YangS_team at SemEval-2026 Task 3: Transformer-Based Aspect-Aware Regression for Dimensional Sentiment and Stance Analysis

Tsung-Hsien Yang
Advanced Technology Laboratory
Telecommunication Laboratories
Chunghwa Telecom Co., Ltd.
yasamyang@cht.com.tw

Shu-Fei Yang
Independent Researcher
shufeiyangtw@gmail.com

Abstract

This paper describes our system for the SemEval-2026 Task 3: Dimensional Aspect-Based Sentiment Analysis (DimABSA). We participate in Track A (DimABSA) and Track B (DimStance), both of which involve Subtask 1 – predicting continuous valence–arousal (VA) scores for given text–aspect pairs in English and Chinese. Our system combines pre-trained multilingual transformers with aspect-marker input encoding and dual regression heads for VA prediction, trained with a 5-fold cross-validation ensemble. We select XLM-RoBERTa-large as the backbone for Track A and mDeBERTa-v3-base for Track B based on systematic model comparison on the development sets. On the official test sets, our system substantially outperforms the organizer-provided baselines across all language domain settings. On the unofficial post-evaluation leaderboard, the system achieves strong results on Chinese subsets, ranking 1st on `zho-env` (Track B) and 2nd on `zho-fin` (Track A).

1 Introduction

Aspect-Based Sentiment Analysis (ABSA) is a fundamental task that identifies sentiments toward specific aspects in text (Zhang et al., 2023; Pontiki et al., 2016). Traditional ABSA adopts categorical sentiment labels (positive, negative, neutral), which cannot capture the nuanced nature of human emotional expression. Dimensional sentiment analysis, grounded in the valence–arousal (VA) model from affective psychology (Russell, 1980), represents sentiment as continuous scores along two dimensions: *valence* (negative to positive) and *arousal* (calm to excited), allowing finer-grained distinctions.

SemEval-2026 Task 3 (Yu et al., 2026) introduces DimABSA, which integrates dimensional sentiment into the ABSA framework. Track A extends ABSA with VA scores for consumer reviews, while Track B (DimStance) reformulates stance detection as a Stance-as-DimABSA task, treating stance targets as aspects and replacing categorical stance labels with continuous VA scores. We participate in Subtask 1 of both tracks—Dimensional Aspect Sentiment Regression (DimASR)—which requires predicting real-valued VA scores for given text–aspect pairs.

We frame this paper as an *empirical system description*: the building blocks—transformer fine-tuning, aspect-marker encoding, dual regression heads, and k-fold ensembling—are all standard. Our goal is to document which off-the-shelf choices work well and to report calibrated numbers against official baselines and the unofficial post-evaluation leaderboard. We fine-tune pre-trained multilingual transformers with aspect markers around target terms and predict VA independently from the marker token's hidden state, selecting XLM-RoBERTa-large (Conneau et al., 2020) for Track A and mDeBERTa-v3-base (He et al., 2023) for Track B, both with 5-fold ensemble. Our main findings are: (i) the system outperforms organizer baselines on every test set; (ii) backbone choice is the dominant factor; and (iii) According to the unofficial leaderboard, our system ranks 1st on Track B `zho-env`, 2nd on Track A `zho-fin`, 4th on Track A `zho-res`, 5th on Track B `eng-env`, 8th on Track A `zho-lap`, and 11th–14th on the English Track A subsets, achieving top-5 placement on 4 of 7 subsets.

2 Background

2.1 Task Description

SemEval-2026 Task 3 comprises two complementary tracks (Yu et al., 2026). **Track A (DimABSA)** extends traditional ABSA by replacing categorical sentiment polarities with continuous VA scores. Given a text and one or more aspect terms, systems predict a VA score pair for each aspect. The datasets span two languages—English (restaurant,

Proceedings of the The 20th International Workshop on Semantic Evaluation (2026), pages 1533–1538
July 3-4, 2026 ©2026 Association for Computational Linguistics

laptop domains) and Chinese (restaurant, laptop, finance domains) (Lee et al., 2026). **Track B (DimStance)** reformulates stance detection under the DimABSA schema, where stance targets are treated as aspects and discrete stance labels are replaced with continuous VA scores (Becker et al., 2026). The datasets include English and Chinese, both in the environmental protection domain.

For both tracks, VA scores range from 1.00 to 9.00, where 5.00 is neutral. The official evaluation metric is $\mathrm{RMSE_{VA}}$:

$$\mathrm{RMSE_{VA}} = \sqrt{\frac{\sum_{i=1}^{N}[(V_p^i - V_g^i)^2 + (A_p^i - A_g^i)^2]}{N}} \quad (1)$$

where N is the number of instances, V_p^i and A_p^i are predicted values, and V_g^i and A_g^i are gold values.

2.2 Related Work

Aspect-Based Sentiment Analysis. ABSA has traditionally been formulated with categorical polarity labels (positive/negative/neutral), as in the SemEval ABSA shared tasks (Pontiki et al., 2016). Zhang et al. (2023) survey recent neural ABSA approaches, most of which output discrete labels and therefore cannot represent fine-grained intensity differences between, e.g., “acceptable” and “excellent”.

Dimensional Sentiment Analysis. The valence–arousal model from affective psychology (Russell, 1980) represents sentiment as two continuous dimensions. Mohammad et al. (2018) introduced VA regression for tweets, and Lee et al. (2022) released the Chinese EmoBank resource for VA-based sentiment analysis. SemEval-2026 Task 3 (Yu et al., 2026) unifies these two lines of work by introducing aspect-level VA regression for both review (DimABSA, Lee et al., 2026) and stance (DimStance, Becker et al., 2026) data.

Pre-trained Multilingual Encoders. XLM-RoBERTa (Conneau et al., 2020) is pre-trained on 100 languages and remains a strong cross-lingual baseline; DeBERTa (He et al., 2021) and DeBERTa-v3 (He et al., 2023) introduce disentangled attention and ELECTRA-style pre-training, with mDeBERTa-v3-base offering competitive multilingual performance at base size. For Chinese, MacBERT and Chinese RoBERTa-wwm-ext (Cui et al., 2020) are widely used Chinese-specific encoders. We treat all of these as candidate backbones and choose between them empirically.

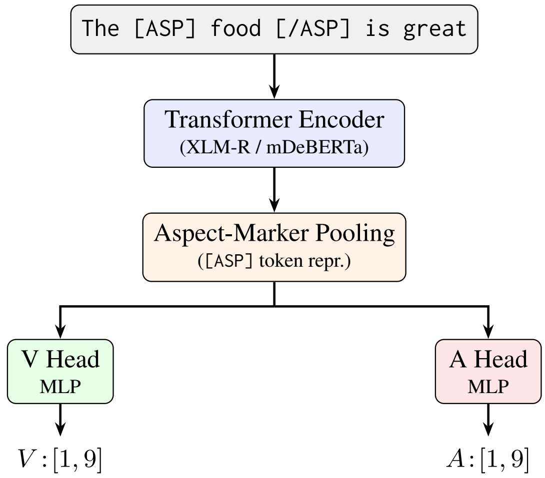

Figure 1: System architecture. Text with aspect markers is encoded, the [ASP] representation is pooled, and dual MLP heads predict valence and arousal.

Aspect-Marker Input Encoding. Wrapping the target span with special tokens to make the aspect explicit to a transformer is a common practice in aspect- and entity-level tasks: it requires no architectural changes and lets self-attention condition representations on the marked span. We adopt this technique unchanged and use the marker token (rather than [CLS]) as the regression input.

3 System Overview

Our system follows a unified architecture for both tracks, consisting of three components: (1) aspect-aware input encoding, (2) a pre-trained transformer encoder with aspect-marker pooling, and (3) dual regression heads for VA prediction. Figure 1 illustrates the overall architecture.

3.1 Aspect-Aware Input Encoding

Given an input text T and an aspect term a, we construct the model input by wrapping the aspect with special marker tokens [ASP] and [/ASP]. If the aspect term appears in the text, we insert markers at the corresponding position; otherwise, we append it as: [CLS] T [SEP] [ASP] a [/ASP] [SEP]. This marking strategy enables the model to explicitly attend to the aspect position within its self-attention mechanism. Unicode normalization (NFC) is applied to all text inputs, and URLs and excessive whitespace are cleaned, while casing is preserved.

3.2 Transformer Encoder

We explored multiple pre-trained models in our experiments. For Track A, **XLM-RoBERTa-large** (Conneau et al., 2020) is selected as the final backbone due to its strong cross-lingual transfer ability, especially for English data. We also experimented with Chinese-specific models including Chinese-MacBERT-large and Chinese-RoBERTa-wwm-ext-large (Cui et al., 2020), which are competitive on Chinese subsets. For Track B, **mDeBERTa-v3-base** (He et al., 2023) is selected as it achieves better performance than larger alternatives on the stance data.

The aspect markers `[ASP]` and `[/ASP]` are added to the tokenizer's vocabulary as special tokens, and the encoder's embedding layer is resized accordingly. The maximum sequence length is set to 256 tokens.

3.3 Aspect-Marker Pooling

Rather than using the standard `[CLS]` token representation, we employ *aspect-marker pooling*, which extracts the contextual representation of the `[ASP]` token from the encoder's last hidden layer. Since this token is positioned immediately before the aspect term, its hidden state encodes aspect-conditioned contextual information through the transformer's self-attention layers. If the marker is absent (edge cases), the system falls back to `[CLS]` pooling.

3.4 Dual Regression Heads

Two separate MLP-based regression heads predict valence and arousal independently. Each head consists of two fully-connected hidden layers with GELU activation and dropout of 0.1:

$$\hat{y} = W_2 \cdot \text{GELU}(\text{Drop}(W_1 \cdot h_{\text{asp}} + b_1)) + b_2 \quad (2)$$

where h_{asp} is the pooled aspect representation and the hidden size is 256. In preliminary experiments, separate heads consistently outperformed a joint 2-output head. The model is trained with MSE loss averaged over the two dimensions:

$$\mathcal{L} = \frac{1}{2}\Big(\text{MSE}(\hat{V}, V) + \text{MSE}(\hat{A}, A)\Big) \quad (3)$$

During inference, predictions are clipped to $[1.0, 9.0]$ and rounded to two decimal places.

3.5 K-Fold Ensemble

To improve prediction stability, we apply 5-fold cross-validation during training. At inference time, we average the predictions from all five fold models as the final output. This ensemble strategy reduces variance and consistently improves over single-model performance across all language–domain settings.

4 Experimental Setup

4.1 Data

For Track A (Lee et al., 2026), we use English (restaurant, laptop) and Chinese (restaurant, laptop, finance) datasets. For Track B (Becker et al., 2026), we use English and Chinese environmental protection datasets. All data is provided in JSONL format with train/dev/test splits.

4.2 Training Details

We fine-tune with the HuggingFace Transformers library (Wolf et al., 2020), using differential learning rates: 2×10^{-5} for the pre-trained encoder and 1×10^{-4} for the randomly initialised regression heads. We use a cosine learning rate scheduler with 10% warmup, weight decay 0.01 (Loshchilov and Hutter, 2019), per-device batch size 16 with gradient accumulation of 2 steps (effective batch size 32), maximum 30 epochs, and FP16 mixed precision. Early stopping with patience 3 monitors validation RMSE_{VA}; maximum gradient norm is clipped to 1.0.

4.3 Model Selection

For Track A, we evaluated six single models and three cross-model ensembles on the development set. For Track B, we compared four models of different sizes. Models are selected based on development set RMSE_{VA}.

5 Results

5.1 Track A: DimABSA — Subtask 1

Table 1 presents development-set results for Track A across different backbone models and cross-model ensembles; we treat this comparison as our main *ablation* over the choice of pre-trained encoder (see Section 5.3). XLM-RoBERTa-large achieves the best overall performance on English subsets, while Chinese-specific models (MacBERT-large, Chinese-RoBERTa-large) provide marginally better scores on some Chinese domains. Cross-model ensembles (averaging predictions from different backbones) are competitive but do not consistently outperform the best single model.

Model	E-Res	E-Lap	C-Res	C-Lap	C-Fin
Single Models					
XLM-R-L	**1.029**	**0.969**	0.711	0.678	0.503
DeBERTa-v3-L	1.064	1.007	0.782	0.723	0.528
MacBERT-L	1.284	1.190	0.734	0.658	0.514
C-RoBERTa-L	1.321	1.294	0.728	0.658	0.525
CKIP-BERT	1.509	1.362	0.779	0.740	0.580
Cross-Model Ensembles					
Mac.+XLM-R	1.044	1.026	0.706	**0.639**	**0.494**
C-RoB.+XLM-R	1.054	**0.969**	**0.700**	0.649	0.503
Mac.+C-RoB.	1.259	1.190	0.719	0.644	0.511

Table 1: Track A Subtask 1 dev $\mathrm{RMSE}_{\mathrm{VA}}$ (↓). Abbrev.: E/C = English/Chinese, Res = Restaurant, Lap = Laptop, Fin = Finance, L = Large.

System	E-Res	E-Lap	C-Res	C-Lap	C-Fin
Ours	1.277	1.346	0.943	0.687	0.486
rank	*14*	*11*	*4*	*8*	*2*
Kimi-K2	2.146	2.189	1.896	1.644	1.965
Qwen-3 14B	2.643	2.809	2.007	1.771	1.471

Table 2: Track A Subtask 1 test $\mathrm{RMSE}_{\mathrm{VA}}$ (↓) and our position on the unofficial post-evaluation leaderboard. Our system uses XLM-RoBERTa-large; baselines are organizer-provided LLMs.

Table 2 reports the official test-set $\mathrm{RMSE}_{\mathrm{VA}}$ together with our rank on the unofficial post-evaluation leaderboard for each language–domain pair.[1] Our submitted system (XLM-RoBERTa-large) achieves substantially lower $\mathrm{RMSE}_{\mathrm{VA}}$ than both LLM-based baselines on every domain. Among other participants, the system is more competitive on Chinese (rank 4/8/2 across `zho-res`/`zho-lap`/`zho-fin`) than on English (rank 14/11 on `eng-res`/`eng-lap`).

5.2 Track B: DimStance — Subtask 1

Table 3 reports the development set results for Track B. mDeBERTa-v3-base achieves the best performance on both English and Chinese environmental protection datasets. Larger models (XLM-R-large, InfoXLM-large, RemBERT) surprisingly underperform, possibly due to overfitting on the smaller stance datasets.

Table 4 shows the official test-set $\mathrm{RMSE}_{\mathrm{VA}}$ together with our unofficial leaderboard ranks. Our system outperforms both the Mistral-3 14B and mBERT baselines on both languages, with a particularly large margin on Chinese, where it ranks 1st on the unofficial leaderboard.

[1]Ranks are taken from the unofficial post-evaluation leaderboard published by the organizers; team identifier: YangS_team. The leaderboard excludes the two organizer baselines.

Model	E-Env	C-Env
mDeBERTa-v3-B	**2.062**	**0.638**
XLM-R-L	2.291	0.646
InfoXLM-L	2.414	0.652
RemBERT	2.473	0.987

Table 3: Track B Subtask 1 dev set $\mathrm{RMSE}_{\mathrm{VA}}$ (↓). E-Env = English Environmental Protection, C-Env = Chinese Environmental Protection. B = Base, L = Large.

System	E-Env	C-Env
Ours	1.573	0.547
rank	*5*	*1*
Mistral-3 14B	1.643	0.740
mBERT	2.699	1.276

Table 4: Track B Subtask 1 test $\mathrm{RMSE}_{\mathrm{VA}}$ (↓) and our position on the unofficial post-evaluation leaderboard. Our system uses mDeBERTa-v3-base; baselines are organizer-provided.

5.3 Ablation Study

Because our system is a composition of standard components, we use the development-set comparisons in Tables 1 and 3 as a controlled ablation over the two design choices that most affect submitted scores: the *backbone encoder* and *cross-model ensembling*. All other settings (aspect-marker encoding, dual regression heads, training hyperparameters, 5-fold ensembling within a single backbone) are held fixed.

Backbone (Track A). Replacing XLM-RoBERTa-large with a Chinese-specific encoder degrades English $\mathrm{RMSE}_{\mathrm{VA}}$ sharply (e.g., E-Res 1.029 → 1.284 for MacBERT-large, → 1.321 for Chinese-RoBERTa-large, → 1.509 for CKIP-BERT) while only marginally changing Chinese scores. Conversely, switching to DeBERTa-v3-large worsens Chinese subsets (e.g., C-Res 0.711 → 0.782). XLM-RoBERTa-large is therefore the most robust single backbone across all five Track A settings.

Cross-Model Ensembling (Track A). Averaging XLM-RoBERTa-large with a Chinese-specific encoder slightly improves Chinese subsets (best C-Res 0.700 vs. 0.711; best C-Lap 0.639 vs. 0.678; best C-Fin 0.494 vs. 0.503) but does *not* reliably help English (E-Res 1.044 vs. 1.029, E-Lap 0.969 unchanged or worse). Combining two Chinese-only encoders is dominated by the XLM-R-containing ensembles. Given the inconsistent gain and roughly doubled inference cost, we submit the single XLM-RoBERTa-large model for Track A.

Backbone (Track B). For DimStance, the smaller mDeBERTa-v3-base outperforms all three larger candidates we tried on *both* languages (E-Env 2.062 vs. 2.291 for XLM-R-large, 2.414 for InfoXLM-large, 2.473 for RemBERT; C-Env 0.638 vs. 0.646/0.652/0.987). The Track B training sets are smaller and noisier than Track A's, and larger encoders appear to overfit. This is the main empirical reason we use different backbones in the two tracks.

Pilot runs also showed that aspect-marker pooling improved over `[CLS]` pooling and that two separate regression heads outperformed a single 2-output head; these were used to set defaults and are not the focus of the reported ablation.

5.4 Analysis

Domain and Language Gap. Chinese domains consistently show lower RMSE_{VA} than English ones on both dev and test sets. This may reflect differences in annotation variance or text complexity. The Finance domain in Track A has the lowest test RMSE_{VA} (0.486), indicating potentially more predictable VA distributions in financial texts. For Track B, English environmental protection (1.573) is notably harder than Chinese (0.547), perhaps due to more diverse and polarised stances expressed in English texts.

Dev–Test Consistency. Comparing Tables 1 and 2, test-set RMSE_{VA} is consistently higher than dev-set RMSE_{VA} for English (E-Res $1.029 \rightarrow 1.277$; E-Lap $0.969 \rightarrow 1.346$) but remains comparable for Chinese (C-Fin $0.503 \rightarrow 0.486$; C-Lap $0.678 \rightarrow 0.687$). Interestingly, Track B presents a different trend, where both English and Chinese models yield lower RMSE_{VA} on the test set(E-Env $2.062 \rightarrow 1.573$; C-Env $0.638 \rightarrow 0.547$). The English laptop domain shows the largest degradation, suggesting that the test set may contain more challenging or out-of-distribution aspect–text pairs for this domain.

Fine-Tuning vs. LLM Prompting. Our fine-tuned encoders substantially outperform the LLM baselines (Kimi-K2, Qwen-3 14B, Mistral-3 14B) on every test set, suggesting that for fine-grained regression tasks like VA prediction, task-specific fine-tuning of smaller encoder models is more effective than zero/few-shot prompting of larger decoder-only LLMs, likely because continuous VA scores require calibrated numerical outputs that benefit from direct supervised training.

Model Capacity vs. Data Size. An interesting finding across both tracks is that larger models do not always yield better results. In Track B, mDeBERTa-v3-base (86M parameters) consistently outperforms XLM-RoBERTa-large (550M) and other large models. We attribute this to the relatively small size of the stance datasets: with fewer training examples, the additional parameters of larger models become a liability, leading to overfitting. This observation reinforces the importance of matching model capacity to the available training data, rather than defaulting to the largest available pre-trained model.

6 Conclusion

We described our submission to SemEval-2026 Task 3 (Track A and Track B Subtask 1, English and Chinese). The contribution is empirical rather than methodological: we combine standard ingredients—multilingual transformers, aspect-marker encoding, dual regression heads, and 5-fold ensembling—and report controlled comparisons. Our ablation shows that XLM-RoBERTa-large is the most robust single encoder for Track A, cross-model ensembling helps only marginally on Chinese subsets, and the smaller mDeBERTa-v3-base outperforms larger encoders on Track B. The submission beats all organizer baselines on every test set, ranks 1st on `zho-env` (Track B) and 2nd on `zho-fin` (Track A) on the unofficial leaderboard, and is more competitive on Chinese than on English overall. Future work includes exploring cross-lingual transfer strategies and multi-task learning across subtasks.

Limitations

Our system has several limitations. First, we only participate in Subtask 1 (regression) and do not address the extraction subtasks (Subtasks 2 and 3), which require identifying aspect terms and their boundaries. Second, our model selection was conducted primarily on the development set, which may not fully represent the test distribution—as evidenced by the dev–test gap on certain English domains. Third, while our backbone comparison covers multiple pre-trained encoders, we do not provide a full controlled ablation isolating the individual contributions of the core design choices—namely aspect-marker pooling versus standard [CLS] pooling, dual regression heads versus a single joint 2-output head, and 5-fold ensembling versus single-model inference. We leave such a full design ablation as future work. Finally, we only evaluate on English and Chinese, leaving other languages in the shared task unexplored.

References

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.

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. Association for Computational Linguistics.

Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, Shijin Wang, and Guoping Hu. 2020. Revisiting pre-trained models for Chinese natural language processing. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 657–668. 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 *Proceedings of the 11th International Conference on Learning Representations*.

Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. 2021. DeBERTa: Decoding-enhanced BERT with disentangled attention. In *Proceedings of the 9th International Conference on Learning Representations*.

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. 2026. DimABSA: Building multilingual and multidomain datasets for dimensional aspect-based sentiment analysis. *Preprint*, arXiv:2601.23022.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. In *Proceedings of the 7th International Conference on Learning Representations*.

Saif M. 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 (SemEval-2018)*, pages 1–17. Association for Computational Linguistics.

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeny Kotelnikov, Nuria Bel, Salud María Jiménez-Zafra, and Gülşen Eryiğit. 2016. SemEval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30. Association for Computational Linguistics.

James A. Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, and 3 others. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45. Association for Computational Linguistics.

Liang-Chih Yu, Jonas Becker, Shamsuddeen Hassan Muhammad, Idris Abdulmumin, Lung-Hao Lee, Ying-Lung Lin, Jin Wang, Jan Philip Wahle, Terry Ruas, Alexander Panchenko, Ilseyar Alimova, Kai-Wei Chang, Lilian Wanzare, Nelson Odhiambo, Bela Gipp, and Saif M. Mohammad. 2026. SemEval-2026 task 3: Dimensional aspect-based sentiment analysis (DimABSA). In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*. Association for Computational Linguistics.

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(11):11019–11038.