PlurVA-LLM-2026 Shared Task Track-1 and 2: Preference Learning for Pluralistic Value Alignment

Joydeb Mondal
Independent Researcher
Team Modgenix

Abstract

We describe the Modgenix submissions to both tracks of the PlurVA-LLM 2026 Shared Task on pluralistic value alignment across China, Indonesia, and Sri Lanka. For the resource-constrained track, we use one dense 7.94B-parameter Gemma 4 E4B model for all three datasets. We recast each option as a binary candidate judgment and train rank-8 LoRA parameters using classification and pairwise preference losses. Deterministic augmentation permutes answer positions, balances languages, and creates a small Chinese core-action view. For the open track, we additionally evaluate a single GPT-5.4 model with versioned multilingual prompts; its best macro accuracy is 0.795. The compact preference-trained system is stronger, so we submit it unchanged to both tracks. It obtains accuracies of 0.788, 0.683, and 0.954 on the Chinese, Indonesian, and Sri Lankan hidden sets, respectively, for a macro-average of 0.808 and first place on both final leaderboards. Error analysis associates candidate preference learning with fewer downside-salience errors, while an objective mismatch between Indonesian annotator votes and consensus evaluation remains a key limitation.

1 Introduction

Value alignment across cultures cannot be reduced to one universal normative target. Pluralistic alignment instead asks systems to represent or adapt to different perspectives and value distributions (Sorensen et al., 2024). The PlurVA-LLM shared task operationalizes this challenge with locally grounded datasets from China, Indonesia, and Sri Lanka. A single system must handle different languages, value taxonomies, and label semantics while being ranked by macro-average country accuracy.

Track 1 requires one model with at most eight billion total parameters for all three datasets and prohibits commercial APIs, agents, retrieval, and other external inference resources. Track 2 retains the one-model requirement but permits open or closed models and external resources. We therefore develop a compact open-weight system for Track 1 and compare it with commercial-model prompting under the open-track rules.

Initial prompt-only experiments revealed an asymmetric problem. The model was already strong on Sri Lankan binary-statement judgments, but Chinese options were difficult because each often paired a plausible core action with a severe stated consequence. Longer decision-rule prompts fixed some examples but created new errors through over-application. Our final method moves the burden from prompt prose to supervised candidate preference learning.

Our contributions are:

- a unified candidate-scoring formulation that accommodates all three country-specific label protocols;
- a parameter-efficient objective combining candidate classification and pairwise ranking, with deterministic language balancing and label-preserving augmentation; and
- a documented system-development comparison between preference training and versioned GPT-5.4 prompting, followed by an analysis of Chinese transfer and the mismatch between Indonesian training positives and official consensus.

2 Task and Data

The shared task combines three datasets while retaining their native task formats. Table 1 reports the released splits used in our system. The development set contains 1,359 labeled items and the hidden evaluation set contains 5,475 items. Accuracy is computed separately for each country and

Proceedings of the First Workshop on Pluralistic Value Alignment of LLMs, pages 98–103
November 10, 2026 ©2026 Association for Computational Linguistics

Dataset	Dev	Test	Options	Gold protocol
Chinese	790	3,210	A–D	one label
Indonesian	366	1,468	A–D	top-vote label(s)
Sri Lankan	203	797	A/B	A, B, Both, or 0
Total	1,359	5,475		

Table 1: Released data used in our experiments.

the official ranking metric is the unweighted mean of the three accuracies.

Chinese. D2VBench contains first-person daily dilemmas grounded in 158 fine-grained value concepts (Hao et al., 2026). Each item asks for one of four actions. The development split has five examples for each fine-grained English value leaf.

Indonesian. Pancasila-Dilemmas grounds four-option scenarios in Religion, Humanity, Unity, Democracy, and Social Justice (Supryadi et al., 2026). Five annotator votes are supplied per development item. Official evaluation accepts the option or tied options with maximum vote count.

Sri Lankan. LKValues is grounded in survey-derived Sri Lankan societal values and Sinhala questions (Muthugala et al., 2026). The model judges statements A and B, producing A, B, Both, or 0 when neither is acceptable.

We use no additional task-specific corpus or teacher-generated predictions; only the official development data are used for task-specific training and prompt development. The additional training views described below are deterministic transformations of official rows. The pretrained models' original training corpora are outside the scope of this statement, and the closed model's pretraining data are not publicly known.

3 Track 1 System

3.1 Base Model and Eligibility

We use the instruction-tuned `google/gemma-4-E4B-it` checkpoint (Gemma Team, 2026) at pinned revision `ee0ef6023621`; the complete hash is recorded in the released configuration. The model configuration disables expert blocks; the exact text-generation model loaded by `AutoModelForCausalLM` has 7,941,100,832 parameters. It is therefore a dense model below the Track 1 limit. LoRA (Hu et al., 2022) adds 17,440,768 temporary trainable parameters, yielding 7,958,541,600 parameters during training. Before inference, the low-rank deltas are merged into existing linear weights and the temporary modules are removed, returning the answer-generation count to 7,941,100,832.

We compute each count as `sum(p.numel() for p in model.parameters())`, including embedding parameters. No vision or audio encoder is loaded. The runtime code repeats this count after loading and after adapter merging and aborts if the answer-generation model exceeds eight billion parameters.

The same checkpoint, adapter, prompt template, and decoding program process all three datasets. Inference is local and uses no commercial endpoint, external agent, retrieval system, knowledge base, model ensemble, or manual test-item intervention.

3.2 Candidate Reframing

Instead of generating A–D directly, we render one prompt per candidate. The prompt includes the dataset and country context, native and English value labels, scenario, question, all options, and the candidate under examination. The model must answer exactly YES or NO. For candidate c, we define

$$s(c) = z_{\text{YES}}(c) - z_{\text{NO}}(c), \tag{1}$$

where z is the next-token logit. This shared interface converts heterogeneous label spaces into a common value-fit judgment.

For Chinese, the single gold option is positive and all others are negative. For Indonesian, our training implementation marks every distinct option chosen by at least one of the five annotators as positive. For Sri Lankan, A and B are independent binary candidates: Both maps to two positives and 0 maps to none.

At inference, Chinese and Indonesian select $\arg\max_c s(c)$. For Sri Lankan, A and B are independently accepted if their score exceeds a scalar threshold; the two decisions map to A, B, Both, or 0. The final threshold, 0.75, is fitted on full development scores after model training.

3.3 Preference Objective

For candidate target $y \in \{0, 1\}$, we apply binary cross-entropy to $s(c)$:

$$\mathcal{L}_{\text{cls}} = \text{BCEWithLogits}(s(c), y). \tag{2}$$

When an item has at least one positive and one negative option, we sample a pair (c^+, c^-) and add

$$\mathcal{L}_{\text{rank}} = \text{softplus}\big(m - [s(c^+) - s(c^-)]\big). \quad (3)$$

The total loss is $\mathcal{L} = \mathcal{L}_{\text{cls}} + \lambda\mathcal{L}_{\text{rank}}$, with margin $m = 0.5$ and $\lambda = 0.5$. This adapts the standard pairwise logistic ranking formulation (Burges et al., 2005) to candidate judgments with heterogeneous label spaces.

3.4 Augmentation and Balancing

We use four deterministic training transformations. First, option texts are permuted each epoch and labels are remapped, discouraging answer-position memorization. Second, we sample 790 items from each language pool per epoch; the smaller Indonesian and Sri Lankan pools are resampled. Third, one positive and one negative candidate are sampled when possible. Fourth, with probability 0.10 for Chinese items, text after a downside contrast marker is removed to create a core-action view. This last transformation targets a development failure in which severe consequence language dominated the exact value fit. The original option remains present in the other 90% of views.

3.5 Optimization

We inject rank-8 LoRA modules into the q, k, v, o, gate, up, and down projections of all language-model layers. Alpha is 16 and dropout is 0.05. We train for two epochs with micro-batch size 2, gradient accumulation 8, maximum length 1,536, and bfloat16 precision. AdamW uses learning rate 10^{-4}, weight decay 0.01, 5% warmup, cosine decay, and gradient clipping at 1.0. The run has 298 optimizer steps and seed 2026.

Training on one NVIDIA A100 80GB took 32 minutes 25 seconds. Scoring all 5,475 test items (20,306 candidate prompts) took 14 minutes 51 seconds on the same GPU.

4 Track 2 Systems

4.1 GPT-5.4 Prompting

The open-track experiments use `openai.gpt-5.4` through an OpenAI-compatible commercial endpoint. Each run uses that one model and one versioned prompt configuration for all three datasets. No outputs from different models or prompt versions are combined. The prompt includes the native and English value labels, scenario, question, all competing options, shared decision rules, and language-specific guidance within the same configuration. It requires the valid task label and a structured justification that evaluates the selected and rejected options. For Sri Lankan items, it requires independent judgments of statements A and B before mapping them to A, B, Both, or 0.

We use temperature 0, reasoning mode, a maximum of 700 completion tokens, and JSON output. Four prompt stages are evaluated. Version 1 establishes shared instructions; v2 adds task examples and clearer value grounding; v3 adds rules derived from Chinese development errors; and v5 refines the independent Sri Lankan A/B decision. We perform no GPT-5.4 fine-tuning. Reasoning is retained in experiment logs but removed from the submission JSONL.

4.2 Final Open-Track Selection

The strongest GPT-5.4 variants reach 0.795 macro accuracy. Because Track 2 permits an eligible Track 1 system and the preference-trained Gemma system reaches 0.808, we select that system for the final open-track submission. Its model, prompt, inference code, and predictions are identical to the Track 1 submission; GPT and Gemma outputs are never aggregated.

5 Experiments

5.1 Development Protocol

We assign five deterministic folds. Chinese folds are stratified so each fine-grained value leaf contributes one held-out item per fold; the other datasets are grouped by value and accepted-label signature. Each out-of-fold prediction is produced by a model that did not train on that row.

Sri Lankan thresholds can otherwise leak held-out labels. Our strict estimate therefore selects a threshold for each fold using only the other four folds. For Indonesian reporting, we recompute correctness using the official top-frequency consensus rule. This is important because the broader any-annotator-positive proxy used during training and early analysis is optimistic.

In the result tables, ZH, ID, and SI denote the Chinese, Indonesian, and Sri Lankan datasets, respectively.

Table 2 compares the final preference method with our strongest recorded prompt-only development baseline. Candidate learning adds eight Chinese points and 2.6 macro points. The strict development macro estimate (0.806) is close to the

System	ZH	ID	SI	Macro
Prompt v20	.715	.699	.926	.780
Preference OOF	.795	.702	.921	.806
Δ	+.080	+.003	-.005	+.026

Table 2: Development accuracy under official consensus semantics. The Sri Lankan preference result uses cross-fitted thresholds.

Track	System	ZH	ID	SI	Macro
1	Prompt v20†	.724	.688	.950	.787
1, 2	Final preference*	**.788**	.683	.954	**.808**
2	GPT-5.4 v1†	.632	.689	.951	.757
2	GPT-5.4 v2†	.651	.681	.942	.758
2	GPT-5.4 v3†	.752	.681	.951	.795
2	GPT-5.4 v5†	.760	.665	**.960**	.795

Table 3: Final results and leaderboard development history. *Official final submission; †development-stage leaderboard observation. Each row independently uses one model and one prompt configuration for all datasets.

hidden result (0.808), although country-level generalization differs.

An early pooled analysis accepted every Indonesian option receiving any annotator vote and selected one Sri Lankan threshold on all OOF scores. It gave 0.795/0.926/0.936 (macro 0.886). We do not present that number as official accuracy. In particular, the 22.4-point Indonesian gap between the any-vote proxy and consensus accuracy quantifies the objective mismatch.

5.2 Hidden Evaluation

Table 3 shows that the final candidate-based system improves macro accuracy from 0.787 to 0.808. The gain is concentrated in Chinese (+6.4 points), while Sri Lankan performance remains high and Indonesian decreases slightly. Prompt variants were iterated with development analysis and leaderboard feedback; the GPT scores and prompt-v20 comparison should therefore be read as system-development history rather than controlled test-set ablations. The official final leaderboard ranks the 0.808 preference system first in both Track 1 and Track 2 (PlurVA-LLM 2026 Organizers, 2026).

6 Analysis and Discussion

Why Chinese improves. Chinese options frequently combine a value-relevant action with retaliation, privacy, livelihood, or safety costs. Prompt-only models alternated between over-caution and over-correction toward dramatic institutional action. The candidate objective requires each option to be compared against the exact value and against competing actions; position permutation and the core-action view are intended to reduce reliance on surface risk intensity. In paired development analysis against prompt v20, preference OOF predictions corrected 129 Chinese items and regressed on 66, a net gain of 63 (two-sided exact McNemar $p = 7.6 \times 10^{-6}$). Net gains occur in Social Order (+49), Survival Support (+7), and Civilization and Progress (+7). For example, on a Personal Safety item, the prompt baseline chose covert recording followed by confrontation; the preference model instead chose to leave under a pretext while sending a coded request for help, matching the gold option. These paired results establish a robust association, but do not isolate the contribution of the ranking loss, balancing, permutation, or core-action augmentation.

Why Indonesian remains difficult. The development file contains five raw annotator choices, but official scoring uses only maximum-frequency consensus. Treating every observed choice as a positive teaches recall of minority-acceptable responses rather than selection of the modal social judgment. This target came from our initial interpretation of the raw vote list and was retained in the frozen final training run; recomputation under the official consensus rule exposed the mismatch. This design produces a high any-vote proxy but only 0.702 consensus OOF accuracy and 0.683 hidden accuracy. A future system should train with vote counts, for example through soft targets or a listwise distributional loss, while preserving ties at the maximum frequency.

Sri Lankan calibration. Independent A/B scoring naturally represents all four labels and transfers well. Remaining errors are concentrated near the acceptance threshold, especially Both versus a single accepted statement. The cross-fitted estimate is lower than the pooled-threshold estimate, illustrating why threshold selection must be separated from held-out evaluation.

Pluralism versus label matching. The task evaluates agreement with country-specific annotations, not whether a model captures all views within a country. This distinction is especially visible in Indonesian vote distributions. Country-

aware prompts should be understood as conditioning on the benchmark context, not as claims that a country has one homogeneous value system.

Compact preference learning versus closed-model prompting. GPT-5.4 prompting substantially improves Chinese accuracy from v1 to v3, but the best GPT macro score remains 1.3 points below preference-trained Gemma. The closed model provides a Chinese–Sri Lankan trade-off across v3 and v5, whereas Indonesian accuracy remains below the strict development estimate for the preference system. These observed scores suggest that task-specific learning from annotated preferences can remain competitive when a stronger general-purpose model is permitted; because prompt versions were selected with leaderboard feedback, this is not a controlled model comparison.

6.1 Reproducibility

The submitted materials contain the exact 69.8MB adapter, its configuration and checksum, the versioned prompts, training and inference code, dependency versions, fold-level scores, and submission validators. The Gemma inference script loads the pinned base revision, merges 258 LoRA modules, asserts the final parameter count, scores all candidates, and creates the required JSONL archive. The Track 2 code also includes the GPT-5.4 client, prompts v3 and v5, fixed decoding settings, and credential-free launch scripts. Endpoint credentials are supplied only through environment variables. Reproducing the final Track 2 result invokes the same included Gemma system as Track 1. All documented paths and commands are repository-relative.

7 Conclusion

We presented Modgenix systems for both shared-task tracks. The resulting binary candidate system substantially improves Chinese accuracy while preserving strong Sri Lankan performance. The resulting single 7.94B-parameter model achieves 0.808 hidden macro accuracy and ranks first in both tracks, outperforming our best GPT-5.4 prompt-only run at 0.795. The analysis identifies a concrete next step: align Indonesian training targets with consensus vote frequencies rather than collapsing all observed votes into binary acceptance.

Limitations

The development sets are small, particularly for Sri Lankan threshold selection. Repeated leaderboard experiments during prompt development create a risk of adaptive overfitting, so hidden prompt comparisons are not controlled ablations. The Chinese core-action transformation assumes that removing a downside clause preserves the labeled action; this is plausible for the target error pattern but may alter nuance. The Indonesian binary target discards vote frequency and is mismatched to official consensus, as quantified in our analysis. We evaluate one model family and one random seed, and do not report statistical significance or calibration beyond accuracy. Finally, benchmark agreement cannot establish normative correctness or represent all value perspectives within China, Indonesia, or Sri Lanka. The GPT-5.4 experiments depend on a hosted model whose implementation may change, so exact long-term reproduction requires provider-side version stability. Their hidden-set scores were observed during system development and are not controlled ablations.

Ethical Considerations

The system learns from human-annotated, culturally situated judgments. Such labels should not be used to essentialize national or cultural groups, and the model should not be interpreted as an authority on moral decisions. We use only the released shared-task data and do not introduce personal data. The method's errors could reinforce majority views while suppressing minority preferences; the Indonesian analysis demonstrates this tension directly.

Acknowledgments

We thank the PlurVA-LLM organizers and dataset creators. Generative AI assistance was used to organize existing experiment records and to draft and edit portions of this manuscript. The author verified the methodology, numerical results, and citations and remains responsible for the paper.

References

Chris Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Greg Hullender. 2005. Learning to rank using gradient descent. In *Proceedings of the 22nd International Conference on Machine Learning*, pages 89–96.

Gemma Team. 2026. Gemma 4 Technical Report. *Preprint*, arXiv:2607.02770.

Siyi Hao, Yidi Cao, Linhao Yu, Yuqi Ren, and Deyi Xiong. 2026. D2VBench: Benchmarking large language models with value dilemmas in daily scenarios. *Preprint*, arXiv:2607.19834.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*.

Nethmi Muthugala, Supryadi, Surangika Ranathunga, Nisansa de Silva, Ruijie Tao, Ovindu Gunatunga, Pengyun Zhu, Shaowei Zhang, Jingting Zheng, and Deyi Xiong. 2026. LKValues: Aligning large language models with sri lankan societal values. *Preprint*, arXiv:2607.20410.

PlurVA-LLM 2026 Organizers. 2026. PlurVA-LLM Shared Task 2026: Registration and Leaderboard. Accessed 2026-08-11.

Taylor Sorensen, Jared Moore, Jillian Fisher, Mitchell Gordon, Niloofar Mireshghallah, Christopher Michael Rytting, Andre Ye, Liwei Jiang, Ximing Lu, Nouha Dziri, Tim Althoff, and Yejin Choi. 2024. A roadmap to pluralistic alignment. In *Proceedings of the 41st International Conference on Machine Learning*.

Supryadi, Irfan, Julianti, Darren Keanly Martin, Jayvin Fernando, Yuqi Ren, and Deyi Xiong. 2026. Pancasila-Dilemmas: Evaluating large language models on indonesian human value dilemmas grounded in pancasila. *Preprint*, arXiv:2607.18066.