3.2 preference labels, shows 68.5% pairwise consistency with human annotators overall and 80.6% for large score gaps (Appendix E); marginal comparisons remain noisier. Its scores should therefore not be interpreted as definitive judgments of title quality. We use it only for *data filtering*, cross-validate results with two independent LLM judges and a blind human study (Section 5).

Scope of algorithmic comparison. We do not fully reproduce SIPO (Li et al., 2025), whose pipeline adds response refinement and regeneration to preference-data construction. Adapting it to title generation would require a task-specific prompt, another generation stage, and new candidates, changing the candidate pool, data construction, and computational budget. These changes would confound attribution to the alignment method, while omitting them would not represent the complete SIPO pipeline. By contrast, we evaluate ParetoHqD under matched candidate, reward, and evaluation settings because it operates directly on preference-data selection. We therefore discuss SIPO as a structurally different method and leave full reproduction to future work.

Ethical Considerations

Artifact Usage Statement

Datasets. We use paper-title and abstract records sourced from arXiv, a publicly accessible preprint repository. arXiv classifies titles and abstracts as descriptive metadata and makes such metadata available under the Creative Commons CC0 1.0 Universal Public Domain Dedication.¹ This designation applies to the metadata used in our study, not automatically to article PDFs, source files, or other full-text content, whose reuse conditions vary by submission. We use these records for research and follow the applicable arXiv access and API terms.

Pre-trained model. We fine-tune Qwen3-8B (Qwen Team, 2025), an open-weight model released by the Qwen Team. The official Qwen3-8B distribution is provided under the Apache

License 2.0.² We use the model as a backbone for supervised and preference fine-tuning and retain the corresponding attribution and license notices.

Training framework. We implement the training pipeline with LLaMA-Factory (Zheng et al., 2024), whose source code is distributed under the Apache License 2.0.³ Our use consists of configuring and running its supported fine-tuning workflows; any redistribution of code or modified components remains subject to the license’s notice and attribution requirements.

Human Participants and Ethics Review

Recruitment and compensation. We invited three computer science researchers from an ongoing research collaboration to evaluate 200 randomly selected test samples. Each participant ranked three anonymized candidate titles per sample. Participation was voluntary, and no monetary compensation was provided; this arrangement was disclosed before consent was obtained. Because the evaluation was conducted as voluntary research collaboration rather than paid employment or crowdsourced labor, no wage rate was set and a country-specific wage comparison is not applicable.

Consent and data use. Before the evaluation, participants were told that their rankings would be used exclusively to assess AI-generated academic titles for a research publication. We obtained explicit consent from all participants for this use.

Ethics review. The human evaluation involved voluntary ranking of anonymized academic titles by adult researchers and did not collect sensitive personal information. Given the minimal-risk nature of the task, we did not seek formal institutional ethics review. All participants were informed of the purpose and intended use of the evaluation data and provided consent before participation.

Potential Risks

Optimizing title attractiveness may encourage sensational, overly broad, or insufficiently qualified wording. Such titles could exaggerate a paper’s findings, obscure important limitations, or contribute to misleading scientific communication even when they receive high proxy-reward scores. Uncritical automation may also weaken author accountability or facilitate high-volume production

¹<https://info.arxiv.org/help/api/tou.html>

²<https://huggingface.co/Qwen/Qwen3-8B>

³<https://github.com/hiyouga/LLaMA-Factory>

of superficially appealing, low-quality submissions. Generated titles should therefore be treated as suggestions rather than publication-ready outputs and must remain subject to author or editor review for factual support, appropriate qualification, and disciplinary conventions.

Environmental and Computational Impact

Our study requires non-trivial computation. Supervised fine-tuning processes up to 800,000 abstract-title pairs and the experimental analysis includes multiple fine-tuning variants. For the final alignment stage, MODA applies DPO to approximately 12,000 Pareto-dominant pairs selected from 30,000 candidate pairs.

We limit the number of trainable parameters by applying LoRA with rank 8 to an 8B-parameter backbone instead of full-parameter fine-tuning, and conduct training in BF16. Filtering also reduces the number of preference pairs processed during the final DPO stage relative to the candidate pool. These choices are intended to reduce computational requirements and improve accessibility.

AI Assistant

We used AI assistants solely for translation and linguistic refinement of the manuscript.

Acknowledgments

We gratefully acknowledge the China Science and Technology Cloud (CSTCloud) MaaS platform for providing access to the large language model services used in this work.

References

- Iz Beltagy, Kyle Lo, and Arman Cohan. 2019. [SciBERT: A pretrained language model for scientific text](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3615–3620, Hong Kong, China. Association for Computational Linguistics.
- Thulasi Bikku, Nirmala Rani Narimalla, Keerthi Konda, Anusha Nakkala, Avanti Yarlagadda, and B. Sachuthananthan. 2025. Generating accurate and engaging research paper titles using NLP techniques. In *Bio-Inspired Computing*, pages 428–437, Cham. Springer Nature Switzerland.
- 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.
- Elena Callegari, Peter Vajdecka, Desara Xhura, and Anton Karl Ingason. 2023. [Enhancing academic title generation using SciBERT and linguistic rules](#). In *Proceedings of the Second Workshop on Information Extraction from Scientific Publications*, pages 74–83, Bali, Indonesia. Association for Computational Linguistics.
- Yanran Chen and Steffen Eger. 2023. [Transformers go for the LOLs: Generating \(humorous\) titles from scientific abstracts end-to-end](#). In *Proceedings of the 4th Workshop on Evaluation and Comparison of NLP Systems*, pages 62–84, Bali, Indonesia. Association for Computational Linguistics.
- DeepSeek-AI, Aixin Liu, Aoxue Mei, Bangcai Lin, Bing Xue, Bingxuan Wang, Bingzheng Xu, Bochao Wu, Bowei Zhang, Chaofan Lin, Chen Dong, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenhao Xu, Chong Ruan, Damai Dai, Daya Guo, Dejian Yang, and 245 others. 2025. [DeepSeek-V3.2: Pushing the frontier of open large language models](#). *Preprint*, arXiv:2512.02556.
- Ananna Datta Dristy, MD. Tanhab Hossain, and Mohammad Shahidur Rahman. 2023. [Harnessing the power of pretrained models for scientific paper title generation](#). In *2023 26th International Conference on Computer and Information Technology (ICCIT)*, pages 1–6.
- Haoran Gu, Handing Wang, Yi Mei, Mengjie Zhang, and Yaochu Jin. 2026. [ParetoHqD: fast offline multi-objective alignment of large language models using pareto high-quality data](#). In *Proceedings of the Fortieth AAAI Conference on Artificial Intelligence and Thirty-Eighth Conference on Innovative Applications of Artificial Intelligence and Sixteenth Symposium on Educational Advances in Artificial Intelligence, AAAI’26/IAAI’26/EAAI’26*. AAAI Press.
- Stephen B. Heard, Chloe A. Cull, and Easton R. White. 2023. [If this title is funny, will you cite me? citation impacts of humour and other features of article titles in ecology and evolution](#). *FACETS*, 8:1–15.
- Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. [LoRA: Low-rank adaptation of large language models](#). In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net.
- M. G. Kendall and B. Babington Smith. 1939. [The problem of m rankings](#). *The Annals of Mathematical Statistics*, 10(3):275–287.
- Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. [BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.

- Moxin Li, Yuantao Zhang, Wenjie Wang, Wentao Shi, Zhuo Liu, Fuli Feng, and Tat-Seng Chua. 2025. [Self-improvement towards Pareto optimality: Mitigating preference conflicts in multi-objective alignment](#). In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 11010–11031, Vienna, Austria. Association for Computational Linguistics.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. [Roberta: A robustly optimized BERT pretraining approach](#). *CoRR*, abs/1907.11692.
- Prakhar Mishra, Chaitali Diwan, Srinath Srinivasa, and G. Srinivasaraghavan. 2021. [Automatic title generation for text with pre-trained transformer language model](#). In *2021 IEEE 15th International Conference on Semantic Computing (ICSC)*, pages 17–24.
- Tetsuro Morimura, Mitsuki Sakamoto, Yuu Jinnai, Ken-shi Abe, and Kaito Ariu. 2024. [Filtered direct preference optimization](#). In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 22729–22770, Miami, Florida, USA. Association for Computational Linguistics.
- Ramesh Nallapati, Bowen Zhou, Cicero dos Santos, Çağlar Gulçehre, and Bing Xiang. 2016. [Abstractive text summarization using sequence-to-sequence RNNs and beyond](#). In *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning*, pages 280–290, Berlin, Germany. Association for Computational Linguistics.
- Pulkit Pattnaik, Rishabh Maheshwary, Kelechi Ogueji, Vikas Yadav, and Sathwik Tejaswi Madhusudhan. 2024. [Enhancing alignment using curriculum learning & ranked preferences](#). In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 12891–12907, Miami, Florida, USA. Association for Computational Linguistics.
- Qwen Team. 2025. [Qwen3 technical report](#). *CoRR*, abs/2505.09388.
- Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. 2023. Direct preference optimization: your language model is secretly a reward model. In *Proceedings of the 37th International Conference on Neural Information Processing Systems, NIPS ’23*, Red Hook, NY, USA. Curran Associates Inc.
- 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](#). *J. Mach. Learn. Res.*, 21:140:1–140:67.
- Mathieu Ravaut, Shafiq Joty, and Nancy Chen. 2022. [SummaReranker: A multi-task mixture-of-experts re-ranking framework for abstractive summarization](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4504–4524, Dublin, Ireland. Association for Computational Linguistics.
- Tohida Rehman, Debarshi Kumar Sanyal, and Samiran Chattopadhyay. 2024. [Can pre-trained language models generate titles for research papers?](#) In *Sustainability and Empowerment in the Context of Digital Libraries: 26th International Conference on Asia-Pacific Digital Libraries, ICADL 2024, Bandar Sunway, Malaysia, December 4–6, 2024, Proceedings, Part I*, page 154–170, Berlin, Heidelberg. Springer-Verlag.
- Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter J. Liu. 2020a. PEGASUS: pre-training with extracted gap-sentences for abstractive summarization. In *Proceedings of the 37th International Conference on Machine Learning, ICML’20*. JMLR.org.
- Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020b. [BERTScore: Evaluating text generation with BERT](#). In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26–30, 2020*. OpenReview.net.
- Ting Zhang, Ivana Clairine Irsan, Ferdian Thung, Dong-Gyun Han, David Lo, and Lingxiao Jiang. 2022. iTiger: an automatic issue title generation tool. In *Proceedings of the 30th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, pages 1637–1641.
- Yaowei Zheng, Richong Zhang, Junhao Zhang, Yanhan Ye, and Zheyang Luo. 2024. [LlamaFactory: Unified efficient Fine-Tuning of 100+ language models](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations), ACL 2024, Bangkok, Thailand, August 11–16, 2024*, pages 400–410. Association for Computational Linguistics.
- Zhanhui Zhou, Jie Liu, Jing Shao, Xiangyu Yue, Chao Yang, Wanli Ouyang, and Yu Qiao. 2024. [Beyond one-preference-fits-all alignment: Multi-objective direct preference optimization](#). In *Findings of the Association for Computational Linguistics, ACL 2024, Bangkok, Thailand and virtual meeting, August 11–16, 2024*, volume ACL 2024 of *Findings of ACL*, pages 10586–10613. Association for Computational Linguistics.

A Experiment Details

A.1 Training Details

To complement the training procedure described in Section 4.1, Table 4 reports the hyperparameters used for the SFT and DPO stages.

Table 4: Key hyperparameters for SFT and DPO stages.

Hyperparameter	SFT	DPO
Shared Configurations		
BASE_MODEL	Qwen3-8B	
LORA_RANK	8	
LORA_ALPHA	16	
LORA_DROPOUT	0.1	
OPTIMIZER	adamw_torch	
LR_SCHEDULER_TYPE	cosine	
BF16	True	
MAX_GRAD_NORM	1.0	
Stage-Specific Configurations		
LEARNING_RATE	1×10^{-5}	5×10^{-6}
NUM_TRAIN_EPOCHS	1	3

A.2 Metrics

(1) Attractiveness and Faithfulness: We evaluate the faithfulness and attractiveness of the generated titles by prompting **Gemma-4-31B**, as detailed in the Appendix J. To ensure the stability of the results, we repeated the evaluation three times and used the mean score as the result.

(2) BT-Arena Score: We conduct pairwise preference judgments using LLM-based evaluators and fit a Bradley-Terry model to the resulting win/loss outcomes. Let $\hat{\beta}_i$ denote the estimated BT log-strength parameter of model i . We convert the BT coefficients to an Elo-style scale relative to the ORIGIN baseline as

$$\text{BT-Arena}_i = 1000 + \frac{400}{\ln 10} \left(\hat{\beta}_i - \hat{\beta}_{\text{ORIGIN}} \right). \quad (7)$$

Under this transformation, ORIGIN receives a score of 1000, and a 400-point difference corresponds to an approximately 91% predicted win probability under the Bradley-Terry model. The transformation preserves the ordering and relative statistical comparisons implied by the fitted BT coefficients while expressing them on an intuitive Elo-style scale.

A.3 Baselines

To evaluate the effectiveness of our method, we compared it with a range of different baseline methods. These baseline methods included three representative models finetuned on the same data (totaling 800,000 samples), original titles used as heuristic baselines, and three LLM models.

- **BART** (Lewis et al., 2020) : BART is widely recognized for its excellent performance in summarization and text generation tasks
- **T5 (Text-to-Text Transfer Transformer)** (Raffel et al., 2020) : T5 is a unified framework that can transform all natural language processing problems into text-to-text formats. T5 is known for its strong generalization capabilities and consistent performance across a variety of benchmarks.
- **Pegasus** (Zhang et al., 2020a) : Pegasus is a model designed specifically for abstract text summarization. By employing empty sentence generation as the objective during the pre-training phase, Pegasus achieves outstanding performance in capturing key information from documents.
- **Mishra et al. (2021)**: This method decomposes academic title generation into three sequential stages: generation, selection, and refinement. Originally built on GPT-2, we re-implement it using **Qwen3-8B** as the backbone to ensure a fair comparison.
- **Callegari et al. (2023)**: This approach enhances title quality through a post-processing pipeline that combines SciBERT-based semantic matching with linguistic rule filtering. We replace the generation backbone with **Qwen3-8B** for a fair comparison, while retaining the original post-processing module.
- **Chen and Eger (2023)**: This method frames humor-controlled title generation as a sequence-to-sequence task, fine-tuning transformers by prepending humor constraints to the abstract and further training on pseudo-data to enforce these conditions. For a fair comparison with their primary supervised approach, we fine-tune **Qwen3-8B** using an equivalent constraint-based protocol.

- **ParetoHqD** (Gu et al., 2026): ParetoHqD is a multi-objective alignment method that selects high-quality training data along directional rays in the reward space and performs supervised fine-tuning on the selected samples. We include it as the closest Pareto-based data-selection baseline to MODA.
- **DeepSeek-V3.2** (DeepSeek-AI et al., 2025): A leading open-source LLM, renowned for powerful inference capabilities, it is a competitive benchmark for general text generation.
- **GPT-5.1-High**: GPT-5.1-High represents the leading performance of LLMs, where “high” refers to reasoning effort. Due to its superior language generation capabilities, it provides a strong benchmark for zero-shot performance.
- **Human Author**: We use the original title written by human authors as a simple yet competitive benchmark.

B Reward-Model Backbone Robustness

To further examine whether the observed conflict between faithfulness and attractiveness is an artifact of using different reward-model encoders, we conduct an additional robustness analysis with matched-backbone configurations. Specifically, we consider RoBERTa-base, RoBERTa-large, and SciBERT. For each configuration, the faithfulness and attractiveness reward scorers use the same backbone architecture.

We evaluate the consistency of different backbone configurations on 60,000 examples from the preference-data construction process. Specifically, we measure: (1) Spearman’s rank correlation coefficient ρ between reward rankings across backbone configurations for faithfulness and attractiveness, respectively; and (2) the overlap coefficient between the corresponding Pareto-selected subsets. We also report 95% confidence intervals based on 10,000 bootstrap resamples.

As shown in Table 5, the reward rankings remain broadly consistent across backbone choices. For faithfulness, the Spearman correlations range from 0.7226 to 0.8203, while attractiveness exhibits even stronger agreement, with correlations ranging from 0.8746 to 0.9005. The Pareto-selected subsets also retain substantial overlap, with overlap coefficients between 0.6342 and 0.6891.

These results indicate that changing the encoder affects the exact reward scores and selected sam-

ples to some extent, but does not substantially alter their overall rankings or the resulting Pareto-based data selection. Although this analysis does not rule out backbone effects, it suggests that MODA’s filtering behavior is not specific to the original reward-model encoder configuration. In particular, the observed conflict between faithfulness and attractiveness is unlikely to arise solely from encoder mismatch and is consistent with tension between the two objectives under our evaluation setting.

C Research Question 2 Experiment Settings

Table 6 summarizes the six data compositions used in the RQ2 ablation study. *ParetoSelect* uses only Pareto-dominance pairs. *BalancedMix* balances the two objective-specific categories (AO and FO) while retaining a majority of Dom samples. *Attr-Biased* and *Faith-Biased* each up-weight one orientation-specific type to 60%. *Unfiltered-Full* and *Unfiltered-Sub* reflect the unfiltered natural distribution at full scale (30K) and subsampled scale (12K), respectively, serving as baselines to isolate the effect of filtering from that of data volume.

D Weight Tuning and Data Ratio Distribution

Table 7 reports the distribution of the three sample categories for $\alpha \in \{0.2, 0.4, 0.5, 0.6, 0.8\}$, based on 30K candidate sample pairs. Sample pairs with identical weighted scores are excluded. Consequently, the number of valid pairs varies slightly across different values of α . The number of Dom samples remains unchanged, whereas their proportion exhibits minor variation due to the different numbers of valid pairs. As α , the weight assigned to attractiveness, increases, the proportion of AO samples increases monotonically, while the proportion of FO samples decreases monotonically.

E Attractiveness Reward Model Consistency

To validate the effectiveness of our trained attractiveness reward model, we conduct a human evaluation on 200 samples. Evaluators are asked to select the more attractive title from a given pair, after which we analyze the agreement between human choices and model scores.

As shown in Table 8, the overall agreement between the model and human judgments reaches

Table 5: Reward-model robustness across matched backbones. We report Spearman’s ρ for reward rankings and overlap coefficients for Pareto-selected subsets, with 95% bootstrap confidence intervals.

Backbone Pair	Faithfulness ρ	Attractiveness ρ	Overlap Coef.
RoBERTa-base vs. RoBERTa-large	0.8203 [0.8166, 0.8238]	0.9005 [0.8986, 0.9022]	0.6891 [0.6833, 0.6951]
RoBERTa-base vs. SciBERT	0.7478 [0.7430, 0.7525]	0.8746 [0.8723, 0.8769]	0.6342 [0.6275, 0.6409]
RoBERTa-large vs. SciBERT	0.7226 [0.7174, 0.7276]	0.8751 [0.8727, 0.8773]	0.6355 [0.6289, 0.6421]

Table 6: Data compositions for the DPO data-mix ablation. All settings share the same hyperparameters and differ only in sample-type proportions or scale.

Setting	Dom	AO	FO	Pairs
ParetoSelect	1.00	0.00	0.00	12K
BalancedMix	0.60	0.20	0.20	12K
Attr-Biased	0.20	0.60	0.20	12K
Faith-Biased	0.20	0.20	0.60	12K
Unfiltered-Full	0.42	0.35	0.23	30K
Unfiltered-Sub	0.42	0.35	0.23	12K

Table 7: Distribution of sample categories under different reward weights α .

α (attr)	$1 - \alpha$ (faith)	Dom	AO	FO
0.2	0.8	41.67%	14.24%	44.08%
0.4	0.6	41.67%	27.90%	30.42%
0.5	0.5	41.68%	34.97%	23.35%
0.6	0.4	41.67%	41.22%	17.10%
0.8	0.2	41.67%	51.01%	7.32%

68.5%. Furthermore, this agreement improves significantly as the score gap between the titles widens. When the score gap is less than 2, the agreement is 57.1%, slightly above the random baseline (50%). This is expected: for title pairs of similar quality, even human annotators struggle to reach a consensus. Such divergence may reflect the subjectivity of the task rather than reward-model error alone. In contrast, when the score gap exceeds 4, the agreement reaches 78.5%, suggesting that the reward model is more reliable when distinguishing titles with clear quality differences.

Crucially, the attractiveness reward model serves only as a data filtering tool in this paper, rather than a final evaluation metric. To avoid over-reliance on a single reward model, all core conclusions are cross-validated through two independent LLM judges and blind human evaluations.

Table 8: Attractiveness scorer accuracy by score difference interval. $|\Delta s|$ denotes the absolute difference in attractiveness scores assigned by the scorer to the two candidate titles.

$ \Delta s $	Samples	Accuracy
[0, 2)	56	57.1%
[2, 4)	51	62.7%
[4, 6)	31	74.2%
[6, $+\infty$)	62	80.6%
Overall	200	68.5%

F Pairwise Comparison

Tables 9 and 10 report pairwise win rates for a representative subset of systems, including strong proprietary and open-source models, the human-authored baseline, ParetoHqD, and our SFT and MODA models. We use the same head-to-head preference protocol as BT-Arena and report results separately for each automatic judge. Table 9 uses **Gemma-4-31B** as the judge, while Table 10 uses **Qwen3.5-27B**. Each cell shows the percentage of comparisons in which the *row* model is preferred over the *column* model. Results for the remaining models are omitted for clarity, and diagonal entries are not reported.

G Inter-Annotator Agreement Details

For each of the 200 samples, we built a 3×3 rank matrix and calculated Kendall’s W as follows:

$$W = \frac{12S}{m^2(n^3 - n)}, \quad (8)$$

where $m = 3$ is the number of annotators, $n = 3$ is the number of titles, and $S = \sum_j (R_j - \bar{R})^2$ is the sum of squared differences between the column rank sums and their mean. A score of $W = 0$ means no agreement, while $W = 1$ means perfect agreement.

The W scores for all samples take six specific values, and their frequencies are shown in Table 11.

Table 9: Pairwise win rates (%) among selected models with Gemma-4-31B acting as the pairwise judge. Each cell reports the share of comparisons where the row model beats the column model.

Model	DeepSeek-V3.2	DeepSeek-V3.2-Thinking	GPT-5.1-High	Human Author	ParetoHqD	OURS (SFT)	OURS (MODA)
DeepSeek-V3.2	—	49.5%	28.1%	72.0%	76.8%	80.6%	54.5%
DeepSeek-V3.2-Thinking	50.5%	—	31.6%	67.0%	67.9%	69.8%	48.5%
GPT-5.1-High	71.9%	68.4%	—	76.8%	87.0%	81.8%	76.0%
Human Author	28.0%	33.0%	23.2%	—	48.2%	49.5%	32.0%
ParetoHqD	23.2%	32.1%	13.0%	51.8%	—	45.7%	30.3%
OURS (SFT)	19.4%	30.2%	18.2%	50.5%	54.3%	—	28.9%
OURS (MODA)	45.5%	51.5%	24.0%	68.0%	69.7%	71.1%	—

Table 10: Pairwise win rates (%) among selected models with Qwen3.5-27B acting as the pairwise judge. Each cell reports the share of comparisons where the row model beats the column model.

Model	DeepSeek-V3.2	DeepSeek-V3.2-Thinking	GPT-5.1-High	Human Author	ParetoHqD	OURS (SFT)	OURS (MODA)
DeepSeek-V3.2	—	56.1%	29.3%	67.7%	83.1%	67.0%	61.0%
DeepSeek-V3.2-Thinking	43.9%	—	25.5%	80.8%	77.6%	70.4%	57.0%
GPT-5.1-High	70.7%	74.5%	—	83.0%	80.5%	84.8%	68.0%
Human Author	32.3%	19.2%	17.0%	—	36.1%	45.9%	25.3%
ParetoHqD	16.9%	22.4%	19.5%	63.9%	—	54.0%	27.9%
OURS (SFT)	33.0%	29.6%	15.2%	54.1%	46.0%	—	26.8%
OURS (MODA)	39.0%	43.0%	32.0%	74.7%	72.1%	73.2%	—

Table 11: Distribution of per-sample Kendall’s W scores across 200 annotated samples.

W	S	Samples
0.000	0	8
0.111	2	61
0.333	6	34
0.444	8	26
0.778	14	56
1.000	18	15
Mean	—	0.44

We observed a bimodal distribution in the results: 61 samples scored $W = 0.11$, and 56 scored $W = 0.78$. This suggests that annotators strongly agreed on about half of the items, while the remaining titles were naturally difficult to rank.

H Case Study

Figure 6 presents two representative case studies comparing titles generated by the original authors, GPT-5.1-High, and our MODA model. In Case 1, the author’s title emphasizes technical fidelity,

while GPT-5.1-High produces a descriptively accurate but verbose alternative. Our model generates a concise, reader-oriented title that foregrounds the practical value without sacrificing informativeness. In Case 2, the author’s title is technically exhaustive but lengthy. GPT-5.1-High condenses it while preserving technical terminology. Our model takes a bolder approach with a metaphorical opening, achieving both memorability and conciseness — qualities consistently rewarded in our attractiveness reward model.

I Human Evaluation Protocol

Evaluators were provided with clear guidelines before commencing the evaluation. All titles were anonymized and presented in a random order. The exact text of the instructions provided to evaluators is shown in Figure 7.

Case1 abstract: Black box large language models (LLMs) make information extraction (IE) easy to configure, but hard to trust. Unlike traditional information extraction ... Human Author SafePassage: High-Fidelity Information Extraction with Black Box LLMS GPT-5.1-High SafePassage: Grounded Contexts for Trustworthy Information Extraction with Black-Box LLMS OUR MODA SafePassage: Making Information Extraction from Black Box LLMs Trustworthy	Case2 abstract: Accurate intraday solar irradiance forecasting is crucial for optimizing dispatch planning and electricity trading. For this purpose, we introduce a novel and ... Human Author Solar Multimodal Transformer: intraday Solar Irradiance Predictor using Public Cameras and Time Series GPT-5.1-High Solar Multimodal Transformer: Single-Frame Camera and Normalized Time Series Fusion for Short-Term Irradiance Forecasting OUR MODA Seeing the Sun: A Lightweight Multimodal Approach for Accurate Solar Irradiance Forecasting
--	---

Figure 6: Case Study.

Welcome to this task of manually evaluating title quality. This task aims to assess the performance of different models in generating academic/article titles by comparing titles generated by various models. 1. Task Overview You will see an abstract of an academic paper and three randomly arranged title options (A, B, C). Your task is to rank these three titles based on the content of the abstract. Core Rules: Blind Testing Principle: The generation models for options A/B/C are randomly shuffled. Please evaluate objectively based entirely on title quality. Mandatory Ranking: A, B, and C must be ranked in a complete order: 1 (best), 2 (second best), 3 (worst). Ties are not allowed. 2. Evaluation Criteria When ranking, please consider the following priorities from highest to lowest: Accuracy: Does the title accurately reflect the core research content of the abstract? Are there any factual errors or over-inferences? Attractiveness: Given the above conditions, is the title distinctive and attractive? Conciseness: Is the title concise and free of redundant words?
--

Figure 7: Annotation Instruction.

J Prompts

Preference Annotator. Figure 8 illustrates the prompts for constructing our attractiveness preference samples. The model is instructed to select the most and least attractive titles from the candidate list, favoring bold, simple, and unexpected titles.

Title Generation. Figure 9 illustrates how we prompt the LLM to generate titles from abstracts, assigning it the role of a computer science reviewer to balance faithfulness and attractiveness.

Faithfulness Judge. Figure 10 illustrates the prompts for assessing faithfulness. The model performs rigorous checks to ensure generated content is fully supported by the abstract without introducing external assumptions.

Attractiveness Judge. Figure 11 shows the prompt for the attractiveness metric, focusing on stylistic quality to measure how well a title resonates with readers.

Holistic Preference Judge. Figure 12 illustrates the prompt for comparing two candidate titles, instructing the model to act as a senior academic editor and determine the better title by jointly evaluating rigor and attractiveness.

```

# Role
You are an expert in the field of computer science.
You can now only choose titles that immediately
grab attention.

# Input Data
{titles}

# Rules
Before outputting JSON, you should strictly follow
the rules:
Chosen: Select the boldest title. Features: simple
and unexpected.
Reject: Select the most long-winded and boring
title. Features: Long sentences strung together and
lack of clear 'hooks'.
Output: You MUST output in this strict JSON
structure.
Exactness: Strictly prohibited from modifying any
characters in the original titles.

# Output Format
{
  "chosen": "Exact title string from the list",
  "reject": "Exact title string from the list"
}

```

Figure 8: Preference Annotator Prompt.

```

# Role
You are an expert in the field of computer science,
specializing in reviewing papers for top-tier
conferences.

# Task
Please read the provided paper abstract and help
me come up with the best possible title for the
paper. The title must balance faithfulness
and attractiveness.

# Abstract
{abstract}

# Return Format (json)
```json
{
 "title": "your title"
}
```

```

Figure 9: Generating Title.

```

# Role
You are an expert in the field of computer science.
You have a high academic taste and care a lot
about how well the title reflects the abstract.

# Task
Please read the provided [Abstract] and perform a
rigorous quantitative evaluation of the [Candidate
Title List] based on its Faithfulness to the
original text.

# Input Data
## 1. Abstract :
{abstract}
## 2. Candidate Title List :
{titles}

# Evaluation Criteria: Faithfulness
Core Definition: Measures how well the title reflects
the abstract. It demands semantic alignment. The
title must accurately reflect the task, method, and
scope defined in the abstract without overclaiming.
* 0-2 (Low): Factual deviation: core keywords
lack basis in the abstract or introduces hallucinated
benefits not mentioned in the text.
* 2.5-3.5 (Mid): Covers the primary task, but
evidence is weak: uses words not defined in the
abstract or loses specific nuances of the
methodology.
* 4-5 (High): Absolute alignment: every claim
and keyword is anchored in direct, verifiable
evidence within the abstract.

# Rules
Please execute the following steps:
1. Textual Grounding Analysis: For each title,
identify which specific
phrases or concepts from the [Abstract] support it.
2. Independent Scoring: Assign a score from 0-
5 (0.5 increments)
strictly following the Faithfulness criteria above.
3. Formatted Output: You MUST output in this
strict JSON structure:

# Output Format
```json
[
 {
 "title": "Title 1",
 "analysis": "Brief critique focusing on textual
grounding and factual accuracy.",
 "scores": {
 "faithfulness": 0
 }
 }
]
```

```

Figure 10: Faithfulness Judge Prompt.

```

# Role
You are an expert in the field of computer science.
You should understand which titles are
attractive and can enhance the citation potential of
a paper.

# Task
Please perform a quantitative evaluation of the
[Candidate Title List] only based on their
Attractiveness. You will evaluate the titles focusing
on their linguistic appeal rather than the context of
the abstract.

# Input Data
## Candidate Title List :
{titles}

# Evaluation Criteria: Attractiveness
Please provide a score from 0 to 5 (in 0.5
increments). You should follow the rules.
Core Definition: Attractiveness
Measures whether the title appeal to the aesthetic
of most people and whether it is highly unique.
0-2 (Low): Overly ordinary: lacks any tension or
narrative hook; Grammatically correct but lacking
appeal, too lengthy
2.5-3.5 (Mid): Standard academic phrasing:
difficult to stand out among thousands of papers.
4-5 (High): Use precise vocabulary, elegant
structures, thought-provoking or insightful.

# Rules
1. Independent Scoring: Assign a score from 0-5
(0.5 increments) strictly following the
Attractiveness criteria above.
2. Actionable Critique: Provide a brief explanation
of why the title succeeded or failed.
3. Output Format: You MUST output in this strict
JSON structure:
```json
[
 {
 "title": "Title 1",
 "scores": {
 "attractiveness": 0
 },
 "critique": "Briefly describe the strengths or
weaknesses."
 }
]
```

```

Figure 11: Attractiveness Judge Prompt.

```

# Role
You are a senior academic editor with peer-review
experience at top-tier venues (e.g., ACL, EMNLP,
NeurIPS).

# Task
Given a paper abstract and two candidate titles (A
and B), determine which title is better by
considering both rigor and attractiveness.
Output **strict JSON only** — no preamble, no
markdown fences.

# Evaluation Criteria
## Rigor
Does the title accurately capture the core problem
and overarching objective? Prioritize titles that
abstract complex mechanisms into clear, high-
level contributions. Penalize dense strings of
methodological jargon that obscure the primary
focus.
## Attractiveness
Is the title elegant, readable, and designed to
maximize dissemination? Reward narrative-driven
phrasing that explicitly communicates the paper's
value proposition. Prioritize active, engaging
aesthetics over purely descriptive model
nomenclature.

# Judgment Rules
- Evaluate both titles on Rigor and Attractiveness,
then compare holistically.
- **A wins**: A is meaningfully better overall on
both criteria.
- **B wins**: B is meaningfully better overall on
both criteria.
- **Tie**: Both titles are of comparable quality with
no clear winner on either criterion.

# Input
Abstract
{{ABSTRACT}}
Title A
{{TITLE_A}}
Title B
{{TITLE_B}}

# Output Format
{
  "reason": "<one concise sentence explaining the
decision>",
  "winner": "A" | "B" | "tie"
}

```

Figure 12: Holistic Preference Judge Prompt.