SemEval-2026 Task 9: Detecting Multilingual, Multicultural and Multievent Online Polarization

Usman Naseem[1], Robert Geislinger[2], Juan Ren[1], Sarah Kohail[3], Rudy Garrido Veliz[2], P Sam Sahil[2,4], Yiran Zhang[1], Marco Antonio Stranisci[5,6], Idris Abdulmumin[7], Özge Alacam[8], Cengiz Acartürk[9], Aisha Jabr[3], Saba Anwar[2], Abinew Ali Ayele[10], Elena Tutubalina[11,12,13], Aung Kyaw Htet[1], Xintong Wang[2], Surendrabikram Thapa[14], Tanmoy Chakraborty[15], Dheeraj Kodati[16], Sahar Moradizeyveh[1], Firoj Alam[17,18], Ye Kyaw Thu[19], Shantipriya Parida[20], Ihsan Ayyub Qazi[21], Lilian Wanzare[22], Nelson Odhiambo Onyango[22], Clemencia Siro[23], Ibrahim Said Ahmad[24,25], Adem Chanie Ali[2,10], Martin Semmann[2], Chris Biemann[2], Shamsuddeen Hassan Muhammad[26], Seid Muhie Yimam[2]

[1]Macquarie University, [2]University of Hamburg, [3]Zayed University, [4]HKBK College of Engineering, [5]University of Turin, [6]aequa-tech, [7]University of Pretoria, [8]Bielefeld University, [9]Jagiellonian University, [10]Bahir Dar University, [11]AIRI, [12]KFU, [13]HSE University, [14]Virginia Tech, [15]IIT Delhi, [16]ABV-IIITM, [17]Qatar Computing Research Institute, [18]Hamad Bin Khalifa University, [19]Language Understanding Lab., Myanmar, [20]AMD Silo AI, [21]Lahore University of Management Sciences, [22]Maseno University, [23]Centrum Wiskunde & Informatica, [24]Bayero University Kano, [25]Northeastern University, [26]Imperial College London,

Contact: usman.naseem@mq.edu.au and seid.muhie.yimam@uni-hamburg.de

Abstract

We present SemEval-2026 Task 9, a shared task on online polarization detection, covering 22 languages and comprising over 110K annotated instances. Each data instance is multi-labeled with the presence of polarization, polarization type, and polarization manifestation. Participants were asked to predict labels in three subtasks: (1) detecting the presence of polarization, (2) identifying the type of polarization, and (3) recognizing the polarization manifestation. The three tasks attracted over 1,000 participants worldwide and more than 10k submissions on Codabench. We received final submissions from 67 teams and 69 system description papers. We report the baseline results and analyze the performance of the best-performing systems, highlighting the most common approaches and the most effective methods across different subtasks and languages. The dataset and other resources for this task are publicly available.[1]

1 Introduction

Online polarization, defined as sharp division and antagonism between social, political, or identity groups, has become a pervasive threat to democratic institutions, civil discourse, and social cohesion worldwide (Waller and Anderson, 2021). It is often fueled by biased or inflammatory content in digital media, strengthening echo chambers and undermining mutual understanding (Garimella, 2018). Polarized discourse amplifies ideological divides and can escalate into hate speech, harassment, and real-world violence (Piazza, 2023; Martínez-España et al., 2024). Therefore, early detection of polarization is essential for designing interventions that promote healthier online ecosystems.

In this shared task, we provide participants with POLAR, a large-scale, multilingual, multicultural, and multi-event dataset for fine-grained polarization detection (Naseem et al., 2026). The task challenges participants to develop systems that can automatically detect and classify polarized content across multiple languages, cultural contexts, and event types. POLAR covers 22 languages spanning seven language families and comprises over 110,000 annotated instances (see Figure 1 for the geographic and linguistic diversity represented). Table 1 presents the data distribution across the train, development, and test splits. This shared task supports three complementary subtasks:

[1] https://polar-semeval.github.io/

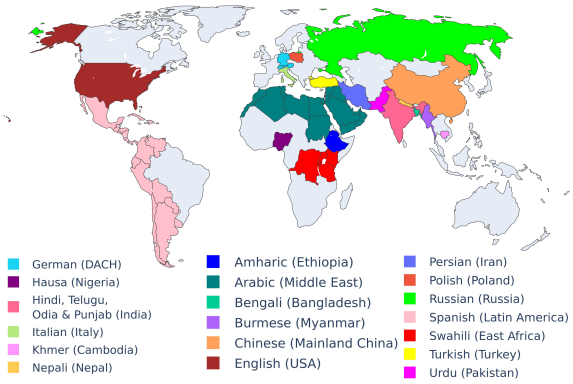

Figure 1: Languages represented in a world map covered by POLAR, covering diverse linguistic and regional contexts. The language and societal context can present themselves across varied areas. Language assignments to countries and regions are approximate.

Lang.	Train	Dev	Test	Total	Inner Agr. (κ)
amh	3,332	166	1,501	4,999	0.59
arb	3,380	169	1,521	5,070	0.25
ben	3,333	166	1,501	5,000	0.59
deu	3,180	159	1,432	4,771	0.10*
eng	3,222	160	1,452	4,834	0.39
fas	3,295	164	1,484	4,943	0.78
hau	3,651	182	1,644	5,477	0.48
hin	2,744	137	1,236	4,117	0.49
ita	3,334	166	1,538	5,038	0.39
khm	6,640	332	2,988	9,960	0.83
mya	2,889	144	1,301	4,334	0.13
nep	2,005	100	903	3,008	0.79
ori	2,368	118	1,066	3,552	0.46
pan	1,700	100	809	2,609	0.55*
pol	2,391	119	1,077	3,587	0.46
rus	3,348	167	1,508	5,023	0.39
spa	3,305	165	1,488	4,958	0.26
swa	6,991	349	3,147	10,487	0.56
tel	2,366	118	1,066	3,550	0.7
tur	2,364	115	1,093	3,572	0.46
urd	3,563	177	1,606	5,346	0.29 / 0.70*
zho	4,280	214	1,927	6,421	0.64
Total	**73,681**	**3,687**	**33,288**	**110,656**	

Table 1: Data distribution across the train, development, and test splits, along with inner agreement. Inner Agre. denotes inter-annotator agreement per language (Fleiss's κ unless otherwise noted). * denotes exceptions: German uses Krippendorff's α; Punjabi reports identical Krippendorff's α and Cohen's κ; Urdu reports Fleiss's κ / Cohen's κ.

- **Binary Polarization Detection:** Determine whether a given text expresses polarization. We refer to this task as **POLARDETECT**.
- **Polarization Type Classification:** Identify the social dimension underlying polarization (e.g., political, religious, racial). We refer to this task as **POLARTYPE**.
- **Manifestation Identification:** Detect how polarization is rhetorically manifested, including strategies such as stereotyping, deindividuation, vilification, dehumanization, extreme language, and other rhetorical devices. We refer to this task as **POLARMANIFEST**.

  Each team could submit results for subtask 1, 2, 3, or all three subtasks in one or more languages. Our official evaluation metrics were the average Macro F 1. Our tasks attracted over 1000 participants, with 548 final submissions in the test phase and 69 system description papers. Subtask 1 received the most submissions (267), followed by subtask 2 with 161, and subtask 3 with 120.

2 Related Work

Online polarization poses a threat to social cohesion, exacerbated by social media echo chambers and biased content (Waller and Anderson, 2021; Iandoli et al., 2021; Garimella, 2018). As social media and other online platforms become key arenas for political and cultural discourse, the need for early detection and nuanced understanding of polarization has grown significantly. Polarization detection is important for content moderation, peace building, responsible digital governance, and healthy democracy. Foundational research has defined polarization as both intergroup hostility and blind ingroup cohesion (Arora et al., 2022), and has highlighted its relationship with hate speech, fragmentation, and incivility (Mathew et al., 2021).

A growing body of research has documented the role of online spaces in intensifying polarization across different regions (Kubin and von Sikorski, 2021; Barberá, 2020; Gitlin, 2016; Soares and Recuero, 2021). However, most computational work focuses on high-resource languages and event- or region-specific datasets, limiting generalizability (Kubin and von Sikorski, 2021). This leaves a significant gap in our ability to generalize findings across cultures, languages, and events, especially for low-resource languages or multilingual regions.

The lack of standardized datasets across languages has hindered progress in developing and evaluating polarization detection models with cross-lingual or cross-cultural capabilities. Recent shared tasks on hate speech and toxicity (Basile et al., 2019; Pamungkas et al., 2020) have expanded the language and domain coverage, yet remain less fine-grained regarding polarization's diverse types and rhetorical manifestations. This shared task addresses this gap by presenting a comprehensive, fine-grained dataset benchmark for multilingual, multicultural, and multievent online polarization, enabling robust cross-lingual and context-aware

modeling.

3 POLAR Dataset Construction

3.1 Operational Definitions

Our work (Naseem et al., 2026) defines polarization as the increasing extremity of opinions, beliefs, or behaviors, resulting in heightened inter-group divisions and conflict. Besides, we defined polarization types including political, racial or ethnic, religious, gender or sexual, ant other. We further distinguish polarization by its rhetorical manifestations, containing stereotype, vilification, dehumanization, extreme language, lack of Empathy, and Invalidation

3.2 Data Collection

We collected data from a range of online platforms, including major social media sites, local news, and commentary forums. For several languages, including Burmese, Polish, and Chinese, we sampled and re-annotated instances from existing toxic or hate speech datasets.

The curated dataset covers diverse real-world events, grounding event selection in the sociopolitical and socioeconomic contexts specific to each language and cultural setting. The data span a broad range of events and issues, including armed conflicts, elections and party politics, public health crises, large-scale migration, climate change, and broader socioeconomic debates. The dataset also includes discussions related to gender and indigenous rights, religion, and ideology.

We provide more detailed information about the definitions of the categories, annotation guidelines, collected events, and data processing in detail in Naseem et al. (2026).

3.3 Annotation Process and Guidelines

We used a hybrid annotation strategy, leveraging crowd-sourced annotators and trained community annotators for low-resource languages where crowd-sourced annotation support is limited. For the crowd-sourced setting, we used Mechanical Turk[2] and Prolific[3], and annotators were selected based on their prior experience and annotation quality. Specifically, we filtered candidates using historical annotation agreement scores and conducted pilot rounds to identify those with consistent performance.

[2] https://www.mturk.com

[3] https://www.prolific.com

Given the cultural and linguistic breadth of POLAR, we developed detailed, multilingual annotation guidelines in English, and then translated and culturally adapted them for each target language.

> Annotators were instructed to:
>
> - Identify whether a text is polarized
> - If the text is classified as polarized, tag the type of polarization (political, racial/ethnic, religious, gender/sexual identity, other)
> - If the text is classified as polarized, tag its manifestations or rhetorical tactics (stereotyping/deindividuation, vilification, dehumanization, extreme language, lack of empathy, invalidation).

Multiple labels were allowed due to the conceptual and contextual overlap often observed in polarized content. The details about the guidelines, annotation process, and annotator reliability are described in (Naseem et al., 2026).

3.4 Annotators' Reliability

To evaluate annotation quality, we report Fleiss' Kappa, Cohen's kappa, and Krippendorff's alpha as inter-annotator agreement (IAA) metrics. Different metrics are used because the annotation setups differ across languages, having different numbers of annotators per instance (Artstein and Poesio, 2008). As shown in Table 1, the IAA scores vary between languages, with the majority showing moderate agreement and a few, such as "khm" and "tel" achieving good agreement. Although guidelines were standardized, their interpretation was influenced by cultural and political context, especially in languages with lower agreement, where some terms may not have direct equivalents across cultures. Latent content or sarcasm often required annotators to draw on their own socio-political knowledge, highlighting the perspectivist nature of polarization (Cabitza et al., 2023). Thus, low agreement can indicate socio-pragmatic complexity rather than error, signaling that polarization markers may not have universal meanings and that divergences can reveal inherent ambiguity in stimuli or interpretation (Aroyo and Welty, 2015).

4 Task Description

The participants received the data of texts from different sources and different lengths. They were

instructed to classify the texts on polarization and its components. The task comprised three subtasks, of which the participants could choose to participate in one or more.

4.1 Subtasks

Subtask 1: POLARDETECT The participants had to correctly assign whether the text was polarized or not polarized, a straightforward binary decision based on the definition of polarization used. All 22 languages were available in this subtask.

Subtask 2: POLARTYPE Based on a polarized text selected in **POLARDETECT**, the participants were asked to assign the text into a type of polarization: political, racial or ethnic, religious, gender or sexual, or other (based on economic class, media, etc.). All 22 languages were available in this subtask as well.

Subtask 3: POLARMANIFEST Given the polarized text, and the type(s) of polarization of that text (i.e., political, racial or ethnic, religious, gender or sexual, or other), participants had to correctly predict the label of manifestation(s) of the polarized text: stereotype, vilification, dehumanization, extreme language and absolutism, lack of empathy, or invalidation. The languages: Burmese (mya), Italian (ita), Polish (pol), and Russian (rus) were not present in this subtask. Resulting in the data available for 18 languages.

4.2 Task Organisation

We used Codabench as the competition platform, setting up three different competitions, one for each subtask, to allow individual participation.

We released pilot datasets before the start of the shared task to help participants better understand the task, such as data structure, the language involved, and the labels. We provided participants with a starter kit on GitHub, resources for beginners, and organized a Q&A session along with a writing tutorial for junior researchers. Participants were also supported with more details on each task, and their concerns were answered throughout the Discord server of the task and through emails forwarded to organizers. Our participants were based in different parts of the world, as shown in Figure 2. The task consisted of two phases: (1) the development phase and (2) the evaluation phase. During the development phase, the leaderboard was open, allowing a maximum of 999 submissions per participant. In the evaluation phase, the leaderboard was closed, and each participant was allowed up to five submissions. Only the last submission is being considered for the official ranking.

Figure 2: Participants came from 28 unique countries and regions: Australia, Bangladesh, China, Egypt, France, Germany, Greece, India, Ireland, Italy, Japan, Malaysia, Mexico, Nigeria, Pakistan, Portugal, Romania, Saudi Arabia, Slovakia, South Africa, South Korea, Spain, Syria, Taiwan, United Kingdom, United States, Uruguay, and Vietnam.

4.3 Evaluation Metrics and Baselines

Evaluation Metrics We use the average macro F1 score for participants' results by comparing predicted and the gold-standard labels.

Our Baselines We provide our baseline for each language by applying LaBSE (Liu et al., 2019). We finetuned LaBSE using the training data for each language for all three subtasks. Table 4, 5, and 6 show the average macro F1 of top-performing systems compared to our baseline in all three subtasks.

5 Participating Systems and Results

POLAR was the most popular SemEval competition on Codabench in 2026. Our three subtasks rank 1st, 3rd, and 7th in popularity among the 18 subtasks across the 12 shared tasks in SemEval 2026[4]. Our shared task attracted over 1,000 participants from 28 countries and regions worldwide, as illustrated in Figure 2. Specifically, POLAR attracted 533 participants in Subtask 1, 344 in Subtask 2, and 185 participants in Subtask 3 (see Table 2). In the development phase, more than 5.7K submissions were made to Subtask 1, more than 2.5K to Subtask 2, and over 1k to Subtask 3. In the test phase, 267 submissions were made to Subtask 1, 161 for Subtask 2, and 120 for Subtask 3. The official results included 67 teams and 69 system description papers; one team submitted three papers, one for each subtask. Overall, 43% of

[4] https://www.codabench.org/competitions/public

teams participated in only one subtask, 16% in two subtasks, and 41% in all three subtasks. Participants generally preferred to submit systems for all languages rather than a subset of languages. Specifically, 41% of teams in subtask 1 submitted predictions for all languages, compared to 56% and 63% in subtask 2 and subtask 3.

Subtask	Participants	Dev Submissions	Test Submissions	Teams in Results
1	533	5,764	267	56
2	344	2,555	161	39
3	185	1,029	110	25
Total	**1,061**	**9,886**	**548**	**69**

Table 2: Participation statistics for the POLAR shared task on Codabench."Teams in Results" are those that submitted system description papers. In total, 67 teams with 69 papers appear on the leaderboard across all subtasks.

We report results only for teams that submitted a system description paper. Table 3 summarizes the distribution of top-3 placements across subtasks. Table 4 presents the results for Subtask 1, which had 79 participating teams. Table 5 shows the results for Subtask 2, with 47 participating teams, while Table 6 reports the results for Subtask 3, which had 30 participating teams.

5.1 Subtask 1: POLARDETECT

5.1.1 Best Performing Systems

Team UTokyo Tsuruoka Lab achieved one of the strongest performances in the competition, ranking first in 8 out of 22 languages. Their system is based on the instruction-tuned Gemma-3-12B-IT (Team et al., 2025a) model and introduces an efficient one-forward-pass strategy for both training and inference. To enable memory-efficient fine-tuning of the large language model, they utilized Unsloth (Daniel Han and team, 2023), which reduces GPU memory requirements during training. A key aspect of their approach is a selective-token training method, where the model predicts labels through one-token inference rather than using a traditional multi-label classification head. This formulation simplifies the prediction process and improves inference efficiency (Tangkulung and Tsuruoka, 2026).

Team NYCU-NLP proposed a system based on instruction-tuned small language models, including Gemma-3 (27B) (Team et al., 2025a), Mistral Small 3.2 (24B) (AI, 2025), and Phi-4 (14B) (Abdin et al., 2024). Their approach leverages parameter-efficient fine-tuning techniques, such as LoRA (Hu et al., 2021) and adapters, allowing the models to be adapted to the task without updating all parameters. The models were trained using task-specific prompts, which were iteratively refined to improve performance across the tasks. To combine the strengths of different models, the team employed a stacking-based ensemble strategy, aggregating predictions from multiple small language models to capture complementary signals (Lin et al., 2026) .

5.1.2 Takeaways

A first general trend emerging from results is the challenge of achieving consistent results on polarization detection across all languages (Table 4). While the best systems for each language often reach an F1-score above 0.8, performances are significantly lower for two low-resourced languages (Khmer, Burmese) and two high-resourced languages (Italian and German). This suggests that the intrinsic challenges in polarization detection could be caused by the lack of models' knowledge about local contexts rather than linguistic factors. Observing models that submitted results for all languages (55 out of 104), this trend seems to be confirmed. The two best performing systems ranked 10th or below on 5 languages, both struggling with Farsi, Hausa, Khmer, and Italian.

Language-specific approaches (28 out of 104) did not perform well, though. Only two systems managed to be ranked among the top-5 in Bengali leaderboard: **CUET823** (Mallik and Dhar, 2026) (2) and **transformer_1376** (Saha and Saha, 2026) (5). Finally, it is worth mentioning teams focused on specific macro-regions. It is the case of **PolAR Bears** (Ulli and Kumari, 2026), which submitted runs for languages spoken in Southern Asia (Bengali, Hindi, Odia, and Telugu). Results achieved by this team were considerably subpar, though, as they consistently ranked below 10. This once again demonstrates that handling cultural variation in the computational understanding of polarization is still a major challenge in NLP research.

5.2 Subtask 2: POLARTYPE

5.2.1 Best Performing Systems

Team UTokyo Tsuruoka Lab managed to score first place in Subtask 2 in 7 out of the 22 languages, making them the best-scoring team. They used the same models and fine-tuning tools as their efforts

Subtask 1				
Team	**Total**	**1st**	**2nd**	**3rd**
UTokyo Tsuruoka Lab	12	8	4	0
NYCU-NLP	12	3	5	4
PSK	9	2	4	3
CYUT	4	2	0	2
SMASH	7	1	2	4
Lingo Research Group	5	1	1	3
taien	3	1	1	1
OZemi	2	1	0	1
Sagarmatha	1	1	0	0
mdok-style	1	1	0	0
PhatThachDau	1	1	0	0
MKJ	2	0	2	0
StanceLab	2	0	2	0
CUET-823	1	0	1	0
PolDeck	1	0	1	0
Projet Fil Rouge 821	1	0	1	0
UMUSP	1	0	1	0
PolaFusion	1	0	0	1
YEZE	1	0	0	1
MoMo	1	0	0	1
Semantic Vectors	1	0	0	1
Tralaleros	1	0	0	1

Subtask 2				
Team	**Total**	**1st**	**2nd**	**3rd**
UTokyo Tsuruoka Lab	13	7	5	1
NYCU-NLP	15	6	5	4
SMASH	13	4	6	3
Lingo Research Group	7	2	1	4
PolaFusion	4	1	0	3
Sagarmatha	2	1	0	1
AIvengers	1	0	1	0
ShefFriday	1	0	1	0
Stochastic Gradient Descenders	1	0	1	0
MSqrd	1	0	1	0
CYUT	1	0	0	1
YEZE	1	0	0	1
mdok-style	1	0	0	1
YNU-HPCC	1	0	0	1
PolarMind	1	0	0	1

Subtask 3				
Team	**Total**	**1st**	**2nd**	**3rd**
SMASH	16	9	4	3
NYCU-NLP	11	7	3	1
Sagarmatha	4	2	0	2
Ping An	4	0	4	0
PolaFusion	7	0	2	5
OZemi	3	0	2	1
AIvengers	4	0	1	3
CYUT	1	0	1	0
ShefFriday	1	0	1	0
YEZE	2	0	0	2
Lingo Research Group	1	0	0	1

Table 3: Top-3 placements achieved by teams across the three subtasks. For each task, the table reports the total number of top-3 finishes achieved by each team and their breakdown into 1st, 2nd, and 3rd places.

in Subtask 1, they performed key modifications to account for the multilabel setup. They used JSON finetuning as an auto-regressive baseline, where they instructed the model to generate JSON objects with a binary decision for each label, to later use it during training and inference with cross-entropy loss and greedy rule-based approach, respectively. Finally, they adapted SALSA for a multi-label classification (Tangkulung and Tsuruoka, 2026).

Team NYCU-NLP found difficulty in the "Other" category, and therefore implemented a heuristic based on the prediction made in Subtask 1, using it as an auxiliary signal during inference. With this modification to their initial approach, the team managed to land in first place for 6 of the 22 languages, close second to the best performing team (Lin et al., 2026).

5.2.2 Takeaways

Results of Subtask 2 (Table 5) are significantly lower than Subtask 1, with 7 languages in which the highest F-score was below 0.6 and only 3 languages with a score above 0.8. No specific trends about language families and macro-regions emerge, though.

Similarly to what has been observed in Section 5.1.2, the highest ranked models exhibit a drop in performance related to specific languages, even if they are not the same from the previous task. E.g., team **UTokyo Tsuruoka Lab**, which ranked 25th in Subtask 1 for Italian, ranked first in Subtask 2.

Additional insights from the task emerge from model performances across different languages and topics. Table 7 reports the percentage of polarization types correctly predicted by all the models that participated in the tasks (true positives). As it can be observed, a strong cultural variation seems to emerge across languages. E.g., the percentage of correct prediction of Gender/Sexual polarity types is 0.239 for Amharic and 0.825 for Chinese. Such oscillation is also present in languages from the same macro-region. For instance, only 0.365 Religious polarity types are correctly identified in Telugu; 0.905 in Hindi. Therefore, the generalization of polarity types across different languages and local contexts remains an open issue for the NLP research community.

5.3 Subtask 3: POLARMANIFEST

5.3.1 Best Performing Systems

Team SMASH achieved strong performance in the competition, ranking first in 9 out of 18 languages. Their system relies on full model fine-tuning and uses 5-fold cross-validation with three random seeds for each language. Logits are averaged across seeds and folds to obtain out-of-fold predictions, which are then used to tune per-language ensemble weights and label-specific thresholds that maximize macro-F1. For final predictions, the model is retrained on all training data, logits are averaged across seeds, and the optimized weights and thresholds are applied to generate the final labels (Bokaei et al., 2026).

Team NYCU-NLP changed little in their approach from their Subtask 2. However, they still manage to come in first place for multiple languages, a total of 7 from the 18 languages pool

Lang	Team	Score	Lang	Team	Score	Lang	Team	Score	Lang	Team	Score
amh	PSK	0.800	hau	PhatThachDau	0.834	pan	UTokyo Tsuruoka Lab	0.826	rus	UTokyo Tsuruoka Lab	0.830
	UTokyo Tsuruoka Lab	0.795		Projet Fil Rouge 821	0.832		PSK	0.812		NYCU-NLP	0.823
	Lingo Research Group	0.793		OZemi	0.831		NYCU-NLP	0.811		CYUT	0.814
	baseline	0.764		*baseline*	0.821		*baseline*	0.749		*baseline*	0.748
arb	UTokyo Tsuruoka Lab	0.849	hin	CYUT	0.828	tel	Sagarmatha	0.905	spa	UTokyo Tsuruoka Lab	0.803
	PSK	0.848		PSK	0.824		SMASH	0.901		NYCU-NLP	0.800
	NYCU-NLP	0.843		Lingo Research Group	0.821		Tralaleros	0.897		SMASH	0.798
	baseline	0.812		*baseline*	0.782		*baseline*	0.889		*baseline*	0.750
ben	UTokyo Tsuruoka Lab	0.863	khm	SMASH	0.774	tur	NYCU-NLP	0.833	swa	PSK	0.811
	CUET-823	0.858		StanceLab	0.761		UTokyo Tsuruoka Lab	0.830		SMASH	0.810
	NYCU-NLP	0.854		Semantic Vectors	0.755		PSK	0.809		taien	0.799
	baseline	0.825		*baseline*	0.737		*baseline*	0.750		*baseline*	0.790
ita	mdok-style	0.730	fas	OZemi	0.835	mya	taien	0.891	pol	Lingo Research Group	0.843
	StanceLab	0.672		taien	0.831		MKJ	0.887		NYCU-NLP	0.835
	PolaFusion	0.671		MKJ	0.831		NYCU-NLP	0.887		PSK	0.835
	MoMo	0.671		PSK	0.828		SMASH	0.885		SMASH	0.828
	baseline	0.564		*baseline*	0.835		*baseline*	0.861		*baseline*	0.773
deu	NYCU-NLP	0.761	nep	NYCU-NLP	0.924	urd	UTokyo Tsuruoka Lab	0.820			
	UTokyo Tsuruoka Lab	0.753		Lingo Research Group	0.918		NYCU-NLP	0.817			
	CYUT	0.747		SMASH	0.914		Lingo Research Group	0.816			
	baseline	0.686		*baseline*	0.883		*baseline*	0.742			
eng	UTokyo Tsuruoka Lab	0.825	ori	UTokyo Tsuruoka Lab	0.826	zho	CYUT	0.932			
	PolDeck	0.819		UMUSP	0.814		UTokyo Tsuruoka Lab	0.929			
	PSK	0.818		YEZE	0.812		NYCU-NLP	0.927			
	baseline	0.773		*baseline*	0.776		*baseline*	0.864			

Table 4: Top three performing systems for each language in subtask 1 evaluated using macro-F1 score.

Lang	Team	Score	Lang	Team	Score
amh	PolaFusion	0.670	nep	NYCU-NLP	0.810
	SMASH	0.650		Lingo Research Group	0.805
	YEZE	0.649		mdok-style	0.803
	baseline	0.471		*baseline*	0.664
arb	NYCU-NLP	0.670	ori	UTokyo Tsuruoka Lab	0.603
	UTokyo Tsuruoka Lab	0.668		AIvengers	0.594
	SMASH	0.658		NYCU-NLP	0.578
	baseline	0.559		*baseline*	0.423
ben	Lingo Research Group	0.422	pol	UTokyo Tsuruoka Lab	0.650
	NYCU-NLP	0.401		NYCU-NLP	0.640
	SMASH	0.378		Lingo Research Group	0.625
	baseline	0.268		*baseline*	0.416
deu	UTokyo Tsuruoka Lab	0.620	rus	NYCU-NLP	0.630
	NYCU-NLP	0.616		SMASH	0.619
	Lingo Research Group	0.599		UTokyo Tsuruoka Lab	0.617
	baseline	0.533		*baseline*	0.409
eng	UTokyo Tsuruoka Lab	0.532	spa	NYCU-NLP	0.681
	Stochastic Gradient Descenders	0.516		UTokyo Tsuruoka Lab	0.674
	NYCU-NLP	0.514		SMASH	0.673
	baseline	0.347		*baseline*	0.593
fas	SMASH	0.644	swa	SMASH	0.569
	MSqrd	0.609		UTokyo Tsuruoka Lab	0.540
	PolaFusion	0.605		NYCU-NLP	0.522
	baseline	0.525		*baseline*	0.402
hau	NYCU-NLP	0.480	tel	Sagarmatha	0.465
	SMASH	0.454		SMASH	0.458
	Sagarmatha	0.427		PolaFusion	0.446
	baseline	0.216		*baseline*	0.426
hin	SMASH	0.807	tur	UTokyo Tsuruoka Lab	0.652
	NYCU-NLP	0.801		NYCU-NLP	0.646
	YNU-HPCC	0.793		Lingo Research Group	0.624
	baseline	0.700		*baseline*	0.484
ita	UTokyo Tsuruoka Lab	0.551	urd	Lingo Research Group	0.798
	ShefFriday	0.538		SMASH	0.790
	CYUT	0.484		NYCU-NLP	0.789
	baseline	0.261		*baseline*	0.739
khm	UTokyo Tsuruoka Lab	0.705	zho	NYCU-NLP	0.844
	SMASH	0.702		UTokyo Tsuruoka Lab	0.835
	PolaFusion	0.699		Lingo Research Group	0.825
	baseline	0.586		*baseline*	0.631
mya	SMASH	0.736			
	UTokyo Tsuruoka Lab	0.708			
	PolarMind	0.702			
	baseline	0.551			

Table 5: Top three performing systems for each language in subtask 2 evaluated using macro-F1 score.

for this Subtask (Lin et al., 2026).

5.3.2 Takeaways

As with the previous Subtask, a decrease in performance is noticeable but in a more dramatic tone. Table 6 shows the best performing systems, and for only one language, Urdu, the score was above 0.8. Furthermore, a score above 0.7 was achieved by only three more languages, and for five languages the score was below 0.5. A similar trend can be seen in Table 7, where correct labeling did not improve much from the previous subtask.

It is worth noting that the best-performing languages are from the region of Southern Asia: Hindi, Nepali, and Urdu. In these languages, the SMASH (Bokaei et al., 2026) and NYCU-NLP (Lin et al., 2026) teams either tied or are very close in their score. It is assumed that their approaches perform specifically well for these languages, as their scores in other languages fall drastically behind.

6 Discussion

6.1 Popular Methods

The most common methods include ensemble prediction, fine-tuning, threshold tuning for language or class labels, and data augmentation.

Model Families The Qwen family (Bai et al., 2023) is the most frequently used model (31%), followed by the LLaMA family (Touvron et al., 2023) (20%) and the Gemma/Gemini (Team et al.,

Lang	Team	Score	Lang	Team	Score
amh	SMASH	0.579	nep	NYCU-NLP	0.713
	NYCU-NLP	0.559		SMASH	0.712
	AIvengers	0.554		Lingo Research Group	0.669
	baseline	0.512		*baseline*	0.602
arb	NYCU-NLP	0.646	ori	SMASH	0.330
	SMASH	0.641		Ping An	0.328
	YEZE	0.610		NYCU-NLP	0.297
	baseline	0.568		*baseline*	0.240
ben	SMASH	0.281	pan	NYCU-NLP	0.544
	Ping An	0.255		SMASH	0.541
	PolaFusion	0.249		AIvengers	0.529
	baseline	0.258		*baseline*	0.484
deu	NYCU-NLP	0.518	spa	SMASH	0.541
	ShefFriday	0.515		NYCU-NLP	0.520
	SMASH	0.513		PolaFusion	0.507
	baseline	0.471		*baseline*	0.480
eng	Sagarmatha	0.511	swa	SMASH	0.584
	Ping An	0.507		AIvengers	0.565
	SMASH	0.507		OZemi	0.562
	baseline	0.466		*baseline*	0.565
fas	SMASH	0.493	tel	SMASH	0.445
	OZemi	0.476		PolaFusion	0.429
	Sagarmatha	0.461		Sagarmatha	0.424
	baseline	0.395		*baseline*	0.392
hau	Sagarmatha	0.207	tur	NYCU-NLP	0.538
	OZemi	0.206		Ping An	0.537
	PolaFusion	0.204		PolaFusion	0.515
	baseline	0.206		*baseline*	0.449
hin	SMASH	0.771	urd	NYCU-NLP	0.821
	NYCU-NLP	0.770		SMASH	0.821
	PolaFusion	0.759		YEZE	0.815
	baseline	0.701		*baseline*	0.771
khm	SMASH	0.437	zho	NYCU-NLP	0.719
	PolaFusion	0.400		CYUT	0.700
	AIvengers	0.377		SMASH	0.677
	baseline	0.343		*baseline*	0.461

Table 6: Top three performing systems for each language in subtask 3 evaluated using macro-F1 score.

2025a) family (19%). Several teams also employed GPT (OpenAI et al., 2024), Mistral (AI, 2025), and BERT-based encoder models (each 7%), while Deepseek (DeepSeek-AI et al., 2025), Phi (Abdin et al., 2024), GLM (Team et al., 2025b), and Nemotron (Nvidia et al., 2024) are used in a only small number of systems.

Ensemble models Model ensembling is one of the commonly used techniques. Methods include ensembling multiple transformer models, combining transformer encoders with LLMs, or integrating models from different architectural families. Teams adopted various strategies to determine ensemble weights, including learning weights from out-of-folder logits, soft-voting ensembles, and weighted or average fusion.

Fine tuning Approximately 39% of teams reported applying fine-tuning, while half of them employ parameter-efficient fine-tuning (PEFT) such as LoRA (Hu et al., 2021).

Loss optimization Because the multi-label subtasks involved heavily imbalanced distributions, standard cross-entropy was frequently replaced with more robust loss optimization techniques. Popular alternatives included Asymmetric Loss (ASL), Weighted Binary Cross-Entropy, and Focal Loss.

Data augmentation Approximately 38% of teams reported using data augmentation to mitigate class or language imbalance. Common techniques include back-translation, cross-lingual translation, extending instances with generated explanation, paraphrasing, hard-negative generation, or easy data augmentation like lowercasing, uppercasing, shuffling words, or replacing them with synonyms.

Per-label and per-language threshold calibration Most systems report using per-label or per-language threshold tuning to address underrepresented label or language distribution, often improving performance in imbalanced settings.

6.2 Best performing Systems

Based on the overall ranking statistics (See Table 3) across languages and subtasks, we highlight three teams that demonstrated particularly strong performance in the shared task: UTokyo Tsuruoka Lab (Tangkulung and Tsuruoka, 2026), NYCU-NLP (Lin et al., 2026), and SMASH (Bokaei et al., 2026). UTokyo Tsuruoka Lab achieved the most first-place rankings across Subtask 1 and Subtask 2, indicating strong peak performance. Team NYCU-NLP obtained 38 top-3 placements across all subtasks, the highest among participating teams. Team SMASH also achieved competitive results, ranking first in Subtask 3 and obtaining 36 top-3 placements across the evaluation.

The strategies behind these strong results differ across teams. UTokyo fine-tuned Gemma-3-12B-IT and Gemma-3-27B-IT-bnb-4bit (Team et al., 2025a) using LoRA (Hu et al., 2021). They attribute their performance to a single-forward-pass inference paradigm rather than JSON-format inference. NYCU-NLP employed a stacking-based ensemble strategy using three LLMs: Gemma-3 (27B) (Team et al., 2025a), Mistral Small 3.2 (24B) (AI, 2025), and Phi-4 (14B) (Abdin et al., 2024), and also introduced a heuristic method for predicting the "other" category in Subtask 2. SMASH adopted an ensemble approach that combines monolingual and multilingual encoder-based transformers, including mDeBERTa (He

et al., 2023), XLM-R (Conneau et al., 2020), and mBERT (Devlin et al., 2019). In addition, they attribute their performance to out-of-fold ensemble weight tuning and per-class threshold calibration.

7 Conclusion

This shared task has attracted over 1,000 participants, with 69 system description papers submitted, making it the most popular task among all 12 SemEval/-2026 tasks. While most participants adopted commonly used strategies, such as data augmentation, fine-tuning, ensemble models, and per-class or per-language threshold calibration, the top-performance teams employed different approaches. No single method dominates, and strong performance can be achieved through multiple strategies.

We created a successful and challenging experience for the computational linguistics community. Through an engaging team and willing organizers, the task was the most involved task for the SemEval-2026. Bringing forward the pressing issue of polarization that occurs in many cultures and languages, through multiple events, many interesting approaches emerged, and this communal effort has fostered research opportunities and collaboration. As a byproduct, the remarkable dataset has been created and made public to help future research on polarization.

Limitations

We consider our task an important step towards multilingual, multicultural, and multi-event polarization analysis, but several limitations remain. Particularly on the quality assurance of the labeling, as some of the languages utilized crowdsource, and it often comes with a disadvantage in the full grasp and understanding of the task at hand from the annotators. Quality assessments, like pilots and control questions, were placed to help with this, but some inconsistencies may linger.

For some of the languages in our task, the available data was limited, which could limit the generalizability of the resulting models or may have resulted in abstention from participation. Future efforts or iterations should strive for wider size and diversity, and possibly explore language- or region-specific data points to platform underrepresented communities, as well as NLP researchers.

Ethics Statement

Throughout the task, we strive to have an open channel of communication and support for the participants, as well as warn them of the potential triggering content they attempted to classify.

The annotators of the dataset were also warned and given resources to handle the stress the annotation may have brought forward. They were fairly paid, in accordance with their local laws and, where valid, the standard of the crowdsourcing tools. All the annotators were either native speakers of their respective languages, hoping to capture their unique insight into polarization for their culture and language.

Polarization is a sensitive and, up to a degree, subjective topic. The resulting dataset was made public, and the data was anonymized, which therefore begs for its further usage to be in a responsible and ethical manner.

7.1 Acknowledgments

We thank the SemEval-2026 organizers for the opportunity to take our tasks into the international research community, and the participants for taking part in it in a meaningful and eager way.

We thank all annotators who participated in annotating each language for their contributions and efforts.

The University of Hamburg (UHH) team acknowledges the grant from the Google Award for Inclusion Research Program, which supports AI-MAP[5] project that results in the extension of POLAR project.

Tanmoy Chakraborty acknowledges the financial support of Anusandhan National Research Foundation (CRG/2023/001351).

Ö. Alacam received funding through the project SAIL: SustAInable Life-cycle of Intelligent Socio-Technical Systems (Grant ID NW21059A), funded by the Ministry of Culture and Science of the State of North Rhine-Westphalia (Germany).

Shamsuddeen Hassan Muhammad acknowledges the support of Google DeepMind, whose funding made this work possible.

Cengiz Acartürk thank Jagiellonian University Strategic Programme Excellence Initiative ID.UJ for providing partial financial support, and Anna Maria Wilkosz and Jakub Romanowski for the annotations.

Usman Naseem acknowledges the support of the DAAD Research Fellowship, which supported the initiation of this work.

The work of Elena Tutubalina was supported within the framework of the HSE University Basic Research Program, and the computational resources of HSE University's HPC facilities are acknowledged.

References

Marah Abdin, Jyoti Aneja, Harkirat Behl, Sébastien Bubeck, Ronen Eldan, Suriya Gunasekar, Michael Harrison, Russell J. Hewett, Mojan Javaheripi, Piero Kauffmann, James R. Lee, Yin Tat Lee, Yuanzhi Li, Weishung Liu, Caio C. T. Mendes, Anh Nguyen, Eric Price, Gustavo de Rosa, Olli Saarikivi, and 8 others. 2024. Phi-4 Technical Report. *Preprint*, arXiv:2412.08905.

Faisal Muhammad Adam, Sani Aji, Lukman Jibril Aliyu, and Abdulhamid Abubakar. 2026. Team HausaNLP at SemEval-2026 Task 9: Tackling Class Imbalance in Low-Resource Hausa Polarization Detection. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Mistral AI. 2025. mistralai/mistral-small-3.2-24b-instruct-2506. https://huggingface.co/mistralai/Mistral-Small-3.2-24B-Instruct-2506. Hugging Face model card, accessed 2026-03-13.

Jacob Altamirano, Mario Leon Pérez, Bruno Ruiz-Juarez, Luis Chiruzzo, Helena Gomez-Adorno, and

[5] https://www.hcds.uni-hamburg.de/en/research/ai-map.html

Fazlourrahman Balouchzahi. 2026. ServSocIA at Semeval-2026 Task 9: Evaluating Prompt Strategies for Polarization Detection. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Aaron Bundi Anampiu. 2026. Aaron at SemEval-2026 Task 9: Multilingual Polarization Detection using Transformer-Based Models with Class Weighting and Threshold Tuning. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Georgios Arampatzis and Avi Arampatzis. 2026. DUTH at SemEval-2026 Task 9: Joint Multilingual Fine-Tuning for Online Polarization Detection. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Pushkar Arora. 2026. CoPol at SemEval-2026 Task 9: Modeling Polarization Type Co-occurrence with Label Correlation Networks. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Swapan Deep Arora, Guninder Pal Singh, Anirban Chakraborty, and Moutusy Maity. 2022. Polarization and Social Media: A Systematic Review and Research Agenda. *Technological Forecasting and Social Change*, 183:121942.

Lora Aroyo and Chris Welty. 2015. Truth is a lie: Crowd truth and the seven myths of human annotation. *AI Magazine*, 36(1):15–24.

Ron Artstein and Massimo Poesio. 2008. Inter-coder agreement for computational linguistics. *Comput. Linguist.*, 34(4):555–596.

Surangana Aryal and Saurav K. Aryal. 2026. AI4PC-Howard University at SemEval-2026 Task 9: Evaluating Teacher-Student Weak Supervision and Direct LLM Prompting for Multilingual Political Polarization Detection. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, Binyuan Hui, Luo Ji, Mei Li, Junyang Lin, Runji Lin, Dayiheng Liu, Gao Liu, Chengqiang Lu, Keming Lu, and 29 others. 2023. Qwen Technical Report. *Preprint*, arXiv:2309.16609.

Di Bao, Jin Wang, and Xuejie Zhang. 2026. YNU-HPCC at SemEval-2026 Task 9: Hybrid Augmentation and Regularization Strategies for Multilingual Polarization Type Classification. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Pablo Barberá. 2020. *Social Media and Democracy: The State of the Field, Prospects for Reform*, chapter Social Media, Echo Chambers, and Political Polarization. Cambridge University Press Cambridge.

Arup Baruah. 2026. ABARUAH at SemEval-2026 Task 9: Multilingual Polarization Detection across Seven Indic Languages using Qwen3. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

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, Minneapolis, Minnesota, USA. Association for Computational Linguistics.

Abdulkadir Shehu Bichi and Jyoti Shekhawat. 2026. VGU-M.Tech-AI at SemEval-2026: Multilingual Multi-Label Classification of Online Polarization Types via Weighted Transformer Fine-Tuning and Adaptive Per-Label Threshold Optimization. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Zahra Bokaei, Alessandra Terranova, Yi Zheng, Tom Bidewell, and Bjorn Ross. 2026. SMASH at SemEval-2026 Task 9: Detecting Multilingual Polarisation with Encoder Ensembles and Calibrated Decision Thresholds. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Declan Booth, Gavin Abercrombie, and Simona Frenda. 2026. ILab-NLP at SemEval-2026 Task 9: Comparing XLM-RoBERTa and LLaMA-2 for Multilingual Polarization Detection. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Federico Cabitza, Andrea Campagner, and Valerio Basile. 2023. Toward a Perspectivist Turn in Ground Truthing for Predictive Computing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37-6, pages 6860–6868, Washington, DC, USA.

Miriam Calderon-Reyes, Fernando Sanchez-Vega, and Adrian Pastor Lopez-Monroy. 2026. NLP-CIMAT at SemEval-2026 Task 9: LLM-Based One-Shot and Cross-Lingual Data Augmentation for Polarization Detection. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco

Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Owen S. Cook, Meredith A. Gibbons, and Xingyi Song. 2026. ShefFriday at SemEval-2026 Task 9: LLM-Based Annotation Methods for Detecting Multilingual, Multicultural and Multievent Online Polarisation. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Adrian Dahl, Bado Völckers, and Adam Jakub Mierzwa. 2026. Tralaleros at SemEval-2026 Task 9: Multilingual Polarization Detection with Transformer-based Models. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Michael Han Daniel Han and Unsloth team. 2023. Unsloth.

Syeda Samah Daniyal, Muneeba Badar, Manal Hasan, Shifa Shah, Sandesh Kumar, and Abdul Samad. 2026. MSqrd at SemEval-2026 Task 9: Detecting Multilingual, Multicultural and Multievent Online Polarization. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Ankit Dash, Priyanshu Mittal, Piyush Prashant, and Sunil Saumya. 2026. Semantic Vectors at SemEval-2026 Task 9: Robust Multilingual Polarization Detection via Dual-Encoder Fusion and Expert Ensembling. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

DeepSeek-AI, Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, Damai Dai, Daya Guo, Dejian Yang, Deli Chen, Dongjie Ji, Erhang Li, Fangyun Lin, Fucong Dai, and 181 others. 2025. DeepSeek-V3 Technical Report. *Preprint*, arXiv:2412.19437.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota.

Abdelbasset Djamai, Sahara Hussam Al-Madi, Norah Naji Al-Zaid, Khloud Al Jallad, and Mona A. Azim. 2026. NAMAA at SemEval-2026 Task 9: Comparing Generative, Retrieval-Augmented, and Discriminative Methods for Arabic Online Polarization Detection and Type Classification. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Antoine Durand, Rémi Hamon, Matthieu Pereira, Nathan Boucneau, and Paul Cintra. 2026. pfr812 at SemEval-2026 Task 9: Multilingual Polarization Detection via Hybrid XLM-RoBERTa with Targeted Data Augmentation and Imbalance-Aware Training. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Boon Elschenbroich and Lars Britz. 2026. AIvengers at SemEval-2026 Task 9: Utilizing Language Specific Encoders for Multilingual Text Classification. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Ahmed Megahed Fetouh, Mariam Labib Francies, Nsrin Ashraf, Hamada Nayel, and Rahmath Mohammed. 2026. REGLAT at SemEval-2026 Task 9: Enhancing Arabic Online Polarization Detection Using AraBERT and Synonym Replacement Augmentation. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Julio Cesar Alves Araujo Fuganti, Tulio Ferreira Leite Da Silva, Adelino Gala, Francisco S. Marcondes, José Machado, and Paulo Novais. 2026. UMUSP at SemEval-2026 Task 9: Mitigating Cross-Lingual Interference via Selective Multilingual and Multitask Specialization. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Kiran Garimella. 2018. *Polarization on Social Media*. Ph.D. thesis, Aalto University, Finland.

Todd Gitlin. 2016. The Outrage Industry: Political Opinion Media and the New Incivility By Jeffrey M. Berry and Sarah Sobieraj Oxford University Press. *Social Forces*, 95(1):e26–e26.

Ben Grandy and Daniel Khir. 2026. PolDeck at SemEval-2026 Task 9: Multilingual Online Polarization Detection via Hybrid Model Ensembling and Data Augmentation. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Fengze Guo and Yue Chang. 2026. YEZE at SemEval-2026 Task 9: Detecting Multilingual, Multicultural and Multievent Online Polarization via Heterogeneous Ensembling. In *Proceedings of the*

20th International Workshop on Semantic Evaluation (SemEval-2026), San Diego, California. Association for Computational Linguistics.

Atharva Gupta, Dhruv Kumar, and Yash Sinha. 2026. BITS Pilani at SemEval-2026 Task 9: Structured Supervised Fine-Tuning with DPO Refinement for Polarization Detection. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Pengcheng He, Jianfeng Gao, and Weizhu Chen. 2023. DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing. In *The Eleventh International Conference on Learning Representations*.

Eduardo C. C. Hernandez-Garcia, Guillermo Ruiz, and Mario Graff. 2026. INFOTEC-NLP at SemEval-2026 Task 9: Comparing Regional Transformers and Bag-of-Words Approaches for Polarization Detection in Spanish. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. LoRA: Low-Rank Adaptation of Large Language Models. *Preprint*, arXiv:2106.09685.

Luca Iandoli, Simonetta Primario, and Giuseppe Zollo. 2021. The impact of group polarization on the quality of online debate in social media: A systematic literature review. *Technological Forecasting and Social Change*, 170:1–12.

Angelo Iannielli, Samuele Maroli, Marco Roberto, Stefano Sammartino, Valentino Vacirca, Claudio Savelli, Riccardo Coppola, and Flavio Giobergia. 2026. MINDS at SemEval-2026 Task 9: A Multi-Paradigm Approach to Cross-Lingual Polarization Detection. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Teodor Ivanusca, Dan Dodun-Des-Perrieres, and Stefana Gheorghita. 2026. StanceLab at SemEval-2026 Task 9: Addressing Class Imbalance in Multilingual Polarization Detection. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Teagan Johnson. 2026. The Counterfactuals at SemEval-2026 Task 9: Can Counterfactually-Inspired Preprocessing Help Detect Polarization? In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Maziar Kianimoghadam Jouneghani. 2026. MKJ at SemEval-2026 Task 9: A Comparative Study of Generalist, Specialist, and Ensemble Strategies for Multilingual Polarization. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Pritam Kadasi, Anuj Tiwari, and Mayank Singh. 2026. Lingo_Research_Group at SemEval-2026 Task 9: Evaluating Prompt Variants for Polarization Detection. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Inzmam Khadam, Zaufishan Mahmood, Junaid Rashid, Shamaila Hayat, and Samira Kanwal. 2026. UPR at SemEval-2026 Task 9: Multi-Label Classification of Polarization in Urdu Text Across Multiple Social Dimensions. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Wang Kongqiang and Tan Qingli. 2026. wangkongqiang at SemEval-2026 Task 9: Detecting Multilingual, Multicultural and Multievent Online Polarization. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Emily Kubin and Christian von Sikorski. 2021. The Role of (Social) Media in Political Polarization: A Systematic Review. *Annals of the International Communication Association*, 45(3):188–206.

Sandeep S. Kumar, Mothish M, and Sachin Sundar. 2026. PolarMind at SemEval-2026 Task 9: Leveraging LaBSE with Progressive Curriculum Learning for Multicultural Polarization. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Minh-Hoang Le and Khoan Khac Phung. 2026. AlphaLyrae at SemEval-2026 Task 9: Metric Learning and Asymmetric Loss for Chinese Polarization Analysis. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Joshua Lee. 2026. Joshualee2 at SemEval-2026 Task 9: Cross-Lingual Transformer-Based Polarization Detection. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Ding-Xiang Lin, Po-Chun Chu, and Lung-Hao Lee. 2026. NYCU-NLP at SemEval-2026 Task 9: Stacking Small Language Models for Multilingual, Multicultural and Multievent Polarization Detection. In

Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026), San Diego, California. Association for Computational Linguistics.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *Preprint*, arXiv:1907.11692.

Dominik Macko, Alok Debnath, and Jakub Simko. 2026. mdok-style at SemEval-2026 Task 9: Finetuning LLMs for Multilingual Polarization Detection. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Sujal Maharjan, Astha Shrestha, and Pratikshya Shrestha. 2026. Sagarmatha at SemEval-2026 Task 9: Heterogeneous Ensembling and Hierarchical Task Conditioning for Multilingual Latent Distributional Divergence Modeling. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Sha Newaz Mahmud, Sajib Bhattacharjee, Md. Refaj Hossan, Kawsar Ahmed, and Mohammed Moshiul Hoque. 2026. The Argonauts at SemEval-2026 Task 9: Multilingual Polarization Detection and Classification Using LLM Prompting and Transformer Fine-Tuning. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Soumadip Majumder, Arjun Mukherjee, Krishna Tewari, Sanjaya Kumar Lenka, and Sukomal Pal. 2026. IReL_IIT(BHU) at SemEval-2026 Task 9: Detecting Multilingual, Multicultural and Multievent Online Polarization. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Arpita Mallik and Ratnajit Dhar. 2026. CUET-823 at SemEval-2026 Task 9: LoRA-Based Instruction Fine-Tuning of LLMs vs. Transformer Models for Bengali Polarization Detection. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Raquel Martínez-España, Julio Fernández-Pedauye, José Giner-Pérez de Lucía, Jose Miguel Rojo-Martínez, Kaoutar Bakdid-Albane, and Juan José García-Escribano. 2024. Methodology for Measuring Individual Affective Polarization Using Sentiment Analysis in Social Networks. *IEEE Access*.

Binny Mathew, Punyajoy Saha, Seid Muhie Yimam, Chris Biemann, Pawan Goyal, and Animesh Mukherjee. 2021. HateXplain: A Benchmark Dataset for Explainable Hate Speech Detection. In *Proceedings of the AAAI Conference on Artificial Intelligence 2021*, volume 35, pages 14867–14875, online.

Aleksandra Matkowska, Taya Lin, and Yu-Chun Chao. 2026. Team JAT at SemEval-2026 Task 9: Enhancing Polarization Detection with Cross-Lingual Transfer and Feature Fusion. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Abdullah Mohammad. 2026. PolaFusion at SemEval-2026 Task 9: Ensemble Transformers with Targeted Augmentation for Multilingual Polarization Detection. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. 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. 2026. POLAR: A Benchmark for Multilingual, Multicultural, and Multi-Event Online Polarization. *Preprint*, arXiv:2505.20624.

Maria Nestor, Maroan Al Shrafat, Ioana Peşte, Daniela Gîfu, and Diana Trandabăţ. 2026. PolarizedTeam at SemEval-2026 Task 9: Detecting Multilingual, Multicultural and Multievent Online Polarization. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Nvidia, Bo Adler, Niket Agarwal, Ashwath Aithal, Dong H. Anh, Pallab Bhattacharya, Annika Brundyn, Jared Casper, Bryan Catanzaro, Sharon Clay, Jonathan Cohen, Sirshak Das, Ayush Dattagupta, Olivier Delalleau, Leon Derczynski, Yi Dong, Daniel Egert, Ellie Evans, Aleksander Ficek, and 63 others. 2024. Nemotron-4 340B Technical Report. *Preprint*, arXiv:2406.11704.

OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, and 262 others. 2024. GPT-4 Technical Report. *Preprint*, arXiv:2303.08774.

Veronica Orsanigo, Alan Ramponi, and Elisa Leonardelli. 2026. DigiS-FBK at SemEval-2026 Task 9: Multi-task Learning for Multilingual and Cross-cultural Polarization Classification. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Endang Wahyu Pamungkas, Valerio Basile, and Viviana Patti. 2020. Misogyny Detection in Twitter: a Multilingual and Cross-Domain Study. *Information Processing & Management*, 57(6):102360.

Zhang Peng and Lu Gehao. 2026. zhangpeng at SemEval-2026 Task 9: Detecting Multilingual, Multicultural and Multievent Online Polarization. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Phan Phat. 2026. Phatthachdau at SemEval-2026 Task 9: A Multi-Stage Augment-Judge-Train Pipeline for Multilingual Online Polarization Detection. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Huynh Nguyen Phu and Dang Van Thin. 2026. Stochastic Gradient Descenders at SemEval-2026 Task 9: Few-Shot LLM Prompting for Polarization Type Classification. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

James A. Piazza. 2023. Political polarization and political violence. *Security Studies*, 32(3):476–504.

Srikar Kashyap Pulipaka. 2026. PSK at SemEval-2026 Task 9: Multilingual Polarization Detection Using Ensemble Gemma Models with Synthetic Data Augmentation. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Sushant Kr. Ray and Rakshita Saksainaa. 2026. MoMo at SemEval-2026 Task 9: Inference-Only Prompting vs. Fine-Tuning for Multilingual Polarization Detection. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Neel Sabhahit, Sanjeevan Selvaganapathy, and Mehwish Nasim. 2026. NASIM_Lab at SemEval-2026 Task 9: A Comparative Analysis of Fine-Tuned Small Language Models vs. Generative Large Language Models on Low and High-Resource Languages for Polarization Type Detection. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Shuvodwip Saha and Pritha Saha. 2026. transformer_1376 at SemEval-2026 Task 9: A Multi-Stage Pipeline with Calibrated Ensembles and Lexical Post-Processing for Online Polarization Detection in Bengali. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Victor Manuel García Sanabria, Álvaro Rodrigo, and Roberto Centeno. 2026. UNED at SemEval-2026 Task 9: Sentiment-Aware Transformer Models with Back-Translation Augmentation for Online polarisation Detection. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Mtayyaba Shahzad, Inzmam Khadam, Zaufishan Mahmood, Junaid Rashid, Shamaila Hayat, and Fakhar Ayub. 2026. UPR at SemEval-2026 Task 9: Multi-Label Classification of Polarization Across Social Dimensions and Manifestation Identification in Urdu. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Shivam Singh, Manish Kumar, and Anupam Jamatia. 2026. NIT-Agartala-NLP-Team at SemEval-2026 Task 9: A Weighed Soft-Voting Ensemble Framework of Fine-Tuned LLMs for Binary and Multi-Class Polarization Detection. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Minh Smith and Cheryl Seals. 2026. Seals-NLP at SemEval-2026 Task 9: A Comparative Study of Transformer Architectures for Polarization Detection. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Felipe Bonow Soares and Raquel Recuero. 2021. Hashtag Wars: Political Disinformation and Discursive Struggles on Twitter Conversations During the 2018 Brazilian Presidential Campaign. *Social Media+ Society*, 7(2):1–13.

Tanisha Sriram, Sathvika Kamali Shankar, Sowmya Anand, Rajalakshmi Sivanaiah, Angel Deborah S, and Mirnalinee TT. 2026. DataBees at SemEval-2026 Task 9: Detecting Multilingual, Multicultural and Multievent Online Polarization. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Jan Strich, Adeline Scharfenberg, Chris Biemann, and Martin Semmann. 2025. EncouRAGe: Evaluating RAG Local, Fast, and Reliable. *Preprint*, arXiv:2511.04696.

Saida Islam Taien and Palash Hossen. 2026. Taien at SemEval-2026 Task 9: Multilingual Polarization Detection Using Transformer-based Models. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Hidetsune Takahashi, Eleale Nusi Tee, Aika Yu, Ruri Furukawa, Sooeun Kim, Shuta Niinomi, Dingyu Zhang,

and Emily Sofi Ohman. 2026. OZemi at SemEval-2026 Task 9: A Cross-Lingual Approach to Online Text Polarization Classification Using Multilingual Models and Adaptive Loss Formulation. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Howard Tangkulung and Yoshimasa Tsuruoka. 2026. UTokyo Tsuruoka Lab at SemEval-2026 Task 9: Efficient Single Forward Pass Inference for Multi-Label Polarization Classification. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Gemma Team, Aishwarya Kamath, Johan Ferret, Shreya Pathak, Nino Vieillard, Ramona Merhej, Sarah Perrin, Tatiana Matejovicova, Alexandre Ramé, Morgane Rivière, Louis Rouillard, Thomas Mesnard, Geoffrey Cideron, Jean bastien Grill, Sabela Ramos, Edouard Yvinec, Michelle Casbon, Etienne Pot, Ivo Penchev, and 197 others. 2025a. Gemma 3 Technical Report. *Preprint*, arXiv:2503.19786.

GLM-4.5 Team, Aohan Zeng, Xin Lv, Qinkai Zheng, Zhenyu Hou, Bin Chen, Chengxing Xie, Cunxiang Wang, Da Yin, Hao Zeng, Jiajie Zhang, Kedong Wang, Lucen Zhong, Mingdao Liu, Rui Lu, Shulin Cao, Xiaohan Zhang, Xuancheng Huang, Yao Wei, and 152 others. 2025b. GLM-4.5: Agentic, Reasoning, and Coding (ARC) Foundation Models. *Preprint*, arXiv:2508.06471.

Eliasse Tiao, Josue Romaric Edou, and Mahugnon Aime Loick Gohouede. 2026. DeepSemantics at SemEval-2026 Task 9: Label-Wise Optimization with Adaptive Focal Loss for Polarization Manifestation Identification. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. LLaMA: Open and Efficient Foundation Language Models. *Preprint*, arXiv:2302.13971.

Hoàn Trân. 2026. UIT-Polar at SemEval-2026 Task 9: Detecting Multilingual, Multicultural and Multievent Online Polarization. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Vinay Babu Ulli and Jyoti Kumari. 2026. PolAR Bears at SemEval-2026 Task 9: Parameter-Efficient Fine-Tuning and Cross-Lingual Augmentation for Multilingual Polarization Detection. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Aatman Vaidya. 2026. Aatman at SemEval-2026 Task 9: Transfer Learning for Multilingual Polarization Detection. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Ilinca Vandici, Ådne Skjæveland Jøssing, and Lukas Viestädt. 2026. Multi-Label Polarization Classification with twHIN-BERT and SCUT Threshold Optimization. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Isaac Waller and Ashton Anderson. 2021. Quantifying social organization and political polarization in online platforms. *Nature*, 600(7887):264–268.

Alishba Wazir, Muhammad Asad Khan, Junaid Rashid, Shamaila Hayat, and Samira Kanwal. 2026. UPR at SemEval-2026 Task 9: Polarization Detection in Urdu with Language-Specific Transformer and Data Augmentation. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Shih-Hung Wu, Yun-Kuang Liao, Shih-Siang Su, and Yi-Min Jian. 2026. CYUT at SemEval-2026 Task 9: Monolingual vs. Multilingual LoRA Tuning for Multicultural and Multievent Polarization Detection. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Lang.	Total	Subtask 1	Subtask 2					Subtask 3					
		Polarized (%)	Political	Racial / Ethnic	Religious Polarization	Gender / Sexual	Other	Stereo-type	Vilifi-cation	Dehuman-ization	Extreme Language	Lack of Empathy	Invalid-ation
eng	4,834	37%	36%	9%	3%	2%	4%	15%	26%	12%	24%	11%	18%
deu	4,771	48%	41%	19%	11%	6%	14%	36%	30%	15%	22%	27%	16%
urd	5,346	69%	67%	54%	55%	51%	51%	62%	65%	56%	62%	56%	57%
hin	4,117	85%	74%	12%	59%	11%	13%	50%	65%	18%	51%	57%	66%
ben	5,000	43%	34%	1%	2%	1%	10%	6%	24%	11%	5%	2%	2%
ori	3,552	29%	21%	5%	6%	3%	4%	10%	11%	1%	13%	2%	3%
pan	2,609	49%	31%	6%	8%	11%	9%	16%	40%	22%	24%	12%	24%
nep	3,008	50%	17%	14%	8%	5%	12%	27%	31%	7%	27%	11%	15%
fas	4,943	74%	44%	2%	10%	6%	24%	13%	58%	4%	17%	10%	8%
ita	5,038	43%	8%	18%	7%	9%	4%	-	-	-	-	-	-
spa	4,958	50%	27%	19%	16%	13%	13%	27%	31%	9%	24%	24%	11%
rus	5,023	30%	14%	10%	4%	6%	2%	-	-	-	-	-	-
pol	3,587	42%	37%	9%	4%	5%	6%	-	-	-	-	-	-
arb	5,070	45%	24%	17%	8%	11%	17%	33%	37%	11%	30%	17%	8%
amh	4,999	75%	67%	26%	2%	1%	25%	55%	48%	13%	31%	18%	16%
hau	5,477	11%	5%	3%	3%	1%	0%	4%	1%	4%	3%	1%	0%
zho	6,421	50%	6%	23%	2%	17%	9%	30%	19%	5%	8%	8%	5%
mya	4,334	58%	25%	5%	3%	11%	45%	-	-	-	-	-	-
khm	9,960	91%	18%	1%	3%	2%	66%	68%	2%	1%	2%	11%	7%
tel	3,550	53%	22%	17%	9%	13%	24%	11%	22%	2%	13%	26%	23%
swa	10,487	50%	3%	35%	4%	2%	8%	40%	41%	13%	24%	30%	23%
tur	3,566	50%	44%	16%	16%	6%	5%	41%	33%	11%	44%	10%	4%
Total	110,650	53%	28%	16%	10%	8%	19%	28%	26%	10%	18%	16%	14%

Table 7: Proportion of correct label = 1 for each topic across languages (ISO codes).

Team Name	Attended Tasks	Affiliation	Publication
Aaron	1	African Institute for Mathematical Sciences	(Anampiu, 2026)
Aatman	1	University of Tübingen	(Vaidya, 2026)
ABARUAH	1,2,3	Assam Don Bosco University	(Baruah, 2026)
AI4PC-Howard University	1	Howard University	(Aryal and Aryal, 2026)
AIvengers	1,2,3	University of Augsburg	(Elschenbroich and Britz, 2026)
AlphaLyrae	1,2,3	University of Information Technology, Ho Chi Minh City; Vietnam National University	(Le and Phung, 2026)
BITS Pilani	1,2,3	University of Information Technology; Vietnam National University	(Gupta et al., 2026)
CoPol	2	Delhi Skill and Entrepreneurship University	(Arora, 2026)
CUET-823	1,2	Chittagong University of Engineering and Technology	(Mallik and Dhar, 2026)
CYUT	1,2,3	Chaoyang University of Technology	(Wu et al., 2026)
DataBees	1	Sri Sivasubramaniya Nadar College of Engineering	(Sriram et al., 2026)
DeepSemantics	3	African Institute for Mathematical Sciences (AIMS)	(Tiao et al., 2026)
DigiS-FBK	1	Fondazione Bruno Kessler; University of Trento	(Orsanigo et al., 2026)
DUTH	1	Democritus University of Thrace	(Arampatzis and Arampatzis, 2026)
Gradient Descenders	2	University of Information Technology; National University	(Strich et al., 2025)
HausaNLP	2	ACETEL; HausaNLP; Gombe State University; Nasarawa State University	(Adam et al., 2026)
ILA-POLAR	2	Universität Tübingen	(Vandici et al., 2026)
ILab-NLP	1	Heriot-Watt University Heriot-Watt University Heriot-Watt University	(Booth et al., 2026)
INFOTEC-NLP	1	INFOTEC; SECIHTI	(Hernandez-Garcia et al., 2026)
IReL_IIT(BHU)	1,2,3	Indian Institute of Technology (BHU) Varanasi	(Majumder et al., 2026)
JAT	1	Universität Tübingen	(Matkowska et al., 2026)
Joshualee2	1	De Anza College	(Lee, 2026)
Lingo Research Group	1,2,3	Noida Institute of Engineering and Technology; IIT Gandhinagar	(Kadasi et al., 2026)
mdok-style	1,2,3	Kempelen Institute of Intelligent Technologies; ADAPTCentre,Trinity College Dublin	(Macko et al., 2026)
MINDS	1,2	Politecnico di Torino	(Iannielli et al., 2026)
MKJ	1	University of Turin	(Jouneghani, 2026)
MoMo	1	University of Delhi Delhi Skill and Entrepreneurship University	(Ray and Saksainaa, 2026)
MSqrd	1,2,3	Habib University	(Daniyal et al., 2026)
NAMAA	1,2	NAMAA Community; Datategy	(Djamai et al., 2026)
NASIM_Lab	2	The University of Western Australia	(Sabhahit et al., 2026)
NIT-Agartala-NLP-Team	1,2	National Institute of Technology Agartala	(Singh et al., 2026)
NLP-CIMAT	1	CIMAT; SECIHTI	(Calderon-Reyes et al., 2026)
NYCU-NLP	1,2,3	National Yang Ming Chiao Tung University	(Lin et al., 2026)
OZemi	1,2,3	Waseda University	(Takahashi et al., 2026)
pfr821	1	Télécom Paris; Airbus Defence & Space	(Durand et al., 2026)
Phatthachdau	1	VNUHCM-University of Information Technology	(Phat, 2026)
PolaFusion	1,2,3	Delhi Skill and Entrepreneurship University (DSEU)	(Mohammad, 2026)
PolAR Bears	1,2,3	Oogwai Analytics; Banaras Hindu University	(Ulli and Kumari, 2026)
PolarizedTeam	1,2	"Alexandru Ioan Cuza" University of Iasi; Romanian Academy- Iasi Branch	(Nestor et al., 2026)
PolarMind	1,2	Indian Institute of Technology	(Kumar et al., 2026)
PolDeck	1,2	University of Augsburg	(Grandy and Khir, 2026)
PSK	1	Independent Researcher	(Pulipaka, 2026)
REGLAT	1	Benha University; College of Engineering; University of Al-Kharj; SUTech	(Fetouh et al., 2026)
Sagarmatha	1,2,3	IIMSCollege; PCPSCollege	(Maharjan et al., 2026)
Seals-NLP	1	Auburn University	(Smith and Seals, 2026)
Semantic Vectors	1	IIT Dharwad	(Dash et al., 2026)
ServSocIA	1	Universidad de la República; Aplicadas y en Sistemas	(Altamirano et al., 2026)
ShefFriday	1,2,3	The University of Sheffield	(Cook et al., 2026)
SMASH	1,2,3	University of Edinburgh	(Bokaei et al., 2026)
StanceLab	1	University of Iasi	(Ivanusca et al., 2026)
Stochastic Gradient Descenders	2	University of Information Technology; Vietnam National University	(Phu and Thin, 2026)
Taien	1	BGC Trust University; University of Chittagong	(Taien and Hossen, 2026)
The Argonauts	1,2	Chittagong University of Engineering and Technology	(Mahmud et al., 2026)
The Counterfactuals	1,2,3	University of Colorado	(Johnson, 2026)
Tralaleros	1	Kiel University; University of Hamburg	(Dahl et al., 2026)
transformer_1376	1	Chittagong University of Engineering & Technology	(Saha and Saha, 2026)
UIT-Polar	1	University of Information Technology; Vietnam National University	(Trân, 2026)
UMUSP	1,2,3	University of Minho	(Fuganti et al., 2026)
UNED	1	NLP & IR group at UNED	(Sanabria et al., 2026)
UPR	1,2,3	Sejong University	(Khadam et al., 2026; Wazir et al., 2026; Shahzad et al., 2026)
UTokyo Tsuruoka Lab	1,2	The University of Tokyo	(Tangkulung and Tsuruoka, 2026)
VGU-M.Tech-AI	2	Vivekananda Global University Jaipur	(Bichi and Shekhawat, 2026)
wangkongqiang	1,2,3	Yunnan University	(Kongqiang and Qingli, 2026)
YEZE	1,2,3	University of Tübingen	(Guo and Chang, 2026)
YNU-HPCC	2	Yunnan University	(Bao et al., 2026)
zhangpeng	1,2,3	Yunnan University; Yunnan Province Smart Tourism Engineering Research Center	(Peng and Gehao, 2026)

Table 8: Overview of participating teams, their attended subtasks, affiliations, and publications.