SemEval-2026 Task 3: Dimensional Aspect-Based Sentiment Analysis (DimABSA)

Liang-Chih Yu[1,*], **Jonas Becker**[2,*], **Shamsuddeen Hassan Muhammad**[3], **Idris Abdulmumin**[4], **Lung-Hao Lee**[5,*], **Ying-Lung Lin**[6], **Jin Wang**[7], **Jan Philip Wahle**[2], **Terry Ruas**[2], **Natalia Loukachevitch**[8], **Alexander Panchenko**[9,10], **Ilseyar Alimova**[9], **Lilian Wanzare**[11], **Nelson Odhiambo**[11], **Bela Gipp**[2], **Kai-Wei Chang**[12], **and Saif M. Mohammad**[13]

[1]Yuan Ze University, [2]University of Göttingen, [3]Imperial College London, [4]University of Pretoria, [5]National Yang Ming Chiao Tung University, [6]Central Police University, [7]Yunnan University, [8]Lomonosov Moscow State University, [9]Skoltech, [10]AIRI, [11]Maseno University, [12]UCLA, [13]National Research Council Canada

***Equal contribution**

Contact: lcyu@saturn.yzu.edu.tw, jonas.becker@uni-goettingen.de

Abstract

We present the SemEval-2026 shared task on Dimensional Aspect-Based Sentiment Analysis (DimABSA), which improves traditional ABSA by modeling sentiment along valence–arousal (VA) dimensions rather than using categorical polarity labels. To extend ABSA beyond consumer reviews to public-issue discourse (e.g., political, energy, and climate issues), we introduce an additional task, Dimensional Stance Analysis (DimStance), which treats stance targets as aspects and reformulates stance detection as regression in the VA space. The task consists of two tracks: Track A (DimABSA) and Track B (DimStance). Track A includes three subtasks: (1) dimensional aspect sentiment regression, (2) dimensional aspect sentiment triplet extraction, and (3) dimensional aspect sentiment quadruplet extraction, while Track B includes only the regression subtask for stance targets. We also introduce a continuous F1 (cF1) metric to jointly evaluate structured extraction and VA regression.

The task attracted more than 400 participants, resulting in 112 final submissions and 42 system description papers. We report baseline results, discuss top-performing systems, and analyze key design choices to provide insights into dimensional sentiment analysis at the aspect and stance-target levels. All resources are available on our GitHub repository.[1]

1 Introduction

Aspect-Based Sentiment Analysis (ABSA) is a widely used technique for analyzing opinions and sentiments at the aspect level. It is formulated as the extraction of sentiment elements, including aspect terms, aspect categories, opinion terms, and sentiment polarity, individually or jointly. For example, given the sentence *The food was excellent.*, an ABSA system is expected to extract the aspect term *food*, the opinion term *excellent*, assign the aspect category `FOOD#QUALITY` from a predefined set, and predict `Positive` sentiment polarity. Following the success of prior SemEval tasks (Pontiki et al., 2014, 2015, 2016), ABSA has attracted substantial attention, providing deeper insights into user opinions across various applications (D'Aniello et al., 2022; Zhang et al., 2023; Hua et al., 2024).

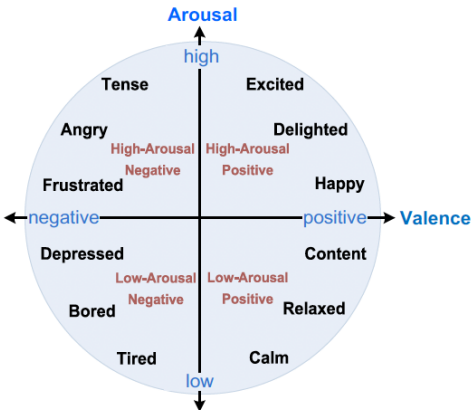

Figure 1: Valence–Arousal (VA) space.

However, current ABSA research adopts a coarse-grained, categorical sentiment representation (e.g., positive, negative, and neutral). This approach contrasts with long-established theories in psychology and affective science (Russell, 1980, 2003), where sentiment is represented along fine-grained, real-valued dimensions of **valence** (from negative to positive) and **arousal** (from sluggish to

[1] https://github.com/DimABSA/DimABSA2026.

excited), as illustrated in Figure 1. This valence–arousal (VA) representation has motivated research on dimensional sentiment analysis (Yu et al., 2016; Buechel and Hahn, 2017a; Mohammad et al., 2018; Lee et al., 2022, 2024; Mohammad, 2025), enabling more nuanced distinctions in emotional expression and supporting broader applications.

To move beyond categorical sentiment labels, we introduce a SemEval shared task that integrates the dimensional VA representation into the traditional ABSA framework. We refer to this task as Dimensional ABSA (DimABSA). To this end, we construct multilingual, multi-domain datasets by annotating traditional ABSA elements (aspect terms, aspect categories, and opinion terms) together with continuous VA scores (Lee et al., 2026b).

Furthermore, stance detection and ABSA are conceptually related, as stance targets can be treated as aspects. Building on this connection, we introduce an additional task, Dimensional Stance Analysis (DimStance), which requires systems to predict VA scores for given targets. For this task, we annotate stance targets with VA scores to construct multilingual, multi-domain datasets (Becker et al., 2026). The DimStance formulation not only extends ABSA beyond consumer reviews to public-issue discourse (e.g., political, energy, and climate issues) but also generalizes stance analysis from categorical labels to the VA representation.

We organize the SemEval task into two tracks: Track A (DimABSA) and Track B (DimStance). We further design three subtasks that combine VA scores with different ABSA elements: (1) Dimensional Aspect Sentiment Regression (DimASR), predicting VA scores for each aspect in a sentence; (2) Dimensional Aspect Sentiment Triplet Extraction (DimASTE), jointly extracting aspect and opinion terms and predicting their associated VA scores; and (3) Dimensional Aspect Sentiment Quadruplet Extraction (DimASQP), extending DimASTE by additionally predicting aspect categories. Track A includes all three subtasks, while Track B includes only Subtask 1 (DimASR).

Our task attracted over 400 participants, resulting in 112 final submissions from 44 teams and 42 system description papers. Track A (DimABSA) was the most popular, with over 300 participants and 84 final submissions, while Track B (DimStance) attracted over 100 participants with 28 final submissions. Notably, most teams participated in multilingual and multidomain settings, covering an average of ~4.5 languages and ~3.4 domains.

Analysis of participating systems reveals that most approaches leverage pretrained transformers or large language models (LLMs). These models are typically trained with supervised fine-tuning and enhanced with various training and prompting strategies. Evaluation results show that dimensional sentiment analysis at the aspect and stance-target levels remains challenging.

2 Related Work

Categorical ABSA. Most existing ABSA datasets are English-centric and primarily focus on customer review applications (Chebolu et al., 2023). SemEval-2014 (Pontiki et al., 2014) introduced the first ABSA shared task for English restaurant and laptop reviews, followed by extensions to additional subtasks and languages (Pontiki et al., 2015, 2016). Subsequent datasets further enriched the annotation schema, introducing triplets of aspect, opinion, and polarity (Xu et al., 2020; Peng et al., 2020), and quadruples by adding an aspect category (Zhang et al., 2021; Cai et al., 2021). The domain coverage has also been broadened to areas such as finance (Kubo and Nakayama, 2018), COVID-19 (Aygün et al., 2022; Hou et al., 2025), and education (Hua et al., 2025). Moreover, M-ABSA (Wu et al., 2025) extended this line of work to the multilingual setting by constructing a parallel benchmark through automatic translation.

Categorical Stance. Prior work on stance detection has expanded primarily along three axes: language coverage, scale, and domain specificity. Early benchmarks focused on English Twitter data, such as the SemEval stance dataset (Mohammad et al., 2016)). Multilingual extensions followed, including X-Stance (Vamvas and Sennrich, 2020) for German, French, and Italian, and the Catalonia Independence Corpus (CIC) for Catalan and Spanish (Zotova et al., 2020). Large-scale English resources, such as P-Stance (Li et al., 2021) and COVID-19-Stance (Glandt et al., 2021), further increased dataset size and target diversity. Recent work has extended stance detection to zero-shot (Allaway and McKeown, 2020; Zhao et al., 2023; Zhao and Caragea, 2024), multimodal (Zhou et al., 2025; Zhang et al., 2025), and conversational (Ding et al., 2025; Marreddy et al., 2025) settings.

Dimensional Sentiment Analysis. Previous studies have developed resources with single- or

Task	Input	Output	Prediction Type	Metric
DimASR	text + aspects	V#A	Regression	RMSE
	1. The food was excellent. 2. It's great that after coal plants shut down there are transmission lines to connect solar and wind resources too.	8.00#8.12 7.75#7.62		
DimASTE	text	(A, O, V#A)	Extraction	cF1
	Service at the bar was a little slow	(Service, a little slow, 4.10#4.30)	Regression	
DimASQP	text	(A, C, O, V#A)	Extraction	cF1
	Their sodas are usually expired and flat	(sodas, DRINKS#QUALITY, usually expired, 1.90#7.20) (sodas, DRINKS#QUALITY, flat, 2.40#6.80)	Classification Regression	

Table 1: Overview of DimABSA subtasks with input–output structure, prediction types, and evaluation metrics, along with illustrative examples.

combined-dimensional representations across lexical, phrasal, and sentential granularities. Sentiment lexicons assign affective scores to individual words, such as SentiWordNet (Baccianella et al., 2010), SO-CAL (Taboada et al., 2011), SentiStrength (Thelwall et al., 2012), and NRC-VAD (Mohammad, 2018, 2025). Phrase-level datasets formulate sentiment composition through modifiers, including SemEval-2015 Task 10 (Rosenthal et al., 2015) and SemEval-2016 Task 7 (Kiritchenko et al., 2016). At the sentence level, affective scores are provided for texts of varying lengths (Preoţiuc-Pietro et al., 2016; Buechel and Hahn, 2017b; Mohammad and Bravo-Marquez, 2017; Mohammad et al., 2018; Muhammad et al., 2025). The Stanford Sentiment Treebank (Socher et al., 2013) and Chinese EmoBank (Lee et al., 2022) provide cross-granularity resources, bridging phrase- and sentence-level representations and covering all three granularities.

3 Task Description

3.1 Track A: Dimensional Aspect-Based Sentiment Analysis (DimABSA)

This track involves three traditional ABSA elements and VA scores, described as follows.

Aspect Term (A): a word or phrase indicating an opinion target, such as *service*, *screen*, *profit*.

Aspect Category (C): a predefined Entity#Attribute label associated with an aspect term (e.g., `FOOD#QUALITY`, `SERVICE#GENERAL`) (Pontiki et al., 2015, 2016). The full list of aspect categories is presented in Appendix A.

Opinion Term (O): a sentiment-bearing word or phrase associated with a specific aspect term. The opinion term includes sentiment modifiers to support fine-grained sentiment representation (e.g., *very good*, *extremely bad*, *a little slow*).

Valence-Arousal (VA): a pair of real-valued scores, each ranging from 1 to 9, where 1 denotes extreme negative valence or low arousal, 9 denotes extreme positive valence or high arousal, and 5 denotes neutral valence or medium arousal.

Based on these elements, we define three subtasks that adapt traditional ABSA formulations to the dimensional sentiment paradigm. We present the input/output formats and an example in Table 1.

- **Subtask 1 - Dimensional Aspect Sentiment Regression (DimASR):** Given a sentence and one or more aspects, predict VA scores for each aspect. This task generalizes traditional Aspect Sentiment Classification (ASC) (Pontiki et al., 2014, 2015, 2016) to VA regression.
- **Subtask 2 - Dimensional Aspect Sentiment Triplet Extraction (DimASTE):** Given a sentence, extract all `(A, O, VA)` triplets. This task jointly extracts aspect and opinion terms and predicts their associated VA scores, extending traditional ASTE (Peng et al., 2020) by incorporating VA regression.
- **Subtask 3 - Dimensional Aspect Sentiment Quadruplet Prediction (DimASQP):** Given a sentence, extract all `(A, C, O, VA)` quadruplets. Compared to DimASTE, this task additionally incorporates aspect category classification, extending traditional ASQP (Cai et al., 2021; Zhang et al., 2021) to include VA regression.

3.2 Track B: Dimensional Stance Analysis (DimStance)

Given an utterance or post and a target entity, stance detection is formulated as determining whether the speaker is in favor of the target, against the target, or neither inference is likely (Mohammad et al., 2017). This track reformulates stance detection by treating stance targets as aspects and generalizes categorical stance classification to VA regression.

Track	Dataset	Source(s)	Subtask	Number of texts / instances			
				Train	Dev	Test	Total
Track A	eng-rest	ACOS; Yelp Open Dataset	ST1	2284 / 3659	200 / 340	1000 / 1504	3484 / 5503
			ST2–3		200 / 408	1000 / 2129	3484 / 6196
	eng-lap	ACOS; Amazon Reviews 2023	ST1	4076 / 5773	200 / 275	1000 / 1421	5276 / 7469
			ST2–3		200 / 317	1000 / 1975	5276 / 8065
	jpn-hot	Rakuten Travel	ST1	1600 / 2846	200 / 284	800 / 1092	2600 / 4222
			ST2–3		200 / 364	800 / 1443	2600 / 4653
	jpn-fin	chABSA; EDINET	ST1	1024 / 1672	200 / 319	800 / 1302	2024 / 3293
	rus-rest	SemEval'16 Task 5 (Restaurant)	ST1	1240 / 2487	56 / 81	1072 / 1637	2368 / 4205
			ST2–3		48 / 102	630 / 1310	1918 / 3899
	tat-rest	SemEval'16 (MT)	ST1	1240 / 2487	56 / 81	1072 / 1637	2368 / 4205
			ST2–3		48 / 102	630 / 1310	1918 / 3899
	ukr-rest	SemEval'16 (MT)	ST1	1240 / 2487	56 / 81	1072 / 1637	2368 / 4205
			ST2–3		48 / 102	630 / 1310	1918 / 3899
	zho-rest	SIGHAN-2024; Google Reviews; PTT	ST1	6050 / 8523	225 / 416	1000 / 1929	7275 / 10868
			ST2–3		300 / 761	1000 / 2861	7350 / 12145
	zho-lap	Mobile01	ST1	3490 / 6502	261 / 431	1000 / 2662	4751 / 9595
			ST2–3		300 / 551	1000 / 2798	4790 / 9851
	zho-fin	MOPS	ST1	1000 / 2633	200 / 563	842 / 2354	2042 / 5550
Track B	eng-env	EZ-STANCE; Reddit	ST1	922 / 2059	200 / 339	1020 / 1813	2142 / 4211
	deu-pol	Wahl-O-Mat Archive	ST1	683 / 1335	34 / 75	263 / 438	980 / 1848
	zho-env	Threads Platform	ST1	683 / 1091	34 / 49	263 / 898	980 / 2038
	pcm-pol	X Platform	ST1	1049 / 1118	119 / 122	331 / 343	1499 / 1583
	swa-pol	X Platform	ST1	1375 / 1622	123 / 145	266 / 299	1764 / 2066

Table 2: **Dataset statistics for Track A (DimABSA) and Track B (DimStance).** For each dataset (language–domain), we report the source(s), subtask type (ST1 vs. ST2–3), and the number of texts and instances in the train/dev/test splits, using the format `#texts/#instances`. There can be multiple instances per text.

Unlike Track A, which is based on aspect-based sentiment data, this track focuses on stance data. Specifically, we adopt the formulation of Subtask 1 (DimASR) in Track A, where the input is a text and one or more targets, and the task is to predict VA scores for each target.

4 Datasets

We construct multilingual, multi-domain datasets for both Track A (DimABSA) and Track B (DimStance), as shown in Table 2. Detailed descriptions of the data sources, annotation process, and annotation agreement are provided in our dataset papers (Lee et al., 2026b; Becker et al., 2026). Key information is summarized below.

Track A covers six languages: English (`eng`), Japanese (`jpn`), Russian (`rus`), Tatar (`tar`), Ukrainian (`ukr`), and Chinese (`zho`). These datasets span four domains: restaurant (`rest`), laptop (`lap`), hotel (`hot`), and finance (`fin`). The finance datasets are used exclusively for Subtask 1, while the other domains support all three subtasks. In total, Track A provides 76,958 aspect instances (aspect pairs, triplets, and quadruplets) across 42,590 sentences.

Track B comprises five language-specific datasets: English (`eng`), German (`deu`), Chinese (`zho`), Nigerian Pidgin (`pcm`), and Swahili (`swa`). These datasets cover two domains: environmental protection (`env`) and politics (`pol`). In total, Track B contains 11,746 stance targets across 7,365 texts.

4.1 Data Collection

4.1.1 Track A: DimABSA

We collect data from multiple sources, including existing labeled ABSA datasets and newly curated unlabeled data. The existing labeled datasets are used solely for training, while the newly curated data are annotated and split into training, development, and test sets. The data sources for each language are described below.

English. We use the training split of the ACOS dataset (Cai et al., 2021), manually annotating the restaurant and laptop quadruplets with VA scores to replace the sentiment polarity labels. For the development and test sets, we collect restaurant reviews from Yelp Open Dataset[2] and laptop reviews from Amazon Reviews 2023 (Hou et al., 2024).

Japanese. For the finance domain, the training set

[2] https://business.yelp.com/data/resources/open-dataset

is sampled from the chABSA dataset.[3] We manually annotate VA scores for each aspect in these samples, replacing the original sentiment polarity labels. The development and test sets are collected from the same EDINET[4] sources as chABSA, with samples involving the same companies removed to avoid overlap. For the hotel domain, we crawl reviews from Rakuten Travel.[5]

Russian. The SemEval-2016 restaurant review dataset (Pontiki et al., 2016) serves as the data source. The labeled portion contains annotated aspects, their categories, and sentiment polarity. For the remaining instances, opinion terms and VA values are annotated. The unlabeled portion of reviews is used for the development and test sets.

Tatar. We automatically translate the Russian dataset into Tatar using Yandex Translate. All translations are reviewed by a native speaker, with 45.5% of instances manually corrected.

Ukrainian. Similar to Tatar, we translated the Russian dataset into Ukrainian. All translations are reviewed by native speakers, with 35.6% of instances manually corrected.

Chinese. For the restaurant domain, we use the SIGHAN-2024 dataset (Lee et al., 2024) for training, and construct the development and test sets from Google Reviews[6] and the PTT platform[7]. For the laptop domain, we crawl reviews from Mobile01[8]. For the finance domain, we collect annual reports of Taiwanese companies from MOPS[9].

4.1.2 Track B: DimStance

English. Data for the training split is collected from the environmental protection domain of EZ-STANCE (Zhao and Caragea, 2024). The dev and test splits are obtained from Reddit texts[10], using the same keywords as in EZ-STANCE.

German. Sampled from Wahl-O-Mat Archive, provided by the Federal Agency for Civic Education of Germany[11]. The data contains responses by political parties to political statements.

Chinese. Collected messages from the Threads platform[12]. Crawling is performed using a predefined set of Chinese query keywords about environmental protection.

Nigerian Pidgin. Sampled posts and comments from the X platform[13]. The discussions concern Nigerian elections, i.e., politics, ranging from January 1st to March 8th, 2023.

Swahili. Combined data from Afrisenti (Muhammad et al., 2023), HateSpeech_Kenya (Ombui, 2022), and Politikweli (Amol et al., 2024), covering political tweets from the X platform.

4.2 Annotation Process

The annotated elements vary across datasets depending on the subtask configuration. We annotate `(A, VA)` pairs for datasets exclusive to Subtask 1 (DimASR), specifically the finance datasets in Track A and all datasets in Track B. For datasets that support all subtasks, we annotate full `(A, C, O, VA)` quadruplets. This design facilitates a shared training set across subtasks, as indicated by the dataset splits in Table 2. However, we do not use a shared development/test set for all subtasks, as Subtask 1 assumes aspect terms are provided as input. Instead, we create a dedicated development/test set for Subtask 1, and a shared set for Subtask 2 (DimASTE) and Subtask 3 (DimASQP).

The annotation process is conducted in two phases: we first extract categorical tuples from sentences, identifying the element `A` for each `(A, VA)` pair and the triplet `(A, C, O)` for each quadruplet, followed by the assignment of VA scores. For Track A, two annotators independently extract tuples from each sentence, with a third adjudicator resolving disagreements. For Track B, we use LLMs to extract candidate stance targets, which are then validated by five annotators through majority voting. For VA rating, both tracks rely on five annotators, and the final VA score is computed by averaging the ratings.

4.3 Annotation Quality

We evaluate the agreement at the tuple level using the F1 score, following prior work (Chebolu et al., 2024; Wu et al., 2025). The F1 score is computed between two annotators by treating one as the prediction and the other as the gold standard. To assess VA agreement, we use Root Mean Square Error (RMSE) separately for valence and arousal. RMSE is calculated by comparing each annotator's rating against the mean of all five annotators. The

[3] https://github.com/chakki-works/chABSA-dataset
[4] https://disclosure2.edinet-fsa.go.jp
[5] https://travel.rakuten.co.jp
[6] https://customerreviews.google.com
[7] https://www.pttweb.cc
[8] https://www.mobile01.com/category.php?id=2
[9] https://emops.twse.com.tw/server-java/t58query
[10] `https://reddit.com/`. Version: 2025-07-01
[11] `https://www.bpb.de/themen/wahl-o-mat/556865/datensaetze-des-wahl-o-mat/`. Version: 2026-03-25.
[12] `https://www.threads.com/`. Version: 2025-10-15
[13] `https://x.com/`. Version: 2023-12-31

final agreement score is the average RMSE across all five annotators.

5 Evaluation

5.1 Metrics

Subtask 1: RMSE. DimASR is formulated as a regression task, and its performance is evaluated by measuring prediction error in the VA space using RMSE, defined as

$$\mathrm{RMSE}_{\mathrm{VA}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(V_p^{(i)} - V_g^{(i)}\right)^2 + \left(A_p^{(i)} - A_g^{(i)}\right)^2} \quad (1)$$

where N is the total number of instances; $V_p^{(i)}$ and $A_p^{(i)}$ denote the predicted valence and arousal values for an instance; and $V_g^{(i)}$ and $A_g^{(i)}$ denote the corresponding gold values.

Subtask 2 & 3: Continuous F1 (cF1). DimASTE and DimASQP are hybrid tasks that require both categorical prediction and VA regression. Therefore, the standard F1 score, widely used in ABSA, is insufficient to jointly assess these components. To address this limitation, we propose the *continuous F1 (cF1)* metric, which incorporates VA prediction error into the F1 formulation.

Following the standard F1, a predicted tuple is counted as a true positive (TP) only if all its categorical elements exactly match the gold annotation. This categorical TP is then extended as a *continuous true positive (cTP)* by incorporating the VA prediction error. Formally, let P denote the set of predicted triplets `(A, O, VA)` or quadruplets `(A, C, O, VA)`. For any prediction $t \in P$, its cTP is defined as

$$\mathrm{cTP}^{(t)} = \begin{cases} 1 - \mathrm{dist}(\mathrm{VA}_p^{(t)}, \mathrm{VA}_g^{(t)}), & t \in P_{\mathrm{cat}} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where $P_{cat} \subseteq P$ denotes the set of predictions in which all categorical elements, `(A, O)` for a triplet or `(A, C, O)` for a quadruplet, exactly match the gold annotation for the same sentence. The distance function is defined as

$$\mathrm{dist}(\mathrm{VA}_p, \mathrm{VA}_g) = \frac{\sqrt{(V_p - V_g)^2 + (A_p - A_g)^2}}{\mathrm{D}_{\max}} \quad (3)$$

where $\mathrm{dist}(\cdot)$ denotes the normalized Euclidean distance between the predicted $\mathrm{VA}_p = (V_p, A_p)$ and gold $\mathrm{VA}_g = (V_g, A_g)$ in the VA space, and $\mathrm{D}_{\max} = \sqrt{8^2 + 8^2} = \sqrt{128}$ is the maximum possible Euclidean distance in the VA space on the [1, 9] scale, ensuring that $\mathrm{dist} \in [0, 1]$.

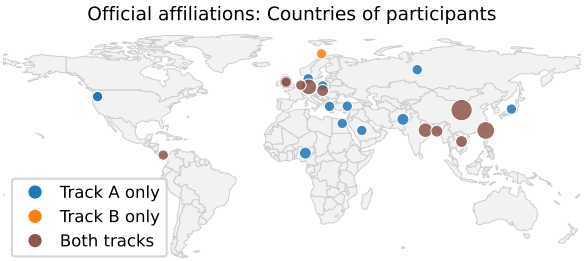

Figure 2: **Countries of official affiliations of participants.** Larger dots indicate more participants. A total of 24 countries are represented.

Building on per-prediction $\mathrm{cTP}^{(t)}$, *cRecall* and *cPrecision* are defined as the total cTP divided by the number of gold and predicted triplets/quadruplets, respectively. The cF1 is computed as their harmonic mean. An illustrative example is given in Appendix B. An official evaluation script is available on the task GitHub repository.

5.2 Baselines

We provide two baseline systems for each track.

Track A. We employ the closed-source LLM Kimi K2 Thinking (MoonshotAI, 2025) with one-shot prompting and Qwen3-14B (Alibaba, 2025) across all subtasks. Qwen3-14B is separately fine-tuned for each dataset using QLoRA (Dettmers et al., 2023) on the official training split.

Track B. We adopt the multilingual pretrained Transformer mBERT (Devlin et al., 2019) and Mistral-3-14B (MistralAI, 2025). mBERT is fully fine-tuned, while Mistral-3-14B is fine-tuned using QLoRA. Both models are trained separately on the official training split for each dataset.

Implementation details and additional baseline results are provided in our dataset papers (Lee et al., 2026b; Becker et al., 2026).

5.3 Task Organization

We used Codabench[14] as the competition platform and released pilot data before the shared task to help participants understand the task. We also provided a starter kit on GitHub, beginner resources, and organized a Q&A session and a writing tutorial for junior researchers. Participants came from different parts of the world, as shown in Figure 2.

The task consisted of two phases: (1) a development phase and (2) an evaluation phase. During

[14]Track A: https://www.codabench.org/competitions/10918/; Track B: https://www.codabench.org/competitions/11139/

the development phase, the leaderboard was open, allowing up to 999 submissions per participant. During the evaluation phase, the leaderboard was closed, and each participant could submit up to four runs, with the last used for the official ranking.

6 Participating Systems and Results

6.1 Overview

The task attracted over 300 participants in Track A (DimABSA) and over 100 in Track B (DimStance). During the development phase, 2664 submissions were made to Track A and 357 to Track B. In the evaluation phase, 177 submissions were made to Track A and 67 to Track B. While the English datasets received the most submissions, all languages had at least 20 submissions in each track.

We report results only for teams that submitted a system description paper. In total, 39 teams with 84 submissions participated in Track A, and 13 teams with 28 submissions participated in Track B, resulting in 112 submissions from 42 unique teams, including 10 teams that participated in both tracks. Participant information is listed in Table 5.

6.2 Track A: DimABSA

Track A includes three subtasks. Subtask 1 (DimASR) attracted the most teams (33), followed by Subtask 2 (DimASTE) with 20 teams and Subtask 3 (DimASQP) with 19 teams. Table 3 presents the top three systems for each dataset across all subtasks, together with the baselines. The complete results for each subtask are reported in Table 6, Table 7, and Table 8.

The DimASR results for sentiment regression show that systems achieve lower RMSE on the Chinese and Japanese datasets, whereas the highest RMSE is observed on the low-resource Tatar dataset. DimASTE and DimASQP report results for joint structured extraction and regression. In DimASTE, systems achieve the highest cF1 on the English datasets and the lowest on the Tatar dataset. DimASQP is more difficult than DimASTE due to the additional classification of domain-dependent aspect categories. The laptop and hotel domains show a noticeable performance drop, likely due to the larger number of aspect categories and their long-tailed distribution (Lee et al., 2026b).

6.2.1 Best-Performing Systems

Team PAI (1st on $\texttt{rus-rest}_{\text{ST1–3}}$, $\texttt{tat-rest}_{\text{ST1}}$, $\texttt{ukr-rest}_{\text{ST1–3}}$, $\texttt{zho-rest}_{\text{ST2}}$). They propose a distributional adaptation method to align predicted VA scores with the training set distribution while preserving the inter-dimensional correlation between valence and arousal. Initial predictions are generated by Qwen3-32B (Yang et al., 2025) fine-tuned with LoRA (Hu et al., 2021) and subsequently calibrated using the Sinkhorn algorithm.

Team TeleAI. (1st on $\texttt{jpn-hot}_{\text{ST1–2}}$, $\texttt{jpn-fin}_{\text{ST1}}$, $\texttt{zho-lap}_{\text{ST1}}$). Their system is based on Qwen2.5-7B (Qwen Team, 2025) fine-tuned with LoRA. To improve generalization, they train a single multilingual, multi-domain model on all task training sets. They apply robust training, including Smooth L1 loss with R-Drop consistency, embedding-level PGD adversarial training, and post-hoc linear calibration.

Team Takoyaki. (1st on $\texttt{eng-rest}_{\text{ST2–3}}$, $\texttt{eng-lap}_{\text{ST2–3}}$, $\texttt{tat-rest}_{\text{ST3}}$). They adopt retrieval-based in-context learning, where multiple BM25 variants retrieve similar training examples for the Gemini 3.0 Pro (Gemini Team, 2023) to generate quadruplet predictions. An agreement-based ensemble strategy is then applied to retain quadruplets with high agreement scores across variants. Finally, LLM-mined correction rules are applied to fix extraction and category errors. The VA scores are averaged across duplicate quadruplets after ensembling and correction.

6.3 Track B: DimStance

Track B had 13 participating teams. Table 4 presents the top three systems together with our baselines. The complete results are reported in **??**. The DimASR results for stance targets show that systems achieve the lowest RMSE on the Chinese dataset, whereas the highest RMSE is observed on the low-resource Swahili dataset.

6.3.1 Best-Performing Systems

Team LogSigma (1st on Track A: $\texttt{eng-rest}_{\text{ST1}}$, $\texttt{eng-lap}_{\text{ST1}}$; Track B: $\texttt{eng-env}_{\text{ST1}}$, $\texttt{deu-pol}_{\text{ST1}}$, $\texttt{swa-pol}_{\text{ST1}}$). They treat VA prediction as two regression tasks for valence and arousal and focus on balancing them. Instead of fixing loss weights, the model learns task-specific log-variance parameters that down-weight noisier objectives during training, allowing it to balance valence and arousal losses based on their prediction difficulty. The model uses a language-specific transformer encoder that produces a shared representation, which is then passed

Dataset	Team	Score	Dataset	Team	Score	Dataset	Team	Score	Dataset	Team	Score
Subtask 1											
eng-rest	LogSigma	1.1035	eng-lap	LogSigma	1.2408	jpn-hot	TeleAI	0.5561	jpn-fin	TeleAI	0.6581
	Bert Kittens	1.1812		TeleAI	1.2425		PALI	0.6237		PALI	0.7532
	PICT	1.1958		Bert Kittens	1.2769		ICT-NLP	0.6289		PAI	0.7584
	Baseline (Kimi-K2 Thinking)	2.1461		*Baseline (Kimi-K2 Thinking)*	2.1893		*Baseline (Kimi-K2 Thinking)*	1.7553		*Baseline (Kimi-K2 Thinking)*	1.6396
	Baseline (Qwen-3 14B)	2.6427		*Baseline (Qwen-3 14B)*	2.8089		*Baseline (Qwen-3 14B)*	2.2906		*Baseline (Qwen-3 14B)*	1.8964
rus-rest	PAI	1.2190	tat-rest	PAI	1.5294	ukr-rest	PAI	1.1888	zho-rest	ICT-NLP	0.9256
	TeleAI	1.2456		Habib University	1.6041		TeleAI	1.3234		TeleAI	0.9265
	HUS@NLP-VNU	1.3075		PALI	1.7121		HUS@NLP-VNU	1.3538		YangS_team	0.9433
	Baseline (Kimi-K2 Thinking)	1.7768		*Baseline (Kimi-K2 Thinking)*	1.9380		*Baseline (Kimi-K2 Thinking)*	1.7805		*Baseline (Kimi-K2 Thinking)*	1.8959
	Baseline (Qwen-3 14B)	2.1528		*Baseline (Qwen-3 14B)*	2.6367		*Baseline (Qwen-3 14B)*	2.2121		*Baseline (Qwen-3 14B)*	2.0073
zho-lap	TeleAI	0.6103	zho-fin	HUS@NLP-VNU	0.4841						
	ICT-NLP	0.6553		YangS_team	0.4864						
	HUS@NLP-VNU	0.6663		TeleAI	0.4866						
	Baseline (Kimi-K2 Thinking)	1.6440		*Baseline (Kimi-K2 Thinking)*	1.9652						
	Baseline (Qwen-3 14B)	1.7706		*Baseline (Qwen-3 14B)*	1.4707						
Subtask 2											
eng-rest	Takoyaki	0.7021	eng-lap	Takoyaki	0.6366	jpn-hot	TeleAI	0.5837	rus-rest	PAI	0.5793
	nchellwig	0.6985		PALI	0.6242		TeamLasse	0.5694		TeleAI	0.5736
	PALI	0.6928		PAI	0.6169		PAI	0.5682		PALI	0.5724
	Baseline (Kimi-K2 Thinking)	0.4920		*Baseline (Kimi-K2 Thinking)*	0.4424		*Baseline (Kimi-K2 Thinking)*	0.3464		*Baseline (Kimi-K2 Thinking)*	0.4242
	Baseline (Qwen-3 14B)	0.4483		*Baseline (Qwen-3 14B)*	0.3827		*Baseline (Qwen-3 14B)*	0.1622		*Baseline (Qwen-3 14B)*	0.3341
tat-rest	nchellwig	0.5119	ukr-rest	PAI	0.5787	zho-rest	PAI	0.5638	zho-lap	PALI	0.5308
	Takoyaki	0.5092		TeleAI	0.5712		PALI	0.5634		PAI	0.5306
	PAI	0.4908		PALI	0.5671		nchellwig	0.5488		TeleAI	0.5292
	Baseline (Kimi-K2 Thinking)	0.3577		*Baseline (Kimi-K2 Thinking)*	0.4220		*Baseline (Kimi-K2 Thinking)*	0.3529		*Baseline (Kimi-K2 Thinking)*	0.2494
	Baseline (Qwen-3 14B)	0.2020		*Baseline (Qwen-3 14B)*	0.3099		*Baseline (Qwen-3 14B)*	0.2509		*Baseline (Qwen-3 14B)*	0.2099
Subtask 3											
eng-rest	Takoyaki	0.6514	eng-lap	Takoyaki	0.4227	jpn-hot	PALI	0.4252	rus-rest	PAI	0.5599
	nchellwig	0.6403		nchellwig	0.4006		Takoyaki	0.4086		PALI	0.5496
	PALI	0.6395		PALI	0.3793		TeamLasse	0.3992		Takoyaki	0.5130
	Baseline (Kimi-K2 Thinking)	0.3746		*Baseline (Kimi-K2 Thinking)*	0.2795		*Baseline (Kimi-K2 Thinking)*	0.1943		*Baseline (Kimi-K2 Thinking)*	0.2963
	Baseline (Qwen-3 14B)	0.2673		*Baseline (Qwen-3 14B)*	0.1529		*Baseline (Qwen-3 14B)*	0.0400		*Baseline (Qwen-3 14B)*	0.1682
tat-rest	Takoyaki	0.4736	ukr-rest	PAI	0.5437	zho-rest	NYCU Speech Lab	0.5521	zho-lap	NYCU Speech Lab	0.4824
	nchellwig	0.4557		PALI	0.5307		PAI	0.5360		PALI	0.4319
	PAI	0.4523		ALPS-Lab	0.5163		PALI	0.5357		PAI	0.4316
	Baseline (Kimi-K2 Thinking)	0.2380		*Baseline (Kimi-K2 Thinking)*	0.2971		*Baseline (Kimi-K2 Thinking)*	0.2859		*Baseline (Kimi-K2 Thinking)*	0.1900
	Baseline (Qwen-3 14B)	0.0954		*Baseline (Qwen-3 14B)*	0.1641		*Baseline (Qwen-3 14B)*	0.1605		*Baseline (Qwen-3 14B)*	0.1124

Table 3: **Track A (DimABSA) results across all subtasks.** Subtask 1 is evaluated using RMSE, while Subtasks 2 and 3 are evaluated using cF1. The top three teams and the official baseline systems are reported for each dataset.

Dataset	Team	Score	Dataset	Team	Score
Subtask 1					
eng-env	LogSigma	1.4734	deu-pol	LogSigma	1.3417
	hllwan	1.5122		NTNU-SMIL	1.3467
	NTNU-SMIL	1.5207		HUS@NLP-VNU	1.4108
	Baseline (Mistral-3 14B)	1.6430		*Baseline (Mistral-3 14B)*	1.5910
	Baseline (mBERT)	2.6985		*Baseline (mBERT)*	2.3294
zho-env	YangS_team	0.5468	pcm-pol	CYUT	1.1024
	NTNU-SMIL	0.5561		LogSigma	1.1269
	HUS@NLP-VNU	0.5826		PAI	1.1399
	Baseline (Mistral-3 14B)	0.7400		*Baseline (Mistral-3 14B)*	1.7390
	Baseline (mBERT)	1.2756		*Baseline (mBERT)*	3.2152
swa-pol	LogSigma	1.7959			
	HUS@NLP-VNU	1.8713			
	Scmhl5	1.9391			
	Baseline (Mistral-3 14B)	2.2990			
	Baseline (mBERT)	2.7835			

Table 4: **Track B (DimStance) results.** Evaluation is based on RMSE. The top three teams and the official baseline systems are reported for each dataset.

to separate regression heads. Final predictions are stabilized using a multi-seed ensemble.

Team YangS_team (1st on zho-env$_{ST1}$). They fine-tune mDeBERTa-v3-base (He et al., 2021) with aspect-aware marker encoding to predict VA scores. The contextual representation of the aspect marker is pooled and passed to dual regression heads to jointly estimate valence and arousal, and the prediction stability is further improved through a 5-fold ensemble.

Team CYUT (1st on pcm-pol$_{ST1}$). They introduce a geometry-informed multi-task framework to fine-tune Qwen2-7B (Bai et al., 2023) with LoRA for VA regression. The framework incorporates auxiliary geometry-derived signals (polarity, intensity, quadrant, and directional prototypes), derived from the VA annotations, to stabilize training.

7 Analysis and Discussion

Model Architecture. Figure 3 summarizes the architectures adopted by participating systems and shows a trend consistent with recent SemEval tasks (Muhammad et al., 2025), where systems are primarily based on pretrained transformers (e.g., RoBERTa-family models) and LLMs (e.g., Qwen), as shown in Figure 4. Another popular approach is model ensembling. Teams constructed ensembles from models trained with different random seeds (LogSigma), cross-validation folds (YangS), and hyperparameters (ICT-NLP, 1st on zho-rest$_{ST1}$, Track A), as well as heterogeneous model architectures (NYCU Speech Lab, 1st on zho-rest$_{ST3}$ and zho-lap$_{ST3}$, Track A). In addition, Team HUS@NLP-VNU (1st on zho-fin$_{ST1}$, Track A) uses a syntax-aware Graph Convolutional Network (GCN) model.

Training Techniques. Figure 5 summarizes the training techniques used by participating systems. Most systems rely on fine-tuning pretrained models, typically via full fine-tuning or parameter-efficient adaptation, as shown in Figure 6. Beyond standard fine-tuning, some systems improve training stabil-

ity, such as using Smooth L1 loss (TeleAI) and log-variance loss weighting (LogSigma). Introducing auxiliary learning signals (CYUT) and adjusting the prediction distribution (PAI) can also improve performance. Team PALI (1st on zho-lap$_{ST2}$, Track A) further employs per-language adapters to capture language-specific VA distributions while reducing the number of required models.

Prompting Strategies. Figure 7 summarizes the prompting strategies adopted by participating systems, showing that instruction prompting with few-shot demonstrations is widely used. Beyond random sampling of demonstrations, in-context retrieval can improve prediction consistency by retrieving semantically similar training instances (Takoyaki). Meanwhile, Team nchellwig (1st on tat-rest$_{ST2}$, Track A) adopts a self-consistency strategy that executes the model multiple times with stochastic decoding and aggregates the resulting predictions via majority voting, retaining only tuples that achieve consensus to improve reliability.

8 Conclusions

This paper presents the SemEval-2026 shared task, which extends categorical ABSA and stance detection by incorporating a dimensional valence–arousal representation. We organize the task into the DimABSA and DimStance tracks and introduce three subtasks, ranging from pure regression to hybrid structured extraction with regression. We also introduce a new cF1 metric that unifies categorical and continuous evaluation.

We report results on systems evaluated on our multilingual and multidomain datasets, discuss top-performing systems, and summarize key design choices. These findings highlight challenges and opportunities for advancing dimensional sentiment analysis at the aspect and stance-target levels.

Limitations

Although DimABSA and DimStance datasets are multilingual, interpretations of valence and arousal can vary across cultures, thereby affecting cross-lingual comparability. We mitigate this by using five native-speaker annotators per language and sample, consistent 1–9 VA scales, and shared guidelines; nonetheless, results should be interpreted as comparisons across language-community-domain settings. Expanding language coverage and testing measurement invariance are important directions for future work.

Some datasets (e.g., Nigerian Pidgin) include more samples of negative or positive valence, which may bias models during training and inflate performance in the majority regions of the VA space. We document these differences and encourage explicit handling (e.g., reweighting or stratified sampling) when training or comparing models across languages (Chawla et al., 2002).

Ethical Considerations

People express attitudes, opinions, and sentiment towards entities and their aspects in complex and nuanced ways. Further, there can be considerable person-to-person variation. It should be noted that human annotated labels capture **perceived** sentiment and attitudes, and that in several cases this may be different from the speaker's true attitudes. Nonetheless, since language is a key mechanism to communicate, at an aggregate-level perceived opinions tend to correlate with actual opinions. Thus, even perceived opinions are useful at an aggregate level. However, caution must be employed when using individual inferred opinions to make decisions about individuals, especially high-stakes decisions.

ABSA and stance detection, like many technologies, can be abused and misused. For example, it can be used to identify likes and dislikes and to manipulate people into behaviours that may not be in their best interests (e.g., purchasing products or availing services that they cannot afford or that are not particularly useful to them). This is especially concerning for vulnerable populations such as children and the elderly. We expressly forbid any commercial use of our data.

For a detailed discussion of a large number of ethical considerations associated with automatic sentiment and emotion detection, we refer the reader to Mohammad (2022, 2023).

Acknowledgments

Liang-Chih Yu and Lung-Hao Lee acknowledge support from the National Science and Technology Council, Taiwan, under grants NSTC 113-2221-E-155-046-MY3 and NSTC 114-2221-E-A49-059-MY3.
Jonas Becker acknowledges the support of the Landeskriminalamt NRW.
The work of Alexander Panchenko was supported by the RSF project 25-71-30008 "Laboratory for reliable, adaptive, and trustworthy Artificial

Intelligence".
Ilseyar Alimova gratefully acknowledges Dina Abdullina for the Tatar data annotation and AIRI for financial support.
Jan Philip Wahle, Terry Ruas, and Bela Gipp acknowledge the support of the Lower Saxony Ministry of Science and Culture, and the VW Foundation.
Shamsuddeen Hassan Muhammad acknowledges the support of Google DeepMind, whose funding made this work possible.

References

Faisal Muhammad Adam, Lukman Jibril Aliyu, Sani Aji, Abdulhamid Abubakar, and Aliyu Rabiu Shuaibu. 2026. Team faisalm3at SemEval-2026 Task 3: From Standard Regression to Distributional Alignment in Dimensional Sentiment Analysis. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Muhammad Affan, M Hassan Shahzad, Mikaal Imam, Moiz Zulfiqar, Sandesh Kumar, and Abdul Samad. 2026. Habib University at SemEval-2026 Task 3: A Pipeline Approach for Dimensional Aspect-Based Sentiment Analysis. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Alibaba. 2025. Qwen3 technical report. *arXiv preprint*, page arXiv:2505.09388.

Emily Allaway and Kathleen McKeown. 2020. Zero-shot stance detection: A dataset and model using generalized topic representations. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8913–8931. Association for Computational Linguistics.

Rafif Alshawi, Amit Raj, Aleksey Kudelya, and Alexander Shirnin. 2026. The Classics at SemEval-2026 Task 3: Combining Transformer Models and LLM-Generated Annotations for Dimensional Aspect-Based Sentiment Analysis. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Cynthia Amol, Lilian Wanzare, and James Obuhuma. 2024. Politikweli: A swahili-english code-switched twitter political misinformation classification dataset. In *Speech and Language Technologies for Low-Resource Languages*, pages 3–17, Cham. Springer Nature Switzerland.

Georgios Arampatzis and Avi Arampatzis. 2026. DUTH at SemEval-2026 Task 3: Multilingual Transformer Models for Dimensional Stance Prediction Across Tracks. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

İrfan Aygün, Buket Kaya, and Mehmet Kaya. 2022. Aspect based twitter sentiment analysis on vaccination and vaccine types in COVID-19 pandemic with deep learning. *IEEE Journal of Biomedical and Health Informatics*, 26(5):2360–2369.

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, pages 2200–2204.

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.

Jonas Becker, Liang-Chih Yu, Shamsuddeen Hassan Muhammad, Jan Philip Wahle, Terry Ruas, Idris Abdulmumin, Lung-Hao Lee, Nelson Odhiambo, Lilian Wanzare, Wen-Ni Liu, Tzu-Mi Lin, Zhe-Yu Xu, Ying-Lung Lin, Jin Wang, Maryam Ibrahim Mukhtar, Bela Gipp, and Saif M. Mohammad. 2026. Dimstance: Multilingual datasets for dimensional stance analysis. *Preprint*, arXiv:2601.21483.

Aditya Praful Bhalgat, Omkar Dnyaneshwar Jagtap, and Anupama Phakatkar. 2026. PICT at SemEval-2026 Task 3: A Transformer-Based System for Dimensional Aspect-Aware Sentiment Regression with Weighted Layer Pooling. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Sven Buechel and Udo Hahn. 2017a. EmoBank: A Corpus of AnalyzeD Emotions on a Dimensional Level. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 572–578, Valencia, Spain. Association for Computational Linguistics.

Sven Buechel and Udo Hahn. 2017b. Emobank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pages 578–585.

Hongjie Cai, Rui Xia, and Jianfei Yu. 2021. Aspect-category-opinion-sentiment quadruple extraction with implicit aspects and opinions. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, pages 340–350.

An Cao-Hai, Lam Hoang-Thiet, Toan Le-Ngoc, and Linh Ha-My. 2026. HUS@NLP-VNU at SemEval-2026 Task 3: Dual-Stream Syntax-Aware Modeling and Direct Preference Optimization for Dimensional ABSA. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. 2002. Smote: synthetic minority over-sampling technique. *J. Artif. Int. Res.*, 16(1):321–357.

Siva Uday Sampreeth Chebolu, Franck Dernoncourt, Nedim Lipka, and Thamar Solorio. 2023. A review of datasets for aspect-based sentiment analysis. In *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 611–628, Nusa Dua, Bali. Association for Computational Linguistics.

Siva Uday Sampreeth Chebolu, Franck Dernoncourt, Nedim Lipka, and Thamar Solorio. 2024. Oats: A challenge dataset for opinion aspect target sentiment joint detection for aspect-based sentiment analysis. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation*, pages 12336–12347.

Cheng Chen. 2026. PALI at SemEval-2026 Task 3: LoRA Fine-Tuning with Validation for DimABSA. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Haohuan Chen and Han Liu. 2026. Scmhl5 at SemEval-2026 Task 3: Uncertainty-Aware Adversarial Learning for Embedding Enhancement in Dimensional Aspect-Based Sentiment Analysis. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Songqian Dai and Wei Lin. 2026. ALPS-Lab at SemEval-2026 Task 3: A Multilingual Generative LLM Approach for Dimensional Aspect Sentiment Analysis. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Giuseppe D'Aniello, Matteo Gaeta, and Ilaria La Rocca. 2022. Knowmis-absa: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis. *Artificial Intelligence Review*, 55(7):5543–5574.

A.J.W. De Vink, Filippos Karolos Ventirozos, Natalia Amat-Lefort, and Lifeng Han. 2026. QuadAI at SemEval-2026 Task 3: Ensemble Learning of Hybrid RoBERTa and LLMs for Dimensional Aspect-Based Sentiment Analysis. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. 2023. QLoRA: Efficient finetuning of quantized LLMs. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*.

Harshal Dharpure and Nicolay Rusnachenko. 2026. hdharpure at SemEval-2026 Task 3: BERT-Based Modeling and Prediction Behavior Analysis for Multilingual Valence–Arousal Scoring. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Yuzhe Ding, Kang He, Bobo Li, Li Zheng, Haijun He, Fei Li, Chong Teng, and Donghong Ji. 2025. Zero-shot conversational stance detection: Dataset and approaches. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 3221–3235, Vienna, Austria. Association for Computational Linguistics.

Anatolii Aleksanfrovich Frolov and Elisei Rykov. 2026. ssurface3 at SemEval-2026 Task 3: Efficient Methods for Multilingual Dimensional Aspect-Based Sentiment Analysis. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Stavros Gazetas, George Filandrianos, Maria Lymperaiou, Paraskevi Tzouveli, Athanasios Voulodimos, and Giorgos Stamou. 2026. AILS-NTUA at SemEval-2026 Task 3: Efficient Dimensional Aspect-Based Sentiment Analysis. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Gemini Team. 2023. Gemini: A family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.

Kyle Glandt, Sarthak Khanal, Yingjie Li, Doina Caragea, and Cornelia Caragea. 2021. Stance Detection in COVID-19 Tweets. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1596–1611, Online. Association for Computational Linguistics.

Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. 2021. Deberta: Decoding-enhanced bert with disentangled attention. *Preprint*, arXiv:2006.03654.

Qimao He and Xiaobing Zhou. 2026. YNU-ABSA at SemEval-2026 Task 3: A Unified Framework for Continuous and Structured Dimensional ABSA. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Nils Constantin Hellwig, Jakob Fehle, Udo Kruschwitz, and Christian Wolff. 2026. nchellwig at SemEval-2026 Task 3: Self-Consistent Structured Generation (SCSG) for Dimensional Aspect-Based Sentiment Analysis using Large Language Models. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Baraa Hikal, Jonas Becker, and Bela Gipp. 2026. LogSigma at SemEval-2026 Task 3: Uncertainty-Weighted Multitask Learning for Dimensional Aspect-Based Sentiment Analysis. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Linlin Hou, Wenhui Tu, Ting Yu, Ting Jiang, Mohamed Bah, Zenghui Xu, Yu Zhang, Gaoming Yang, and Ji Zhang. 2025. Aspect-based sentiment analysis for covid-19: A heterogeneous graph convolutional network approach. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 24(6):1–26.

Yupeng Hou, Jiacheng Li, Zhankui He, An Yan, Xiusi Chen, and Julian McAuley. 2024. Bridging language and items for retrieval and recommendation. *arXiv preprint arXiv:2403.03952*.

Hao-Chun Hsieh, Cheng-En Wu, and Yuan-Fu Liao. 2026. NYCU Speech Lab at SemEval-2026 Task 3: Heterogeneous Model Ensemble with Adaptive Weighted Voting for Dimensional Aspect Sentiment Quadruplet Extraction. 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, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.

Shuangjin Hu. 2026. kirito at SemEval-2026 Task 3: Dimensional Aspect-Based Sentiment Analysis via Sentence Structure Parsing Preprocessing and Prompt-Enhanced Instruction Tuning. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Yan Cathy Hua, Paul Denny, Jörg Wicker, and Katerina Taskova. 2024. A systematic review of aspect-based sentiment analysis: domains, methods, and trends. *Artificial Intelligence Review*, 57:296.

Yan Cathy Hua, Paul Denny, Jörg Wicker, and Katerina Taskova. 2025. Edurabsa: An education review dataset for aspect-based sentiment analysis tasks. *Preprint*, arXiv:2508.17008.

Liyuan Huang, Jiawei He, Wutao Shen, Lin Li, and Jin Zhang. 2026. ICT-NLP at SemEval-2026 Task 3: Less Is More - Multilingual Encoder with Joint Training and Adaptive Ensemble for Dimensional Aspect Sentiment Regression. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Wardat Shams Iqbal, Ruwad Naswan, and Swakkhar Shatabda. 2026. CLRG at SemEval-2026 Task 3: One Size Does Not Fit All: A Resource Adaptive Framework for Dimensional Sentiment Regression. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Athlene Jones, Vishwaa Shah, and Indika Kahanda. 2026. UNF-BMI at SemEval-2026 Task 3: Research Domain Criteria-Guided Large Language Models for Dimensional Aspect-Based Sentiment Analysis. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Lars Kaesberg, Terry Ruas, Jan Philip Wahle, and Bela Gipp. 2024. CiteAssist: A system for automated preprint citation and BibTeX generation. In *Proceedings of the Fourth Workshop on Scholarly Document Processing (SDP 2024)*, pages 105–119, Bangkok, Thailand. Association for Computational Linguistics.

Svetlana Kiritchenko, Saif Mohammad, and Mohammad Salameh. 2016. SemEval-2016 task 7: Determining sentiment intensity of english and arabic phrases. In *Proceedings of the10th International WorkshoponSemantic Evaluation*, pages 42–51.

Takahiro Kubo and Hiroki Nakayama. 2018. chabsa: Aspect-based sentiment analysis dataset.

Denis Laschenko and Albert Korotyk. 2026. SokraTUM at SemEval-2026 Task 3: A hybrid cascade of Label Distribution Learning, RAG supported generative extraction and contrastive metric learning for dimensional sentiment analysis. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Chia-Yun Lee, Matus Pleva, Daniel Hladek, and Ming-Hsiang Su. 2026a. SCU_Mesclab at SemEval-2026 Task 3: An Adaptive Dual-Track Framework for Dimensional Aspect-Based Sentiment Analysis. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Lung-Hao Lee, Jian-Hong Li, and Liang-Chih Yu. 2022. Chinese emobank: Building valence-arousal resources for dimensional sentiment analysis. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 21(4):65.

Lung-Hao Lee, Liang-Chih Yu, Natalia Loukashevich, Ilseyar Alimova, Alexander Panchenko, Tzu-Mi Lin, Zhe-Yu Xu, Jian-Yu Zhou, Guangmin Zheng, Jin Wang, Sharanya Awasthi, Jonas Becker, Jan Philip Wahle, Terry Ruas, Shamsuddeen Hassan Muhammad, and Saif M. Mohammad. 2026b. Dimabsa: Building multilingual and multidomain datasets for dimensional aspect-based sentiment analysis. *Preprint*, arXiv:2601.23022.

Lung-Hao Lee, Liang-Chih Yu, Suge Wang, and Jian Liao. 2024. Overview of the sighan 2024 shared task for chinese dimensional aspect-based sentiment analysis. In *Proceedings of the 10th SIGHAN Workshop on Chinese Language Processing*, pages 165–174.

Hongyu Li. 2026. SRCB at SemEval-2026 Task 3: Boosting DimASR via Contrastive LLM-Based Data Augmentation. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Jinglong Li and Yang Yang. 2026. hllwan at SemEval-2026 Task 3: Dimensional Aspect-Based Sentiment Analysis via LLM Feature Fusion and Test-Time Adaptation. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Yingjie Li, Tiberiu Sosea, Aditya Sawant, Ajith Jayaraman Nair, Diana Inkpen, and Cornelia Caragea. 2021. P-Stance: A Large Dataset for Stance Detection in Political Domain. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2355–2365, Online. Association for Computational Linguistics.

Siang-Ting Lin, Tien-Hong Lo, Yun-Ting Sun, Jhih-Rong Guo, Tung-Yen Hao, Fong-Chun Tsai, and Berlin Chen. 2026. NTNU-SMIL at SemEval-2026 Task 3: Logistic-Loss Regression with Same-Language Transfer for Valence–Arousal Stance Prediction in Dimensional Stance Analysis (DimStance). In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Mounika Marreddy, Subba Reddy Oota, Venkata Charan Chinni, Manish Gupta, and Lucie Flek. 2025. USDC: A dataset of *U*ser *S*tance and *D*ogmatism in long *C*onversations. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 23715–23759, Vienna, Austria. Association for Computational Linguistics.

MistralAI. 2025. Introducing mistral 3. *mistral.ai*, pages Accessed: 2025–12–31.

Mohammed Shahid Modi and Boleslaw Szymanski. 2026. RPI Team at SemEval-2026 Task 3: An LLM-Encoder Ensemble for Coarse-to-Fine Valence-Arousal Sentiment Prediction. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Saif Mohammad. 2018. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (volume 1: Long papers)*, pages 174–184.

Saif Mohammad. 2023. Best practices in the creation and use of emotion lexicons. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 1825–1836, Dubrovnik, Croatia. Association for Computational Linguistics.

Saif Mohammad and Felipe Bravo-Marquez. 2017. Emotion intensities in tweets. In *Proceedings of the 6th Joint Conference on Lexical and Computational Semantics*, pages 65–77.

Saif Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018. SemEval-2018 task 1: Affect in tweets. In *Proceedings of the 12th International Workshop on Semantic Evaluation*, pages 1–17.

Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. A Dataset for Detecting Stance in Tweets. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3945–3952, Portorož, Slovenia. European Language Resources Association (ELRA).

Saif M. Mohammad. 2022. Ethics sheet for automatic emotion recognition and sentiment analysis. *Computational Linguistics*, 48(2):239–278.

Saif M. Mohammad. 2025. NRC VAD Lexicon v2: Norms for Valence, Arousal, and Dominance for over 55k English Terms. *arXiv*, abs/2503.23547. ArXiv:2503.23547.

Saif M. Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. 2017. Stance and sentiment in tweets. *ACM Transactions on Internet Technology*, 17(3):26:1–26:23.

MoonshotAI. 2025. Kimi k2: Open agentic intelligence. *arXiv preprint*, page arXiv:2507.20534.

Shamsuddeen Hassan Muhammad, Idris Abdulmumin, Abinew Ali Ayele, Nedjma Ousidhoum, David Ifeoluwa Adelani, Seid Muhie Yimam, Ibrahim Sa'id Ahmad, Meriem Beloucif, Saif M. Mohammad, Sebastian Ruder, Oumaima Hourrane, Pavel Brazdil, Alipio Jorge, Felermino Dário Mário António Ali, Davis David, Salomey Osei, Bello Shehu Bello, Falalu Ibrahim, Tajuddeen Gwadabe, and 8 others. 2023. AfriSenti: A Twitter sentiment analysis benchmark for African languages. In *Proceedings of the*

2023 Conference on Empirical Methods in Natural Language Processing, pages 13968–13981, Singapore. Association for Computational Linguistics.

Shamsuddeen Hassan Muhammad, Nedjma Ousidhoum, Idris Abdulmumin, Jan Philip Wahle, Terry Ruas, Meriem Beloucif, Christine de Kock, Nirmal Surange, Daniela Teodorescu, Ibrahim Said Ahmad, David Ifeoluwa Adelani, Alham Fikri Aji, Felermino D. M. A. Ali, Ilseyar Alimova, Vladimir Araujo, Nikolay Babakov, Naomi Baes, Ana-Maria Bucur, Andiswa Bukula, and 29 others. 2025. BRIGHTER: BRIdging the gap in human-annotated textual emotion recognition datasets for 28 languages. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8895–8916, Vienna, Austria. Association for Computational Linguistics.

Edward Ombui. 2022. HateSpeech_kenya.

Haiyun Peng, Lu Xu, Lidong Bing, Fei Huang, Wei Lu, and Luo Si. 2020. Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, 34(5):8600–8607.

Maria Pontiki, Dimitrios Galanis, Harris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad Al-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, and 1 others. 2016. SemEval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation*, pages 19–30.

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. 2015. SemEval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation*, pages 486–495.

Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation*, pages 27–35.

Daniel Preoţiuc-Pietro, H. Andrew Schwartz, Gregory Park, Johannes Eichstaedt, Margaret Kern, Lyle Ungar, and Elisabeth Shulman. 2016. Modelling valence and arousal in Facebook posts. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 9–15, San Diego, California. Association for Computational Linguistics.

Qwen Team. 2025. Qwen2.5 technical report. *arXiv preprint arXiv:2412.15115*.

Oskar Riewe-Perła and Agata Filipowska. 2026. PUEB-DimASR at SemEval-2026 Task 3: Escaping the Mean Regression Trap with Graph-Enhanced Transformers for Dimensional Aspect-Based Sentiment Regression. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Sara Rosenthal, Preslav Nakov, Svetlana Kiritchenko, Saif Mohammad, Alan Ritter, and Veselin Stoyanov. 2015. SemEval-2015 task 10: Sentiment analysis in twitter. In *Proceedings of the 9th International WorkshoponSemantic Evaluation*, pages 451–463.

Zhihao Ruan, Kaifeng Yang, Cheng Chen, Wenwen Dai, and Wenjia Mao. 2026. PAI at SemEval-2026 Task 3: An LLM and Data Redistribution Adaptation-Based Predictive Strategy for Valence-Arousal Scores. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161–1178.

James A Russell. 2003. Core affect and the psychological construction of emotion. *Psychological review*, 110(1):145.

Michal Rynowiecki and Rob Van Der Goot. 2026. Team BOBW (Best Of Both Worlds) at SemEval-2026 Task 3: Modular Cross-Attention Encoders for Dimensional Aspect-Based Sentiment Analysis. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Jithu Morrison S and Abisha Rose S. 2026. Pixel Phantoms at SemEval-2026 Task 3: Language-Specific Transformer Regression for Dimensional Aspect-Based Sentiment Analysis. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1631–1642.

Lasse Strothe, Shaghayegh Sha Kolli, and Jana Diesner. 2026. TeamLasse at SemEval-2026 Task 3: A Hybrid Generative-Discriminative Framework for Dimensional Aspect-Based Sentiment Analysis. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Arseny Sukhodolsky, Ruslan Salimgareev, and Tatiana Ianshina. 2026. BertKittens at SemEval-2026 Task 3: Multi-Domain Aspect Sentiment with BERT/DeBERTa Ensembles for VA Regression and Aspect–Opinion–VA Triplets. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2):267–307.

Mike Thelwall, Kevan Buckley, and Georgios Paltoglou. 2012. Sentiment strength detection for the social web. *Journal of the Association for Information Science and Technology*, 63(1):163–173.

Vishal Thenuwara, Widanalage Mario Yomal De Mel, and Nisansa De Silva. 2026. Team VYN at SemEval-2026 Task 3: Dimensional Aspect-Based Sentiment Analysis. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Jannis Vamvas and Rico Sennrich. 2020. X-Stance: A Multilingual Multi-Target Dataset for Stance Detection. *arXiv preprint*. ArXiv:2003.08385 [cs].

ChengYan Wu, Bolei Ma, Yihong Liu, Zheyu Zhang, Ningyuan Deng, Yanshu Li, Baolan Chen, Yi Zhang, Yun Xue, and Barbara Plank. 2025. M-absa: A multilingual dataset for aspect-based sentiment analysis. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 2530–2557.

Shih-Hung Wu, Xian-Yan Chen, and Yi-Min Jian. 2026a. CYUT at SemEval-2026 Task 3: Multi-Task Dimensional Aspect Sentiment Regression with Polar Multi-Zone Labeling in VA Space. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Tong Wu, Nicolay Rusnachenko, and Huizhi(elly) Liang. 2026b. NCL-BU at SemEval-2026 Task 3: Fine-tuning XLM-RoBERTa for Multilingual Dimensional Sentiment Regression. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Lu Xu, Hao Li, Wei Lu, and Lidong Bing. 2020. Position-aware tagging for aspect sentiment triplet extraction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 2339–2349.

Kosuke Yamada, Sho Takase, and Ryosuke Kohita. 2026. Takoyaki at SemEval-2026 Task 3: Ensembling LLM Predictions using Demonstration Retrieval for Dimensional Aspect-based Sentiment Analysis. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, and 41 others. 2025. Qwen3 technical report. *Preprint*, arXiv:2505.09388.

Liu Yang, Gang Hu, and Jing Li. 2026. looploop at SemEval-2026 Task 3: A Dimensional Aspect-Based Sentiment System with DeBERTa Regression and Qwen3 Instruction Fine-Tuning. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Tsung-Hsien Yang and Shu-Fei Yang. 2026. YangS_team at SemEval-2026 Task 3: Transformer-Based Aspect-Aware Regression for Dimensional Sentiment and Stance Analysis. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Kuanlin Yu and Wen-Ni Liu. 2026. kevinyu66 at SemEval-2026 Task 3: A Retrieval-Augmented LLM System for Aspect–Opinion Triplet Extraction. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Liang-Chih Yu, Lung-Hao Lee, Shuai Hao, Jin Wang, Yunchao He, Jun Hu, K. Robert Lai, and Xuejie Zhang. 2016. Building chinese affective resources in valence-arousal dimensions. In *Proceedings of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 540–545.

Wenxuan Zhang, Yang Deng, Xin Li, Yifei Yuan, Lidong Bing, and Wai Lam. 2021. Aspect sentiment quad prediction as paraphrase generation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9209–9219.

Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing, and Wai Lam. 2023. A survey on aspect-based sentiment analysis: tasks, methods, and challenges. *IEEE Transactions on Knowledge and Data Engineering*, 35(11019–11038):101436.

ZhaoDan Zhang, Jin Zhang, Xueqi Cheng, and Hui Xu. 2025. T-MAD: Target-driven multimodal alignment for stance detection. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 580–595, Suzhou, China. Association for Computational Linguistics.

Chenye Zhao and Cornelia Caragea. 2024. EZ-STANCE: A large dataset for English zero-shot stance detection. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15697–15714, Bangkok, Thailand. Association for Computational Linguistics.

Chenye Zhao, Yingjie Li, and Cornelia Caragea. 2023. C-STANCE: A large dataset for Chinese zero-shot

stance detection. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13369–13385, Toronto, Canada. Association for Computational Linguistics.

Shijia Zhou, Siyao Peng, Simon M. Luebke, Jörg Haßler, Mario Haim, Saif M. Mohammad, and Barbara Plank. 2025. What media frames reveal about stance: A dataset and study about memes in climate change discourse. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 5337–5356.

Yan Zhou, Wangshicheng Shicheng Wang, Shiquan Wang, Mengjiao Bao, Ruiyu Fang, Shuangyong Song, Yongxiang Li, and Xuelong Li. 2026a. TeleAI at SemEval-2026 Task 3: Large Language Models for Dimensional Aspect-Based Sentiment Analysis. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Ziang Zhou, Xiangmei He, and Chenhongyi Bai. 2026b. SCUZANE at SemEval-2026 Task 3: Dimension Aspect-based Sentiment Analysis with Supervised Contrastive Regression and R-Drop Regularization. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*, San Diego, California. Association for Computational Linguistics.

Elena Zotova, Rodrigo Agerri, Manuel Nuñez, and German Rigau. 2020. Multilingual Stance Detection in Tweets: The Catalonia Independence Corpus. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 1368–1375, Marseille, France. European Language Resources Association.

A Aspect Category List

- Laptop

Entity Labels
LAPTOP, DISPLAY, KEYBOARD, MOUSE, MOTHERBOARD, CPU, FANS_COOLING, PORTS, MEMORY, POWER_SUPPLY, OPTICAL_DRIVES, BATTERY, GRAPHICS, HARD_DISK, MULTIMEDIA_DEVICES, HARDWARE, SOFTWARE, OS, WARRANTY, SHIPPING, SUPPORT, COMPANY
Attribute Labels
GENERAL, PRICE, QUALITY, DESIGN_FEATURES, OPERATION_PERFORMANCE, USABILITY, PORTABILITY, CONNECTIVITY, MISCELLANEOUS

- Restaurant

Entity Labels
RESTAURANT, FOOD, DRINKS, AMBIENCE, SERVICE, LOCATION
Attribute Labels
GENERAL, PRICES, QUALITY, STYLE_OPTIONS, MISCELLANEOUS

- Hotel

Entity Labels
HOTEL, ROOMS, FACILITIES, ROOM_AMENITIES, SERVICE, LOCATION, FOOD_DRINKS
Attribute Labels
GENERAL, PRICE, COMFORT, CLEANLINESS, QUALITY, DESIGN_FEATURES, STYLE_OPTIONS, MISCELLANEOUS

B Example Calculation of cF1

Prediction/Gold	TP_{cat} (A)	**VA error distance**		**cTP** (A)-(C)
		Raw (B)	Normalized $(C) = (B)/\sqrt{128}$	
P: (food, good, 8.00#8.00) G: (food, good, 7.00#7.00)	1	$\sqrt{2}$	$\frac{\sqrt{2}}{\sqrt{128}} = 0.125$	0.875
P: (soup, spicy, 7.50#7.50) G: (soup, spicy, 3.50#3.50)	1	$\sqrt{32}$	$\frac{\sqrt{32}}{\sqrt{128}} = 0.5$	0.5
P: (staff, friendly, 7.00#7.00) G: (staff, always friendly, 7.50#7.50)	0	-	-	0
P: (staff, good, 7.00#7.00) G: N/A	0	-	-	0
			Total cTP	1.375
$cRecall = 1.375/3 = 0.458$				
$cPrecision = 1.375/4 = 0.344$				
$cF1 = (2 * 0.458 * 0.344)/(0.458 + 0.344) = 0.393$				

Note: The VA scores lie in the range [1, 9]. When the VA prediction is perfect (i.e., dist=0), cRecall/cPrecision reduces to the standard recall/precision.

C Teams

Tracks	Team	Affiliation	Paper
A	AILS-NTUA	Artificial Intelligence and Learning Systems Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, Greece	(Gazetas et al., 2026)
A	ALPS-Lab	Fujian University of Technology, China	(Dai and Lin, 2026)
A	Bert Kittens	Individual researcher	(Sukhodolsky et al., 2026)
B	CLRG	Bangladesh University of Engineering and Technology, Bangladesh; BRAC University, Bangladesh	(Iqbal et al., 2026)
B	CYUT	Chaoyang University of Technology, Taiwan	(Wu et al., 2026a)
A, B	DUTH	Department of Electrical & Computer Engineering, Democritus University of Thrace, Greece	(Arampatzis and Arampatzis, 2026)
A, B	HUS@NLP-VNU	Faculty of Mathematics, Mechanics and Informatics, VNU University of Science, Hanoi, Vietnam	(Cao-Hai et al., 2026)
A	Habib university	Habib University, Pakistan	(Affan et al., 2026)
A	ICT-NLP	Institute of Computing Technology, Chinese Academy of Sciences, China	(Huang et al., 2026)
A, B	LogSigma	University of Göttingen, Germany	(Hikal et al., 2026)
A	NCL-BU	Bournemouth University, UK; Newcastle University, UK	(Wu et al., 2026b)
A, B	NTNU-SMIL	Speech and Machine Intelligence Laboratory (SMIL), Department of Computer Science and Information Engineering, National Taiwan Normal University, Taiwan	(Lin et al., 2026)
A	NYCU Speech Lab	Institute of Artificial Intelligence Innovation, National Yang Ming Chiao Tung University, Taiwan	(Hsieh et al., 2026)
A, B	PAI	Ping An Life Insurance Company of China, Ltd.	(Ruan et al., 2026)
A, B	PALI	none	(Chen, 2026)
A	PICT	Pune Institute of Computer Technology, India	(Bhalgat et al., 2026)
A	PUEB-DimASR	Poznan University of Economics and Business, Poland	(Riewe-Perła and Filipowska, 2026)
A, B	Pixel Phantoms	Sri Sivasubramaniya Nadar College of Engineering, India; Loyola-ICAM College of Engineering and Technology, India	(S and S, 2026)
A	QuadAI	Leiden University, Netherlands; Leiden University Medical Center (LUMC), Netherlands; Manchester Metropolitan University, UK	(De Vink et al., 2026)
A	RPI Team	Rensselaer Polytechnic Institute, Troy NY, USA	(Modi and Szymanski, 2026)
A	SCUZANE	Sichuan University, China	(Zhou et al., 2026b)
A, B	SCU_Mesclab	Department of Data Science, Soochow University, Taiwan; Department of Computer Networks, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovakia	(Lee et al., 2026a)
A	SRCB	Ricoh Software Research Center (Beijing) Co., Ltd	(Li, 2026)
A, B	Scmhl5	College of Computer Science and Software Engineering, Shenzhen University, China	(Chen and Liu, 2026)
A	SokraTUM	Technical University of Munich, Germany	(Laschenko and Korotyk, 2026)
A	Takoyaki	CyberAgent, Japan	(Yamada et al., 2026)
A	Team BOBW (Best Of Both Worlds)	IT University of Copenhagen, Denmark	(Rynowiecki and Van Der Goot, 2026)
A	Team HausaNLP	National Open University of Nigeria, Nigeria; Gombe State University, Nigeria; Nassarawa State University Keffi, Nigeria; Nile University Abuja, Nigeria	(Adam et al., 2026)
A	Team VYN	Department of Computer Science & Engineering, University of Moratuwa, Sri Lanka	(Thenuwara et al., 2026)
A	TeamLasse	Technical University of Munich, Germany	(Strothe et al., 2026)
A	TeleAI	Institute of Artificial Intelligence (TeleAI), China Telecom	(Zhou et al., 2026a)
A	The Classics	HSE University, Russia	(Alshawi et al., 2026)

(Continued on next page)

(Continued from previous page)

Tracks	Team	Affiliation	Paper
A	UNF-BMI	University of North Florida, USA	(Jones et al., 2026)
A	YNU-ABSA	Yunnan University, China	(He and Zhou, 2026)
A, B	YangS_team	Chunghwa Telecom Co., Ltd., Taiwan	(Yang and Yang, 2026)
A	hdharpure	Indian Institute of Technology Patna, India	(Dharpure and Rusnachenko, 2026)
B	hllwan	Nanjing University of Science and Technology, China	(Li and Yang, 2026)
A	kevinyu66	National Cheng Kung University, Taiwan	(Yu and Liu, 2026)
A	kirito	Yunnan University, China	(Hu, 2026)
A	looploop	Yunnan University, China	(Yang et al., 2026)
A	nchellwig	Media Informatics Group, University of Regensburg, Germany	(Hellwig et al., 2026)
A	ssurface3	Skoltech, Russia	(Frolov and Rykov, 2026)

Table 5: Participants information (tasks, affiliations, and papers).

D Scores

S/N	Team	eng-rest	eng-lap	jpn-hot	jpn-fin	rus-rest	tat-rest	ukr-rest	zho-rest	zho-lap	zho-fin
1	AILS-NTUA	1.3933	1.4401	0.7484	0.9635	1.7236	2.1144	1.6724	1.0023	0.7457	0.5425
2	Bert Kittens	1.1812	1.2769	0.7267	0.9675	1.5828	2.2118				
3	DUTH		1.5924								
4	HUS@NLP-VNU	1.2745	1.4109	0.6386	0.8296	1.3075	1.8220	1.3538	0.9595	0.6663	**0.4841**
5	Habib university	1.3049	1.3654	0.6680	0.8907	1.4344	1.6041	1.4661	0.9898	0.7311	0.5333
6	ICT-NLP	1.2676	1.3098	0.6289	0.8331	1.4420	1.8596	1.4750	**0.9256**	0.6553	0.4892
8	LogSigma	**1.1035**	**1.2408**								
9	NCL-BU	1.4861	1.4562						0.9553	0.7510	0.5391
11	NTNU-SMIL	1.2846	1.3501	0.6378	0.9278	1.4430	2.1785	1.4655	0.9841	0.6695	0.5115
12	PAI	1.2141	1.4394	0.6508	0.7584	**1.2190**	**1.5294**	**1.1888**	0.9766	0.6800	0.5977
13	PALI	1.2866	1.3612	0.6237	0.7532	1.3642	1.7121	1.4030	0.9805	0.6681	0.6042
14	PICT	1.1958	1.3261								
15	PUEB-DimASR	1.7011	1.7587	1.2827	1.4505	2.2749	2.3347	2.2589	1.2405	1.1343	0.8179
16	Pixel Phantoms	1.3656	1.4190	0.7297	1.0242	1.7686	2.0729	1.5937	0.9823	0.7438	0.7259
17	QuadAI	1.3632	1.4062								
18	RPI Team	1.2006	1.2833	0.6413	0.8254	1.4849	1.7837	1.5485	0.9599	0.7005	0.5398
19	SCUZANE	1.3483	1.4242	0.7129	0.9580	1.5572	2.3199	1.5730	0.9636	0.6981	0.5117
20	SCU_Mesclab	1.2277	1.3946						1.1210	0.9222	0.6692
21	SRCB	1.2270									
22	Scmhl5	1.3168		0.6811	0.9292	1.4609	2.0142	1.4732	0.9838	0.7165	
23	SokraTUM	1.3011	1.2942								
24	Team HausaNLP	1.4936	1.5143								
25	Team VYN	1.7978									
26	TeamLasse	1.4265			0.9982	1.5991	2.0212	1.6039	1.1601	1.0931	
27	TeleAI	1.2139	1.2425	**0.5561**	**0.6581**	1.2456	1.7662	1.3234	0.9265	**0.6103**	0.4866
28	The Classics	1.2324	1.3283			1.6390					
29	UNF-BMI	1.3920	1.4336								
30	YNU-ABSA	1.4001	1.4198	0.7554	1.0026	1.5967	2.0104		0.9945		
31	YangS_team	1.2772	1.3455						0.9433	0.6867	0.4864
32	hdharpure	1.5003	1.5412	0.8378	1.0292	1.6515	2.0463	1.7172	0.9847	0.7902	0.5704
33	kirito	1.3966	1.5010								
34	looploop	1.2048	1.3021								
36	ssurface3	1.9115	1.8486	1.1509	1.4514	1.7572	1.9471	1.7793	1.0870	0.9482	0.8329
	Average	1.3508	1.4110	0.7408	0.9531	1.5471	1.9482	1.5467	1.0008	0.7628	0.5805
	Baseline (Kimi-K2 Thinking)	2.1461	2.1893	1.7553	1.6396	1.7768	1.9380	1.7805	1.8959	1.6440	1.9652
	Baseline (Qwen-3 14B)	2.6427	2.8089	2.2906	1.8964	2.1528	2.6367	2.2121	2.0073	1.7706	1.4707

Table 6: DimABSA results (Track A, Subtask 1).

S/N	Team	eng-rest	eng-lap	jpn-hot	rus-rest	tat-rest	ukr-rest	zho-rest	zho-lap
1	AILS-NTUA	0.6518	0.5311	0.5021	0.4988	0.3874	0.4725	0.5042	0.4646
2	ALPS-Lab	0.0000	0.0000	0.0000	0.5414	0.4798	0.5613	0.5247	0.4935
3	Bert Kittens	0.5628	0.4469	0.4202	0.3137	0.1692			
5	HUS@NLP-VNU	0.6391	0.5304						
6	Habib university	0.5202	0.4770	0.3311	0.5492	0.4839	0.5324	0.4622	0.4159
7	ICT-NLP	0.6174	0.5622	0.3152	0.4622	0.3088	0.4355	0.2756	0.3019
9	PAI	0.6903	0.6169	0.5682	**0.5793**	0.4908	**0.5787**	**0.5638**	0.5306
10	PALI	0.6928	0.6242	0.5666	0.5724	0.4828	0.5671	0.5634	**0.5308**
11	Pixel Phantoms	0.0265							
12	Scmhl5	0.6127	0.5136	0.3357	0.3960	0.3649	0.4267	0.3955	0.3111
13	SokraTUM	0.6326	0.5635						
14	Takoyaki	**0.7021**	**0.6366**	0.5340	0.5564	0.5092	0.5438	0.5382	0.4758
15	TeamLasse	0.6391	0.5513	0.5694	0.5253	0.4496	0.5270	0.5320	0.4807
16	TeleAI	0.6294	0.5345	**0.5837**	0.5736	0.4863	0.5712	0.5448	0.5292
17	The Classics	0.5650	0.4763						
18	YNU-ABSA	0.5240	0.4952						
19	kevinyu66	0.6707	0.5503	0.5366	0.5117	0.3731	0.4865	0.5089	0.4802
20	kirito	0.5676	0.4733						
21	looploop	0.5799	0.4799						
22	nchellwig	0.6985	0.6092	0.5518	0.5640	**0.5119**	0.5285	0.5488	0.5110
	Average	0.5686	0.5123	0.4570	0.5106	0.4223	0.5183	0.4973	0.4606
	Baseline (Kimi-K2 Thinking)	0.4920	0.4424	0.3464	0.4242	0.3577	0.4220	0.3529	0.2494
	Baseline (Qwen-3 14B)	0.4483	0.3827	0.1622	0.3341	0.2020	0.3099	0.2509	0.2099

Table 7: DimABSA results (Track A, Subtask 2).

S/N	Team	eng-rest	eng-lap	jpn-hot	rus-rest	tat-rest	ukr-rest	zho-rest	zho-lap
1	AILS-NTUA	0.5988	0.2694	0.3747	0.4369	0.3306	0.4154	0.4544	0.3703
2	ALPS-Lab	0.6202	0.3395	0.3617	0.5042	0.4404	0.5163	0.4853	0.3968
3	Bert Kittens	0.5162	0.2578	0.2845		0.1479			
4	HUS@NLP-VNU	0.5871	0.2587						
5	Habib university	0.0000	0.0000	0.1853	0.3029	0.2500	0.2938	0.4199	0.3139
7	NYCU Speech Lab							**0.5521**	**0.4824**
8	PAI		0.3758		**0.5599**	0.4523	**0.5437**	0.5360	0.4316
9	PALI	0.6395	0.3793	**0.4252**	0.5496	0.4443	0.5307	0.5357	0.4319
10	Scmhl5	0.5119	0.2752	0.2195	0.3138	0.2629	0.3384	0.3309	0.1996
11	SokraTUM	0.5612	0.2512						
12	Takoyaki	**0.6514**	**0.4227**	0.4086	0.5130	**0.4736**	0.5019	0.4966	0.3745
13	Team BOBW	0.5317	0.2317						
14	TeamLasse	0.5937	0.3049	0.3992	0.4991	0.4113	0.4879	0.5026	0.3478
15	TeleAI	0.5487	0.3281	0.1258	0.3357	0.2512	0.3245	0.3979	0.1885
16	The Classics		0.3072						
17	YNU-ABSA	0.5183							
18	kirito	0.5201	0.2480						
19	looploop	0.5562	0.2781						
20	nchellwig	0.6403	0.4006	0.3974	0.5083	0.4557	0.4746	0.4966	0.4016
	Average	0.5398	0.2908	0.3259	0.4526	0.3581	0.4446	0.4728	0.3602
	Baseline (Kimi-K2 Thinking)	0.3746	0.2795	0.1943	0.2963	0.2380	0.2971	0.2859	0.1900
	Baseline (Qwen-3 14B)	0.2673	0.1529	0.0400	0.1682	0.0954	0.1641	0.1605	0.1124

Table 8: DimABSA results (Track A, Subtask 3).

E System Statistics

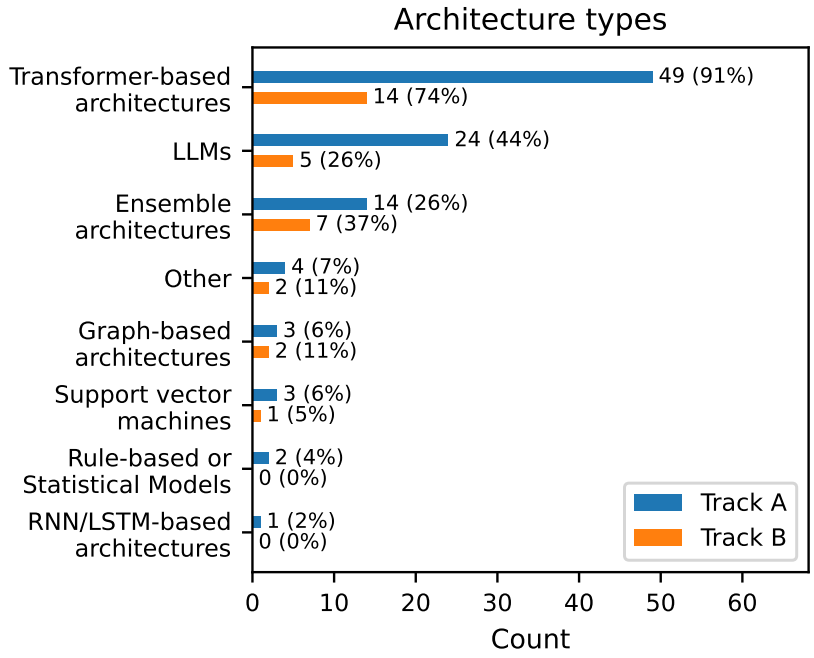

Figure 3: **Model architectures used by participants.** One team can be eligible for multiple categories.

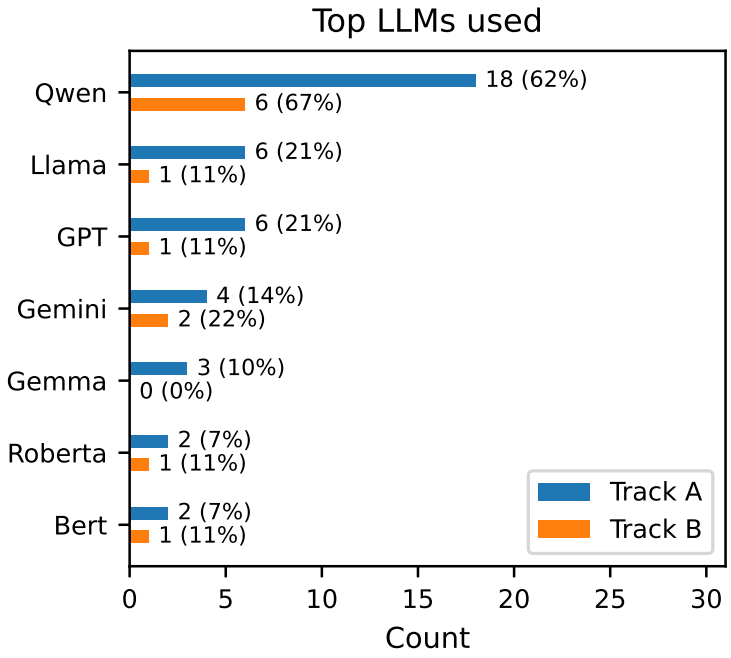

Figure 4: **LLMs used by participants.** One team can be eligible for multiple categories.

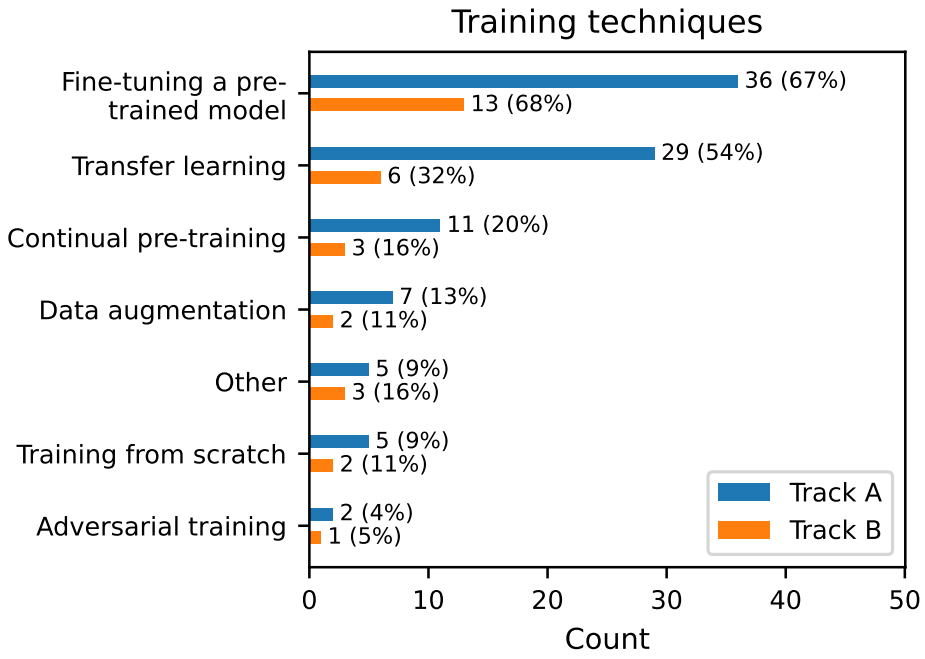

Figure 5: **Training techniques used by participants.** One team can be eligible for multiple categories.

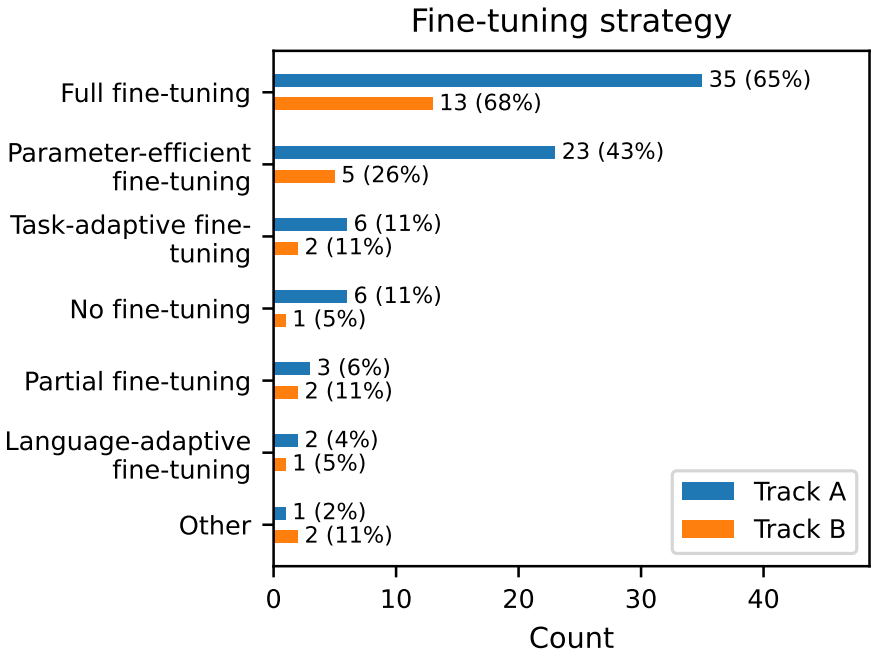

Figure 6: **Fine-tuning strategies used by participants.** One team can be eligible for multiple categories.

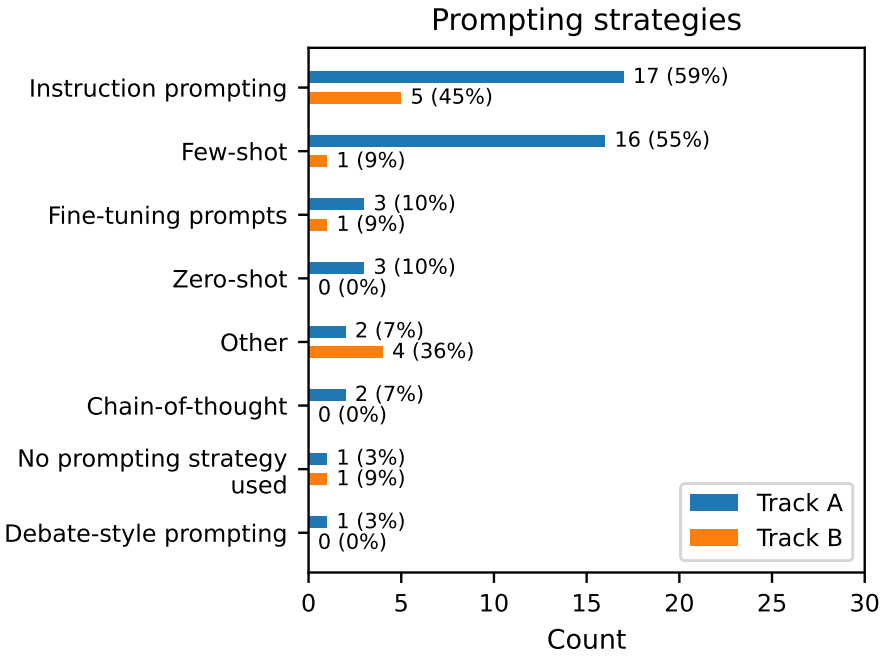

Figure 7: **Prompting strategies used by participants.** One team can be eligible for multiple categories.

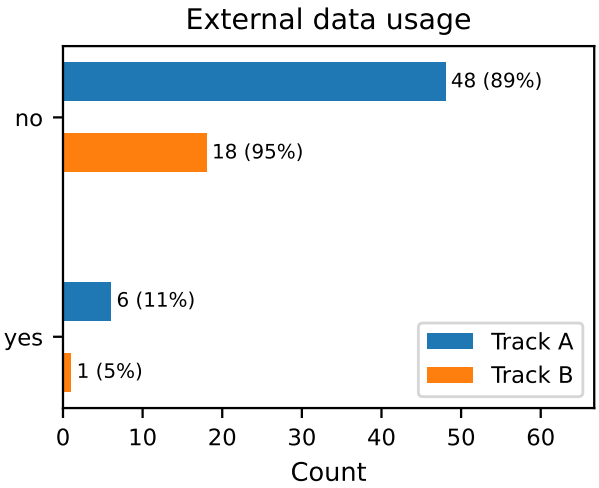

Figure 8: **External data used by participants.**

CiteAssist
CITATION SHEET

Generated with citeassist.uni-goettingen.de
(Kaesberg et al., 2024)

BibTeX Entry

```
@inproceedings{yu-etal-2026-semeval,
  author={Yu, Liang-Chih and Becker, Jonas and Muhammad, Shamsuddeen Hassan and Abdulmumin,
      Idris and Lee, Lung-Hao and Lin, Ying-Lung and Wang, Jin and Wahle, Jan Philip and Lima
      Ruas, Terry and Loukachevitch, Natalia and Panchenko, Alexander and Alimova, Ilseyar and
       Wanzare, Lilian Diana Awuor and Odhiambo, Nelson and Gipp, Bela and Chang, Kai-Wei and
      Mohammad, Saif},
  title={{S}em{E}val-2026 Task 3: Dimensional Aspect-Based Sentiment Analysis ({D}im{ABSA})},
  abstract={We present the SemEval-2026 shared task on Dimensional Aspect-Based Sentiment
      Analysis (DimABSA), which improves traditional ABSA by modeling sentiment along valence
      {--}arousal (VA) dimensions rather than using categorical polarity labels. To extend
      ABSA beyond consumer reviews to public-issue discourse (e.g., political, energy, and
      climate issues), we introduce an additional task, Dimensional Stance Analysis (DimStance
      ), which treats stance targets as aspects and reformulates stance detection as
      regression in the VA space. The task consists of two tracks: Track A (DimABSA) and Track
       B (DimStance). Track A includes three subtasks: (1) dimensional aspect sentiment
      regression, (2) dimensional aspect sentiment triplet extraction, and (3) dimensional
      aspect sentiment quadruplet extraction, while Track B includes only the regression
      subtask for stance targets. We also introduce a continuous F1 (cF1) metric to jointly
      evaluate structured extraction and VA regression.The task attracted more than 400
      participants, resulting in 112 final submissions and 42 system description papers. We
      report baseline results, discuss top-performing systems, and analyze key design choices
      to provide insights into dimensional sentiment analysis at the aspect and stance-target
      levels. All resources are available on our GitHub repository.},
  address={San Diego, California, USA},
  booktitle={Proceedings of the 20th {I}nternational {W}orkshop on {S}emantic {E}valuation
      (2026)},
  editor={Kochmar, Ekaterina and Ghosh, Debanjan and North, Kai and Komachi, Mamoru and
      Zampieri, Marcos},
  isbn={979-8-89176-414-9},
  pages={3753--3778},
  publisher={Association for Computational Linguistics},
  year={2026},
  month={07}
}
```

Online Access

ACL Anthology https://aclanthology.org/2026.semeval-1.452/