# Sagarmatha at SemEval-2026 Task 9: Heterogeneous Ensembling and Hierarchical Task Conditioning for Multilingual Latent Distributional Divergence Modeling

**Astha Shrestha** [1*] **and Sujal Maharjan** [1*] **and Pratikshya Shrestha**[2]
[1]IIMS College (Affiliated with Taylor's University, Malaysia), Kathmandu, Nepal
[2]PCPS College (Affiliated with University of Bedfordshire, UK), Lalitpur, Nepal
{sujalmaharjan007, aasthashrestha688, shresthapratikshya857}@gmail.com

## Abstract

The digital public square is increasingly fragmented by affective polarization, requiring computational systems capable of identifying discursive strategies such as dehumanization and vilification. This paper presents Sagarmatha, the system developed for SemEval-2026 Task 9. We propose a heterogeneous ensemble architecture that addresses the limitations of standard transformer fine-tuning across 22 languages. Our approach integrates mDeBERTa-v3, ReMBERT, LaBSE, mmBERT, and XLM-RoBERTa, through two primary architectural pillars: learnable weighted layer pooling and hierarchical task conditioning. While our final submission (a broad ensemble, $R_3$) demonstrated high stability on the leaderboard, our primary architectural configuration (Weighted Polyglot, $R_1$) yielded superior performance in complex multi-label tasks. The system ranked 1st globally in English and Hausa manifestation identification, and 1st in Telugu detection (2nd in categorization). All code and resources are available at https://github.com/SUJAL390/Sagarmatha_at_Semeval_task_9.git.

## 1 Introduction

Online polarization manifests as rhetorical strategies (dehumanization, vilification) that go beyond lexical hostility (Iyengar et al., 2019; Lelkes, 2016). SemEval-2026 Task 9 (Naseem et al., 2026a) formalizes polarization via a three-stage taxonomy consisting of: (S1) presence detection, (S2) type categorization, and (S3) manifestation identification.

A central modeling challenge is that S3 labels are semantically nested under S2, so naive multitask training can suffer from gradient interference and logical inconsistency (Ruder, 2017). We address this with two compact, practical contributions that are broadly applicable: (1) learnable weighted pooling across deep transformer layers to recover mid-layer signals that encode rhetorical nuance, and (2) hierarchical task conditioning (HTC), a light-weight structured-conditioning mechanism that approximates a conditional factorization of the joint label space while preserving end-to-end optimization stability. The overall Sagarmatha architecture is shown in Fig. 1.

**Contributions.** We summarize our contributions succinctly:

1. A practical heterogeneous-inductive-bias fusion pipeline combining five multilingual backbones and lightweight task conditioning.
2. Formalization and operationalization of *Logical Violation Rate* (LVR) to audit taxonomic consistency.
3. A compact set of inference aggregation strategies (mean, power-mean, rank-avg, max) and deterministic post-processing to enforce taxonomy.
4. Forensic analyses showing where weighted pooling and HTC materially change errors (see Table 1, Fig. 2 and Fig. 3).

## 2 Related Work

### 2.1 Socio-Linguistics and Polarization

Linguistic polarization is bifurcated into *Ideological Polarization* and *Affective Polarization* (Iyengar et al., 2019). In online media, this manifests as othering language where discourse delegitimizes the out-group rather than persuading (Rathje et al., 2021). Previous shared tasks like HatEval (Basile et al., 2019) focused on toxicity, but Task 9 requires deeper NLU for rhetorical framing across multicultural contexts.

* The authors contributed equally to this work and are designated as joint first authors. The author order follows alphabetical order by first name.



*Proceedings of the The 20th International Workshop on Semantic Evaluation (2026)*, pages 3045–3051

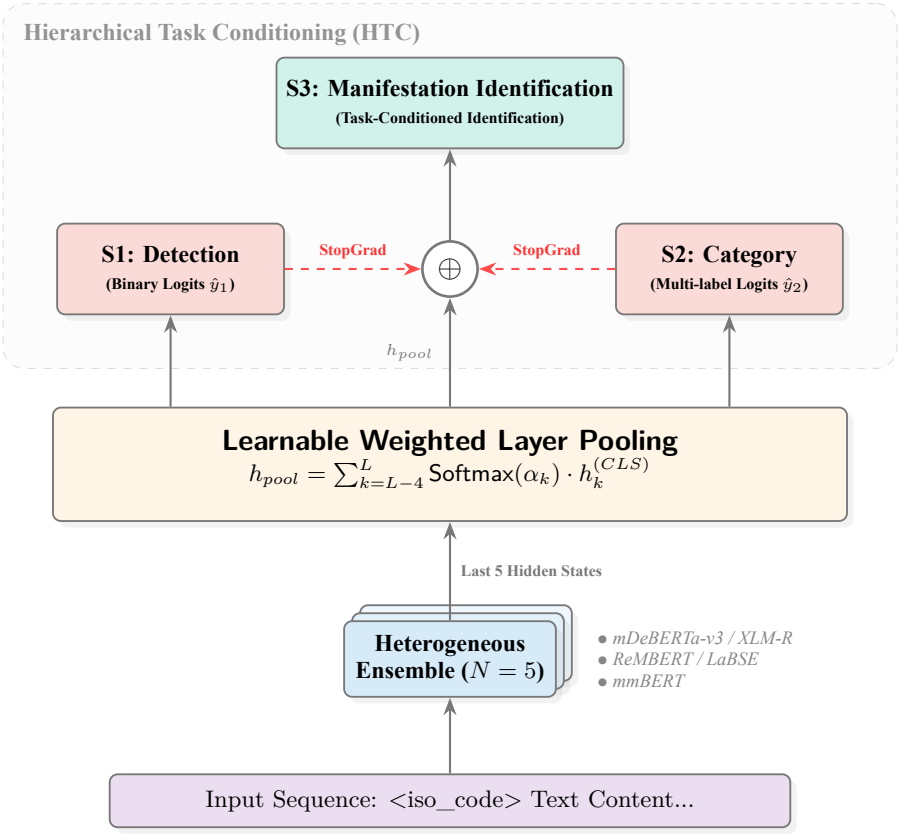


Figure 1: Sagarmatha Architecture. Features are extracted via learnable weighted pooling of the final five hidden states. The hierarchical head conditions Subtask 3 on detached (*StopGrad*) logits from Subtasks 1 and 2.

### 2.2 Architectural Advancements

The practice of extracting features from multiple transformer layers, known as *Weighted Layer Pooling*, was popularized by Kondratyuk and Straka (2019) to capture both syntax and semantics simultaneously. Furthermore, multilingual models such as mDeBERTa-v3 (He et al., 2021), which utilizes disentangled attention, and ReMBERT (Chung et al., 2020), which leverages a significantly expanded vocabulary, have shown varying efficacy across scripts. Our work unifies these backbones into a coherent hierarchical pipeline.

## 3 Sagarmatha System Architecture

To overcome the 16GB VRAM limitation of the Kaggle T4 GPU, the five large transformer backbones were not trained simultaneously. Instead, each model was fine-tuned independently, and its best checkpoint weights were saved. Throughout this paper, we refer to three primary configurations: $R_1$ (Weighted Polyglot), which evaluates the standalone impact of learnable weighted layer pooling on the individual backbones; $R_2$ (Hierarchical), which extends $R_1$ independently by incorporating hierarchical task conditioning (HTC); and $R_3$ (Broad Ensemble), our final leaderboard submission, which represents the late-fusion inference aggregation of these saved best-checkpoint models.

### 3.1 Heterogeneous Ensemble Strategy

The system utilizes an ensemble of five distinct backbones. For our primary models, we focus on backbones selected for unique inductive biases: mDeBERTa-v3 for disentangled attention (He et al., 2021), ReMBERT for its 250k token vocabulary (Chung et al., 2020), and LaBSE for cross-lingual alignment. Additionally, we incorporate XLM-RoBERTa for its robust universal cross-lingual baseline, and mmBERT to provide specialized morphological representations tailored for highly inflected or under-resourced scripts, ensuring balanced coverage across the 22 languages.

#### 3.1.1 Ensemble Aggregation

Let $p^{(m)}$ denote the probability output of the $m$-th individually trained model for a given task. Our primary ensemble ($R_3$) uses mean-probability aggregation:

| Taxonomy | Sequence [Original → Translation] | $R_3$ (Ensemble) | Specialized Config. | Scientific Discovery |
|---|---|---|---|---|
| **Taxonomic** | यो घटनाबाट लक्ष्मीदेवीले प्रतिनिधित्व गर्ने वर्गका महिलाले स्वास्थ्य सेवा लिन नसकिरहेको प्रस्ट पार्छ। *Trans:* Systemic critique of health access for the poor. | $S_1$ : 0 (Neg) $S_3$ : 1 (Pos) | $R_2$: $S_1$ : 0, $S_3$ : 0 *(Consistent)* | $R_3$ suffers from Taxonomic Drift (hallucinated tactic). The Hierarchical variant ($R_2$) enforces logic via task-dependency. |
| **Semantic** | నిర్లక్ష్యంతో వ్యవహరిస్తున్నారు రైతుల సమస్యలు పరిష్కరించకపోవడం బాధాకరం. *Trans:* They are acting with negligence. It is unfortunate that the farmers' issues remain unresolved. | $S_1$ : 1 (Pos) $S_3$ : 0 (Neg) | $R_1$: $S_1$ : 1, $S_3$ : 1 *(Detail Recovered)* | $R_3$ smooths nuanced predictions. The Weighted Polyglot variant ($R_1$) recovers mid-layer signals of "neglect" and "pain." |
| **Script** | عبدالحفیظ رہنما کے سندھ اسلامی جماعت سے ریلی پاکستان جیوے میں اباڑو ۔ ہے۔ وابستہ سے نظام اسلامی سلامتی اور بقا کی ملک کہا میں خطاب نے بجارانی *Trans:* While addressing the 'Jeevay Pakistan' rally in Ubauro, Jamaat-e-Islami Sindh leader Abdul Hafeez Bijarani emphasized that the nation's survival and security are inextricably linked to the implementation of an Islamic system. | $S_2$ : Other $S_3$ : None | $R_1$: $S_2$ : $Religious$ $R_1$: $S_3$ : $Ideological$ | Standard pooling in $R_3$ misses script-specific entities. $R_1$ demonstrates Morphological Resilience in Perso-Arabic scripts. |

Table 1: Comparative Error Taxonomy of Sagarmatha Configurations. We analyze how our specialized variants ($R_1, R_2$) mitigate the logical inconsistencies and semantic smoothing found in our final ranking ensemble ($R_3$).

$$\hat{p} = \frac{1}{M} \sum_{m=1}^{M} p^{(m)} \tag{1}$$

where $M$ is the number of ensemble members.

In addition, we explored alternative aggregation strategies such as Power Mean:

$$\hat{p}_{\text{power}} = \sqrt{\frac{1}{M} \sum_{m=1}^{M} (p^{(m)})^2} \tag{2}$$

and Rank Averaging:

$$\hat{p}_{\text{rank}} = \frac{1}{M} \sum_{m=1}^{M} \text{RankNorm}(p^{(m)}) \tag{3}$$

where RankNorm$(\cdot)$ converts probabilities to normalized rank scores to reduce calibration bias across heterogeneous backbones. Ultimately, mean-probability aggregation was selected for $R_3$ because it acts as a robust regularizer against the overconfidence of individual models, yielding the most stable validation calibration without the disproportionate penalty to minority predictions introduced by rank averaging or the variance amplification seen in power means.

### 3.2 Weighted Layer Pooling

The pooled representation $h_{\text{pool}}$ is a weighted sum of the last $n = 5$ hidden states:

$$h_{\text{pool}} = \sum_{k=L-4}^{L} \text{Softmax}(\alpha_k) \cdot h_k \tag{4}$$

where $\boldsymbol{\alpha}$ are learnable scalar weights. The choice of the final five layers ($n = 5$) optimally balances computational efficiency with the extraction of both high-level semantic abstractions (final layers) and intermediate syntactic and rhetorical features (lower layers) critical for detecting subtle polarization tactics.

### 3.3 Hierarchical Task Conditioning (HTC)

To model the logical taxonomy, the input to the manifestation classifier ($S_3$) is augmented with predicted states of detection ($S_1$) and categorization ($S_2$):

$$X_{S3} = [\text{Drop}(h_{\text{pool}}) \oplus \text{SG}(\hat{y}_{S1}) \oplus \text{SG}(\hat{y}_{S2})] \tag{5}$$

The StopGrad (SG) operator is critical: it prevents the highly sparse and complex loss landscape of the manifestation classifier ($S_3$) from backpropagating into and destabilizing the foundational representation spaces optimized for the primary detection ($S_1$) and categorization ($S_2$) tasks. At inference, we enforce taxonomic consistency via a deterministic constraint:

$$\hat{y}_{S1} = 0 \Rightarrow \hat{y}_{S2} = \mathbf{0}, \quad \hat{y}_{S3} = \mathbf{0} \tag{6}$$

This ensures category or manifestation labels are not predicted for non-polarized sequences.

| Parameter | Value |
|---|---|
| Loss Function | Cross-Entropy w/ *Pos. Weighting* |
| Optimization | AdamW, LR $2 \times 10^{-5}$, 8 Epochs |
| Hyperparams | Batch 16, Max Len 128, Dropout 0.1 |
| Validation | 10% Stratified (Metric: Macro-$F_1$) |
| Hardware | Kaggle T4 GPU (Seed: 42) |

Table 2: Experimental Setup.

## 4 Experimental Setup.

Training details and hyperparameters for all variants are in Table 2; model selection is based on validation macro-$F_1$. We developed our models using

the official POLAR benchmark dataset (Naseem et al., 2026b).

### 4.1 Threshold Optimization

For Subtask $S_1$, decision thresholds were optimized on the validation set via grid search:

$$t^* = \arg \max_{t \in [0.3, 0.7]} F_1^{macro}(t) \quad (7)$$

using a step size of 0.02. Subtasks $S_2$ and $S_3$ used fixed thresholds of 0.35 based on validation calibration.

## 5 Results and Detailed Analysis

The Sagarmatha system achieved elite rankings across all 22 languages. Table 3 provides the comprehensive performance matrix. Our system established a new benchmark for Telugu, ranking 1st in Detection ($F_1 = 0.905$) and 2nd in Categorization ($F_1 = 0.465$). In the most complex task ($S_3$), we achieved the Global Rank 1 in English ($F_1 = 0.511$) and Hausa ($F_1 = 0.208$).

### 5.1 Ablation and Component Analysis

To understand the contribution of each architectural component, we compare the overall averages of our configurations (Table 3). Removing the ensemble aggregation (comparing $R1$ and $R2$ against $R3$) reveals a surprising dynamic: while the broad ensemble ($R3$) provides general leaderboard stability, the standalone Weighted Polyglot architecture ($R1$) actually outperforms the ensemble in Subtask 1 (0.790 vs 0.787) and Subtask 2 (0.575 vs 0.559). This demonstrates that learnable weighted pooling effectively captures nuanced mid-layer signals that standard ensembling tends to over-smooth. Furthermore, evaluating the impact of HTC (comparing $R2$ to $R1$) shows that while HTC introduces a strict structural constraint that slightly limits raw multi-label recall, it is strictly necessary to prevent the cascading taxonomic hallucinations detailed in our LVR audit.

## 6 Interpretability and Forensic Audit

### 6.1 Cross-Lingual Performance Insights

A deeper analysis of language-specific performance reveals distinct cross-lingual behaviors. Our system achieved its highest peaks in Indic languages, notably Hindi ($S1$: 0.951) and Telugu ($S1$: 0.905). This strong performance is largely attributable to the inclusion of ReMBERT and LaBSE in our backbone pool, which possess extensive vocabularies that prevent severe token fragmentation in complex scripts. Conversely, the system struggled with subtask 2 and 3 performance in languages like Italian ($S2$: 0.230) and Bengali ($S3$: 0.228). This degradation correlates with either limited training instances for specific manifestation types or high morphological complexity where the base models failed to align affective nuances. In these under-resourced scenarios, the smoothing effect of the $R3$ ensemble actively hindered performance compared to the localized feature extraction of $R1$.

### 6.2 Taxonomic Consistency (LVR)

A critical contribution of this work is the mitigation of *Logical Violations* instances where a model identifies a manifestation tactic ($S_3$) despite classifying the sequence as non-polarized ($S_1$). As shown in Fig. 2, our *Hierarchical Attention* ($R_2$) configuration significantly reduced the Logical Violation Rate (LVR) compared to the base ensemble ($R_3$). While the ensemble exhibited LVRs as high as 33% in languages like Italian, our hierarchical conditioning constrained the prediction space to logically valid subspaces, ensuring taxonomic integrity.

#### 6.2.1 Formal Definition of Logical Violation Rate (LVR)

We define Logical Violation Rate (LVR) as:

$$\text{LVR} = \frac{\sum_{i=1}^{N} \mathbb{1}\left[\hat{y}_{S1}^{(i)} = 0 \wedge \|\hat{y}_{S3}^{(i)}\|_1 > 0\right]}{N} \quad (8)$$

where $\mathbb{1}[\cdot]$ is the indicator function and $\|\hat{y}_{S3}^{(i)}\|_1$ denotes the number of predicted manifestation labels for sample $i$. LVR captures the proportion of logically inconsistent predictions in which a manifestation is detected despite absence of polarization.

### 6.3 Nuance Sensitivity in Multi-label Detection

Fig. 3 illustrates the Nuance Sensitivity of our architectures. Our *Broad Ensemble* ($R_3$) and *Weighted Polyglot* ($R_1$) configurations consistently identified a high number of manifestations per post in high-context scripts like Urdu and Hindi. While $R_3$ maximizes raw recall, the *Hierarchical* ($R_2$) configuration exhibits a more conservative identification rate, effectively suppressing

| Fam. | ISO | Lang | $S_1$ (Detection) | | | | $S_2$ (Type) | | | | $S_3$ (Manifest.) | | | | Rank |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $R_1$ | $R_2$ | $R_3$ | $B$ | $R_1$ | $R_2$ | $R_3$ | $B$ | $R_1$ | $R_2$ | $R_3$ | $B$ | $(S_1/S_2/S_3)$ |
| dra | te | Telugu | .905 | .884 | .892 | .644 | **.465** | .458 | .446 | .314 | .406 | .385 | **.424** | .220 | **1 / 2 / 3** |
| inc | ur | Urdu | .878 | .812 | .816 | .789 | .780 | .770 | .779 | .712 | **.808** | .800 | .799 | .769 | 39 / 6 / 11 |
| | hi | Hindi | **.951** | .913 | .917 | .737 | .775 | .764 | .753 | .791 | .748 | .740 | .740 | .745 | 8 / 17 / 8 |
| | ne | Nepali | .914 | .914 | .907 | .879 | **.789** | .768 | .778 | — | .667 | .667 | .658 | .609 | 12 / 12 / 6 |
| | pa | Punjabi | .793 | .790 | .792 | .789 | .536 | .520 | .524 | .365 | .529 | **.531** | .523 | .384 | 4 / 2 / 4 |
| | or | Odia | .624 | .809 | **.829** | .776 | .468 | .468 | .468 | .560 | **.280** | .278 | .277 | .131 | 34 / 14 / 5 |
| | bn | Bengali | .833 | .833 | **.836** | .852 | **.367** | .324 | .333 | .288 | **.229** | .213 | .228 | .086 | 24 / 11 / 9 |
| afa | am | Amharic | **.839** | **.839** | .836 | .715 | **.611** | .604 | .593 | .371 | **.532** | .523 | .524 | .443 | 21 / 8 / 8 |
| | ha | Hausa | .934 | .934 | **.941** | .775 | **.427** | .394 | .426 | .203 | **.208** | .170 | .207 | .000 | 10 / 3 / **1** |
| | ar | Arabic | **.827** | .826 | .823 | .795 | **.605** | .598 | .587 | .485 | **.595** | .584 | .577 | .390 | 16 / 17 / 10 |
| ine | en | English | .752 | **.826** | .817 | .780 | **.511** | .506 | .503 | .333 | **.511** | .485 | .510 | .410 | 12 / 4 / **1** |
| | es | Spanish | **.772** | .783 | .763 | .726 | **.637** | .622 | .615 | .593 | **.515** | .508 | .495 | .456 | 21 / 18 / 6 |
| | de | German | **.728** | **.728** | .724 | .671 | .570 | **.577** | .551 | .407 | .498 | .474 | .498 | .348 | 15 / 10 / 4 |
| | it | Italian | **.660** | **.660** | .652 | .677 | **.260** | .246 | .230 | .376 | — | — | — | — | 31 / 19 / — |
| | fa | Persian | **.915** | .861 | .867 | .842 | **.600** | .586 | .579 | .463 | **.470** | .460 | .461 | .200 | 6 / 12 / 4 |
| | pl | Polish | .764 | **.818** | .814 | .724 | **.564** | .551 | .551 | .449 | — | — | — | — | 15 / 9 / — |
| | ru | Russian | .672 | .731 | **.821** | .745 | **.565** | .502 | .523 | .590 | — | — | — | — | 27 / 13 / — |
| sit | zh | Chinese | .883 | **.893** | .879 | .869 | **.750** | .734 | .732 | .669 | .536 | **.538** | .510 | .531 | 30 / 20 / 18 |
| | my | Burmese | **.894** | .873 | .864 | .821 | **.690** | .675 | .674 | .477 | — | — | — | — | 18 / 15 / — |
| oth | km | Khmer | **.918** | **.918** | .913 | .659 | **.629** | .607 | .601 | .626 | .313 | **.341** | .337 | .234 | 42 / 17 / 8 |
| | tr | Turkish | .775 | .776 | **.787** | .696 | .561 | **.577** | .544 | .470 | .501 | **.512** | .502 | .673 | 26 / 20 / 6 |
| | sw | Swahili | **.800** | .788 | .773 | .757 | **.507** | .484 | .493 | .441 | **.553** | .550 | .528 | .508 | 7 / 8 / 10 |
| AVG | | Overall | **.790** | .789 | .787 | .741 | **.575** | .571 | .559 | .446 | **.492** | .486 | .489 | .352 | — |

Table 3: Sagarmatha Performance Matrix. Variants: ($R_1$) Weighted Polyglot, ($R_2$) Hierarchical, ($R_3$) Broad Ensemble, ($B$) Baseline. Language identifiers follow lowercase ISO 639-1 standards to distinguish from ISO 3166-1 country codes. Best scores are **bolded**; global peaks are in **red**.

the hallucinated marginal signals by strictly enforcing the logical taxonomy.

### 6.4 Cross-Lingual Discursive Mapping

By auditing the raw predictions of our best configuration, we generated a global correlation map between polarization types and discursive tactics (Fig. 4). Our analysis reveals that Racial/Ethnic Polarization is uniquely characterized by a 0.91 correlation with *Stereotyping*, whereas Religious Polarization shows the strongest link to *Vilification* (0.85). Interestingly, Political Polarization demonstrated its strongest correlations with *Vilification* (0.84) and *Extreme Language* (0.76), highlighting a highly aggressive discursive profile. This mapping provides a data-driven sociological profile of how polarization manifests across 22 diverse global cultures.

## 7 Conclusion

We present a compact, reproducible pipeline that integrates heterogeneous backbones with learnable layer pooling and hierarchical conditioning. Our contributions are practical and generalizable: HTC offers a light-weight way to enforce structural constraints in multi-label taxonomies, and weighted pooling provides an inexpensive mechanism to recover rhetorical nuance. Together, these design choices yield strong multilingual performance and improved taxonomic integrity (reduced LVR) on the POLAR benchmark.

## 8 Limitations and Ethics

We explicitly note limitations: single-seed experiments (seed 42) due to shared-task time constraints; some Romance languages have lower S2 performance likely because of limited training instances. We caution against deployment without human oversight: labeling is subjective, cultural interpretation varies, and errors could harm targeted groups. Recommended safeguards: human-in-the-loop review, abstention thresholds for high-uncertainty predictions, and further demographic bias audits prior to deployment.

## References

Valerio Basile, Cristina Bosco, Elisabetta Fersini, Debora Nozza, Viviana Patti, Francisco Manuel Rangel Pardo, Paolo Rosso, and Manuela Sanguinetti. 2019. Semeval-2019 task 5: Multilingual detection of hate speech against immigrants and women in twitter. In *Proceedings of the 13th international workshop on semantic evaluation*, pages 54–63.

Hyung Won Chung, Thibault Fevry, Henry Tsai,

Melvin Johnson, and Sebastian Ruder. 2020. Rethinking embedding coupling in pre-trained language models. *arXiv preprint arXiv:2010.12821*.

Pengcheng He, Jianfeng Gao, and Weizhu Chen. 2021. Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing. *arXiv preprint arXiv:2111.09543*.

Shanto Iyengar, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J Westwood. 2019. The origins and consequences of affective polarization in the united states. *Annual review of political science*, 22(1):129–146.

Dan Kondratyuk and Milan Straka. 2019. 75 languages, 1 model: Parsing universal dependencies universally. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, pages 2779–2795.

Yphtach Lelkes. 2016. Mass polarization: Manifestations and measurements. *Public Opinion Quarterly*, 80(S1):392–410.

Usman Naseem, Robert Geislinger, Juan Ren, Sarah Kohail, Rudy Garrido Veliz, P Sam Sahil, Yiran Zhang, Marco Antonio Stranisci, Idris Abdulmumin, "Ozge Alacam, Cengiz Acar"urk, Aisha Jabr, Saba Anwar, Abinew Ali Ayele, Elena Tutubalina, Aung Kyaw Htet, Xintong Wang, Surendrabikram Thapa, Tanmoy Chakraborty, and 15 others. 2026a. SemEval-2026 task 9: Detecting multilingual, multicultural and multievent online polarization. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*. Association for Computational Linguistics.

Usman Naseem, Robert Geislinger, Juan Ren, Sarah Kohail, Rudy Garrido Veliz, P Sam Sahil, Yiran Zhang, Marco Antonio Stranisci, Idris Abdulmumin, Özge Alacam, Cengiz Acartürk, Aisha Jabr, Saba Anwar, Abinew Ali Ayele, Simona Frenda, Alessandra Teresa Cignarella, Elena Tutubalina, Oleg Rogov, Aung Kyaw Htet, and 24 others. 2026b. Polar: A benchmark for multilingual, multicultural, and multi-event online polarization. *Preprint*, arXiv:2505.20624.

Steve Rathje, Jay J Van Bavel, and Sander Van Der Linden. 2021. Out-group animosity drives engagement on social media. *Proceedings of the national academy of sciences*, 118(26):e2024292118.

Sebastian Ruder. 2017. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*.

# A Figures

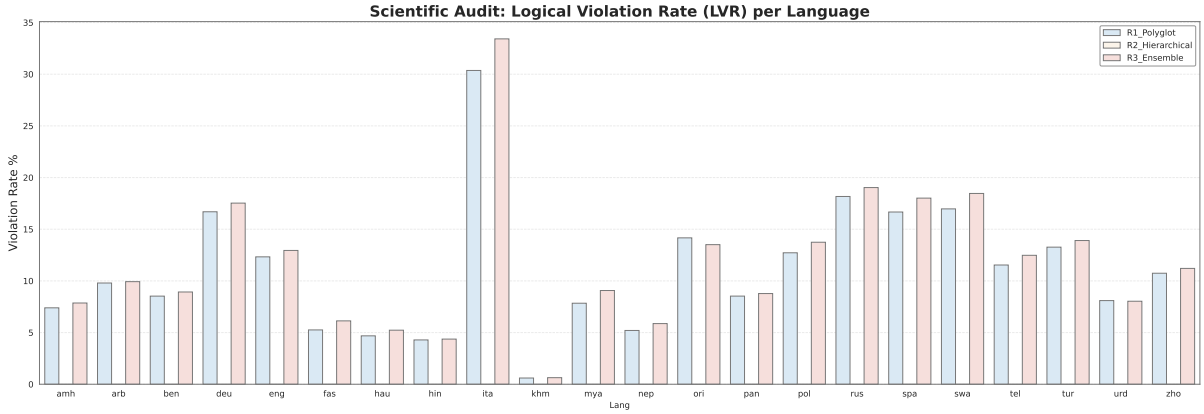


Figure 2: Logical Violation Rates (LVR) by Language. Base ensemble ($R3$) frequently predicts polar manifestations for non-polarized text; Hierarchical ($R2$) and Weighted ($R1$) architectures enforce taxonomic integrity.

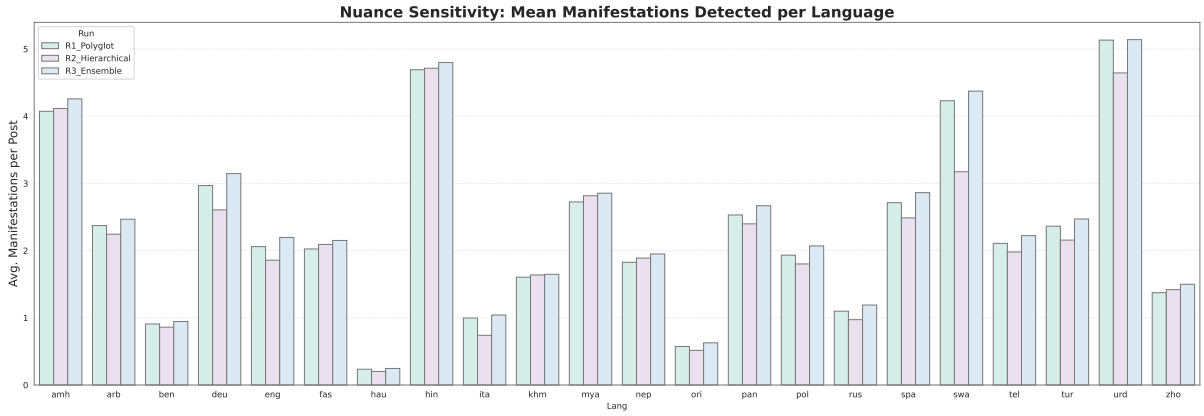


Figure 3: Nuance Sensitivity Analysis. Mean number of manifestations identified per sequence. The Broad Ensemble ($R3$) exhibits the highest sensitivity across high-context scripts (e.g., Urdu, Hindi, Amharic), followed closely by the Weighted Polyglot ($R1$) configuration. The Hierarchical ($R2$) model shows a more conservative identification rate due to its strict taxonomic constraints.

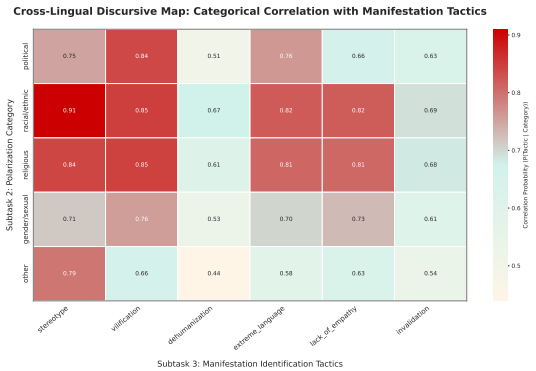


Figure 4: Cross-lingual Discursive Map. Heatmap shows conditional probabilities $P(\text{Tactic}|\text{Category})$. High correlations (e.g., *Racial* $\rightarrow$ *Stereotype*, 0.91) reveal rhetorical profiles of polarized discourse.