CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	LaCoMSA: Language-Consistency Multilingual Self-Alignment with latent representation rewarding
Authors	Tran, Khanh Tung;O'Sullivan, Barry;Nguyen, Hoang D.
Publication date	2026-03-29
Original Citation	Tran, K T, O'Sullivan, B & Nguyen, H D 2026, LaCoMSA: Language-Consistency Multilingual Self-Alignment with latent representation rewarding. in V Demberg, K Inui & L Marquez Villodre (eds), Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers) : Rabat, Morocco, 24–29 March 2026. EACL 2026 - 19th Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of the Conference, Vol. 1 - (Long Papers), vol. 1, Association for Computational Linguistics (ACL), pp. 4839-4853, 19th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2026, Rabat, Morocco, 24/03/26. https://doi.org/10.18653/v1/2026.eacl-long.224
Type of publication	Conference item
Link to publisher's version	10.18653/v1/2026.eacl-long.224
Rights	© 2026, Association for Computational Linguistics. - https://creativecommons.org/licenses/by/4.0/
Download date	2026-10-13 13:21:19
Item downloaded from	https://hdl.handle.net/10468/19019

University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

LaCoMSA: Language-Consistency Multilingual Self-Alignment with Latent Representation Rewarding

Khanh-Tung Tran **Barry O'Sullivan** **Hoang D. Nguyen**
Research Ireland Centre for Research Training in Artificial Intelligence
Insight Research Ireland Centre for Data Analytics
School of Computer Science and Information Technology, University College Cork, Ireland
123128577@umail.ucc.ie b.osullivan@cs.ucc.ie hn@cs.ucc.ie

Abstract

Large Language Models (LLMs) have achieved impressive performance yet remain inconsistent across languages, often defaulting to high-resource outputs such as English. Existing multilingual alignment methods mitigate these issues through preference optimization but rely on external supervision, such as translation systems or English-biased signal. We propose Multilingual Self-Alignment (*MSA*), a targeted preference optimization framework that leverages an LLM's own latent representations as intrinsic supervision signals, rewarding lower-resource language outputs based on their alignment with high-resource (English) counterparts in the "semantic hub". We further introduce Language-Consistency MSA (*LaCoMSA*), which augments MSA with a final-layer language-consistency factor to prevent off-target generation. Integrated with Direct Preference Optimization, LaCoMSA improves a Llama 3 8B-based model multilingual win rates by up to 6.8% absolute (55.0% relatively) on X-AlpacaEval and achieves consistent gains across benchmarks and models. Our findings demonstrate that LaCoMSA can serve as an effective and scalable mechanism, opening a new venue toward multilingual self-alignment.

1 Introduction

Large Language Models (LLMs) have demonstrated remarkable capabilities, yet performance disparities across languages persist (Zhang et al., 2023; Alam et al., 2024). Multilingual generation often faces two recurring problems: (i) inconsistent performance between high-resource languages (HRLs) and lower-resource languages (LRLs) for the same task; and (ii) off-target language problem, where LLMs generate English or other HRL responses even when prompted in a LRL (Gu et al., 2019; Sennrich et al., 2024; Zhang et al., 2024a).

Recent approaches address these multilingual alignment issues through preference optimization. MAPO (She et al., 2024) and LIDR (Yang et al., 2025b) align multilingual outputs by leveraging external translation models or prompting-based translations, while ICR (Yang et al., 2025a) employs an English preference model to score LRL outputs. These methods depend on external supervision, e.g., translation systems, embedding models (Zhang et al., 2025), or English-centric signal, which are costly to build for LRLs, and risk biasing toward limitations of the English-centric model. Moreover, recent analyses reveal that this reliance is problematic: multilingual LLMs often encode the correct semantics internally but fail at the final translation stage, producing fluent but off-target outputs (Holtermann et al., 2024; Tang et al., 2024).

In contrast, there is emerging evidence that the latent representations of LLMs already contain strong cross-lingual alignment signals. For instance, (Tran et al., 2024; Wendler et al., 2024; Wu et al., 2025) demonstrate the existence of a "semantic hub" in the middle layers, where embeddings of semantically equivalent inputs from different languages converge. Similarly, MEXA (Kargaran et al., 2025) and NeuronXA (Huang et al., 2025) show that retrieval scores between parallel sentence representations at these layers correlate strongly with downstream task performance. These findings suggest that LLMs implicitly encode language-independent semantics in their middle layers, even if final outputs fail to reflect them.

This paper operationalizes these insights into latent multilingual semantic representations. We introduce *Multilingual Self-Alignment (MSA)*, a targeted preference optimization framework that exploits the LLM's own latent representations to derive reward signals: an LRL output is rewarded in proportion to its semantic hub middle-layer similarity with HRL (English) outputs generated for the same prompt. This turns the model's internal semantic space into an intrinsic supervision source. We then propose **La**nguage-**Co**nsistency

Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics
Volume 1: Long Papers, pages 4839–4853
March 24-29, 2026 ©2026 Association for Computational Linguistics

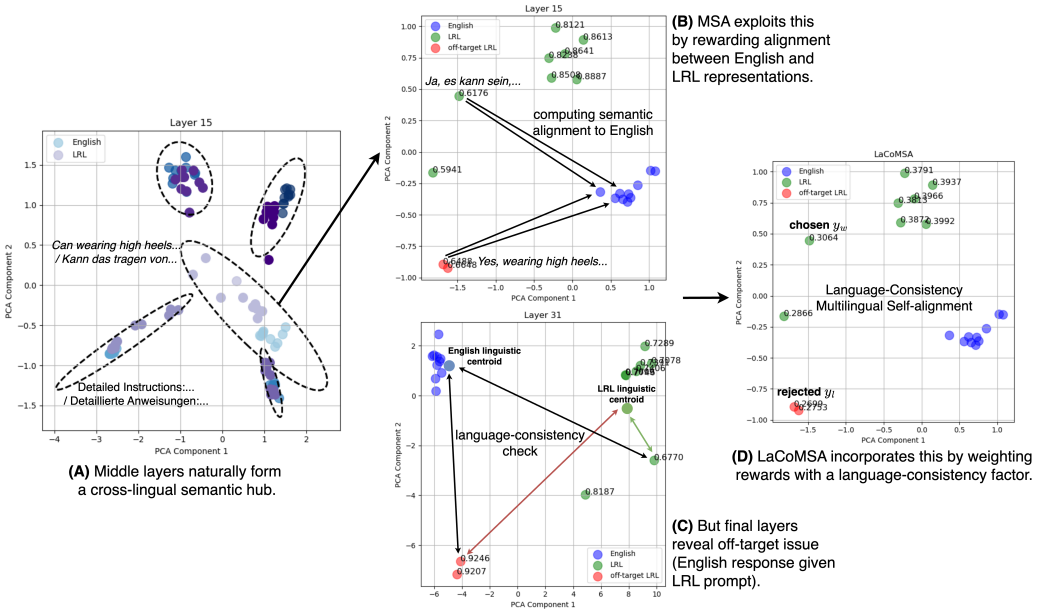

Figure 1: Representation visualizations of responses sampled from Llama-3-8B-SFT-DPO for 5 UltraFeedback prompts. (A) Middle layers form a semantic hub. (B) MSA rewards outputs aligned with English representations. (C) Deeper layers reveal off-target English drift. (D) LaCoMSA incorporates a final-layer consistency weight.

MSA(**LaCoMSA**), addressing the off-target language issue by augmenting the reward with a final-layer language-consistency factor. By comparing each output against centroids of English and the target LRL, LaCoMSA downweights cases where the model drifts back into English despite high semantic similarity at middle layers. Figure 1 illustrates an example, demonstrating LLMs already exhibit cross-lingual semantic alignment by grouping responses across languages by their source questions, LaCoMSA leverages these internal semantic hubs as intrinsic rewards and further enforces language consistency by filtering out such off-target cases. Importantly, LaCoMSA is self-contained: no external annotation, translation, or embedding models are required, as both candidate outputs and preference signals are derived directly from the LLM. LaCoMSA integrates seamlessly with preference optimization methods such as Direct Preference Optimization (DPO) (Rafailov et al., 2023) and Iterative DPO (Yuan et al., 2024), offering a scalable path toward aligning multilingual LLMs.

Our contributions:

- We introduce **Multilingual Self-Alignment (MSA)**, leveraging an LLM's own latent representations as intrinsic supervision signals. This allows the model to generate outputs in multiple languages for the same prompt, compare their semantic similarity, and create preference pairs by itself.
- We propose **Language-Consistency MSA (LaCoMSA)**, which augments MSA with a final-layer language-consistency factor. This extension explicitly addresses the off-target language problem by rewarding outputs that are both semantically aligned and correctly realized in the target LRL.
- We empirically validate the effectiveness of the proposed approaches. Integrated with DPO and Iterative DPO, LaCoMSA improves up to 6.8% length-controlled win-rate (55.0% relative) on X-AlpacaEval, and achieves consistent gains across multilingual benchmarks and model sizes, demonstrating an effective approach to multilingual self-alignment.

Our source code, model weights, and data are publicly available at: https://github.com/ReML-AI/LaCoMSA.

2 Related Works

2.1 Cross-Lingual Semantic Alignment

Cross-lingual alignment and representation learning have been extensively studied in encoder-based transformer models. The comprehensive survey by (Hämmerl et al., 2024) shows that cross-lingual generalization arises from shared latent subspaces across languages. Methods such as LaBSE (Feng et al., 2022) and InfoXLM (Chi et al., 2021) explicitly maximize cross-lingual similarity through

contrastive or translation-based objectives, leveraging parallel data. Recent work extends cross-lingual interpretability to decoder-only LLMs (Tran et al., 2025). (Wendler et al., 2024) employ an early-exiting technique (logit lens) to show that Llama-2 remains "English-centric" until late decoding layers, where surface-level tokens from the target language finally emerge. (Wu et al., 2025) identifies a "semantic hub" in the middle layers, where semantically equivalent sentences from different languages converge in the embedding space. Similarly, MEXA (Kargaran et al., 2025) and NeuronXA (Huang et al., 2025) demonstrate that cross-lingual retrieval scores between parallel representations at these layers strongly correlate with downstream multilingual task performance. These findings suggest that LLMs implicitly encode language-independent semantics in their middle layers, even when their outputs remain language-biased.

However, prior work has largely been diagnostic: studies observe alignment but do not use it as a training signal. Explicitly enforcing alignment through, e.g., contrastive learning, can lead to degeneration or off-target generation (Hämmerl et al., 2024; Li et al., 2024). Our work takes a different direction by *operationalizing* middle-layer alignment into a self-aligned reward function for preference optimization, transforming these latent signals as rewards for reinforcement learning.

2.2 Multilingual Alignment with Preference Optimization

Recent multilingual alignment methods focus on preference optimization frameworks. Multilingual Alignment-as-Preference Optimization (MAPO) (She et al., 2024) constructs preference pairs using a translation model that scores reasoning-chain consistency between low- and high-resource languages. Language Imbalance Driven Reward (LIDR) (Yang et al., 2025b) uses the LLM itself to perform both generation and translation, assuming that a translation of the HRL response into the LRL is superior to the original LRL generation. Similar works also explore the use of translation systems for supervision (Zhang et al., 2024b; Ranaldi et al., 2024). Implicit Cross-Lingual Reward (ICR) (Yang et al., 2025a) extends this idea, leveraging an English-aligned DPO model as a universal preference judge to directly score LRL responses without translation, showing stronger performance than both MAPO and LIDR.

While these approaches effectively improve multilingual performance, they depend on external supervision such as translation systems or English-centric reward models, which are costly and often biased toward dominant languages. In contrast, our proposed approach uses the model's *own latent representations* to generate preference signals, rewarding outputs that are both semantically aligned and linguistically consistent. This approach bridges cross-lingual interpretability and preference optimization, opening a path toward scalable and resource-efficient multilingual alignment.

3 Method

We first introduce **Multilingual Self-Alignment (MSA)**, a targeted preference optimization approach that leverages the LLM's *own latent representations* as a source of preference signals. The key hypothesis underlying MSA is that LLMs naturally encode language-independent semantic content at the *middle layers*, which can serve as an intrinsic measure of quality for LRL outputs. Empirical studies show that hidden representations extracted from middle layers across different languages exhibit high cross-lingual similarity, reflecting the model's representation in a shared semantic space that is relatively invariant to surface-level linguistic variation. We exploit this property by rewarding an LRL output proportionally to how closely its latent representation aligns with the HRL (English) outputs generated for the same prompt.

3.1 Representation-Based Preference Signal

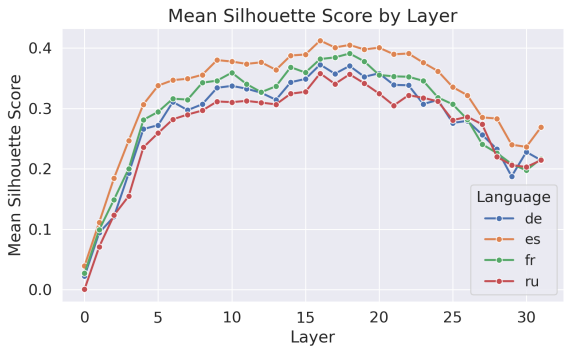

Figure 2: Silhouette scores of latent representations across layers of Llama-3-8B-SFT-DPO. The peak in the middle layers reveals a semantic hub, while later layers diverge by language to realize surface forms.

Formally, given a prompt in an LRL, x^{LRL}, we sample an output y^{LRL} and extract its hidden representation at the middle layer l, denoted as y_l^{LRL}.

For the corresponding parallel prompt in English, x^{en}, we generate N outputs $\{y_{l,i}^{\text{en}}\}_{i=1}^{N}$ and collect their hidden representations at the same layer l. The reward assigned to the LRL output is then defined as the average cosine similarity between its representation and the English counterparts:

$$r_{\text{MSA}}(y_l^{\text{LRL}}) = \frac{1}{N}\sum_{i=1}^{N} \text{sim}(y_l^{\text{LRL}}, y_{l,i}^{\text{en}}), \quad (1)$$

where $\text{sim}(\cdot,\cdot)$ denotes a distance function. In this work, we choose cosine similarity as the distance metric, following prior works on computing semantic alignment (Tran et al., 2024; Hämmerl et al., 2024; Wu et al., 2025). Intuitively, this reward captures how well the semantic representation of the LRL output aligns with that of the HRL outputs in a shared latent space.

To validate this design, we provide new evidence for the semantic hub hypothesis, which posits that middle layers encode language-agnostic semantics. We examine whether hidden representations of semantically equivalent outputs cluster more strongly by question than by language identity. At each layer l, we compute silhouette scores (Rousseeuw, 1987) that measure clustering quality when responses are grouped by source question x. Using 100 random questions from the UltraFeedback dataset, with 10 outputs generated per language, we repeat the procedure five times for robustness. Results (Figure 2) for Llama-3-8B-SFT-DPO show a pronounced clustering peak in the middle layers, with much weaker semantic clustering at the final layer. Unlike earlier studies that rely on parallel corpora, our analysis generates parallel responses from the LLM itself, showing that the semantic hub effect persists even when the parallelism arises from the model's own generative behavior. Motivated by this, in our experiments with Llama 3 8B-based models, we fix the alignment layer to the middle layer, layer 15 ($l_{\text{semantic_hub}} = \lfloor L/2 \rfloor$), and the language centroid layer to the last, layer 31 ($l_{\text{lang_centroid}} = L$). This provides a strong balance of semantic alignment and language consistency, while reducing hyperparameter search and improving reproducibility. We leave adaptive layer selection to future work.

3.2 Language-Consistency Multilingual Self-Alignment

While middle-layer alignment provides a powerful signal, it is not without limitations. Prior analyses have shown that LLMs often fail at the *translation stage* in the final layers (off-target language issue (Gu et al., 2019; Sennrich et al., 2024; Tang et al., 2024)), producing outputs in English even when prompted in an LRL. This phenomenon arises because the model's late layers tend to prioritize surface-level realization in a dominant training language. As a result, if MSA relies solely on middle-layer similarity, the model might incorrectly reward English responses to LRL prompts, thereby reinforcing undesirable behavior.

To mitigate this issue, we introduce **Language-Consistency Multilingual Self-Alignment (LaCoMSA)**, which incorporates a weighting factor based on final-layer representations, as shown in Figure 3. Specifically, we measure whether the final output representation aligns more strongly with the centroid of the LRL than with the centroid of English. For each language λ, we compute a centroid embedding c^{λ} at the final layer L, obtained by averaging pooled representations of monolingual sentences. For an LRL output y^{LRL}, its language-consistency score is:

$$\phi(y_L^{\text{LRL}}) = \text{sim}(y_L^{\text{LRL}}, c^{\text{LRL}}) - \text{sim}(y_L^{\text{LRL}}, c^{\text{en}}). \quad (2)$$

This score is positive when the output is closer to the target lower-resource language and negative when it drifts toward English. We then define a soft weight to adjust the middle-layer MSA reward:

$$w(y_{l,L}^{\text{LRL}}) = \sigma(\alpha[\phi(y_L^{\text{LRL}}) - r_{\text{MSA}}(y_l^{\text{LRL}})])^{\gamma}, \quad (3)$$

where $\sigma(\cdot)$ is the sigmoid function, α controls the sharpness of weighting, and $\gamma \geq 1$ amplifies the effect of high-confidence cases. Finally, the adjusted reward is given by:

$$r_{\text{LaCoMSA}}(y_{l,L}^{\text{LRL}}) = w(y_{l,L}^{\text{LRL}}) \cdot r_{\text{MSA}}(y_l^{\text{LRL}}). \quad (4)$$

This formulation ensures that the reward signal favors outputs that are semantically aligned and consistent with the target language, while penalizing degenerate cases where the model produces English responses to LRL prompts.

3.3 Integration with Reinforcement Learning

The adjusted reward r_{LaCoMSA} can be seamlessly integrated into reinforcement learning fine-tuning frameworks. In practice, reinforcement learning from human or model feedback often faces two challenges: designing a stable reward model and mitigating reward hacking. Direct Preference Optimization (DPO) (Rafailov et al., 2023) provides

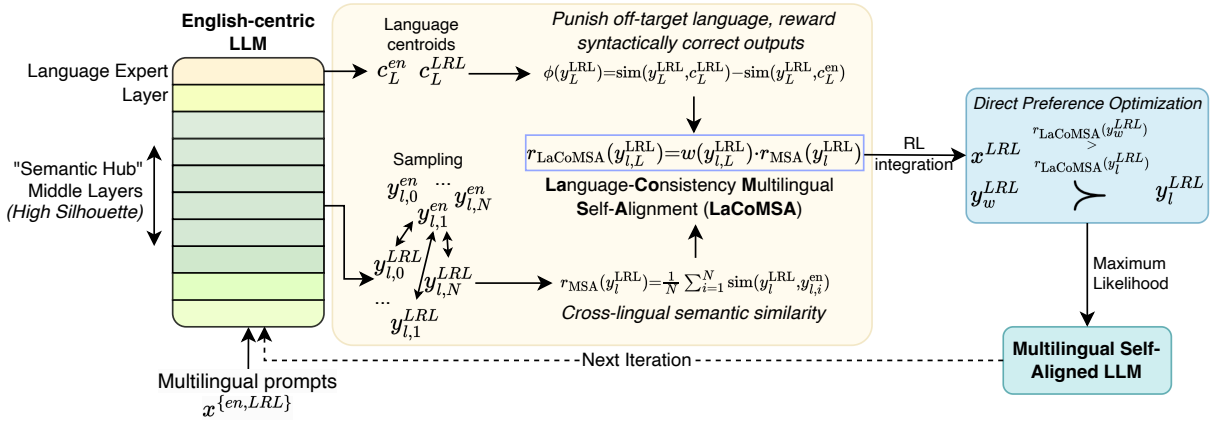

Figure 3: Overview of our proposed framework. MSA computes cross-lingual semantic similarity between middle-layer representations of English and LRL outputs, while LaCoMSA adds a final-layer language-consistency weight to penalize off-target. The weighted reward is optimized iteratively via DPO for self-aligned multilingual LLM.

a parameterized solution by eliminating the need for a separately trained reward model, instead optimizing the policy directly from pairwise preferences. Given a dataset of prompts x and pairs of outputs (y_w, y_l) sampled from the LLM, where y_w is preferred over y_l according to r_{LaCoMSA}, $r_{\text{LaCoMSA}}(y_w) > r_{\text{LaCoMSA}}(y_l)$. DPO directly trains the optimal model by maximizing the likelihood of these pairwise preferences using the following objective:

$$\mathcal{L}_{\text{DPO}}(\pi_\theta) = -\mathbb{E}_{(x,y_w,y_l)\sim\mathcal{D}}\Big[\log\sigma\Big(\beta\big[\log\frac{\pi_\theta(y_w|x)}{\pi_{\text{ref}}(y_w|x)} - \log\frac{\pi_\theta(y_l|x)}{\pi_{\text{ref}}(y_l|x)}\big]\Big)\Big], \quad (5)$$

where π_θ is the current policy, π_{ref} is the reference model, and β controls the temperature of the preference sharpness. To further enhance alignment stability, we incorporate a negative log-likelihood (NLL) term into the standard DPO objective, encouraging the model to maintain high-likelihood generations for preferred responses:

$$\mathcal{L}_{\text{NLL}} = -\mathbb{E}_{(x,y_w)\sim\mathcal{D}}\Big[\log\frac{\pi_\theta(y_w|x)}{|y_w|}\Big]. \quad (6)$$

where $|y_w|$ denotes the token length of the preferred output. The final training objective combines the two components: $\mathcal{L} = \mathcal{L}_{\text{DPO}} + \gamma\mathcal{L}_{\text{NLL}}$, where γ is a hyper-parameter. After DPO training, the updated policy model π^*_θ can be leveraged to sample responses and score with LaCoMSA for subsequent iterations of self-alignment (Yuan et al., 2024).

4 Experimental setup

Base models. We primarily use Llama-3-8B-SFT-DPO (Meng et al., 2024) for multilingual alignment, chosen for its open-weight and reproducible post-training pipeline. It is derived from Meta-Llama-3-8B (AI@Meta, 2024) via supervised fine-tuning on UltraChat-200k (Ding et al., 2023) followed by DPO on UltraFeedback (Cui et al., 2024). To confirm generalization, we also experiment with Llama-3-8B-Instruct (AI@Meta, 2024), where details of training data, including languages, for the post-alignment pipeline are not available.

Datasets. To ensure comparability with prior work, we adopt the same sets of training prompts. For Llama-3-8B-SFT-DPO, we leverage the 3k prompts subset of multilingual UltraFeedback dataset (Yang et al., 2025a), a large-scale instruction set. For Llama3-8B-Instruct, we use the same 1k subsets of prompts from Alpagasus (Chen et al., 2024) as (Yang et al., 2025b). Both datasets consist of 5 languages, ranging from high to low-resources: English, Spanish, Russian, German and French.

Sampling and Reward Computation. For each prompt, we sample $N = 10$ responses using temperature 0.9 and top-p $= 1.0$. Following prior analyses, we select the middle layer, layer 15, for semantic representation and MSA reward computation, and layer 31 (the final layer) to compute language centroids for the language-consistency factor in LaCoMSA, setting α and γ in Eq. 3 to 0.5 and 1.0 in all cases for simplicity.

Training Details. Our code is implemented using the TRL library (von Werra et al., 2020) and DeepSpeed (Rasley et al., 2020) on 1 NVIDIA A100 80GB GPU. Following (She et al., 2024; Yang et al., 2025a), we set the training duration to one epoch with a learning rate of $1e-6$, using AdamW optimizer (Loshchilov and Hutter, 2019) with a cosine scheduler. For DPO, we set $\beta = 0.1$

X-AlpacaEval		en	es	ru	de	fr	Avg (ΔBaseline)
Llama-3-8B-SFT-DPO (Meng et al., 2024)		17.24	11.32	11.05	10.17	11.56	12.27
Iteration 1	LIDR* (Yang et al., 2025b)	18.69	13.99	12.68	11.31	12.86	13.91 (↑1.64)
	ICR (Yang et al., 2025a)	20.46	14.52	**16.00**	14.54	17.08	16.52 (↑4.25)
	MSA (OURs)	25.40	10.63	6.99	8.32	10.65	12.51 (↑0.24)
	LaCoMSA (OURs)	**25.96**	**18.72**	13.72	**16.82**	**19.12**	**18.76 (↑6.49)**
Iteration 2	ICR	21.19	16.88	**18.11**	**17.92**	17.12	18.24 (↑5.97)
	MSA (OURs)	25.96	1.53	1.13	1.85	1.39	6.37 (↓5.90)
	LaCoMSA (OURs)	**22.75**	**18.63**	14.69	**17.92**	**18.61**	**19.08 (↑6.81)**

* Results for LIDR and ICR are from (Yang et al., 2025a), which reports only the best LIDR iteration, since the second degraded in most languages.

Table 1: Result (Length-controlled win rate) on X-AlpacaEval with GPT-4 as the judge. Bold scores indicate the best performance per iteration of DPO training.

and follow (Pang et al., 2024) in setting the weight $\gamma = 1.0$ for the auxiliary NLL term.

Baselines. We compare our method against SoTA multilingual alignment approaches:

- **MSA (ablation).** Ours without language-consistency factor, using only middle-layer semantic alignment as reward signal.
- **LIDR** (Yang et al., 2025b). Constructs preference pairs through LLM translation: translated English responses are treated as preferred over native LRL outputs, while for English prompts the direction is reversed.
- **ICR** (Yang et al., 2025a). Uses an English preference-aligned model as a judge, transferring its learned preference signal to LRLs.

Evaluation datasets.

- **Instruction following.** We use the X-AlpacaEval leaderboard (Zhang et al., 2024c), a multilingual extension of AlpacaEval 2.0 (Li et al., 2023), highly correlated with humans (a spearman correlation of 0.98) using GPT-4 judge. To mitigate length bias, we report the length-controlled Win Rate (LC) against GPT-4-Turbo responses.
- **Conversational quality.** We use Multilingual MT-Bench (Yang et al., 2025b; Zheng et al., 2023) with open-ended dialogue and prompts. Responses are rated by GPT-4o on 1–10 scale.
- **Multilingual NLP Benchmarks.** To measure whether preference alignment harms general abilities, we evaluate on multilingual versions of standard benchmarks: MMLU (Hendrycks et al., 2021), HellaSwag (Zellers et al., 2019), ARC Challenge (Clark et al., 2018), and TruthfulQA (Lin et al., 2022).

5 Results and analysis

5.1 Effective Self-Alignment

Table 1 reports the main evaluation results on X-AlpacaEval. LaCoMSA improves the multilingual capability of the base model by up to 6.81% absolute win rate (12.27% to 19.08%, a relative gain of 55%) without requiring any manual preference annotations. Improvements are consistent and statistically significant (two-tailed t-test, $p < 0.01$) across all five training languages, including the anchor language, English, where performance rises from 17.24% to 25.96% after the first iteration. This shows LaCoMSA effectively improves generative capability across all languages.

Iterative preference optimization further amplifies the gains: average performance increases from 12.27% to 18.76% in the first iteration, and to 19.08% in the second. Notably, compared to other alignment baselines, LaCoMSA, using only latent representations as reward signal, achieves the strongest performance in both training iterations, outperforming ICR (which relies on English-trained preference models) and LIDR (which depends on explicit translation prompts).

The effectiveness of LaCoMSA generalizes across benchmarks and models. On Multilingual MT-Bench (Table 2), LaCoMSA improves by 1.10 points over the base model, nearly double the gain of the strongest baseline ICR (0.57). Furthermore, as shown in Table 3, results also hold across base models: with Llama3-8B-Instruct, LaCoMSA achieves state-of-the-art performance, improving

GPT-4o Judge		en	es	ru	de	fr	Avg (ΔBaseline)
Llama-3-8B-SFT-DPO (Meng et al., 2024)		6.86	5.96	6.01	5.93	6.23	6.20
Iteration 1	ICR (Yang et al., 2025a)	6.93	6.61	6.42	6.76	6.56	6.66 (↑0.44)
	MSA	7.23	6.48	**8.18**	**7.09**	7.00	7.20 (↑1.00)
	LaCoMSA	**7.51**	**6.80**	7.96	7.00	**7.24**	**7.30 (↑1.10)**
Iteration 2	ICR (Yang et al., 2025a)	7.02	**6.96**	6.44	6.75	6.68	6.77 (↑0.57)
	MSA	7.15	6.66	7.78	7.06	6.30	6.99 (↑0.79)
	LaCoMSA	**7.53**	6.69	**7.80**	**7.09**	**6.86**	**7.19 (↑0.99)**

Table 2: Result on MT-bench as judged by GPT-4o. Bold scores indicate the best performance per DPO iteration.

X-AlpacaEval - Avg (ΔBaseline)		Llama-3-8B-SFT-DPO (Meng et al., 2024)	Llama3-8B-Instruct (AI@Meta, 2024)
Base Model		12.27	13.74
Iteration 1	LIDR*	13.91 (↑1.64)	17.59 (↑3.85)
	ICR	16.52 (↑4.25)	20.10 (↑6.36)
	MSA (OURs)	12.51 (↑0.24)	9.85 (↓3.89)
	LaCoMSA (OURs)	**18.76 (↑6.49)**	**21.21 (↑7.47)**
Iteration 2	LIDR*	13.91 (↑1.64)	16.79 (↑3.05)
	ICR	18.24 (↑5.97)	22.02 (↑8.28)
	MSA (OURs)	6.37 (↓5.90)	7.52 (↓6.22)
	LaCoMSA (OURs)	**19.08 (↑6.81)**	**22.07(↑8.33)**

Table 3: Length-controlled win rate on X-AlpacaEval across multiple base LLMs. Bold scores indicate the best performance per model and iteration of DPO training.

up to 8.33% win rate over the baseline. Given that prior work has identified semantic hubs across different LLMs, our results demonstrate that latent representation alignment can be effectively leveraged to improve multilingual capability without external annotation.

5.2 Ablation: MSA vs. LaCoMSA

Ablation studies confirm the necessity of the language-consistency factor in LaCoMSA. While MSA improves standalone evaluations such as MT-Bench in Table 2, it underperforms on X-AlpacaEval (Table 3) because off-target English outputs are penalized when compared against reference responses (from GPT-4-Turbo) in the target language. LaCoMSA resolves this issue by explicitly encouraging fidelity to the target LRL, yielding consistent improvements across all benchmarks. This highlights the expressive advantage of LaCoMSA: generating semantically aligned and high-quality responses in the correct language.

5.3 Language Fidelity

We further assess output language correctness using the NLLB language identification model (Team et al., 2022), following prior work (Bafna et al., 2025; Zhang et al., 2024a). As shown in Table 5, MSA fails to produce responses in the target language, and performance degrades with more iterations. In contrast, LaCoMSA achieves robust improvements, raising the average correct-language rate to 90.78%. For example, the base model produces only 57% Russian responses correctly, while LaCoMSA exceeds 80%. These results confirm our hypothesis: incorporating language-consistency directly addresses off-target failures.

Model	l	en+de	en+fr	en+ru	en+es
Base model	31	0.2166	0.2093	0.1895	0.2286
Base model	15	0.3452	0.3602	0.3208	0.3821
LaCoMSA it.1	15	0.3668	0.3897	0.3473	0.4075
LaCoMSA it.2	15	0.3761	0.4041	0.3639	0.4210

Table 4: Silhouette scores at the middle layer of Llama-3-8B-SFT-DPO across training iterations.

5.4 Semantic Hub Enhancement

Interestingly, LaCoMSA improves the clustering of cross-lingual representations, as measured by Silhouette scores (Table 4). The clustering effect is stronger across all language pairs within each iteration of preference optimization. This occurs even though semantic clustering was not an explicit optimization objective. These findings reinforce

MT-bench		en	es	ru	de	fr	Avg
Llama-3-8B-SFT-DPO		98.75	70.00	57.00	71.25	81.25	75.65
Iteration 1	MSA	98.75	37.50	30.00	31.88	45.00	48.63
	LaCoMSA	**99.38**	83.75	71.00	81.88	89.38	85.08
Iteration 2	MSA	**99.38**	35.63	14.00	26.25	53.75	45.80
	LaCoMSA	98.75	**90.00**	**73.00**	**87.50**	**92.50**	**88.35**
X-AlpacaEval		**en**	**es**	**ru**	**de**	**fr**	**Avg**
Llama-3-8B-SFT-DPO		99.75	75.15	70.93	73.04	84.22	80.62
Iteration 1	MSA	99.75	34.41	30.31	37.52	50.43	50.48
	LaCoMSA	**99.86**	84.35	78.63	82.73	90.19	87.15
Iteration 2	MSA	99.88	0.62	0.12	0.37	1.37	20.47
	LaCoMSA	99.63	**89.57**	**83.98**	**87.20**	**93.54**	**90.78**

Table 5: Language fidelity rate using the NLLB language detection model (Team et al., 2022).

Model		Multilingual ARC (0-shot)	Multilingual HellaSwag (0-shot)	Multilingual MMLU (5-shot)	Multilingual TruthfulQA	
					MC1 (0-shot)	MC2 (0-shot)
Llama-3-8B-SFT-DPO		0.4653	0.5231	0.5349	0.3479	0.5079
Iteration 1	MSA	0.4647	0.5204	0.5525	0.3489	0.5104
	LaCoMSA	0.4626	0.5218	0.5528	0.3465	0.5087
Iteration 2	MSA	0.4615	0.5196	0.5524	0.3482	0.5099
	LaCoMSA	0.4564	0.5210	0.5511	0.3424	0.5034

Table 6: Result on multilingual language understanding benchmarks.

evidence from prior work of a shared semantic hub, suggesting that LaCoMSA can be extended beyond multilingual alignment to other under-represented domains and modalities.

5.5 Multilingual NLP Benchmarks

To examine potential degradation of general knowledge, we evaluate on multilingual versions of 4 language understanding benchmarks. As shown in Table 6, LaCoMSA introduces negligible performance loss compared to the base model, and even improves approximately 2% in the case of Multilingual MMLU. We attribute this stability to our method's focus on reinforcing semantic hub representation alignment rather than forcing instruction following capability in place of general knowledge.

5.6 Varying Number of Training Samples

Figure 4 presents X-AlpacaEval results with varying numbers of training prompts per language. LaCoMSA exhibits positive scaling: with only 1,000 samples per language, it still yields consistent improvements of 4.58% (37.33% relatively) over the baseline, demonstrating robustness in low-resource settings. Performance continues to improve as data increases from 3,000 (default setting) to 5,000 prompts, indicating that further gains are achievable with more training data and compute.

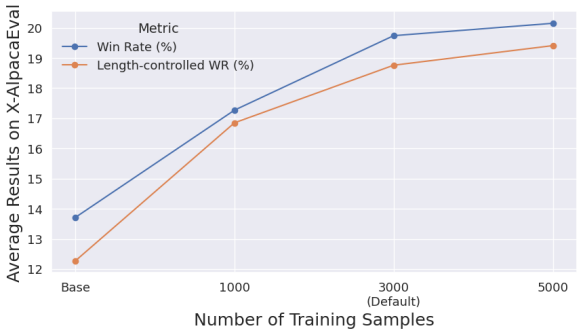

Figure 4: Results on X-AlpacaEval of LaCoMSA with varying number of training samples, highlighting its efficiency in various resource regimes.

6 Conclusion

We propose LaCoMSA, a novel approach that leverages LLM's latent representations for multilingual alignment through preference optimization. Integrated with DPO and Iterative DPO, LaCoMSA improves a Llama 3 8B-based model's multilingual win rates by up to 6.8 points (55.0% relative) on X-AlpacaEval and achieves consistent gains across benchmarks and model architectures. Our findings demonstrate LaCoMSA can serve as an effective and scalable mechanism for multilingual preference optimization, opening a path toward multilingual self-alignment.

Acknowledgments

This publication has emanated from research supported in part by grants from Research Ireland under Grant [12-RC-2289-P2] and [18/CRT/6223] which is co-funded under the European Regional Development Fund. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Limitations

Our experiments focus on DPO and Iterative DPO for fast iteration, though in principle LaCoMSA can be integrated into any reinforcement learning algorithms, such as RLHF. One limitation of our work is its focus on general multilingual alignment: LaCoMSA relies on the model's own internal representations, and such latent semantic alignment may not always capture culturally grounded or context-specific meanings. Extending LaCoMSA to incorporate language-specific, cultural alignment remains an important direction for future research.

References

AI@Meta. 2024. Llama 3 model card.

Firoj Alam, Shammur Absar Chowdhury, Sabri Boughorbel, and Maram Hasanain. 2024. LLMs for low resource languages in multilingual, multimodal and dialectal settings. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: Tutorial Abstracts*, pages 27–33, St. Julian's, Malta. Association for Computational Linguistics.

Niyati Bafna, Tianjian Li, Kenton Murray, David R. Mortensen, David Yarowsky, Hale Sirin, and Daniel Khashabi. 2025. The translation barrier hypothesis: Multilingual generation with large language models suffers from implicit translation failure. *Preprint*, arXiv:2506.22724.

Lichang Chen, Shiyang Li, Jun Yan, Hai Wang, Kalpa Gunaratna, Vikas Yadav, Zheng Tang, Vijay Srinivasan, Tianyi Zhou, Heng Huang, and Hongxia Jin. 2024. Alpagasus: Training a better alpaca with fewer data. In *The Twelfth International Conference on Learning Representations*.

Zewen Chi, Li Dong, Furu Wei, Nan Yang, Saksham Singhal, Wenhui Wang, Xia Song, Xian-Ling Mao, Heyan Huang, and Ming Zhou. 2021. InfoXLM: An information-theoretic framework for cross-lingual language model pre-training. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3576–3588, Online. Association for Computational Linguistics.

Peter Clark, Isaac Cowhey, Oren Etzioni, Tushar Khot, Ashish Sabharwal, Carissa Schoenick, and Oyvind Tafjord. 2018. Think you have solved question answering? try arc, the ai2 reasoning challenge. *Preprint*, arXiv:1803.05457.

Ganqu Cui, Lifan Yuan, Ning Ding, Guanming Yao, Wei Zhu, Yuan Ni, Guotong Xie, Zhiyuan Liu, and Maosong Sun. 2024. Ultrafeedback: Boosting language models with high-quality feedback.

Ning Ding, Yulin Chen, Bokai Xu, Yujia Qin, Shengding Hu, Zhiyuan Liu, Maosong Sun, and Bowen Zhou. 2023. Enhancing chat language models by scaling high-quality instructional conversations. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 3029–3051, Singapore. Association for Computational Linguistics.

Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Arivazhagan, and Wei Wang. 2022. Language-agnostic BERT sentence embedding. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 878–891, Dublin, Ireland. Association for Computational Linguistics.

Jiatao Gu, Yong Wang, Kyunghyun Cho, and Victor O.K. Li. 2019. Improved zero-shot neural machine translation via ignoring spurious correlations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1258–1268, Florence, Italy. Association for Computational Linguistics.

Katharina Hämmerl, Jindřich Libovický, and Alexander Fraser. 2024. Understanding cross-lingual Alignment—A survey. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 10922–10943, Bangkok, Thailand. Association for Computational Linguistics.

Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. 2021. Measuring massive multitask language understanding. *Proceedings of the International Conference on Learning Representations (ICLR)*.

Carolin Holtermann, Paul Röttger, Timm Dill, and Anne Lauscher. 2024. Evaluating the elementary multilingual capabilities of large language models with MultiQ. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 4476–4494, Bangkok, Thailand. Association for Computational Linguistics.

Chongxuan Huang, Yongshi Ye, Biao Fu, Qifeng Su, and Xiaodong Shi. 2025. From neurons to semantics: Evaluating cross-linguistic alignment capabilities of large language models via neurons alignment. In

Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 28956–28974, Vienna, Austria. Association for Computational Linguistics.

Amir Hossein Kargaran, Ali Modarressi, Nafiseh Nikeghbal, Jana Diesner, François Yvon, and Hinrich Schuetze. 2025. MEXA: Multilingual evaluation of English-centric LLMs via cross-lingual alignment. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 27001–27023, Vienna, Austria. Association for Computational Linguistics.

Chong Li, Shaonan Wang, Jiajun Zhang, and Chengqing Zong. 2024. Improving in-context learning of multilingual generative language models with cross-lingual alignment. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 8058–8076, Mexico City, Mexico. Association for Computational Linguistics.

Xuechen Li, Tianyi Zhang, Yann Dubois, Rohan Taori, Ishaan Gulrajani, Carlos Guestrin, Percy Liang, and Tatsunori B. Hashimoto. 2023. Alpacaeval: An automatic evaluator of instruction-following models. `https://github.com/tatsu-lab/alpaca_eval`.

Stephanie Lin, Jacob Hilton, and Owain Evans. 2022. TruthfulQA: Measuring how models mimic human falsehoods. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3214–3252, Dublin, Ireland. Association for Computational Linguistics.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. In *International Conference on Learning Representations*.

Yu Meng, Mengzhou Xia, and Danqi Chen. 2024. Simpo: Simple preference optimization with a reference-free reward. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Richard Yuanzhe Pang, Weizhe Yuan, He He, Kyunghyun Cho, Sainbayar Sukhbaatar, and Jason E Weston. 2024. Iterative reasoning preference optimization. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. In *Thirty-seventh Conference on Neural Information Processing Systems*.

Leonardo Ranaldi, Giulia Pucci, and Andre Freitas. 2024. Empowering cross-lingual abilities of instruction-tuned large language models by translation-following demonstrations. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 7961–7973, Bangkok, Thailand. Association for Computational Linguistics.

Jeff Rasley, Samyam Rajbhandari, Olatunji Ruwase, and Yuxiong He. 2020. Deepspeed: System optimizations enable training deep learning models with over 100 billion parameters. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '20, page 3505–3506, New York, NY, USA. Association for Computing Machinery.

Peter J. Rousseeuw. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.

Rico Sennrich, Jannis Vamvas, and Alireza Mohammadshahi. 2024. Mitigating hallucinations and off-target machine translation with source-contrastive and language-contrastive decoding. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 21–33, St. Julian's, Malta. Association for Computational Linguistics.

Shuaijie She, Wei Zou, Shujian Huang, Wenhao Zhu, Xiang Liu, Xiang Geng, and Jiajun Chen. 2024. MAPO: Advancing multilingual reasoning through multilingual-alignment-as-preference optimization. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10015–10027, Bangkok, Thailand. Association for Computational Linguistics.

Tianyi Tang, Wenyang Luo, Haoyang Huang, Dongdong Zhang, Xiaolei Wang, Xin Zhao, Furu Wei, and Ji-Rong Wen. 2024. Language-specific neurons: The key to multilingual capabilities in large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5701–5715, Bangkok, Thailand. Association for Computational Linguistics.

NLLB Team, Marta R. Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, Anna Sun, Skyler Wang, Guillaume Wenzek, Al Youngblood, Bapi Akula, Loic Barrault, Gabriel Mejia Gonzalez, Prangthip Hansanti, and 20 others. 2022. No language left behind: Scaling human-centered machine translation. *Preprint*, arXiv:2207.04672.

Khanh-Tung Tran, Barry O'Sullivan, and Hoang Nguyen. 2024. Irish-based large language model with extreme low-resource settings in machine translation. In *Proceedings of the Seventh Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2024)*, pages 193–202, Bangkok, Thailand. Association for Computational Linguistics.

Khanh-Tung Tran, Nguyet-Hang Vu, Barry O'Sullivan, and Hoang D. Nguyen. 2025. Disentangling language understanding and reasoning structures in cross-lingual chain-of-thought prompting. In *Findings of the Association for Computational Linguistics:*

EMNLP 2025, pages 12200–12206, Suzhou, China. Association for Computational Linguistics.

Leandro von Werra, Younes Belkada, Lewis Tunstall, Edward Beeching, Tristan Thrush, Nathan Lambert, Shengyi Huang, Kashif Rasul, and Quentin Gallouédec. 2020. Trl: Transformer reinforcement learning. `https://github.com/huggingface/trl`.

Chris Wendler, Veniamin Veselovsky, Giovanni Monea, and Robert West. 2024. Do llamas work in English? on the latent language of multilingual transformers. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15366–15394, Bangkok, Thailand. Association for Computational Linguistics.

Zhaofeng Wu, Xinyan Velocity Yu, Dani Yogatama, Jiasen Lu, and Yoon Kim. 2025. The semantic hub hypothesis: Language models share semantic representations across languages and modalities. In *The Thirteenth International Conference on Learning Representations*.

Wen Yang, Junhong Wu, Chen Wang, Chengqing Zong, and Jiajun Zhang. 2025a. Implicit cross-lingual rewarding for efficient multilingual preference alignment. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 21125–21147, Vienna, Austria. Association for Computational Linguistics.

Wen Yang, Junhong Wu, Chen Wang, Chengqing Zong, and Jiajun Zhang. 2025b. Language imbalance driven rewarding for multilingual self-improving. In *The Thirteenth International Conference on Learning Representations*.

Weizhe Yuan, Richard Yuanzhe Pang, Kyunghyun Cho, Xian Li, Sainbayar Sukhbaatar, Jing Xu, and Jason Weston. 2024. Self-rewarding language models. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org.

Rowan Zellers, Ari Holtzman, Yonatan Bisk, Ali Farhadi, and Yejin Choi. 2019. HellaSwag: Can a machine really finish your sentence? In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4791–4800, Florence, Italy. Association for Computational Linguistics.

Liang Zhang, Qin Jin, Haoyang Huang, Dongdong Zhang, and Furu Wei. 2024a. Respond in my language: Mitigating language inconsistency in response generation based on large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4177–4192, Bangkok, Thailand. Association for Computational Linguistics.

Xiang Zhang, Senyu Li, Bradley Hauer, Ning Shi, and Grzegorz Kondrak. 2023. Don't trust ChatGPT when your question is not in English: A study of multilingual abilities and types of LLMs. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 7915–7927, Singapore. Association for Computational Linguistics.

Xue Zhang, Yunlong Liang, Fandong Meng, Songming Zhang, Yufeng Chen, Jinan Xu, and Jie Zhou. 2025. Cm-align: Consistency-based multilingual alignment for large language models. *Preprint*, arXiv:2509.08541.

Yuanchi Zhang, Yile Wang, Zijun Liu, Shuo Wang, Xiaolong Wang, Peng Li, Maosong Sun, and Yang Liu. 2024b. Enhancing multilingual capabilities of large language models through self-distillation from resource-rich languages. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11189–11204, Bangkok, Thailand. Association for Computational Linguistics.

Zhihan Zhang, Dong-Ho Lee, Yuwei Fang, Wenhao Yu, Mengzhao Jia, Meng Jiang, and Francesco Barbieri. 2024c. PLUG: Leveraging pivot language in cross-lingual instruction tuning. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7025–7046, Bangkok, Thailand. Association for Computational Linguistics.

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. 2023. Judging llm-as-a-judge with mt-bench and chatbot arena. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA. Curran Associates Inc.

A Additional Results

Table 7 illustrates results on X-AlpacaEval for both Length-controlled Win Rate (LC) and Win Rate (WR) metrics. For each base model, all multilingual alignment approaches follow the same training settings (e.g., data, hyperparamaters) as discussed in Section 4.

Detailed results on Multilingual Language Understanding benchmarks are shown in Table 8, including results per language.

B Prompt Template

The prompt template for LLM-as-a-judge evaluation with GPT-4 on X-AlpacaEval is shown in Figure 5. The prompt is taken from the original benchmark paper, and each turn of evalution includes the instruction, the output of the current model being evaluated, and GPT-4-Turbo output for comparison.

X-AlpacaEval		en		es		ru		de		fr		Avg	
		LC	WR	LC	WR	LC	WR	LC	WR	LC	WR	LC	WR
Llama-3-8B-SFT-DPO		17.24	17.35	11.32	12.41	11.05	13.82	10.17	11.87	11.56	13.09	12.27	13.71
Iteration 1	LIDR*	18.69	20.97	13.99	16.69	12.68	16.60	11.31	15.22	12.86	15.54	13.91	17.00
	ICR	20.46	26.40	14.52	19.49	16.00	22.50	14.54	19.69	17.08	21.20	16.52	21.86
	MSA	25.40	23.74	10.63	10.48	6.99	9.56	8.32	9.30	10.65	11.18	12.51	12.85
	LaCoMSA	25.96	23.02	18.72	18.82	13.72	18.31	16.82	18.34	19.12	20.22	18.76	19.74
Iteration 2	ICR	21.19	31.38	16.88	23.37	18.11	25.76	17.92	26.27	17.12	25.35	18.24	26.43
	MSA	25.96	25.77	1.53	2.03	1.27	2.02	1.85	2.72	1.39	1.90	6.37	6.89
	LaCoMSA	22.75	21.00	18.63	18.29	14.69	18.04	17.92	18.03	18.61	18.74	19.08	19.37
Llama3-8B-Instruct		23.48	24.90	17.52	18.08	6.37	7.81	7.74	8.65	13.58	14.18	13.74	14.72
Iteration 1	LIDR	26.10	30.11	18.78	21.82	14.23	18.01	14.36	16.87	14.49	17.51	17.59	20.86
	ICR	26.80	31.76	21.92	24.03	16.03	19.31	16.79	19.37	18.95	21.49	20.10	23.19
	MSA	30.19	35.03	10.98	13.75	1.38	2.67	2.66	4.25	4.04	5.93	9.85	12.33
	LaCoMSA	34.47	34.41	25.92	28.47	10.38	12.79	13.23	14.90	22.07	24.42	21.21	23.00
Iteration 2	LIDR	27.12	34.09	15.91	21.21	13.53	19.25	12.17	16.02	15.24	20.34	16.79	22.18
	ICR	28.12	35.32	23.21	26.37	19.17	23.01	19.12	24.69	20.46	26.82	22.02	27.24
	MSA	29.92	35.29	3.02	4.70	1.36	2.84	1.76	3.21	1.55	2.63	7.52	9.73
	LaCoMSA	34.29	33.97	26.32	28.78	12.24	12.28	14.21	15.58	23.30	25.24	22.07	23.17

* Results for LIDR and ICR are from (Yang et al., 2025a), which reports only the best LIDR iteration on Llama-3-8B-SFT-DPO, since the second degraded in most languages.

Table 7: Result on X-AlpacaEval with GPT-4 as judge.

Multilingual ARC challenge, 0-shot	**en**	**es**	**ru**	**de**	**fr**	**Avg**
Llama-3-8B-SFT-DPO	0.5819	0.4598	0.3995	0.4140	0.4713	0.4653
ICR	0.5742	0.4684	0.4021	0.4234	0.4713	0.4679
MSA iteration 1	0.5785	0.4573	0.3995	0.4149	0.4731	0.4647
MSA iteration 2	0.5725	0.4564	0.4003	0.4063	0.4722	0.4615
LaCoMSA iteration 1	0.5734	0.4547	0.3978	0.4149	0.4722	0.4626
LaCoMSA iteration 2	0.5708	0.4547	0.3944	0.4038	0.4585	0.4564
Multilingual HellaSwag, 0-shot	**en**	**es**	**ru**	**de**	**fr**	**Avg**
Llama-3-8B-SFT-DPO	0.6292	0.5270	0.4624	0.4864	0.5104	0.5231
ICR	0.6301	0.5304	0.4655	0.4899	0.5114	0.5255
MSA	0.6275	0.5249	0.4595	0.4830	0.5071	0.5204
MSA iteration 2	0.6267	0.5232	0.4583	0.4828	0.5070	0.5196
LaCoMSA iteration 1	0.6275	0.5253	0.4617	0.4863	0.5083	0.5218
LaCoMSA iteration 2	0.6261	0.5243	0.4621	0.4856	0.5071	0.5210
Multilingual MMLU, 5-shot	**en**	**es**	**ru**	**de**	**fr**	**Avg**
Llama-3-8B-SFT-DPO	0.6232	0.5301	0.4883	0.5108	0.5223	0.5349
ICR	0.6236	0.5293	0.4853	0.5103	0.5297	0.5356
MSA iteration 1	0.6189	0.5587	0.5111	0.5296	0.5440	0.5525
MSA iteration 2	0.6204	0.5596	0.5090	0.5299	0.5431	0.5524
LaCoMSA iteration 1	0.6205	0.5584	0.5116	0.5299	0.5436	0.5528
LaCoMSA iteration 2	0.6203	0.5576	0.5087	0.5278	0.5413	0.5511
Multilingual TruthfulQA MC1, 0-shot	**en**	**es**	**ru**	**de**	**fr**	**Avg**
Llama-3-8B-SFT-DPO	0.3856	0.3232	0.3452	0.3363	0.3494	0.3479
ICR	0.3966	0.3321	0.3363	0.3350	0.3443	0.3489
MSA iteration 1	0.3880	0.3295	0.3414	0.3363	0.3494	0.3489
MSA iteration 2	0.3868	0.3270	0.3439	0.3312	0.3520	0.3482
LaCoMSA iteration 1	0.3819	0.3295	0.3414	0.3351	0.3444	0.3465
LaCoMSA iteration 2	0.3794	0.3283	0.3312	0.3274	0.3456	0.3424
Multilingual TruthfulQA MC2, 0-shot	**en**	**es**	**ru**	**de**	**fr**	**Avg**
Llama-3-8B-SFT-DPO	0.5354	0.4811	0.5173	0.4913	0.5146	0.5079
ICR	0.5460	0.4848	0.5163	0.4931	0.5094	0.5099
MSA iteration 1	0.5364	0.4841	0.5197	0.4952	0.5166	0.5104
MSA iteration 2	0.5376	0.4841	0.5207	0.4930	0.5143	0.5099
LaCoMSA iteration 1	0.5345	0.4864	0.5169	0.4950	0.5107	0.5087
LaCoMSA iteration 2	0.5304	0.4827	0.5099	0.4896	0.5042	0.5034

Table 8: Result on Multilingual NLP Benchmarks.

```
<|im_start|>system
You are a highly efficient assistant, who evaluates and selects the best large
    language model (LLMs) based on the quality of their responses to a given
    instruction. This process will be used to create a leaderboard reflecting
    the most accurate and human-preferred answers.
<|im_end|>
<|im_start|>user
I require a leaderboard for various large language models. I'll provide you with
     prompts given to these models and their corresponding outputs. Your task is
     to assess these responses, and select the model that produces the best
    output from a human perspective.

## Instruction

{
    "instruction": """{instruction}""",
}

## Model Outputs

Here are the unordered outputs from the models. Each output is associated with a
     specific model, identified by a unique model identifier.

{
    {
        "model_identifier": "m",
        "output": """{output_1}"""
    },
    {
        "model_identifier": "M",
        "output": """{output_2}"""
    }
}

## Task

A good output should be in the same language as the instruction, except when the
     instruction explicitly requests the output in a different language.
    Evaluate the models based on the quality and relevance of their outputs, and
     select the model that generated the best output. Answer by providing the
    model identifier of the best model. We will use your output as the name of
    the best model, so make sure your output only contains one of the following
    model identifiers and nothing else (no quotes, no spaces, no new lines, ...)
    : m or M.

## Best Model Identifier
<|im_end|>
```

Figure 5: Prompt used in X-AlpacaEval with GPT-4-as-a-judge, including comparison to GPT-4 Turbo results.