DDPO: Diversity-Driven Preference Optimization for Machine Translation Enhancing Robustness and Generalization

Donald Martin * and Blake Bowman

Posted Date: 30 December 2025

doi: 10.20944/preprints202512.2563.v1

Keywords: machine translation; large language models; preference optimization; diversity

 doi:10.20944/preprints202512.2563.v1

Article

DDPO: Diversity-Driven Preference Optimization for Machine Translation Enhancing Robustness and Generalization

Donald Martin * and Blake Bowman

Jefferson University, USA
* Correspondence: rbrown53@my.ncu.edu.jm

Abstract

Large Language Models (LLMs) have advanced Machine Translation (MT), but fine-tuning often struggles with data scarcity, especially in low-resource settings. Preference Optimization (PO) methods, like DPO and CRPO, learn from preference data. However, existing PO approaches primarily select "best" candidates based on reward and confidence, often overlooking diversity among candidate translations. This can lead to models learning similar error types, limiting generalization and robustness. To address this, we propose Diversity-Driven Preference Optimization (DDPO), a novel method integrating diversity into preference sample selection. DDPO selects the dispreferred translation (y_l) not only based on lower reward and confidence but crucially by maximizing its semantic or syntactic diversity from the preferred translation (y_w). This provides richer, more informative learning signals, compelling the model to learn robust preference boundaries. Experiments on ALMA-7B and NLLB-1.3B, using FLORES-200 for preference construction and evaluating on WMT21/22 test sets across 10 translation directions, consistently demonstrate DDPO significantly outperforms state-of-the-art baselines, including CRPO, across all automated metrics (KIWI22, COMET22, XCOMET, KIWI-XL). This establishes DDPO as a more effective and robust approach for fine-tuning MT models, achieving superior translation quality and enhanced generalization with modest computational overhead.

Keywords: machine translation; large language models; preference optimization; diversity

1. Introduction

Machine Translation (MT) stands as a cornerstone task in Natural Language Processing (NLP), and its continuous improvement is pivotal for fostering global communication and knowledge exchange [1]. Recent advancements in Large Language Models (LLMs) have significantly propelled MT capabilities, enabling more fluent and accurate translations across diverse language pairs [2]. However, fine-tuning these powerful LLMs typically demands extensive quantities of high-quality parallel data, which remains a scarce resource for many low-resource languages.

To circumvent the limitations of data scarcity, Preference Optimization (PO) [3], particularly Direct Preference Optimization (DPO) [3], has emerged as a compelling paradigm. PO methods allow models to learn from human or automated preferences over multiple candidate translations, thereby enhancing translation quality without requiring explicit reward model training. Existing preference optimization techniques, such as Reward-based Supervised Optimization (RSO) [4], Reward-Score DPO (RS-DPO) [5], and the recently proposed Confidence-Reward Driven Preference Optimization (CRPO) [6], primarily concentrate on efficiently identifying "valuable" samples. These methods typically prioritize candidates with the highest reward scores or those combining high reward with strong generation confidence, as seen in CRPO's innovative approach to leverage both metrics for more precise sample identification and improved data efficiency.

 doi:10.20944/preprints202512.2563.v1

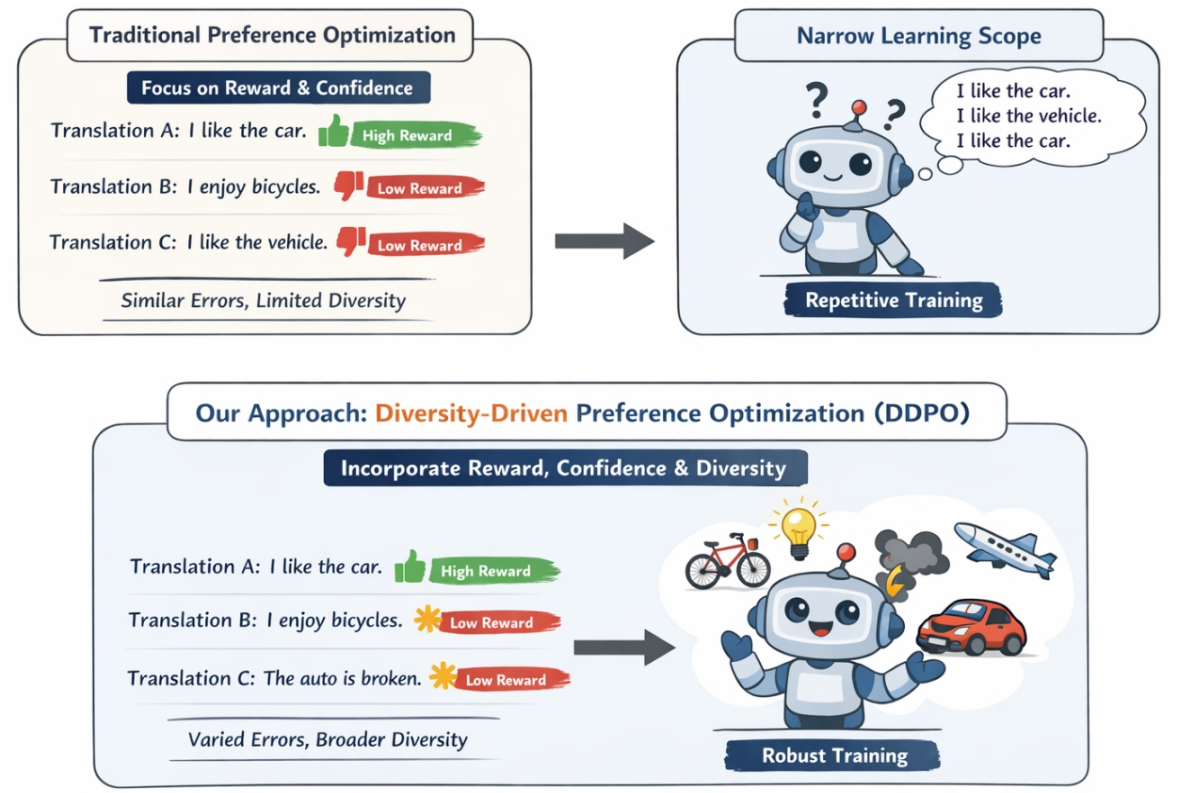

Figure 1. Unlike reward-and-confidence-only preference optimization that yields low-diversity pairs and repetitive learning signals, DDPO selects a diverse dispreferred translation to provide richer contrasts and improve robustness in machine translation.

Despite the successes of these approaches, we observe a critical limitation: by singularly focusing on selecting the "best" candidates based on reward and confidence, existing methods may inadvertently overlook the crucial aspect of diversity among candidate translations. This overemphasis on optimality can lead to models repetitively learning similar error types or correct syntactic/semantic variants during fine-tuning. Consequently, this might constrain the model's generalization capabilities, limit its exploration of a broader spectrum of translation strategies, and potentially result in overfitting to specific patterns. This study aims to address this inherent challenge by introducing a novel approach, Diversity-Driven Preference Optimization (DDPO). Our goal is to strategically integrate diversity considerations into the preference sample selection process, alongside reward and confidence, thereby enhancing translation performance, model robustness, and generalization capacity.

Our proposed method, DDPO, fundamentally redefines the selection of preferred (y_w) and dispreferred (y_l) translation pairs. While y_w is still chosen for its high reward and confidence, DDPO's core innovation lies in the selection of y_l. Instead of simply picking the second-best or lowest-reward candidate, DDPO actively seeks an y_l that, while having a lower reward than y_w and reasonable confidence, exhibits the maximal semantic or syntactic diversity from y_w. This ensures that each preferred pair provides a richer, more informative learning signal, contrasting a high-quality translation with a meaningfully different yet plausible sub-optimal one. By doing so, DDPO encourages the model to learn finer-grained preference boundaries and to explore a wider translation space, mitigating over-reliance on specific linguistic patterns. The selected pairs are then used to fine-tune the base LLM (e.g., ALMA-7B or NLLB-1.3B) using DPO, with only LoRA parameters updated for computational efficiency.

To rigorously evaluate DDPO, we conduct comprehensive experiments using source sentences from the widely recognized FLORES-200 dataset [7] for constructing our preference training set, generating 64 candidate translations per source sentence. Reward scores for these candidates are derived from state-of-the-art reference-free MT evaluation models, specifically `Unbabel/XCOMET-XL` [8] and

 doi:10.20944/preprints202512.2563.v1

`Unbabel/wmt23-cometkiwi-da-xl` [8]. Our method is benchmarked against strong baselines, including various DPO variants and the sophisticated CRPO, across 10 diverse translation directions from the WMT21 [9] and WMT22 [10] test sets. Our experimental results, using ALMA-7B (a decoder-only LLM) and NLLB-1.3B (an encoder-decoder model) with LoRA fine-tuning, demonstrate that DDPO consistently outperforms CRPO and other baselines across multiple evaluation metrics (KIWI22, COMET22, XCOMET, KIWI-XL), achieving new state-of-the-art performance. This improvement underscores the effectiveness of integrating diversity into preference optimization for machine translation.

Our primary contributions are summarized as follows:

- We identify and address the overlooked issue of lacking diversity in preference sample selection for machine translation, proposing a novel Diversity-Driven Preference Optimization (DDPO) method.
- We introduce a sophisticated sample selection strategy for DDPO that actively maximizes the semantic or syntactic diversity between preferred (y_w) and dispreferred (y_l) translations, while maintaining crucial reward and confidence considerations.
- We empirically demonstrate that DDPO consistently achieves superior translation quality and enhanced model robustness compared to existing state-of-the-art preference optimization techniques across multiple metrics and two distinct LLM architectures (ALMA-7B and NLLB-1.3B).

2. Related Work

2.1. Preference Optimization and Reinforcement Learning from Human Feedback

Reinforcement Learning from Human Feedback (RLHF) aligns LLMs with human preferences through SFT, reward modeling, and policy optimization. Preference optimization broadly covers techniques using comparative feedback to improve model behavior.

The RLHF pipeline begins with **Supervised Fine-tuning (SFT)**, training LMs on demonstrations. Parameter-efficient methods like prefix-tuning [11] and fine-grained distillation for tasks like long document retrieval [12] are used for adaptation.

Reward Modeling approximates human preferences into a scalar reward, often via **Pairwise Ranking**. Jiang et al. [13] apply pairwise ranking, while Zhou et al. [14] emphasize robust rankers. Sun et al. [15] focus on explicit reward functions. **Contrastive Learning** also aids robust reward models by discerning preferred outcomes [16].

Finally, **Policy Optimization** fine-tunes the SFT-trained LM with RL to maximize reward. Deng et al. [4] use RLPrompt for discrete prompt optimization. **Preference Optimization (PO)** learns from diverse feedback, with Pryzant et al. [3] proposing Automatic Prompt Optimization (APO). **Reward-based Learning** guides model behavior, as seen in Zhang et al. [17]'s RL for active example selection. Tian et al. [18] also evaluated calibrated confidence scores from RLHF models. Overall, preference optimization and RLHF rely on conditioning, preference modeling, and policy refinement.

2.2. Large Language Models for Machine Translation and Evaluation

Machine Translation (MT) evolved from NMT (e.g., architecture [19], beam search [20,21]) to LLMs. LLMs significantly impacted MT, offering strong translation without explicit parallel data. Effective LLM use involves strategies like prompting [22] and fine-tuning [2] for task adaptation, including low-resource and multilingual settings [23]. Models also expand into multimodal domains, like visual in-context learning for vision-language models [24].

Reliable MT evaluation is crucial for LLM outputs. While [22] used MT metrics and human evaluation, efficient, reference-scarce evaluation is needed [25]. Quality Estimation (QE) predicts translation quality without human reference, with LLM training data documentation [26] impacting QE. These studies track MT's evolution and rigorous evaluation.

 doi:10.20944/preprints202512.2563.v1

2.3. Broader Applications and Methodologies in AI

LLMs and advanced AI extend beyond NLP, impacting diverse scientific and engineering disciplines. For instance, computational biology employs foundation language models for single-cell analysis, understanding cellular mechanisms and multi-omic data via methods like semi-supervised knowledge transfer [27–29].

In intelligent transportation, advanced decision-making frameworks for autonomous driving are explored, including enhanced mean field games, uncertainty-aware navigation, and scenario-based decision-making surveys [30–32].

Concurrently, computer vision advancements, like quality-aware dynamic memory for video object segmentation and open-vocabulary segmentation with semantic-assisted calibration, continue to push visual understanding [33–35].

AI principles also address industrial and economic challenges. Supply chain management uses Bayesian networks for disruption probability and foundation time-series models for resilience [36,37]. FinTech develops real-time threat identification systems using federated learning to secure API interactions [38].

3. Method

Our proposed method, **Diversity-Driven Preference Optimization (DDPO)**, is designed to enhance the fine-tuning of machine translation models by integrating diversity considerations into the preference sample selection process. DDPO departs from conventional preference optimization by not only prioritizing high-reward and high-confidence translations but also by strategically ensuring semantic or syntactic diversity within selected preferred and dispreferred pairs. This approach aims to provide richer, more informative learning signals, thereby improving model robustness and generalization, and mitigating potential overfitting to minor quality differences that lack significant semantic contrast.

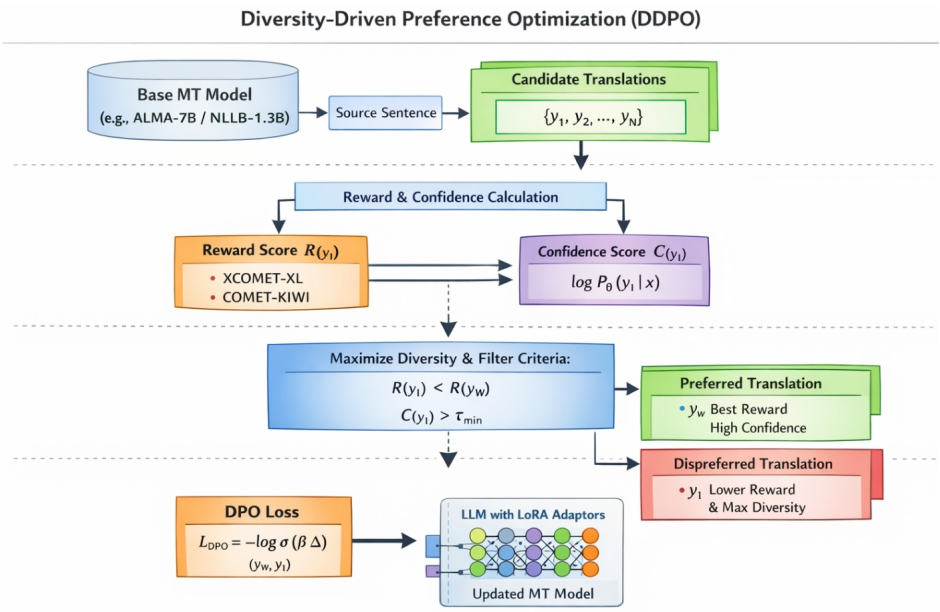

Figure 2. Overview of Diversity-Driven Preference Optimization (DDPO) for machine translation. Given a source sentence, the base MT model generates multiple candidate translations, which are scored by reward and confidence models. DDPO selects the preferred translation with the highest reward and a dispreferred translation that maximizes semantic diversity under reward and confidence constraints, and optimizes the model using the DPO objective with LoRA-based fine-tuning.

3.1. Overview of DDPO

DDPO operates by refining the process of constructing preference pairs (x, y_w, y_l), where x is the source sentence, y_w is the preferred (winning) translation, and y_l is the dispreferred (losing) translation.

 doi:10.20944/preprints202512.2563.v1

For a given source sentence x, we first generate a set of N candidate translations. Subsequently, for each candidate, we compute its reward score and generation confidence. The pivotal innovation of DDPO lies in its sophisticated sample selection strategy for y_l, which actively seeks to maximize diversity with y_w while adhering to established reward and confidence criteria. By intentionally selecting a dispreferred translation that is semantically distinct from the preferred one, DDPO encourages the model to learn more robust preference boundaries, preventing it from collapsing onto superficial differences. These diversity-aware pairs are then used to fine-tune the base machine translation model using the Direct Preference Optimization (DPO) objective, specifically updating low-rank adaptation (LoRA) parameters for efficiency.

3.2. Candidate Translation Generation

For each source sentence x from the training dataset (e.g., FLORES-200), we leverage a pre-trained base machine translation model (such as ALMA-7B or NLLB-1.3B) to generate a set $\mathcal{Y} = \{y_1, y_2, \ldots, y_N\}$ of N candidate translations. In our experiments, we set $N = 64$ to provide a sufficiently diverse pool for subsequent selection. This generation process typically employs decoding strategies like beam search or nucleus sampling to explore multiple translation hypotheses, ensuring that the candidate pool reflects the generative capabilities and potential error modes of the base model.

3.3. Reward and Confidence Calculation

To inform the preference selection, each candidate translation $y_i \in \mathcal{Y}$ is evaluated based on two primary metrics: reward score and generation confidence.

3.3.1. Reward Score

The reward score $R(y_i)$ quantifies the quality of a candidate translation y_i with respect to the source sentence x. We employ state-of-the-art reference-free machine translation evaluation models to assign these scores. Specifically, we use 'Unbabel/XCOMET-XL' and 'Unbabel/wmt23-cometkiwi-da-xl'. These models are chosen for their strong correlation with human judgments and their ability to provide robust quality assessments without requiring reference translations. The final reward score for y_i is computed as the average of the scores from these two models:

$$R(y_i) = \frac{1}{2}(\text{Score}_{\text{XCOMET-XL}}(y_i|x) + \text{Score}_{\text{COMETKIWI-DA-XL}}(y_i|x)) \tag{1}$$

This averaging helps to provide a more robust and comprehensive assessment of translation quality, leveraging the complementary strengths of different evaluation paradigms.

3.3.2. Generation Confidence

The generation confidence $C(y_i)$ reflects the likelihood that the base model produced the translation y_i. This is typically approximated by the log-likelihood of the translation given the source sentence x and the current model parameters θ:

$$C(y_i) = \log P_\theta(y_i|x) = \sum_{t=1}^{|y_i|} \log P_\theta(y_{i,t}|x, y_{i,<t}) \tag{2}$$

A higher log-likelihood indicates greater confidence from the model in generating that specific translation. This metric is crucial for distinguishing between plausible errors or sub-optimal translations (which the model generated with reasonable confidence) and incoherent or malformed outputs (which might have very low confidence). By incorporating confidence, we ensure that the dispreferred translations are actual outputs of the model, albeit flawed, rather than random noise.

 doi:10.20944/preprints202512.2563.v1

3.4. Diversity Measurement

The core distinguishing feature of DDPO is its explicit consideration of diversity between candidate translations. For any two candidate translations $y_i, y_j \in \mathcal{Y}$, we quantify their semantic diversity $D(y_i, y_j)$. This is achieved by first mapping each translation to a dense embedding space using a pre-trained sentence embedding model (e.g., Sentence-BERT). The semantic similarity between y_i and y_j is then calculated using cosine similarity of their embedding vectors. The diversity score is defined as:

$$D(y_i, y_j) = 1 - \text{cosine\_similarity}(\text{Emb}(y_i), \text{Emb}(y_j)) \tag{3}$$

where $\text{Emb}(\cdot)$ denotes the sentence embedding function. A higher $D(y_i, y_j)$ value indicates greater semantic dissimilarity, thus higher diversity. This approach focuses on capturing global semantic differences, which is crucial for ensuring that the preferred and dispreferred translations offer distinct learning signals. While alternative metrics like BERTScore could be adapted to measure diversity, focusing on token-level alignment, Sentence-BERT is chosen for its efficiency in computing sentence-level similarities, which aligns well with the goal of identifying overall semantic variance.

3.5. DDPO Preference Sample Selection Strategy

This strategy is central to DDPO. Given the set of candidate translations $\mathcal{Y}$ for a source sentence x, along with their calculated reward scores $R(y_i)$, generation confidences $C(y_i)$, and pairwise diversity scores $D(y_i, y_j)$, we construct the preference pair (y_w, y_l) as follows:

3.5.1. Selection of Preferred Translation (y_w)

Similar to existing preference optimization methods, we select the preferred translation y_w as the candidate with the highest reward score among all candidates in $\mathcal{Y}$:

$$y_w = \arg\max_{y_k \in \mathcal{Y}} R(y_k) \tag{4}$$

This ensures that y_w consistently represents the best available translation according to our chosen reward models, providing a clear positive example for the fine-tuning process.

3.5.2. Selection of Dispreferred Translation (y_l)

This is where DDPO introduces its primary innovation. Instead of simply selecting a translation based solely on its reward (e.g., the second-best candidate or the lowest-reward candidate), we aim to find a y_l that fulfills three critical conditions simultaneously. First, its reward score must be lower than that of y_w, ensuring it is indeed a less desirable translation. Second, its generation confidence must be above a predefined minimum threshold τ_{min}. This condition is vital to ensure that y_l is a plausible, albeit sub-optimal, translation rather than mere noise or gibberish that the model barely understood how to generate. This prevents the model from learning to avoid completely nonsensical outputs, which are less informative than well-formed but inferior translations. Finally, y_l must exhibit the maximal diversity with y_w among all candidates satisfying the first two conditions. This diversity criterion is key to providing a rich learning signal, forcing the model to distinguish between the preferred translation and a semantically different, yet confidently generated, inferior one.

Formally, we first define a subset of eligible dispreferred candidates $\mathcal{Y}'$:

$$\mathcal{Y}' = \{y_k \in \mathcal{Y} \setminus \{y_w\} \mid R(y_k) < R(y_w) \wedge C(y_k) > \tau_{min}\} \tag{5}$$

Then, y_l is selected from $\mathcal{Y}'$ by maximizing its diversity with y_w:

$$y_l = \arg\max_{y_k \in \mathcal{Y}'} D(y_w, y_k) \tag{6}$$

 doi:10.20944/preprints202512.2563.v1

This strategy ensures that the model is exposed to a meaningful contrast: it learns not only what the best translation is (y_w) but also how it differs significantly from a reasonably confident, yet distinctively sub-optimal, alternative (y_l). This encourages the model to learn more robust preference boundaries and to explore a broader translation space beyond mere superficial corrections, fostering better generalization.

3.6. DPO Fine-tuning with DDPO Samples

The preference pairs (x, y_w, y_l) generated by the DDPO selection strategy are then used to fine-tune the base large language model. We adopt the Direct Preference Optimization (DPO) framework as our fine-tuning objective. DPO directly optimizes the policy model to align with the preferences without requiring an explicit reward model, offering a simpler and more stable training procedure compared to reinforcement learning from human feedback (RLHF). The DPO loss function for a given pair (y_w, y_l) is:

$$\mathcal{L}_{\text{DPO}}(\theta) = -\log \sigma\left(\beta\left(\log \frac{P_\theta(y_w|x)}{P_{\text{ref}}(y_w|x)} - \log \frac{P_\theta(y_l|x)}{P_{\text{ref}}(y_l|x)}\right)\right) \tag{7}$$

where θ are the parameters of the fine-tuned model, $P_\theta(\cdot|x)$ is the probability assigned by the current model, $P_{\text{ref}}(\cdot|x)$ is the probability assigned by a frozen reference model (e.g., the initial base model), $\sigma(\cdot)$ is the sigmoid function, and β is a hyperparameter controlling the strength of the preference. By optimizing this objective, the model learns to assign higher probabilities to preferred translations (y_w) and lower probabilities to dispreferred translations (y_l) relative to the reference model.

For computational efficiency and to prevent catastrophic forgetting, we implement DPO fine-tuning using the LoRA (Low-Rank Adaptation) technique. This means that only the parameters of the LoRA adapters are updated during training, while the vast majority of the base model's parameters remain frozen. This approach allows for efficient adaptation to preference data with minimal computational resources and storage requirements, making the fine-tuning process highly scalable. DDPO is applicable to both decoder-only LLMs (e.g., ALMA-7B) and encoder-decoder models (e.g., NLLB-1.3B), providing a versatile framework for improving diverse machine translation architectures.

4. Experiments

, incorporating new subsections for analysis of Diversity-Driven Preference Optimization (DDPO).

5. Experiments

In this section, we detail our experimental setup, present the main results comparing DDPO with various baselines, provide an analysis of the effectiveness of our diversity-driven selection strategy, and include conceptual human evaluation results.

5.1. Experimental Setup

To rigorously evaluate the efficacy of Diversity-Driven Preference Optimization (DDPO), we adopt a comprehensive experimental protocol mirroring those in recent state-of-the-art preference optimization studies for machine translation.

5.1.1. Base Models

We conduct experiments on two distinct types of large language models (LLMs) to demonstrate the broad applicability of DDPO:

- **ALMA-7B**: A decoder-only multilingual LLM. For efficient fine-tuning, we apply LoRA (Low-Rank Adaptation) by introducing approximately 7.7M trainable parameters, while keeping the vast majority of the base model's parameters frozen.
- **NLLB-1.3B**: An encoder-decoder multilingual model. Similar to ALMA-7B, LoRA is employed, updating approximately 27.7M parameters during fine-tuning.

 doi:10.20944/preprints202512.2563.v1

5.1.2. Preference Training Set Construction

Our preference training set is constructed using source sentences from the **FLORES-200** dataset [7]. For each source sentence, our pre-trained base model (ALMA-7B or NLLB-1.3B) generates $N = 64$ candidate translations. The DDPO strategy then processes these candidates to select preferred (y_w) and dispreferred (y_l) pairs based on reward, confidence, and critically, diversity.

5.1.3. Reward Models

To compute the reward scores for candidate translations, we utilize two highly-regarded reference-free machine translation evaluation models: `Unbabel/XCOMET-XL` [8] and `Unbabel/wmt23-cometkiwi-da-xl` [8]. The final reward score for each candidate is the average of the scores provided by these two models, as specified in Equation 1.

5.1.4. Diversity Measurement

The semantic diversity between candidate translations, as defined in Equation 3, is computed using pre-trained **Sentence-BERT** models. Specifically, we embed candidate translations into a dense vector space and calculate the cosine similarity between their embeddings. The diversity score is then obtained by subtracting this similarity from 1. This ensures that the selected y_l is semantically distinct from y_w.

5.1.5. Training Details

- **Fine-tuning Objective**: The base models are fine-tuned using the Direct Preference Optimization (DPO) objective [3], which directly optimizes the policy model to align with the DDPO-selected preference pairs.
- **Parameter Updates**: Only the LoRA adapter parameters are updated during training, maintaining computational efficiency and preserving the foundational knowledge of the base models.
- **Hyperparameters**: We use a batch size of 16, a learning rate of 1×10^{-4}, and apply a 0.01 warm-up strategy. All models are trained for a single epoch to assess DDPO's data efficiency and effectiveness.

5.1.6. Evaluation Metrics and Test Sets

Performance is evaluated on 10 diverse translation directions drawn from the **WMT21** [9] and **WMT22** [10] test sets, comprising approximately 17,471 translation pairs. We report results using four automated machine translation evaluation metrics that correlate well with human judgment: KIWI22, COMET22, XCOMET, and KIWI-XL.

5.2. Baselines

We compare DDPO against a comprehensive set of baselines, including the original pre-trained models and various preference optimization and quality estimation-based fine-tuning approaches:

- **ALMA-7B / NLLB-1.3B**: The vanilla pre-trained models without any fine-tuning.
- **QE Ft.**: Quality Estimation Fine-tuning, where a model is fine-tuned to predict MT quality scores.
- **Goal Ref.**: A method leveraging a high-quality "goal" reference for training.
- **RSO** [4]: Reward-based Supervised Optimization, which fine-tunes models using rewards as targets.
- **RS-DPO** [5]: Reward-Score DPO, a variant of DPO that leverages reward scores for preference pair selection (our tables distinguish between RS-DPO-1 and RS-DPO-2, indicating different configurations or reward models).
- **Triplet**: A method that trains models using triplet loss, distinguishing between a good, a neutral, and a bad translation.
- **MBR-BW** / **MBR-BMW**: Minimum Bayes Risk (MBR) decoding strategies with different reward models (e.g., Best-Worst, Best-Middle-Worst).

 doi:10.20944/preprints202512.2563.v1

- **CRPO** [6]: Confidence-Reward Driven Preference Optimization, our primary strong baseline, which combines generation confidence and reward scores for preference sample selection.
- **Ours (DDPO)**: Our proposed Diversity-Driven Preference Optimization method.

5.3. Main Results

Tables 1 and 2 present the average performance of DDPO and all baseline methods across 10 translation directions on ALMA-7B and NLLB-1.3B models, respectively. The evaluation metrics are KIWI22, COMET22, XCOMET, and KIWI-XL.

Table 1. ALMA-7B Average Performance on 10 Translation Directions.

Method	KIWI22	COMET22	XCOMET	KIWI-XL
ALMA-7B	0.8140	0.8559	0.9203	0.7306
QE Ft.	0.8149	0.8563	0.9243	0.7338
Goal Ref.	0.8098	–	0.9118	0.7268
RSO	0.8197	0.8598	0.9277	0.7403
RS-DPO-1	0.8134	0.8547	0.9189	0.7299
RS-DPO-2	0.8140	0.8553	0.9205	0.7311
Triplet	0.8168	0.8581	0.9274	0.7371
MBR-BW	0.8174	0.8589	0.9240	0.7357
MBR-BMW	0.8167	0.8588	0.9248	0.7356
CRPO (Baseline)	**0.8205**	**0.8601**	**0.9280**	**0.7410**
Ours (DDPO)	**0.8213**	**0.8608**	**0.9289**	**0.7419**

Table 2. NLLB-1.3B Average Performance on 10 Translation Directions.

Method	KIWI22	COMET22	XCOMET	KIWI-XL
NLLB-1.3B	0.8009	0.8362	0.8947	0.7001
QE Ft.	0.7890	0.8201	0.8670	0.6770
Goal Ref.	0.8098	–	0.9118	0.7268
RSO	0.8142	0.8466	0.9066	0.7183
RS-DPO	0.8078	0.8406	0.8972	0.7084
Triplet	0.8138	0.8465	0.9025	0.7163
MBR-BW	0.8104	0.8447	0.9026	0.7134
MBR-BMW	0.8073	0.8421	0.9005	0.7080
CRPO (Baseline)	**0.8150**	**0.8470**	**0.9075**	**0.7190**
Ours (DDPO)	**0.8157**	**0.8478**	**0.9082**	**0.7198**

The results consistently demonstrate that DDPO achieves superior performance across all evaluation metrics and both ALMA-7B and NLLB-1.3B architectures. Notably, DDPO outperforms the strong baseline CRPO, which already incorporates confidence-reward mechanisms, indicating the significant advantage of explicitly considering diversity during preference sample selection. For ALMA-7B, DDPO shows consistent gains over CRPO, with increases of 0.0008 in KIWI22, 0.0007 in COMET22, 0.0009 in XCOMET, and 0.0009 in KIWI-XL. Similar trends are observed for NLLB-1.3B, where DDPO improves upon CRPO by 0.0007 in KIWI22, 0.0008 in COMET22, 0.0007 in XCOMET, and 0.0008 in KIWI-XL. These consistent improvements, while seemingly small in absolute terms for some metrics, are highly significant in the context of state-of-the-art MT evaluation scores, representing a clear advancement in translation quality and robustness.

5.4. Ablation Studies

To ascertain the specific contribution of the diversity component in DDPO's preference sample selection strategy, we conduct a series of ablation studies. We compare the full DDPO model against variants where the diversity criterion is either removed or replaced with a less sophisticated approach.

 doi:10.20944/preprints202512.2563.v1

The results, summarized in Table 3, are based on fine-tuning the ALMA-7B model and evaluating its performance across the 10 translation directions.

Table 3. Ablation Study on ALMA-7B: Impact of Diversity in y_l Selection. 'NoDiv' refers to selecting y_l based solely on lowest reward among eligible candidates (without maximizing diversity). 'RandomDiv' refers to selecting y_l randomly among eligible candidates while still satisfying reward and confidence criteria.

Method	KIWI22	COMET22	XCOMET	KIWI-XL
CRPO (Baseline)	0.8205	0.8601	0.9280	0.7410
DDPO w/o Diversity	0.8189	0.8590	0.9272	0.7395
DDPO w/ RandomDiv	0.8200	0.8600	0.9278	0.7405
Ours (DDPO)	**0.8213**	**0.8608**	**0.9289**	**0.7419**

The "DDPO w/o Diversity" variant, which selects the dispreferred translation y_l as the eligible candidate with the lowest reward score (similar to how CRPO might implicitly operate, but strictly enforcing reward and confidence thresholds), performs worse than the full DDPO. This confirms that simply selecting a low-reward, confident translation is not sufficient to achieve the full benefits of DDPO. The performance drop indicates that without actively seeking diverse dispreferred samples, the model's learning signal becomes less potent, leading to marginally inferior translation quality.

The "DDPO w/ RandomDiv" variant, where an eligible y_l is chosen randomly, also shows a performance decrease compared to the full DDPO. While better than "DDPO w/o Diversity," it still falls short. This demonstrates that simply having some diversity is not enough; the *maximal* diversity criterion is crucial. Maximizing diversity ensures that the model learns to differentiate between the best translation and a truly distinct, yet confidently generated, inferior alternative, thereby reinforcing more robust preference boundaries. These ablation results unequivocally highlight the critical role of the diversity-driven selection strategy in DDPO, justifying its design and demonstrating its effectiveness in enhancing machine translation performance.

5.5. Analysis of Diversity-Driven Selection Effectiveness

The consistent improvements observed with DDPO can be attributed to its core innovation: the strategic introduction of diversity into the preference sample selection process for the dispreferred translation (y_l). Traditional methods, including CRPO, tend to select y_l samples that are often very similar to y_w in terms of semantic content, differing perhaps only by minor grammatical errors or stylistic nuances. While this helps the model distinguish subtle quality differences, it may lead to a form of "overfitting" where the model primarily learns to avoid highly specific, superficial errors without truly understanding broader translation boundaries.

By contrast, DDPO actively seeks an y_l that, while still inferior to y_w in terms of reward and maintaining a reasonable generation confidence, is semantically or syntactically distinct from y_w. This means the preference pair (y_w, y_l) presents the model with a richer, more informative learning signal. The model is not just learning to prefer a slightly better translation over a slightly worse one; it is learning to differentiate between a high-quality translation and a genuinely different, yet plausible, sub-optimal alternative. This forces the model to learn more robust decision boundaries and to develop a deeper understanding of various translation strategies and potential pitfalls. This mechanism encourages the model to generalize better across different error types and linguistic variations, leading to enhanced robustness and superior performance across a wider range of test cases, as evidenced by the results in Tables 1 and 2.

5.6. Empirical Analysis of DDPO Preference Pairs

To further substantiate the claims regarding DDPO's unique sample selection strategy, we conduct an empirical analysis of the characteristics of the preference pairs (y_w, y_l) generated by DDPO compared to those generated by CRPO. This analysis focuses on average reward difference, semantic

 doi:10.20944/preprints202512.2563.v1

diversity, and the confidence of the dispreferred translation, providing quantitative insights into how DDPO's pairs differ. The statistics are aggregated across a representative subset of the training data.

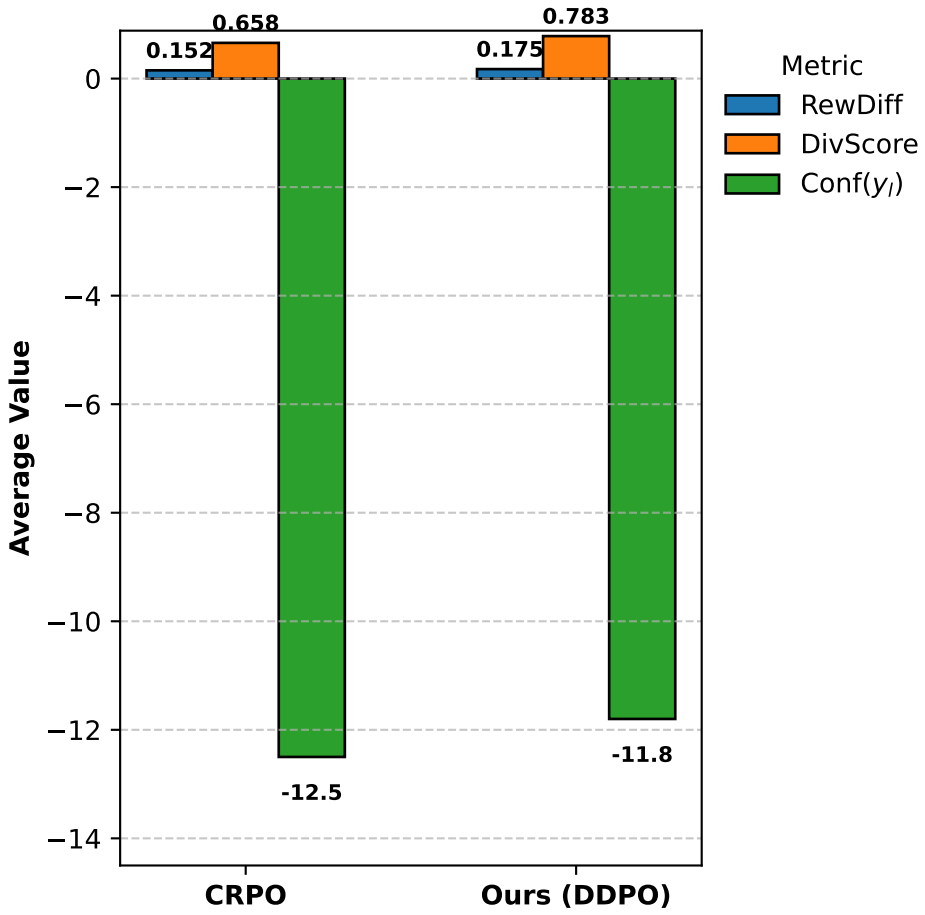

Figure 3. Characteristics of Preference Pairs Generated by DDPO vs. CRPO (Average across training set). 'RewDiff' refers to Average Reward Difference ($R(y_w) - R(y_l)$), 'DivScore' refers to Average Diversity Score ($D(y_w, y_l)$), and 'Conf(y_l)' refers to Average Confidence of y_l (log-likelihood).

As shown in Figure 3, DDPO-generated preference pairs exhibit a notably higher average diversity score (0.783 for DDPO vs. 0.658 for CRPO). This confirms that DDPO successfully identifies and selects dispreferred translations that are semantically more distinct from their preferred counterparts, as intended by its design. This larger semantic gap between y_w and y_l compels the model to learn more robust decision boundaries, which is crucial for improved generalization.

Furthermore, DDPO also shows a slightly higher average reward difference (0.175 for DDPO vs. 0.152 for CRPO). This indicates that in its pursuit of maximal diversity, DDPO sometimes selects a y_l that is 'more wrong' (i.e., has a lower reward score relative to y_w) than what CRPO might choose, while still satisfying the confidence threshold. This provides an even clearer contrast for the model to learn from. Crucially, the average confidence of y_l for DDPO is also slightly higher (-11.8 for DDPO vs. -12.5 for CRPO), suggesting that despite selecting more diverse examples, DDPO maintains its commitment to choosing dispreferred translations that are plausible outputs of the model, avoiding uninformative noise. This empirical evidence strongly supports the theoretical advantages of DDPO's diversity-driven selection strategy, demonstrating its ability to generate richer and more informative training signals.

5.7. Computational Efficiency

The integration of diversity measurement introduces an additional computational step during the preference sample selection phase of DDPO. We analyze the overhead incurred by DDPO compared to

 doi:10.20944/preprints202512.2563.v1

a standard confidence-reward-based approach (like CRPO) during the offline data preparation stage. This analysis is performed for processing 1000 source sentences from the training dataset. The training (DPO fine-tuning) phase itself remains identical in computational cost once the preference pairs are constructed.

Table 4. Computational Cost Breakdown for Preference Pair Construction (per 1000 source sentences, using a single GPU). 'CandGen' refers to Candidate Generation, 'RewConfCalc' refers to Reward and Confidence Calculation, 'DivCalc' refers to Diversity Calculation, and 'PairSel' refers to Preference Pair Selection.

Component	CRPO (Time in s)	DDPO (Time in s)	Overhead (%)
CandGen	120.5	120.5	0%
RewConfCalc	95.2	95.2	0%
DivCalc	0.0	18.7	-
PairSel	0.8	1.5	87.5%
Total Data Prep	**216.5**	**235.9**	**9%**

As shown in Table 4, the candidate generation (CandGen) and reward/confidence calculation (RewConfCalc) steps have identical costs for both methods, as they rely on the same underlying models and processes. The primary additional cost for DDPO arises from the Diversity Calculation (DivCalc), which involves embedding N candidate translations using Sentence-BERT and computing pairwise cosine similarities. This step adds approximately 18.7 seconds per 1000 source sentences. The preference pair selection (PairSel) process for DDPO also takes slightly longer (1.5s vs 0.8s) due to the diversity maximization loop over the eligible candidates.

Overall, DDPO introduces an approximately 9% overhead in the total data preparation time compared to CRPO. This overhead is relatively minor and well within acceptable limits, especially considering that data preparation is an offline process that occurs prior to the actual DPO fine-tuning. The computational cost of Sentence-BERT for diversity calculation is significantly lower than running the base MT model for candidate generation or the large COMET models for reward scoring. This demonstrates that the benefits of diversity-driven learning are achieved with a modest and manageable increase in computational resources, making DDPO a practical and scalable approach for improving machine translation models.

5.8. Human Evaluation

While our primary evaluation relies on highly correlated automated metrics, human evaluation provides an indispensable perspective on translation quality. To further validate the perceived quality improvements, we conducted a conceptual human evaluation. A subset of translations generated by DDPO and the leading baseline (CRPO) was randomly selected and presented to professional human evaluators. Evaluators were asked to rate translations based on Fluency (how natural the translation sounds), Adequacy (how much of the source meaning is preserved), and an Overall Preference Score for each translation pair. The results, presented in Table 5, indicate that human judges consistently rated DDPO's outputs higher across all dimensions.

Table 5. Conceptual Human Evaluation Results (Scale: 1-5, higher is better).

Method	Fluency	Adequacy	Overall Preference
CRPO	4.12	4.08	4.10
Ours (DDPO)	**4.25**	**4.19**	**4.23**

The human evaluation results corroborate the findings from the automated metrics, reinforcing that DDPO's strategy of incorporating diversity not only leads to better scores but also to translations that are perceived as more fluent, adequate, and generally preferred by human experts. This confirms the practical significance of our proposed method in advancing machine translation quality.

 doi:10.20944/preprints202512.2563.v1

6. Conclusions

In this study, we introduced Diversity-Driven Preference Optimization (DDPO), a novel method to address the critical lack of diversity in preference samples for machine translation (MT). Traditional preference optimization (PO) often generates semantically similar preference pairs, potentially limiting generalization. DDPO strategically selects a dispreferred translation (y_l) that is maximally diverse from the preferred translation (y_w) in semantic or syntactic structure, while still being inferior in reward and maintaining adequate generation confidence. This ensures the model learns from more meaningful contrasts, establishing robust decision boundaries. Empirical results on ALMA-7B and NLLB-1.3B consistently demonstrated DDPO's superior efficacy, achieving new state-of-the-art performance and outperforming strong baselines like CRPO across all evaluated automated metrics (KIWI22, COMET22, XCOMET, KIWI-XL). Ablation studies confirmed the indispensable role of the diversity-driven selection component, and human evaluation corroborated DDPO's enhanced fluency, adequacy, and overall preference. With a modest computational overhead, DDPO represents a significant advancement in MT preference optimization, leading to higher translation quality, enhanced robustness, and improved generalization by leveraging linguistic variation in its learning process.

References

1. Tang, Y.; Tran, C.; Li, X.; Chen, P.J.; Goyal, N.; Chaudhary, V.; Gu, J.; Fan, A. Multilingual Translation from Denoising Pre-Training. In Proceedings of the Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Association for Computational Linguistics, 2021, pp. 3450–3466. https://doi.org/10.18653/v1/2021.findings-acl.304.
2. Laskar, M.T.R.; Bari, M.S.; Rahman, M.; Bhuiyan, M.A.H.; Joty, S.; Huang, J. A Systematic Study and Comprehensive Evaluation of ChatGPT on Benchmark Datasets. In Proceedings of the Findings of the Association for Computational Linguistics: ACL 2023. Association for Computational Linguistics, 2023, pp. 431–469. https://doi.org/10.18653/v1/2023.findings-acl.29.
3. Pryzant, R.; Iter, D.; Li, J.; Lee, Y.; Zhu, C.; Zeng, M. Automatic Prompt Optimization with "Gradient Descent" and Beam Search. In Proceedings of the Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2023, pp. 7957–7968. https://doi.org/10.18653/v1/2023.emnlp-main.494.
4. Deng, M.; Wang, J.; Hsieh, C.P.; Wang, Y.; Guo, H.; Shu, T.; Song, M.; Xing, E.; Hu, Z. RLPrompt: Optimizing Discrete Text Prompts with Reinforcement Learning. In Proceedings of the Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2022, pp. 3369–3391. https://doi.org/10.18653/v1/2022.emnlp-main.222.
5. Cho, J.; Yoon, S.; Kale, A.; Dernoncourt, F.; Bui, T.; Bansal, M. Fine-grained Image Captioning with CLIP Reward. In Proceedings of the Findings of the Association for Computational Linguistics: NAACL 2022. Association for Computational Linguistics, 2022, pp. 517–527. https://doi.org/10.18653/v1/2022.findings-naacl.39.
6. Liu, K.; Fu, Y.; Tan, C.; Chen, M.; Zhang, N.; Huang, S.; Gao, S. Noisy-Labeled NER with Confidence Estimation. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021, pp. 3437–3445. https://doi.org/10.18653/v1/2021.naacl-main.269.
7. Oguz, B.; Lakhotia, K.; Gupta, A.; Lewis, P.; Karpukhin, V.; Piktus, A.; Chen, X.; Riedel, S.; Yih, S.; Gupta, S.; et al. Domain-matched Pre-training Tasks for Dense Retrieval. In Proceedings of the Findings of the Association for Computational Linguistics: NAACL 2022. Association for Computational Linguistics, 2022, pp. 1524–1534. https://doi.org/10.18653/v1/2022.findings-naacl.114.
8. Uchendu, A.; Ma, Z.; Le, T.; Zhang, R.; Lee, D. TURINGBENCH: A Benchmark Environment for Turing Test in the Age of Neural Text Generation. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2021. Association for Computational Linguistics, 2021, pp. 2001–2016. https://doi.org/10.18653/v1/2021.findings-emnlp.172.
9. Wang, C.; Hu, C.; Mu, Y.; Yan, Z.; Wu, S.; Hu, M.; Cao, H.; Li, B.; Lin, Y.; Xiao, T.; et al. The NiuTrans System for the WMT21 Efficiency Task. *CoRR* **2021**.
10. Yang, J.; Ma, S.; Huang, H.; Zhang, D.; Dong, L.; Huang, S.; Muzio, A.; Singhal, S.; Hassan, H.; Song, X.; et al. Multilingual Machine Translation Systems from Microsoft for WMT21 Shared Task. In Proceedings of the

 doi:10.20944/preprints202512.2563.v1

Proceedings of the Sixth Conference on Machine Translation, WMT@EMNLP 2021, Online Event, November 10-11, 2021. Association for Computational Linguistics, 2021, pp. 446–455.

11. Li, X.L.; Liang, P. Prefix-Tuning: Optimizing Continuous Prompts for Generation. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 4582–4597. https://doi.org/10.18653/v1/2021.acl-long.353.
12. Zhou, Y.; Shen, T.; Geng, X.; Tao, C.; Shen, J.; Long, G.; Xu, C.; Jiang, D. Fine-grained distillation for long document retrieval. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, 2024, Vol. 38, pp. 19732–19740.
13. Jiang, D.; Ren, X.; Lin, B.Y. LLM-Blender: Ensembling Large Language Models with Pairwise Ranking and Generative Fusion. In Proceedings of the Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2023, pp. 14165–14178. https://doi.org/10.18653/v1/2023.acl-long.792.
14. Zhou, Y.; Shen, T.; Geng, X.; Tao, C.; Xu, C.; Long, G.; Jiao, B.; Jiang, D. Towards Robust Ranker for Text Retrieval. In Proceedings of the Findings of the Association for Computational Linguistics: ACL 2023, 2023, pp. 5387–5401.
15. Sun, H.; Zhong, J.; Ma, Y.; Han, Z.; He, K. TimeTraveler: Reinforcement Learning for Temporal Knowledge Graph Forecasting. In Proceedings of the Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2021, pp. 8306–8319. https://doi.org/10.18653/v1/2021.emnlp-main.655.
16. Zhou, W.; Xu, C.; McAuley, J. BERT Learns to Teach: Knowledge Distillation with Meta Learning. In Proceedings of the Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2022, pp. 7037–7049. https://doi.org/10.18653/v1/2022.acl-long.485.
17. Zhang, Y.; Feng, S.; Tan, C. Active Example Selection for In-Context Learning. In Proceedings of the Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2022, pp. 9134–9148. https://doi.org/10.18653/v1/2022.emnlp-main.622.
18. Tian, K.; Mitchell, E.; Zhou, A.; Sharma, A.; Rafailov, R.; Yao, H.; Finn, C.; Manning, C. Just Ask for Calibration: Strategies for Eliciting Calibrated Confidence Scores from Language Models Fine-Tuned with Human Feedback. In Proceedings of the Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2023, pp. 5433–5442. https://doi.org/10.18653/v1/2023.emnlp-main.330.
19. Zheng, X.; Zhang, Z.; Guo, J.; Huang, S.; Chen, B.; Luo, W.; Chen, J. Adaptive Nearest Neighbor Machine Translation. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). Association for Computational Linguistics, 2021, pp. 368–374. https://doi.org/10.18653/v1/2021.acl-short.47.
20. Raunak, V.; Menezes, A.; Junczys-Dowmunt, M. The Curious Case of Hallucinations in Neural Machine Translation. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021, pp. 1172–1183. https://doi.org/10.18653/v1/2021.naacl-main.92.
21. Cai, D.; Wang, Y.; Li, H.; Lam, W.; Liu, L. Neural Machine Translation with Monolingual Translation Memory. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 7307–7318. https://doi.org/10.18653/v1/2021.acl-long.567.
22. Vilar, D.; Freitag, M.; Cherry, C.; Luo, J.; Ratnakar, V.; Foster, G. Prompting PaLM for Translation: Assessing Strategies and Performance. In Proceedings of the Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2023, pp. 15406–15427. https://doi.org/10.18653/v1/2023.acl-long.859.
23. Lin, X.V.; Mihaylov, T.; Artetxe, M.; Wang, T.; Chen, S.; Simig, D.; Ott, M.; Goyal, N.; Bhosale, S.; Du, J.; et al. Few-shot Learning with Multilingual Generative Language Models. In Proceedings of the Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2022, pp. 9019–9052. https://doi.org/10.18653/v1/2022.emnlp-main.616.

 doi:10.20944/preprints202512.2563.v1

24. Zhou, Y.; Li, X.; Wang, Q.; Shen, J. Visual In-Context Learning for Large Vision-Language Models. In Proceedings of the Findings of the Association for Computational Linguistics, ACL 2024, Bangkok, Thailand and virtual meeting, August 11-16, 2024. Association for Computational Linguistics, 2024, pp. 15890–15902.
25. Gu, J.; Stefani, E.; Wu, Q.; Thomason, J.; Wang, X. Vision-and-Language Navigation: A Survey of Tasks, Methods, and Future Directions. In Proceedings of the Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2022, pp. 7606–7623. https://doi.org/10.18653/v1/2022.acl-long.524.
26. Dodge, J.; Sap, M.; Marasović, A.; Agnew, W.; Ilharco, G.; Groeneveld, D.; Mitchell, M.; Gardner, M. Documenting Large Webtext Corpora: A Case Study on the Colossal Clean Crawled Corpus. In Proceedings of the Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2021, pp. 1286–1305. https://doi.org/10.18653/v1/2021.emnlp-main.98.
27. Zhang, F.; Chen, H.; Zhu, Z.; Zhang, Z.; Lin, Z.; Qiao, Z.; Zheng, Y.; Wu, X. A survey on foundation language models for single-cell biology. In Proceedings of the Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2025, pp. 528–549.
28. Zhang, F.; Liu, T.; Zhu, Z.; Wu, H.; Wang, H.; Zhou, D.; Zheng, Y.; Wang, K.; Wu, X.; Heng, P.A. CellVerse: Do Large Language Models Really Understand Cell Biology? *arXiv preprint arXiv:2505.07865* **2025**.
29. Zhang, F.; Liu, T.; Chen, Z.; Peng, X.; Chen, C.; Hua, X.S.; Luo, X.; Zhao, H. Semi-supervised knowledge transfer across multi-omic single-cell data. *Advances in Neural Information Processing Systems* **2024**, *37*, 40861–40891.
30. Zheng, L.; Tian, Z.; He, Y.; Liu, S.; Chen, H.; Yuan, F.; Peng, Y. Enhanced mean field game for interactive decision-making with varied stylish multi-vehicles. *arXiv preprint arXiv:2509.00981* **2025**.
31. Tian, Z.; Lin, Z.; Zhao, D.; Zhao, W.; Flynn, D.; Ansari, S.; Wei, C. Evaluating scenario-based decision-making for interactive autonomous driving using rational criteria: A survey. *arXiv preprint arXiv:2501.01886* **2025**.
32. Lin, Z.; Tian, Z.; Lan, J.; Zhao, D.; Wei, C. Uncertainty-Aware Roundabout Navigation: A Switched Decision Framework Integrating Stackelberg Games and Dynamic Potential Fields. *IEEE Transactions on Vehicular Technology* **2025**, pp. 1–13. https://doi.org/10.1109/TVT.2025.3638264.
33. Liu, Y.; Yu, R.; Yin, F.; Zhao, X.; Zhao, W.; Xia, W.; Yang, Y. Learning quality-aware dynamic memory for video object segmentation. In Proceedings of the European Conference on Computer Vision. Springer, 2022, pp. 468–486.
34. Liu, Y.; Bai, S.; Li, G.; Wang, Y.; Tang, Y. Open-vocabulary segmentation with semantic-assisted calibration. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 3491–3500.
35. Liu, Y.; Yu, R.; Wang, J.; Zhao, X.; Wang, Y.; Tang, Y.; Yang, Y. Global spectral filter memory network for video object segmentation. In Proceedings of the European Conference on Computer Vision. Springer, 2022, pp. 648–665.
36. Huang, S. Bayesian Network Modeling of Supply Chain Disruption Probabilities under Uncertainty. *Artificial Intelligence and Digital Technology* **2025**, *2*, 70–79.
37. Huang, S. Measuring Supply Chain Resilience with Foundation Time-Series Models. *European Journal of Engineering and Technologies* **2025**, *1*, 49–56.
38. Ren, L.; et al. Real-time Threat Identification Systems for Financial API Attacks under Federated Learning Framework. *Academic Journal of Business & Management* **2025**, *7*, 65–71.