

MGDA-DECOUPLED: GEOMETRY-AWARE MULTI-OBJECTIVE OPTIMISATION FOR DPO-BASED LLM ALIGNMENT

Andor Vári-Kakas, Ji Won Park & Natasa Tagasovska
 Prescient Design, CS CoE, Genentech | Roche

ABSTRACT

Aligning large language models (LLMs) to desirable human values requires balancing multiple, potentially conflicting objectives such as helpfulness, truthfulness, and harmlessness, which presents a multi-objective optimisation challenge. Most alignment pipelines rely on a fixed scalarisation of these objectives, which can introduce procedural unfairness by systematically under-weighting harder-to-optimize or minority objectives. To promote more equitable trade-offs, we introduce MGDA-DECOUPLED, a geometry-based multi-objective optimisation algorithm that finds a shared descent direction while explicitly accounting for each objective’s convergence dynamics. In contrast to prior methods that depend on reinforcement learning (e.g., GAPO) or explicit reward models (e.g., MODPO), our approach operates entirely within the lightweight Direct Preference Optimisation (DPO) paradigm. Experiments on the UltraFeedback dataset show that geometry-aware methods—and MGDA-DECOUPLED in particular—achieve the highest win rates against golden responses, both overall and per objective.

1 INTRODUCTION

In multi-objective LLM alignment, the goal is to fine-tune a language model to satisfy multiple, often conflicting desiderata. While pretraining optimises a sole objective of next-token prediction, aligned models must simultaneously be helpful, truthful, harmless, honest, instruction following, concise, etc. (Bai et al., 2022). These diverse human needs are typically captured by a preference dataset $\mathcal{D} = \{(x, y_w, y_l)_n\}_{n=1}^N$, where annotators select a preferred response y_w over an alternative y_l for a given query x (Christiano et al., 2017). Leveraging such data, models are typically fine-tuned using either Reinforcement Learning from Human Feedback (RLHF) (Ouyang et al., 2022) or Direct Preference Optimisation (DPO) (Rafailov et al., 2023). While RLHF requires training an explicit reward model to guide a reinforcement learning loop, DPO optimises the policy directly on the preference pairs without a separate reward function.

However, representing multiple objectives via a single preference dataset often leads to a collapse of information (Wu et al., 2023). In particular, if the annotator pool is skewed—for example, if a majority prioritises helpfulness while a minority prioritises harmlessness—the resulting “aligned” model will inherently reflect these biases, raising major fairness concerns (Casper et al., 2023). To prevent this, recent approaches advocate for maintaining one preference dataset per objective (Chakraborty et al., 2024). This decomposition provides the distinct signals necessary for fine-tuning algorithms to explicitly manage trade-offs, enabling all objectives to be fairly considered rather than being implicitly weighted by dataset demographics (Sorensen et al., 2024).

Problem setup. We consider a setup with k distinct objectives, represented by individual pairwise preference datasets $\mathcal{D}_1, \dots, \mathcal{D}_k$. Our goal is to design a fine-tuning algorithm that yields a model aligned well across all k objectives, evaluated on held-out prompts via an AI judge (Zheng et al., 2023). Motivated by real-world scenarios, we focus strictly on lightweight algorithms that operate without the computational overhead of reinforcement learning (e.g., GAPO; Li et al., 2025), training auxiliary reward models (MODPO; Zhou et al., 2024), or tuning sensitive hyper-parameters (MOPPO; Agnihotri et al., 2025). Specifically, we find that methods which combine gradients from multiple

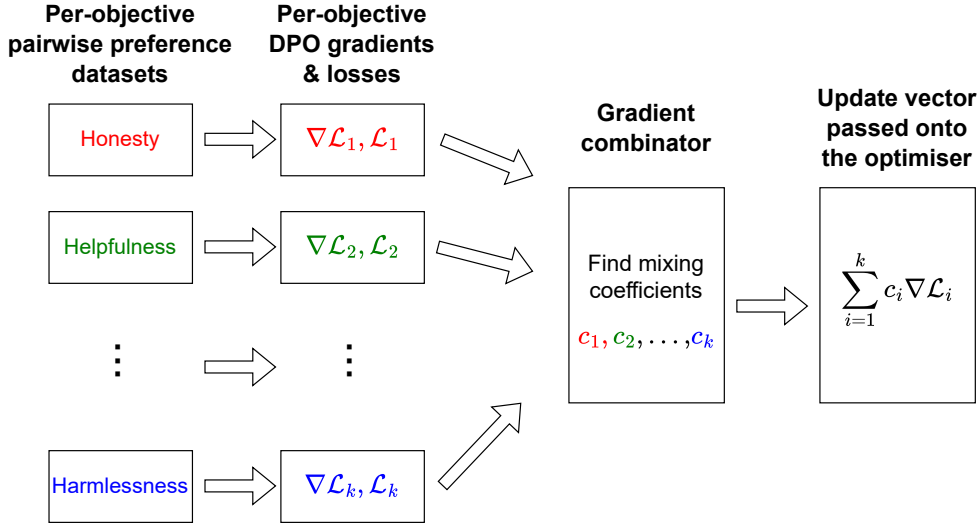


Figure 1: Overview of our DPO-based multi-objective fine-tuning framework. At each training step, independent gradients and losses are computed for each per-objective dataset. Based on these, the key gradient combinator algorithm (e.g., UNIFORM, MGDA-DECOUPLED) dynamically computes mixing coefficients c_i to synthesise a unified update vector, which is then applied to the model weights via the optimiser (e.g., Adam).

per-objective DPO instances provide an efficient and effective solution. See Figure 1 for an overview of the framework.

Formally, let π_θ denote the trainable policy parametrised by weights θ , and π_{ref} be the frozen reference model (e.g., from pretraining). We define the DPO loss (Rafailov et al., 2023) for objective i as:

$$\mathcal{L}_i := -\mathbb{E}_{(x, y_w, y_l) \sim \mathcal{D}_i} \left[\log \sigma \left(\beta \log \frac{\pi_\theta(y_w | x)}{\pi_{\text{ref}}(y_w | x)} - \beta \log \frac{\pi_\theta(y_l | x)}{\pi_{\text{ref}}(y_l | x)} \right) \right],$$

where σ is the sigmoid function and β controls the deviation from π_{ref} . Essentially, the optimisation goal is to simultaneously minimise the losses $\mathcal{L}_1, \dots, \mathcal{L}_k$.

Challenges and Ideas. This formulation constitutes a multi-objective optimisation (MOO) problem. The primary challenge is that alignment objectives are often conflicting (e.g., *Helpfulness* vs. *Harmlessness*). A standard approach is to optimise a fixed scalarisation $\sum_{i=1}^k c_i \mathcal{L}_i$ with $\sum_{i=1}^k c_i = 1$, but this method has critical limitations. First, fixed scalarisation restricts the solution to the convex hull of the Pareto front, making balanced solutions in non-convex regions inaccessible (Sener and Koltun, 2018). Second, determining suitable coefficients c_i *a priori* is difficult, and suboptimal weighting often causes the optimiser to over-prioritise easier objectives while neglecting harder ones (Chen et al., 2018). Furthermore, even with suitable coefficients, Yu et al. (2020) demonstrate that the composite landscape formed by summing the objectives can cause standard optimisers like Adam (Kingma and Ba, 2015) to stall or converge to a suboptimal point far from the Pareto front (surprisingly, even in scenarios where a joint optimum for all objectives is attainable).

To resolve these issues, *dynamic* scalarisation methods adapt the coefficients during training based on real-time loss or gradient information (see Chen et al. 2025 for a recent extensive summary). A particularly promising direction is the family of algorithms rooted in the Multiple Gradient Descent Algorithm (MGDA; Désidéri, 2014), which has demonstrated strong theoretical guarantees in multi-task learning (Sener and Koltun, 2018) and empirical success in RLHF (Li et al., 2025). By leveraging the geometry of the gradient landscape, these methods compute a common descent direc-

tion that improves all objectives simultaneously, effectively automating the selection of scalarisation coefficients that would otherwise require manual tuning. See Figure 2 for an illustration.

We operationalise alignment as a problem of algorithmic fairness in model capacity allocation. When training policies with preference optimisation, we implicitly decide how to distribute this limited capacity across competing normative desiderata (e.g., *Helpfulness*, *Harmlessness*, *Honesty*). Because these desiderata are typically instantiated through datasets and scoring pipelines that reflect annotator demographics, cultural context, and institutional priorities, common scalarisation choices can enact a form of procedural unfairness—systematically over-weighting objectives that are easier to optimise, more represented in data, or more aligned with the base model’s priors. This is especially consequential for downstream agentic systems, where a policy trained under a collapsed objective can repeatedly act in ways that privilege one value regime and reinforce long-run disparities.

To address this, we introduce MGDA-DECOUPLED, a multi-objective optimisation approach for DPO that computes balancing coefficients from loss-normalised per-objective gradients while applying them to the raw gradients, explicitly managing gradient conflicts and mitigating objective dominance. We evaluate across multiple objectives and models and report both aggregate performance and objective-wise outcomes, emphasising worst-objective behaviour as a fairness-relevant criterion. Overall, our results suggest that fairness-aware alignment need not require expensive RLHF pipelines: multi-objective preference optimisation provides a practical lever for enforcing equitable trade-offs among alignment goals.

In summary, our contributions are as follows:

- We propose MGDA-DECOUPLED, a novel algorithm designed to mitigate the scale-sensitivity issues inherent in existing geometry-based methods (Section 2.1).
- We analyse the failure modes of standard loss- and geometry-based baselines and validate the robustness of MGDA-DECOUPLED via a controlled optimisation problem (Section 2.2).
- We empirically evaluate these algorithms on the standard UltraFeedback dataset (Cui et al., 2023), establishing that MGDA-DECOUPLED consistently achieves the highest overall alignment quality (Sections 3 and 4).

2 ALGORITHMS

In what follows, we introduce the specific algorithms used in our comparison. As illustrated in Figure 1, all methods share the same underlying structure: at each step, they aggregate the per-objective gradients $\nabla\mathcal{L}_i$ into a single update vector $\sum_{i=1}^k c_i \nabla\mathcal{L}_i$. The methods differ solely in how they calculate the mixing coefficients c_i .

The most straightforward algorithm, which we refer to as UNIFORM, employs fixed uniform scalarisation, setting $c_i := 1/k$ for all objectives¹ throughout training.

We next consider GROUPDRO (Sagawa et al., 2019), which dynamically upweights objectives with higher losses to focus on the worst-case performance. It employs a multiplicative update rule based on the exponential of the per-objective losses, allowing for a smooth transition between steps:

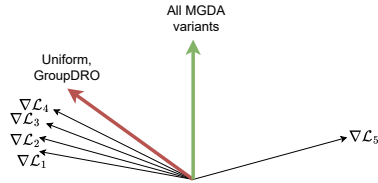


Figure 2: Geometric intuition for multi-objective updates. A majority cluster of aligned gradients $\nabla\mathcal{L}_{1\dots 4}$ (e.g., *Helpfulness*, *Truthfulness*, *Instruction Following*, and *Coherence*) opposes a single conflicting gradient $\nabla\mathcal{L}_5$ (e.g., *Harmlessness*). Assuming equal losses and gradient magnitudes, geometry-agnostic algorithms like UNIFORM and GROUPDRO return a combined gradient biased towards the majority (red), potentially overriding safety constraints. In contrast, MGDA variants (green) identify a common descent direction that improves all objectives simultaneously.

¹While specific applications might necessitate non-uniform weighting (e.g., prioritising safety), we employ uniform scalarisation as a neutral baseline, reflecting that our per-objective datasets originate from a common source with identical prompt distributions and no intrinsic hierarchy.

$$c_i^{(t+1)} := \frac{c_i^{(t)} \cdot e^{\eta \mathcal{L}_i}}{\sum_{j=1}^k c_j^{(t)} \cdot e^{\eta \mathcal{L}_j}}, \quad (1)$$

where η is a hyperparameter governing the sensitivity to loss differences and t denotes the iteration step. While GROUPDRO is dynamic, it remains agnostic to the geometric alignment of gradients.

2.1 MGDA VARIANTS

In contrast, MGDA variants explicitly address gradient conflicts by leveraging the geometry of the optimisation landscape (see Figure 2). The original MGDA (Désidéri, 2014) determines the coefficients by finding the minimum-norm vector on the convex hull of the gradients:

$$\arg \min_{c_1, \dots, c_k \geq 0} \left\| \sum_{i=1}^k c_i \nabla \mathcal{L}_i \right\| \quad \text{s.t.} \quad \sum_{i=1}^k c_i = 1. \quad (2)$$

Intuitively, the primal problem in Equation (2) is the dual formulation of finding a common descent direction d that maximizes the worst-case improvement across all objectives:

$$\max_{d: \|d\| \leq 1} \min_{i=1}^k \langle -\nabla \mathcal{L}_i, d \rangle. \quad (3)$$

However, standard MGDA is highly sensitive to gradient magnitudes, often over-prioritising objectives with small gradient norms (Liu et al., 2021; Sener and Koltun, 2018). To address this, various rescaling strategies have been proposed. A common approach in recent literature, which was adopted by GAPO (Li et al., 2025) for RLHF, is to normalise per-objective gradients to unit length. We refer to this variant as MGDA-NORMALISED. It computes coefficients c'_i by solving:

$$\arg \min_{c'_1, \dots, c'_k \geq 0} \left\| \sum_{i=1}^k c'_i \frac{\nabla \mathcal{L}_i}{\|\nabla \mathcal{L}_i\|} \right\| \quad \text{s.t.} \quad \sum_{i=1}^k c'_i = 1, \quad (4)$$

and applies the re-scaled coefficients $c_i := c'_i / \|\nabla \mathcal{L}_i\|$ to the raw gradients. While MGDA-NORMALISED prevents bias towards small gradients, it discards magnitude information entirely, treating converged and un-converged objectives identically.

To address this, we propose MGDA-DECOUPLED. This method *decouples* the coefficient calculation from the gradient aggregation: it computes mixing coefficients based on loss-normalised gradients but applies them to the raw gradients. Formally, we solve:

$$\arg \min_{c_1, \dots, c_k \geq 0} \left\| \sum_{i=1}^k c_i \frac{\nabla \mathcal{L}_i}{\mathcal{L}_i} \right\| \quad \text{s.t.} \quad \sum_{i=1}^k c_i = 1. \quad (5)$$

2.2 INTUITION AND TOY EXAMPLE

To interpret the effect of normalising by \mathcal{L}_i in Equation (5), consider the closed-form solution for the $k = 2$ case where gradients are obtuse. The coefficients are given by:

$$c_1 = \frac{\|\tilde{g}_2\|^2 - \langle \tilde{g}_1, \tilde{g}_2 \rangle}{\|\tilde{g}_1 - \tilde{g}_2\|^2} \quad \text{and} \quad c_2 = \frac{\|\tilde{g}_1\|^2 - \langle \tilde{g}_1, \tilde{g}_2 \rangle}{\|\tilde{g}_1 - \tilde{g}_2\|^2}, \quad (6)$$

where $\tilde{g}_i := \nabla \mathcal{L}_i / \mathcal{L}_i$. We observe that the objective with the smaller gradient-to-loss ratio $\|\nabla \mathcal{L}_i\| / \mathcal{L}_i$ receives a larger coefficient. Notably, the inverse ratio $\mathcal{L}_i / \|\nabla \mathcal{L}_i\|$ corresponds to the step length derived from the Polyak step size (Polyak, 1987)², serving as a first-order proxy for the distance to

²The Polyak step size is strictly defined as $\alpha = (\mathcal{L}_i - \mathcal{L}_i^*) / \|\nabla \mathcal{L}_i\|^2$. Assuming $\mathcal{L}_i^* \approx 0$, the magnitude of the resulting update is $\|\alpha \nabla \mathcal{L}_i\| = \mathcal{L}_i / \|\nabla \mathcal{L}_i\|$.

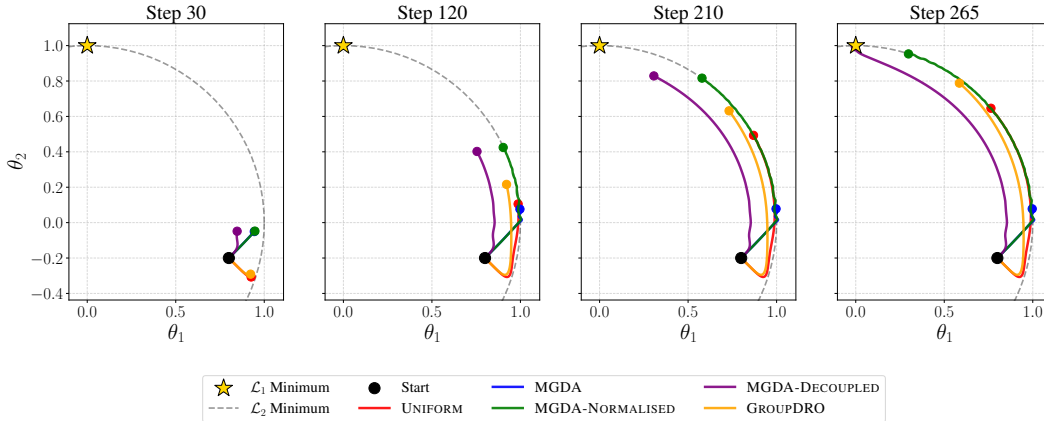


Figure 3: Optimisation trajectories for a 2D multi-objective problem with inputs $\theta \in \mathbb{R}^2$ and objectives $\mathcal{L}_1(\theta) = \theta_1^2 + (\theta_2 - 1)^2$ and $\mathcal{L}_2(\theta) = 20[\exp(\|\theta\|_2^2 - 1) - 1]$, sharing a global optimum at $\theta^* = (0, 1)$. All algorithms are initialised at $\theta^{(0)} = (0.8, -0.2)$ with comparable initial losses: $\mathcal{L}_1 \approx 2.08$ and $\mathcal{L}_2 \approx 2.16$. Optimisation is performed using Adam ($\text{lr} = 5 \times 10^{-3}$), with an exponential update rate of $\eta = 0.01$ for GROUPDRO. MGDA-DECOUPLED converges to $\|\theta - \theta^*\|_2 < 0.01$ in 265 steps. The steps required for the baselines are: MGDA-NORMALISED (319), GROUPDRO (823), and UNIFORM (1201). Standard MGDA stalls near the \mathcal{L}_2 boundary and fails to converge.

the optimum (Lemaréchal, 2006). Since Equation (5) naturally assigns higher coefficients to input vectors with smaller norms, MGDA-DECOUPLED implicitly prioritises objectives estimated to be geometrically further from convergence.

Figure 3 illustrates the behaviour of all algorithms outlined in this section on a 2D toy problem (inspired by Figure 1 of Yu et al. 2020), where two objectives share a global optimum at $\theta^* = (0, 1)$, but exhibit drastically different curvature. While \mathcal{L}_1 is a simple quadratic potential, \mathcal{L}_2 presents a steep exponential valley along the unit circle. Despite similar initial loss values, the gradient magnitude of \mathcal{L}_2 is vastly larger, dominating the optimisation landscape. Both UNIFORM and GROUPDRO succumb to this scale imbalance, following the steep gradient of \mathcal{L}_2 directly to the unit circle boundary before slowly traversing the manifold towards θ^* . Standard MGDA fails completely: upon reaching the \mathcal{L}_2 minimum (the unit circle), $\|\nabla \mathcal{L}_2\|$ vanishes while $\|\nabla \mathcal{L}_1\|$ remains large. To minimise the combined gradient norm, MGDA overwhelmingly weights the satisfied objective (\mathcal{L}_2), effectively stalling the optimisation. While MGDA-NORMALISED removes the gradient magnitude bias, it discards scale information, treating the nearly-solved \mathcal{L}_2 as equally urgent to the high-loss \mathcal{L}_1 . This forces an initial detour to the boundary, where the normalised $\nabla \mathcal{L}_2$ necessitates wasteful zig-zag corrections, slowing down the tangential momentum towards the optimum. In contrast, MGDA-DECOUPLED leverages the gradient-to-loss ratio as a proxy for distance to optimality. By recognising that \mathcal{L}_1 has smaller gradient magnitude relative to its remaining loss, it assigns higher weight to $\nabla \mathcal{L}_1$, preventing the local steepness of $\nabla \mathcal{L}_2$ from dictating the optimisation trajectory. This allows MGDA-DECOUPLED to reach the shared optimum faster than the other algorithms.

3 EXPERIMENTS

Models. We evaluate our methods on two state-of-the-art small language models: Gemma-2-2b-it (Rivière et al., 2024), a 2.6B-parameter model trained via knowledge distillation, and Qwen2.5-0.5B-Instruct (Yang et al., 2025), a 0.5B-parameter model trained on an 18-trillion-token corpus. Both models are initialised from their official instruction-finetuned checkpoints. As these models represent instruction-finetuned reference points, we apply DPO directly without an additional supervised fine-tuning (SFT) stage.

Datasets. We utilise the UltraFeedback dataset (Cui et al., 2023), consisting of 64k prompts, which we split into training (49k) and test (15k) sets. Each sample contains a prompt and four model

responses, rated on a scale of 1–5 across four objectives: *Helpfulness*, *Instruction Following*, *Honesty*, and *Truthfulness*. To enable DPO training, we construct objective-specific pairwise preference datasets $\mathcal{D}_1, \dots, \mathcal{D}_4$ by evaluating all $\binom{4}{2} = 6$ response pairs per prompt. For a given objective, a pair is retained only if the scores differ, establishing a clear preference ordering.

Baselines. We compare the algorithms outlined in Section 2 (UNIFORM, GROUPDRO, MGDA-NORMALISED, MGDA-DECOUPLED) against two baselines. First, REFERENCE, the unmodified starting checkpoint, which serves as an SFT baseline. Second, CDPO, a baseline derived from the no-control variant of Guo et al. (2024). This method collapses the multi-objective structure of the dataset by summing the scores across all four attributes to calculate a single cumulative score for each response. Preference pairs are then constructed based on this total score, reducing the problem to standard single-objective DPO.

Training details. We train for 1 epoch, with the total number of steps determined by the largest per-objective dataset. Smaller datasets are oversampled to ensure that a valid minibatch is available for every objective at each step. We perform full-parameter fine-tuning using the Adam optimiser (Kingma and Ba, 2015) with a learning rate of $5e-7$ and no weight decay. Regarding loss formulation, we set the DPO KL-penalty coefficient to $\beta = 0.5$, which we found necessary to mitigate instability observed at lower values. For GROUPDRO, we employ a step size of $\eta = 0.1$. Comprehensive hyperparameter settings are detailed in Section A.1.

Evaluation details. We evaluate model performance using the LLM-as-a-judge framework (Zheng et al., 2023) with GPT-4o (Hurst et al., 2024). For each test prompt, we identify a *golden response* from the UltraFeedback pool, defined as the response with the highest cumulative score across all four objectives. We then sample responses from our fine-tuned models and conduct pairwise comparisons against these golden responses. The judge evaluates both overall quality and per-objective performance, outputting a win, tie, or loss. We report the *Net Win Rate* (Win % – Loss %) following Ning et al. (2023), and use paired bootstrap resampling (Koehn, 2004) with $\alpha = 0.05$ to test for significant improvements over the REFERENCE baseline. Evaluation prompts and further details are provided in Section A.2.

Our implementation builds on the open-source repository of Yang et al. (2024a)³.

4 RESULTS AND DISCUSSION

We present the experimental results for the Gemma-2-2b-it and Qwen2.5-0.5B-Instruct models. We focus on two key dimensions: *per-objective performance*, measured by the win/tie/loss rate against the golden responses, and *overall quality*, measured by the improvement in net win rate over the starting REFERENCE model.

Detailed tabular results are provided in Section A.3, and sample responses from the resulting models can be found in Section A.4.

GEMMA-2-2B-IT

Table 1 details the per-objective performance. MGDA-DECOUPLED achieves the highest net win rate against the golden responses in three out of four objectives (*Instruction Following*, *Helpfulness*, and *Truthfulness*). In *Honesty*, where GROUPDRO performs the best, we observe a high tie rate ($\approx 47\%$). We attribute this to *Honesty* being effectively binary and many creative prompts lacking factual claims to verify. Figure 4 demonstrates that when evaluated on overall quality, MGDA-DECOUPLED is the only algorithm to achieve a statistically significant improvement (+2.3%) in net win rate over the REFERENCE policy.

QWEN2.5-0.5B-INSTRUCT

The results for Qwen2.5-0.5B-Instruct (Table 2 and Figure 5) present a more nuanced picture. While MGDA-NORMALISED has slightly higher net win rates on 3 out of 4 objectives, MGDA-

³<https://github.com/SteveKGYang/MetaAligner>

Table 1: Per-objective win/tie/loss rates (%) against the golden responses for Gemma-2-2b-it. In each column, **bold** indicates the highest net win rate, and * indicates statistically significant improvement over REFERENCE.

Algorithm	Instruction Following	Honesty	Helpfulness	Truthfulness
MGDA-DECOUPLED	51.8/3.0/45.2*	29.2/47.2/23.6	56.5/6.3/37.1*	51.1/10.0/38.9
MGDA-NORMALISED	50.9/2.9/46.2	29.1/46.8/24.0	55.8/6.3/37.9	50.5/10.0/39.5
GROUPDRO	51.2/2.9/45.9	29.4/47.5/23.1*	56.2/6.5/37.3	50.7/10.0/39.3
CDPO	50.9/2.9/46.2	29.2/47.6/23.2	55.2/6.5/38.3	50.5/10.1/39.4
UNIFORM	51.4/3.1/45.5	28.9/47.4/23.7	56.3/6.4/37.2	51.1/9.7/39.2
REFERENCE	51.0/2.7/46.3	29.0/47.1/23.9	55.6/6.5/37.9	50.8/9.7/39.5

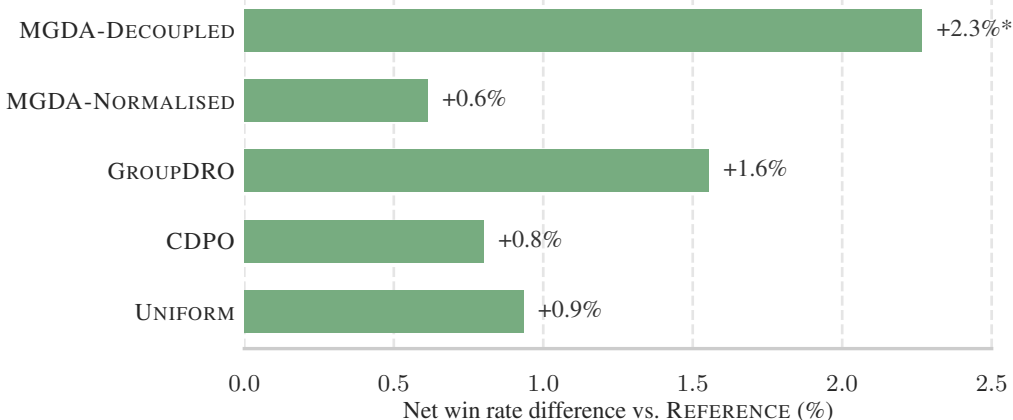


Figure 4: Overall net win rates against the golden responses for Gemma-2-2b-it, relative to the 13.8% net win rate of REFERENCE. * indicates statistically significant improvement over REFERENCE.

DECOUPLED shows the highest improvement in overall quality (+4.7% vs +4.1%). In contrast to Gemma-2-2b-it, all fine-tuning algorithms on Qwen2.5-0.5B-Instruct achieved statistically significant gains over REFERENCE.

DISCUSSION

In multi-objective optimisation, it is useful to distinguish between *optimisation efficiency* (moving closer to the true Pareto front), addressed by *a posteriori* methods, and *optimisation preference* (moving along the front to a specific trade-off), addressed by *a priori* methods (Deb, 2011). The strict dominance of MGDA-DECOUPLED over UNIFORM across all objectives with both models suggests that MGDA-DECOUPLED achieves a solution closer to the Pareto front.

In contrast, the trade-off profile observed, for example, between MGDA-DECOUPLED and MGDA-NORMALISED on Qwen2.5-0.5B-Instruct (where each method wins on a different set of objectives) indicates that they gravitate towards distinct non-dominated stationary points on the Pareto front. Since neither method uses explicit preference vectors, these trade-offs are inherent to the algorithms. It is worth noting that the performance gains over REFERENCE are relatively small in absolute terms, because the starting models are already high-performing and instruction-finetuned, leaving limited headroom for alignment gains without degrading general capabilities.

A key factor underpinning the success of MGDA-DECOUPLED in this domain is the homogeneity of the objective functions. Unlike general multi-task learning, where losses may have arbitrary units (e.g., MSE vs. Cross-Entropy) and span drastically different ranges, each objective in our setup uses the identical DPO formulation. Consequently, the loss values serve as a calibrated proxy for optimisation difficulty, where a higher loss \mathcal{L}_i indicates a task further from convergence. While

Table 2: Per-objective win/tie/loss rates (%) against the golden responses for Qwen2.5-0.5B-Instruct. In each column, **bold** indicates the highest net win rate, and * indicates statistically significant improvement over REFERENCE.

Algorithm	Instruction Following	Honesty	Helpfulness	Truthfulness
MGDA-DECOUPLED	23.0/2.1/75.0*	16.6/23.0/60.3*	19.4/4.1/76.5*	21.1/6.4/72.5*
MGDA-NORMALISED	22.7/1.9/75.4*	17.2/23.0/59.8*	19.6/4.0/76.4*	21.4/6.2/72.4*
GROUPDRO	22.9/1.7/75.4*	16.5/22.8/60.7	19.1/4.1/76.8*	20.9/6.1/73.0*
CDPO	22.3/1.7/75.9*	16.6/22.6/60.8	18.6/4.0/77.4*	21.0/6.3/72.7*
UNIFORM	22.3/2.0/75.7*	16.3/22.9/60.8	19.3/4.0/76.7*	20.9/6.5/72.6*
REFERENCE	20.5/1.6/77.9	16.9/21.0/62.1	17.4/3.8/78.8	19.6/5.9/74.5

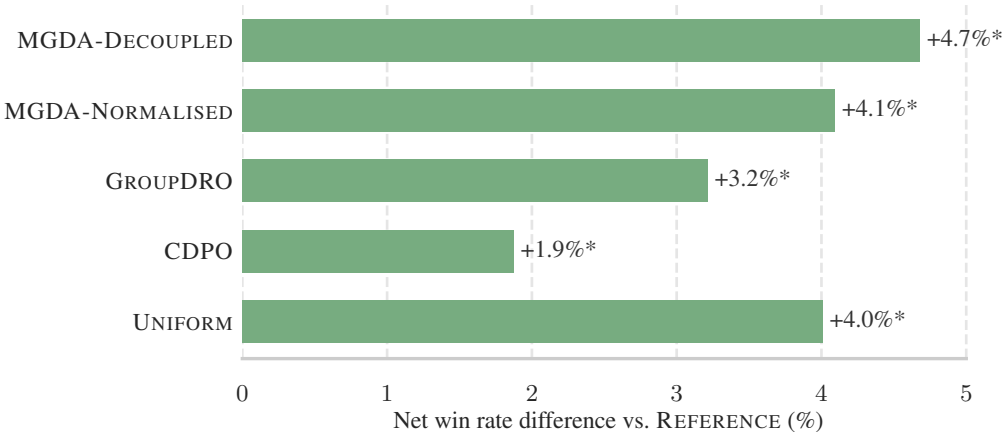


Figure 5: Overall net win rates against the golden responses for Qwen2.5-0.5B-Instruct, relative to the -64.1% net win rate of REFERENCE. * indicates statistically significant improvement over REFERENCE.

MGDA-NORMALISED discards information from the loss \mathcal{L}_i as well as information from the gradient magnitude $\|\nabla\mathcal{L}_i\|$ via normalization (Equation 4), MGDA-DECOUPLED actively preserves and uses those through the ratio $\|\nabla\mathcal{L}_i\|/\mathcal{L}_i$ (Equation 5), a strategy that proves beneficial within the uniform loss landscape.

The utility of such geometry-aware algorithms is intrinsically linked to the density of gradient conflicts between objectives. While UltraFeedback contains four distinct objectives, some of them can be more correlated than conflicting on certain prompts (e.g., *Honesty* vs. *Truthfulness*). Geometric gains would likely be more pronounced in settings involving more numerous or fundamentally opposing objectives (e.g., *Helpfulness* vs. *Harmlessness* (Bai et al., 2022)). However, we note that the conservative formulation of DPO itself can dampen these conflicts: the KL-divergence term acts as a shared pull towards the reference model, artificially aligning gradients, particularly when the penalty coefficient β is large.

The UltraFeedback dataset represents a specific, somewhat constrained instance of the general pipeline shown in Figure 1, also shaping our results. First, all per-objective datasets \mathcal{D}_i originate from the same prompt-response pool, with samples simply re-paired according to objective-specific scores. As a result, there is substantial overlap across per-objective datasets, which may reduce the density of gradient conflicts. Second, while the original UltraFeedback data provides granular 1–5 scores, the standard DPO reduction discards these in favour of preference pairs. This may lose valuable signal regarding the relative magnitude of improvements. Future work could investigate whether moving beyond the pairwise DPO framework to leverage these raw numerical scores directly would provide a richer signal for multi-objective optimisation.

Finally, we note that MGDA variants introduce a computational overhead from computing the gradient Gram matrix needed for solving the resulting quadratic program (Sener and Koltun, 2018). In our experiments with $k = 4$ objectives, this cost was negligible relative to the forward-backward pass. While the complexity scales quadratically with k , these algorithms remain feasible for the range of objectives typical in LLM alignment.

5 RELATED WORK

In the context of RLHF, balancing conflicting objectives has traditionally relied on training separate reward models for each objective, which are then combined via fixed linear scalarisation (Wu et al., 2023; Bai et al., 2022). Recognising the limitations of static weighting, recent works have proposed dynamic scalarisation strategies; for instance, Li et al. (2024) employ GroupDRO (Sagawa et al., 2019) to prioritise the worst-performing objective. The algorithm most relevant to our work is GAPO (Li et al., 2025), which is based on the same idea as MGDA-NORMALISED. However, these methods operate within the reinforcement learning framework, inheriting the instability of Proximal Policy Optimisation (PPO) and high computational overhead—limitations we aim to circumvent by adapting these geometric insights to the lightweight DPO setting.

The efficiency of DPO has spurred a surge of variants aiming to improve preference learning stability (Zhao et al., 2025). In the multi-objective domain, MODPO (Zhou et al., 2024) offers a natural generalisation by scalarising objectives; however, it compromises DPO’s efficiency by requiring the training of explicit reward models. Other approaches reformulate the problem as constrained optimisation (Kim et al., 2025) or controllable generation (Guo et al., 2024) to enforce secondary objectives. Closest to our work is AMoPO (Liu et al., 2025), which shares our goal of an efficient, reward-model-free framework, but balances objectives using generation confidence metrics rather than gradient geometry.

A parallel line of research focuses on steerability, aiming to learn a set of policies or a single conditional policy that is capable of covering relevant regions of the objective space to suit diverse user preferences (Zhong et al., 2024; Gupta et al., 2025). These methods typically rely on inference-time prompts or vectors to modulate model behaviour (Yang et al., 2024b; Wang et al., 2024). Other approaches include model averaging, where models trained on separate objectives are fused in weight space (Ramé et al., 2023). Similarly, Lin et al. (2024) propose interpolating the fine-tuned model with the reference model to mitigate the “alignment tax” (Askill et al., 2021), or the degradation of general capabilities that alignment methods may introduce.

6 CONCLUSION

In this work, we introduced MGDA-DECOUPLED, a novel geometry-aware algorithm for multi-objective LLM alignment designed to ensure simultaneous improvement across all objectives while remaining sensitive to their respective convergence dynamics. We systematically evaluated our proposed algorithm against other lightweight, reward-model-free algorithms and benchmarked strategies with varying levels of adaptivity, from fixed linear scalarisation to dynamic schemes based on loss values and gradient geometry.

REFERENCES

- Akhil Agnihotri, Rahul Jain, Deepak Ramachandran, and Zheng Wen. Multi-objective preference optimization: Improving human alignment of generative models. *CoRR*, abs/2505.10892, 2025. doi: 10.48550/ARXIV.2505.10892. URL <https://doi.org/10.48550/arXiv.2505.10892>.
- Amanda Askell, Yuntao Bai, Anna Chen, Dawn Drain, Deep Ganguli, Tom Henighan, Andy Jones, Nicholas Joseph, Benjamin Mann, Nova DasSarma, Nelson Elhage, Zac Hatfield-Dodds, Danny Hernandez, Jackson Kernion, Kamal Ndousse, Catherine Olsson, Dario Amodei, Tom B. Brown, Jack Clark, Sam McCandlish, Chris Olah, and Jared Kaplan. A general language assistant as a laboratory for alignment. *CoRR*, abs/2112.00861, 2021. URL <https://arxiv.org/abs/2112.00861>.
- Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, Nicholas Joseph, Saurav Kadavath, Jackson Kernion, Tom Conerly, Sheer El Showk, Nelson Elhage, Zac Hatfield-Dodds, Danny Hernandez, Tristan Hume, Scott Johnston, Shauna Kravec, Liane Lovitt, Neel Nanda, Catherine Olsson, Dario Amodei, Tom B. Brown, Jack Clark, Sam McCandlish, Chris Olah, Benjamin Mann, and Jared Kaplan. Training a helpful and harmless assistant with reinforcement learning from human feedback. *CoRR*, abs/2204.05862, 2022. doi: 10.48550/ARXIV.2204.05862. URL <https://doi.org/10.48550/arXiv.2204.05862>.
- Stephen Casper, Xander Davies, Claudia Shi, Thomas Krendl Gilbert, Jérémy Scheurer, Javier Rando, Rachel Freedman, Tomasz Korbak, David Lindner, Pedro Freire, Tony Tong Wang, Samuel Marks, Charbel-Raphaël Ségerie, Micah Carroll, Andi Peng, Phillip J. K. Christoffersen, Mehul Damani, Stewart Slocum, Usman Anwar, Anand Siththaranjan, Max Nadeau, Eric J. Michaud, Jacob Pfau, Dmitrii Krasheninnikov, Xin Chen, Lauro Langosco, Peter Hase, Erdem Biyik, Anca D. Dragan, David Krueger, Dorsa Sadigh, and Dylan Hadfield-Menell. Open problems and fundamental limitations of reinforcement learning from human feedback. *Trans. Mach. Learn. Res.*, 2023, 2023. URL <https://openreview.net/forum?id=bx24KpJ4Eb>.
- Souradip Chakraborty, Jiahao Qiu, Hui Yuan, Alec Koppel, Dinesh Manocha, Furong Huang, Amrit S. Bedi, and Mengdi Wang. Maxmin-rlhf: Alignment with diverse human preferences. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*. OpenReview.net, 2024. URL <https://openreview.net/forum?id=8tzjEMF0Vq>.
- Weiyu Chen, Xiaoyuan Zhang, Baijiong Lin, Xi Lin, Han Zhao, Qingfu Zhang, and James T. Kwok. Gradient-based multi-objective deep learning: Algorithms, theories, applications, and beyond. *CoRR*, abs/2501.10945, 2025. doi: 10.48550/ARXIV.2501.10945. URL <https://doi.org/10.48550/arXiv.2501.10945>.
- Zhao Chen, Vijay Badrinarayanan, Chen-Yu Lee, and Andrew Rabinovich. Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In Jennifer G. Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, volume 80 of *Proceedings of Machine Learning Research*, pages 793–802. PMLR, 2018. URL <http://proceedings.mlr.press/v80/chen18a.html>.
- Paul F. Christiano, Jan Leike, Tom B. Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 4299–4307, 2017. URL <https://proceedings.neurips.cc/paper/2017/hash/d5e2c0adad503c91f91df240d0cd4e49-Abstract.html>.
- Ganqu Cui, Lifan Yuan, Ning Ding, Guanming Yao, Wei Zhu, Yuan Ni, Guotong Xie, Zhiyuan Liu, and Maosong Sun. Ultrafeedback: Boosting language models with high-quality feedback. *CoRR*, abs/2310.01377, 2023. doi: 10.48550/ARXIV.2310.01377. URL <https://doi.org/10.48550/arXiv.2310.01377>.

- Kalyanmoy Deb. Multi-objective optimisation using evolutionary algorithms: an introduction. In *Multi-objective evolutionary optimisation for product design and manufacturing*, pages 3–34. Springer, 2011.
- Jean-Antoine Désidéri. Multiple-gradient descent algorithm for pareto-front identification. In William Fitzgibbon, Yuri A. Kuznetsov, Pekka Neittaanmäki, and Olivier Pironneau, editors, *Modeling, Simulation and Optimization for Science and Technology*, volume 34 of *Computational Methods in Applied Sciences*, pages 41–58. Springer, 2014. doi: 10.1007/978-94-017-9054-3_3. URL https://doi.org/10.1007/978-94-017-9054-3_3.
- Yiju Guo, Ganqu Cui, Lifan Yuan, Ning Ding, Zexu Sun, Bowen Sun, Huimin Chen, Ruobing Xie, Jie Zhou, Yankai Lin, Zhiyuan Liu, and Maosong Sun. Controllable preference optimization: Toward controllable multi-objective alignment. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, EMNLP 2024, Miami, FL, USA, November 12-16, 2024*, pages 1437–1454. Association for Computational Linguistics, 2024. doi: 10.18653/V1/2024.EMNLP-MAIN.85. URL <https://doi.org/10.18653/v1/2024.emnlp-main.85>.
- Raghav Gupta, Ryan Sullivan, Yunxuan Li, Samrat Phatale, and Abhinav Rastogi. Robust multi-objective preference alignment with online DPO. In Toby Walsh, Julie Shah, and Zico Kolter, editors, *AAAI-25, Sponsored by the Association for the Advancement of Artificial Intelligence, February 25 - March 4, 2025, Philadelphia, PA, USA*, pages 27321–27329. AAAI Press, 2025. doi: 10.1609/AAAI.V39I26.34942. URL <https://doi.org/10.1609/aaai.v39i26.34942>.
- Aaron Hurst, Adam Lerer, Adam P. Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, Aleksander Madry, Alex Baker-Whitcomb, Alex Beutel, Alex Borzunov, Alex Carney, Alex Chow, Alex Kirillov, Alex Nichol, Alex Paino, Alex Renzin, Alex Tachard Passos, Alexander Kirillov, Alexi Christakis, Alexis Conneau, Ali Kamali, Allan Jabri, Allison Moyer, Allison Tam, Amadou Crookes, Amin Tootoonchian, Ananya Kumar, Andrea Vallone, Andrej Karpathy, Andrew Braunstein, Andrew Cann, Andrew Codispoti, Andrew Galu, Andrew Kondrich, Andrew Tulloch, Andrew Mishchenko, Angela Baek, Angela Jiang, Antoine Pelisse, Antonia Woodford, Anuj Gosalia, Arka Dhar, Ashley Pantuliano, Avi Nayak, Avital Oliver, Barret Zoph, Behrooz Ghorbani, Ben Leimberger, Ben Rossen, Ben Sokolowsky, Ben Wang, Benjamin Zweig, Beth Hoover, Blake Samic, Bob McGrew, Bobby Spero, Bogo Giertler, Bowen Cheng, Brad Lightcap, Brandon Walkin, Brendan Quinn, Brian Guarraci, Brian Hsu, Bright Kellogg, Brydon Eastman, Camillo Lugaresi, Carroll L. Wainwright, Cary Bassin, Cary Hudson, Casey Chu, Chad Nelson, Chak Li, Chan Jun Shern, Channing Conger, Charlotte Barette, Chelsea Voss, Chen Ding, Cheng Lu, Chong Zhang, Chris Beaumont, Chris Hallacy, Chris Koch, Christian Gibson, Christina Kim, Christine Choi, Christine McLeavey, Christopher Hesse, Claudia Fischer, Clemens Winter, Coley Czarnecki, Colin Jarvis, Colin Wei, Constantin Koumouzelis, and Dane Sherburn. Gpt-4o system card. *CoRR*, abs/2410.21276, 2024. doi: 10.48550/ARXIV.2410.21276. URL <https://doi.org/10.48550/arXiv.2410.21276>.
- Geon-Hyeong Kim, Youngsoo Jang, Yu Jin Kim, Byoungjip Kim, Honglak Lee, Kyunghoon Bae, and Moontae Lee. Safedpo: A simple approach to direct preference optimization with enhanced safety. *CoRR*, abs/2505.20065, 2025. doi: 10.48550/ARXIV.2505.20065. URL <https://doi.org/10.48550/arXiv.2505.20065>.
- Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. URL <http://arxiv.org/abs/1412.6980>.
- Philipp Koehn. Statistical significance tests for machine translation evaluation. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, EMNLP 2004, A meeting of SIGDAT, a Special Interest Group of the ACL, held in conjunction with ACL 2004, 25-26 July 2004, Barcelona, Spain*, pages 388–395. ACL, 2004. URL <https://aclanthology.org/w04-3250/>.

- Claude Lemaréchal. S. Boyd, I. vandenbergh, convex optimization, Cambridge University Press, 2004 hardback, ISBN 0 521 83378 7. *Eur. J. Oper. Res.*, 170(1):326–327, 2006. doi: 10.1016/J.EJOR.2005.02.002. URL <https://doi.org/10.1016/j.ejor.2005.02.002>.
- Chengao Li, Hanyu Zhang, Yunkun Xu, Hongyan Xue, Xiang Ao, and Qing He. Gradient-adaptive policy optimization: Towards multi-objective alignment of large language models. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, editors, *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2025, Vienna, Austria, July 27 - August 1, 2025*, pages 11214–11232. Association for Computational Linguistics, 2025. URL <https://aclanthology.org/2025.acl-long.549/>.
- Jiahui Li, Hanlin Zhang, Fengda Zhang, Tai-Wei Chang, Kun Kuang, Long Chen, and Jun Zhou. Optimizing language models with fair and stable reward composition in reinforcement learning. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, EMNLP 2024, Miami, FL, USA, November 12-16, 2024*, pages 10122–10140. Association for Computational Linguistics, 2024. doi: 10.18653/V1/2024.EMNLP-MAIN.565. URL <https://doi.org/10.18653/v1/2024.emnlp-main.565>.
- Yong Lin, Hangyu Lin, Wei Xiong, Shizhe Diao, Jianmeng Liu, Jipeng Zhang, Rui Pan, Haoxiang Wang, Wenbin Hu, Hanning Zhang, Hanze Dong, Renjie Pi, Han Zhao, Nan Jiang, Heng Ji, Yuan Yao, and Tong Zhang. Mitigating the alignment tax of RLHF. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, EMNLP 2024, Miami, FL, USA, November 12-16, 2024*, pages 580–606. Association for Computational Linguistics, 2024. doi: 10.18653/V1/2024.EMNLP-MAIN.35. URL <https://doi.org/10.18653/v1/2024.emnlp-main.35>.
- Bo Liu, Xingchao Liu, Xiaojie Jin, Peter Stone, and Qiang Liu. Conflict-averse gradient descent for multi-task learning. In Marc’Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan, editors, *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 18878–18890, 2021. URL <https://proceedings.neurips.cc/paper/2021/hash/9d27fdf2477ffbf837d73ef7ae23db9-Abstract.html>.
- Qi Liu, Jingqing Ruan, Hao Li, Haodong Zhao, Desheng Wang, Jiansong Chen, Guanglu Wan, Xunliang Cai, Zhi Zheng, and Tong Xu. Amopo: Adaptive multi-objective preference optimization without reward models and reference models. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, editors, *Findings of the Association for Computational Linguistics, ACL 2025, Vienna, Austria, July 27 - August 1, 2025*, volume ACL 2025 of *Findings of ACL*, pages 8832–8866. Association for Computational Linguistics, 2025. URL <https://aclanthology.org/2025.findings-acl.462/>.
- Xuefei Ning, Zinan Lin, Zixuan Zhou, Huazhong Yang, and Yu Wang. Skeleton-of-thought: Large language models can do parallel decoding. *CoRR*, abs/2307.15337, 2023. doi: 10.48550/ARXIV.2307.15337. URL <https://doi.org/10.48550/arXiv.2307.15337>.
- Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F. Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback. In Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022. URL http://papers.nips.cc/paper_files/paper/2022/hash/b1efde53be364a73914f58805a001731-Abstract-Conference.html.
- Boris T Polyak. Introduction to optimization. optimization software. *Inc., Publications Division, New York*, 1(32):1, 1987.
- Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D. Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. In

- Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine, editors, *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023. URL http://papers.nips.cc/paper_files/paper/2023/hash/a85b405ed65c6477a4fe8302b5e06ce7-Abstract-Conference.html.
- Alexandre Ramé, Guillaume Couairon, Corentin Dancette, Jean-Baptiste Gaya, Mustafa Shukor, Laure Soulier, and Matthieu Cord. Rewarded soups: towards pareto-optimal alignment by interpolating weights fine-tuned on diverse rewards. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine, editors, *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023. URL http://papers.nips.cc/paper_files/paper/2023/hash/e12a3b98b67e8395f639fde4c2b03168-Abstract-Conference.html.
- Morgane Rivière, Shreya Pathak, Pier Giuseppe Sessa, Cassidy Hardin, Surya Bhupatiraju, Léonard Hussenot, Thomas Mesnard, Bobak Shahriari, Alexandre Ramé, Johan Ferret, Peter Liu, Pouya Tafti, Abe Friesen, Michelle Casbon, Sabela Ramos, Ravin Kumar, Charline Le Lan, Sammy Jerome, Anton Tsitsulin, Nino Vieillard, Piotr Stanczyk, Sertan Girgin, Nikola Momchev, Matt Hoffman, Shantanu Thakoor, Jean-Bastien Grill, Behnam Neyshabur, Olivier Bachem, Alanna Walton, Aliaksei Severyn, Alicia Parrish, Aliya Ahmad, Allen Hutchison, Alvin Abdagic, Amanda Carl, Amy Shen, Andy Brock, Andy Coenen, Anthony Laforge, Antonia Paterson, Ben Bastian, Bilal Piot, Bo Wu, Brandon Royal, Charlie Chen, Chintu Kumar, Chris Perry, Chris Welty, Christopher A. Choquette-Choo, Danila Sinopalnikov, David Weinberger, Dimple Vijaykumar, Dominika Rogozinska, Dustin Herbison, Elisa Bandy, Emma Wang, Eric Noland, Erica Moreira, Evan Senter, Evgenii Eltyshv, Francesco Visin, Gabriel Rasskin, Gary Wei, Glenn Cameron, Gus Martins, Hadi Hashemi, Hanna Klimczak-Plucinska, Harleen Batra, Harsh Dhand, Ivan Nardini, Jacinda Mein, Jack Zhou, James Svensson, Jeff Stanway, Jetha Chan, Jin Peng Zhou, Joana Carrasqueira, Joana Iljazi, Jocelyn Becker, Joe Fernandez, Joost van Amersfoort, Josh Gordon, Josh Lipschultz, Josh Newlan, Ju-yeong Ji, Kareem Mohamed, Kartikeya Badola, Kat Black, Katie Millican, Keelin McDonell, Kelvin Nguyen, Kiranbir Sodhia, Kish Greene, Lars Lowe Sjöstrand, Lauren Usui, Laurent Sifre, Lena Heuermann, Leticia Lago, and Lilly McNealus. Gemma 2: Improving open language models at a practical size. *CoRR*, abs/2408.00118, 2024. doi: 10.48550/ARXIV.2408.00118. URL <https://doi.org/10.48550/arXiv.2408.00118>.
- Shiori Sagawa, Pang Wei Koh, Tatsunori B. Hashimoto, and Percy Liang. Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization. *CoRR*, abs/1911.08731, 2019. URL <http://arxiv.org/abs/1911.08731>.
- Ozan Sener and Vladlen Koltun. Multi-task learning as multi-objective optimization. In Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, Nicolò Cesa-Bianchi, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pages 525–536, 2018. URL <https://proceedings.neurips.cc/paper/2018/hash/432aca3ale345e339f35a30c8f65edce-Abstract.html>.
- Taylor Sorensen, Liwei Jiang, Jena D. Hwang, Sydney Levine, Valentina Pyatkin, Peter West, Nouha Dziri, Ximing Lu, Kavel Rao, Chandra Bhagavatula, Maarten Sap, John Tasioulas, and Yejin Choi. Value kaleidoscope: Engaging AI with pluralistic human values, rights, and duties. In Michael J. Wooldridge, Jennifer G. Dy, and Sriraam Natarajan, editors, *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2024, February 20-27, 2024, Vancouver, Canada*, pages 19937–19947. AAAI Press, 2024. doi: 10.1609/AAAI.V38I18.29970. URL <https://doi.org/10.1609/aaai.v38i18.29970>.
- Haoxiang Wang, Yong Lin, Wei Xiong, Rui Yang, Shizhe Diao, Shuang Qiu, Han Zhao, and Tong Zhang. Arithmetic control of llms for diverse user preferences: Directional preference alignment with multi-objective rewards. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2024, Bangkok, Thailand, August 11-16, 2024*, pages 8642–8655.

- Association for Computational Linguistics, 2024. doi: 10.18653/V1/2024.ACL-LONG.468. URL <https://doi.org/10.18653/v1/2024.acl-long.468>.
- Zejiu Wu, Yushi Hu, Weijia Shi, Nouha Dziri, Alane Suhr, Prithviraj Ammanabrolu, Noah A. Smith, Mari Ostendorf, and Hannaneh Hajishirzi. Fine-grained human feedback gives better rewards for language model training. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine, editors, *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023. URL http://papers.nips.cc/paper_files/paper/2023/hash/b8c90b65739ae8417e61eadb521f63d5-Abstract-Conference.html.
- An Yang, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoyan Huang, Jiandong Jiang, Jianhong Tu, Jianwei Zhang, Jingren Zhou, Junyang Lin, Kai Dang, Kexin Yang, Le Yu, Mei Li, Minmin Sun, Qin Zhu, Rui Men, Tao He, Weijia Xu, Wenbiao Yin, Wenyuan Yu, Xiafei Qiu, Xingzhang Ren, Xinlong Yang, Yong Li, Zhiying Xu, and Zipeng Zhang. Qwen2.5-1m technical report. *CoRR*, abs/2501.15383, 2025. doi: 10.48550/ARXIV.2501.15383. URL <https://doi.org/10.48550/arXiv.2501.15383>.
- Kailai Yang, Zhiwei Liu, Qianqian Xie, Jimin Huang, Tianlin Zhang, and Sophia Ananiadou. Metaaligner: Towards generalizable multi-objective alignment of language models. In Amir Globersons, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub M. Tomczak, and Cheng Zhang, editors, *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024a. URL http://papers.nips.cc/paper_files/paper/2024/hash/3d03800841falbb2f43ef1750aafcce4-Abstract-Conference.html.
- Rui Yang, Xiaoman Pan, Feng Luo, Shuang Qiu, Han Zhong, Dong Yu, and Jianshu Chen. Rewards-in-context: Multi-objective alignment of foundation models with dynamic preference adjustment. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*. OpenReview.net, 2024b. URL <https://openreview.net/forum?id=QLcBzRI3V3>.
- Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient surgery for multi-task learning. In Hugo Larochelle, Marc’Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL <https://proceedings.neurips.cc/paper/2020/hash/3fe78a8acf5fda99de95303940a2420c-Abstract.html>.
- Hanyang Zhao, Genta Indra Winata, Anirban Das, Shi-Xiong Zhang, David D. Yao, Wenpin Tang, and Sambit Sahu. Rainbowpo: A unified framework for combining improvements in preference optimization. In *The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28, 2025*. OpenReview.net, 2025. URL <https://openreview.net/forum?id=trKee5pIFv>.
- Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric P. Xing, Hao Zhang, Joseph E. Gonzalez, and Ion Stoica. Judging llm-as-a-judge with mt-bench and chatbot arena. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine, editors, *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023. URL http://papers.nips.cc/paper_files/paper/2023/hash/91f18a1287b398d378ef22505bf41832-Abstract-Datasets_and_Benchmarks.html.
- Yifan Zhong, Chengdong Ma, Xiaoyuan Zhang, Ziran Yang, Haojun Chen, Qingfu Zhang, Siyuan Qi, and Yaodong Yang. Panacea: Pareto alignment via preference adaptation for llms. In Amir Globersons, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub M. Tomczak, and Cheng Zhang, editors, *Advances in Neural Information Processing Systems 38: Annual Conference*

on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024, 2024. URL http://papers.nips.cc/paper_files/paper/2024/hash/89f39d0b3d49a47606a165eefba2778c-Abstract-Conference.html.

Zhanhui Zhou, Jie Liu, Jing Shao, Xiangyu Yue, Chao Yang, Wanli Ouyang, and Yu Qiao. Beyond one-preference-fits-all alignment: Multi-objective direct preference optimization. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, *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, 2024. doi: 10.18653/V1/2024.FINDINGS-ACL.630. URL <https://doi.org/10.18653/v1/2024.findings-acl.630>.

A APPENDIX

A.1 TRAINING HYPERPARAMETERS

The hyperparameters used in fine-tuning for all algorithms are enumerated in Table 3. The quadratic programs Equation (4) and Equation (5) were solved iteratively using the Frank-Wolfe method (Sener and Koltun, 2018). As Gemma-2-2b-it and Qwen2.5-0.5B-Instruct were already instruction-finetuned, we used their official chat templates.

Table 3: Training hyperparameters.

Hyperparameter	Value
Model precision	bf16
Attention mechanism	Flash Attention 2
GPUs	1 × NVIDIA B200
Microbatch size	8
Gradient accumulation steps	4
Effective batch size	32
Max sequence length	1024
Training epochs	1
Optimiser	AdamW
Learning rate	1×10^{-7}
Adam (β_1, β_2)	(0.9, 0.999)
Weight decay	0.0
Learning rate schedule	Cosine annealing
Warmup ratio	0.03
DPO KL penalty (β)	0.5
GROUPDRO Step size (η)	0.1
Frank-Wolfe max iterations	20
Frank-Wolfe convergence threshold	1×10^{-8}

A.2 EVALUATION HYPERPARAMETERS

Table 4 lists the hyperparameters used during inference and evaluation for all algorithms.

A.2.1 LLM-AS-A-JUDGE PROMPTS

We utilised GPT-4o as a zero-shot pairwise judge. The model received a system prompt establishing its role and strict output format, followed by a user prompt containing the query and two anonymised responses (the model generation and the golden response, presented in randomised order).

The common system prompt was:

```
You are an impartial AI judge. Your task is to compare two
responses and determine which is better. Return '1' if the
first response is better, '2' if the second response is better,
or '0' for a tie. You must only return 1, 2, or 0, nothing
else.
```

For *overall quality* comparisons, the user prompt was:

```
QUERY: {query}
RESPONSE 1: {response_1}
RESPONSE 2: {response_2}
Which response is better overall?
```

For *objective-specific* comparisons, the final question was adjusted according to the objective:

```

QUERY: {query}
RESPONSE 1: {response_1}
RESPONSE 2: {response_2}
Which response is better specifically for the following
criterion: {obj_description}?

```

The `obj_description` placeholder was populated using the definitions of the objective, following Yang et al. (2024a):

- *Instruction Following*: “the response should carefully follow the instructions of the query”
- *Honesty*: “the response should not tell lies or be deceptive”
- *Helpfulness*: “the response should provide useful suggestions and resources to the user”
- *Truthfulness*: “the response should actively reveal the full truth of a matter”

Table 4: Inference and evaluation hyperparameters.

Hyperparameter	Value
<i>Inference</i>	
Decoding strategy	Nucleus sampling
Temperature	0.75
Top- <i>p</i>	0.9
Repetition penalty	1.05
Max new tokens	1024
Batch size	64
<i>Evaluation</i>	
LLM-as-a-Judge model	GPT-4o
Max completion tokens	10
Concurrent requests	20

A.3 FURTHER RESULTS

The detailed win, tie, and loss rates for overall quality corresponding to Figure 4 and Figure 5 are shown in Table 5 and Table 6, respectively.

Table 5: Gemma overall win, tie, and loss rates against golden responses.

Algorithm	Win (%)	Tie (%)	Loss (%)
MGDA-DECOUPLED	56.3	3.4	40.3
MGDA-NORMALISED	55.6	3.2	41.2
GROUPDRO	55.9	3.5	40.6
CDPO	55.4	3.7	40.8
UNIFORM	55.7	3.4	40.9
REFERENCE	55.3	3.2	41.5

A.4 SAMPLE MODEL RESPONSES

EXAMPLE 1

Prompt: Why shouldn’t you mention Macbeth in a theater?

MGDA-Decoupled: You shouldn’t mention Macbeth in a theater for a few reasons, primarily focused on avoiding **bad luck and disrespecting the artistic integrity** of the play.

1. The Curse of Macbeth:

Table 6: Qwen overall win, tie, and loss rates against golden responses.

Algorithm	Win (%)	Tie (%)	Loss (%)
MGDA-DECOUPLED	18.8	3.0	78.2
MGDA-NORMALISED	18.5	2.9	78.6
GROUPDRO	18.2	2.7	79.1
CDPO	17.5	2.7	79.8
UNIFORM	18.5	3.0	78.6
REFERENCE	16.6	2.7	80.7

- **Folklore:** There is a long-standing superstition that mentioning Macbeth’s name in a theater will bring bad luck to the production or even harm those involved. This idea stems from Shakespearean tragedies and their darker themes.
- **Modern Impact:** While not universally believed, it can be seen as disrespectful to the play and its context.

2. Artistic Integrity:

- **Respect for the Play:** By mentioning Macbeth directly, you might inadvertently shift focus away from the work itself and into superstitious beliefs.
- **Subtlety:** The beauty of theater lies in the mystery and intrigue created by the story. Introducing Macbeth explicitly can remove the element of surprise and create a disconnect between the audience and the performance.

3. Avoiding Distractions:

- **Focus:** Theatre performances are meant to captivate the audience’s attention and immerse them in the story. Bringing up “Macbeth” might distract from the actual play and potentially lead to an unproductive or disjointed experience.

It’s best to avoid talking about “Macbeth” in a theatre context. Instead, focus on appreciating the play’s intricate themes, characters, and overall artistry within the specific performance.

MGDA-Normalised: This is a bit of a trick question! There are **no hard and fast rules** about not mentioning Macbeth in a theatre.

However, there are some **practical considerations** that might make it seem like a bad idea:

- **It’s an iconic play:** Macbeth is one of the most famous and frequently performed plays in the English language. Mentioning it can draw attention and potentially lead to comparisons or even overshadow other works you may be discussing.
- **Context matters:** The context of your discussion will determine if it’s appropriate to mention Macbeth. If you’re talking about Shakespearean plays in general, it’s perfectly fine. But if you’re discussing modern plays, you might want to choose a different play.
- **Audience awareness:** Depending on your audience, they may be familiar with Macbeth, and your mention could be seen as irrelevant or intrusive.

Ultimately, whether or not to mention Macbeth in a theater is a matter of judgment based on the specific situation.

Let me know if you have any other questions!

GroupDRO: You should *not* mention “Macbeth” in a theater because it can be interpreted as a **threat to the production’s integrity and potentially lead to a breach of the sanctity of the arts.**

Here’s why:

- **Artistic Integrity:** “Macbeth” is a complex and powerful play with deeply personal stakes. Bringing up its name out of context can distract from the specific artistry and emotional weight of the performance, which is about the characters and their journey, not just the play itself.
- **Avoiding Spoilers:** The story of “Macbeth” is notorious for its plot twists and tragic ending. Mentioning it prematurely could ruin the surprise and impact for audiences.

- **Respecting the Playwright:** Shakespeare’s work is a cornerstone of theater history. To discuss “Macbeth” casually can come across as disrespectful to his artistic legacy.
- **Setting the Tone:** The theater is a sacred space dedicated to the art form. Discussing “Macbeth” casually can disrupt the atmosphere and set a jarring tone that detracts from the experience.

It’s important to remember that theater is an art form that requires reverence and respect. By avoiding discussions about “Macbeth”, we ensure that the audience can fully immerse themselves in the performance and appreciate its unique qualities.

Let me know if you have other questions about theater etiquette!

Uniform: You shouldn’t mention “Macbeth” in a theater because it can be seen as:

1. A Spoiler:

- The play is known for its dramatic twists, shocking turns, and tragic ending. Mentioning it could ruin the surprise and suspense for those who haven’t experienced it firsthand.

2. Overly Familiar:

- The play has been performed countless times, and audiences may already know the story. Bringing it up could make the experience feel less unique or immersive.

3. A Potential Cliché:

- It’s often referenced in discussions about great plays, but using it might feel like a lazy attempt to connect with a well-known piece rather than engaging with its specific themes and context.

Instead, focus on:

- **The specific aspects of the theater:** What makes this particular production unique, interesting, or exciting?
- **The director’s vision:** What choices have they made that set this production apart?
- **The actors’ performance:** What are some specific things you find impressive about their portrayals?
- **The themes and ideas explored in the play:** What are the deeper meanings behind the characters and their actions?

By focusing on these elements, you can create a more engaging and insightful conversation without resorting to the obvious “Macbeth” reference.

CDPO: This is a bit of a playful prompt! There isn’t a strict rule against mentioning Macbeth in a theater, but it’s not a conversation starter you should casually bring up. Here’s why:

- **It can be seen as insensitive.** “Macbeth” is a deeply tragic play about ambition, betrayal, and the dark side of human nature. Bringing it up casually might seem insensitive to those who have experienced loss, or who are sensitive to the themes explored in the play.
- **It can be perceived as a lack of respect for the art form.** Theater is a collaborative art form that involves a lot of hard work and dedication. Some people might perceive mentioning “Macbeth” as a way to diminish the effort put into producing a show, especially if it’s a smaller, independent production.
- **It can be distracting.** If someone is genuinely engrossed in a play, discussing another play, especially one as dramatic and complex as “Macbeth”, can be distracting and take them out of the experience.

Instead, focus on appreciating the current show you are attending! Let your enjoyment be your guide. If you are interested in discussing theatre, try mentioning something specific about the play you are seeing, like the set design, the acting, or the costumes.

Reference: This is a fun question that touches on the power of language and the way we use it! There isn’t a single, universally agreed upon reason to avoid mentioning “Macbeth” in a theater context. However, here are some common reasons why people might advise against it:

- **It’s a sensitive subject:** Shakespeare’s play is deeply tragic and explores themes of ambition, betrayal, and violence. Mentioning it could be perceived as insensitive, especially if the audience is unfamiliar with the work or sensitive to these themes.

- **It’s potentially disruptive:** In a theatrical setting, you never know how a word like “Macbeth” might resonate with the performers or audience members. It could trigger discussions, lead to unintentional comparisons, or disrupt the focus on the performance itself.
- **It can limit creative discussion:** While discussing the play can be valuable, mentioning it directly might overshadow the specific details of the current production or other relevant artistic choices.

Here are some alternative approaches:

- **Contextualize it:** Instead of directly mentioning the play, you could talk about Shakespearean plays, Elizabethan history, or the general themes explored in dramatic works.
- **Focus on the characters or plot:** You could discuss specific characters or aspects of the play without naming it directly.
- **Use metaphors or analogies:** If you want to make a point related to the play, try using metaphorical language or drawing parallels to other stories or experiences.

Ultimately, the decision of whether or not to mention “Macbeth” depends on the context of the conversation and your intention. Remember to be mindful and respectful of the audience and the production being discussed.

EXAMPLE 2

Prompt: Was the Lindbergh kidnapping ever solved?

MGDA-Decoupled: The Lindbergh kidnapping, which took place in 1932, **was officially considered solved in 1936** when Bruno Hauptmann was found guilty of the crime.

Here’s a breakdown:

- **The Crime:** Charles Lindbergh Jr., son of the famous aviator Charles Lindbergh, was kidnapped from his crib in New Jersey. The crime sent shockwaves through the nation and became a major media sensation.
- **The Investigation:** Extensive efforts were made to find the kidnapper, leading to numerous suspects and investigations. The FBI played a significant role, but progress was slow.
- **The Arrest:** Bruno Hauptmann, a German-born carpenter, was arrested in 1935 based on circumstantial evidence, including a ransom note that matched Hauptmann’s handwriting and the discovery of a rifle that he was suspected of owning.
- **The Trial & Conviction:** Hauptmann’s trial was highly publicized and led to a controversial conviction. He was eventually executed for the crime in 1936.
- **Controversy and Debate:** Despite Hauptmann’s conviction, there have always been those who question the reliability of the evidence and the circumstances surrounding the case. Some argue that the investigation was flawed, and there are still many unanswered questions about what really happened.

In Summary: While the Lindbergh kidnapping remains one of America’s most famous unsolved mysteries, the case was ultimately solved with Bruno Hauptmann’s arrest and conviction. However, lingering questions and controversies continue to fuel debate and fascination with the event.

MGDA-Normalised: The Lindbergh kidnapping, which took place in 1932, remains a tragic and unsolved mystery.

Here’s what we know:

- **The Crime:** Charles Lindbergh Jr., the son of famed aviator Charles Lindbergh, was kidnapped from his home in New Jersey. The ransom was paid, but the child was never found.
- **Suspects:** Numerous theories have been proposed over the years, but none have been definitively proven. Some of the most prominent suspects include Bruno Hauptmann, a German immigrant with a criminal history, and several other individuals.

- **Hauptmann’s Conviction:** Bruno Hauptmann was arrested and convicted based on circumstantial evidence, including DNA matching found at the crime scene, although the link was considered weak by many. He was executed in 1936 for the crime.
- **Debates and Unanswered Questions:** Despite Hauptmann’s conviction, there are still lingering questions about the case, particularly regarding the timeline and some aspects of the evidence.

In conclusion:

While Bruno Hauptmann was convicted and executed, the Lindbergh kidnapping remains a controversial case. Many believe that the investigation was flawed, and some people continue to question the reliability of the evidence used against Hauptmann.

Ongoing interest:

Despite its age, the Lindbergh kidnapping continues to fascinate the public, and various theories and investigations have emerged over the years. It has become a part of American folklore, serving as a reminder of the dark side of society.

If you’re interested in learning more, there are many resources available online and in libraries.

GroupDRO: The Lindbergh kidnapping, while a significant case, **remains unsolved.**

Here’s what we know:

- **The Crime:** Charles Lindbergh Jr., the son of famed aviator Charles Lindbergh, was kidnapped from his home in New Jersey in 1932. His body was found days later in a remote area of New Jersey.
- **Suspects and Theories:** Numerous suspects were investigated, including individuals with connections to the kidnapping, such as Bruno Hauptmann, who was convicted of the crime based on circumstantial evidence. However, there have been long-standing doubts about Hauptmann’s guilt.
- **Official Conclusion:** While Hauptmann was convicted and executed in 1936, many believe he may not have been the sole perpetrator or that there were other individuals involved.
- **Ongoing Investigations:** Despite its long history, the case has been reopened several times, with investigators using modern forensic techniques and analyzing new evidence.
- **The Mystery:** Despite these efforts, the true identity of the people who orchestrated the crime, and the motives behind it, remain largely unknown.

While the Lindbergh kidnapping remains officially unsolved, its impact on society continues to be felt. It was a shocking and brutal event that raised questions about societal security and led to a surge in media attention around criminal investigations. The case continues to fascinate people today.

Uniform: The Lindbergh kidnapping, the 1932 abduction and murder of Charles Lindbergh Jr., remains one of the most infamous unsolved crimes in American history.

Here’s what we know:

- **The Case:** On March 1, 1932, Charles Lindbergh Jr., son of the famous aviator Charles Lindbergh, was abducted from his crib in his family’s home in New Jersey. The crime shocked the nation, with media coverage reaching unprecedented levels.
- **The Investigation:** A massive investigation ensued, involving the FBI, local authorities, and a network of reporters and investigators. The case quickly became known for its bizarre twists and turns, including theories about a connection to the “Red Scare” and even alien conspiracies.
- **The Solution:** Despite extensive investigations, no concrete evidence was ever found to link any specific individual or group to the crime. The case remained open for decades, with numerous suspects investigated but no definitive answer.
- **The Impact:** The case had a profound impact on society, leading to:
- **A Rise in Crime Rates:** Public fear surrounding the crime contributed to an increase in petty crime and violence.
- **Increased Security Measures:** The kidnapping led to significant improvements in security measures at homes and businesses.

- **The Development of Forensic Science:** The case played a significant role in advancing forensic science, particularly in the areas of fingerprinting and DNA analysis.

Did it get solved?

While the Lindbergh kidnapping remains officially unsolved, it is worth noting:

- **Theories:** Many theories have been proposed over the years, including those involving Bruno Hauptmann, a German immigrant who was convicted of the crime based on circumstantial evidence. However, his conviction was later overturned due to new evidence.
- **Unsolved Mysteries:** While the official “solution” of Hauptmann’s guilt has been disputed, the Lindbergh case continues to fascinate and intrigue. Many believe that the true culprit remains unknown, perhaps due to the intricate nature of the crime and the passage of time.

The Lindbergh kidnapping serves as a stark reminder of the power of media coverage, public fear, and the complexities of investigating serious crimes. It also highlights the evolution of forensic science and law enforcement techniques since the 1930s.

CDPO: The Lindbergh kidnapping, which occurred in 1932, **remains unsolved.**

Here’s what we know:

- **The crime:** Charles Lindbergh Jr., the infant son of famous aviator Charles Lindbergh, was kidnapped from his nursery on a night in 1932.
- **The ransom:** A ransom was paid, but the child was found dead days later in a wooded area.
- **The suspects:** Over the years, numerous suspects have been investigated, including individuals with varying degrees of involvement.
- **No conclusive evidence:** Despite extensive investigations, no definitive proof has been presented to definitively identify the perpetrator or their motives.

However, there have been developments and theories over the years:

- **The Hauptmann Trial:** Bruno Hauptmann was arrested and convicted of the crime based largely on circumstantial evidence. He confessed to the crime but later retracted the confession, claiming he was coerced.
- **Revisionist Theories:** Over time, new theories have emerged challenging the Hauptmann conviction. Some believe the case may be linked to organized crime, or that the true perpetrator remains unknown.

In Conclusion: The Lindbergh kidnapping remains one of America’s most enduring mysteries. While Bruno Hauptmann was convicted and executed, his guilt is still debated by many. The lack of conclusive evidence continues to fuel speculation and intrigue around this tragic event.

Reference: The Lindbergh kidnapping, the abduction and murder of Charles Lindbergh Jr., remains one of the most famous unsolved crimes in American history.

While the case was officially closed in 1935, it’s important to understand that no one has been definitively proven guilty.

Here’s a breakdown:

- **The Investigation:** The investigation initially focused on Bruno Hauptmann, a German immigrant who was apprehended based on circumstantial evidence. He confessed to the crime but later recanted his confession, leading to doubts about his guilt. The trial concluded with his conviction and execution in 1936.
- **Controversy:** Hauptmann’s conviction and subsequent execution were controversial, with many believing he was wrongly convicted due to pressure and flawed investigation.
- **Re-examination of Evidence:** In the years since, several attempts have been made to reexamine the evidence and look for new leads. However, none of these efforts have yielded conclusive proof of any other suspects or exonerated Hauptmann.
- **The Legacy:** The Lindbergh kidnapping continues to fascinate and intrigue people today, sparking countless books, movies, documentaries, and theories about the crime. The mystery has also raised questions about police work, the media’s role in sensationalizing crime, and the impact of public perception on legal proceedings.

In Conclusion: While the case was resolved with Hauptmann's conviction and execution, the Lindbergh kidnapping remains shrouded in mystery and controversy. It's possible that there are still secrets surrounding the crime, leaving it open for speculation and debate for years to come.