wangkongqiang at SemEval-2026 Task 10: PsyCoMark - Psycholinguistic Conspiracy Marker Extraction and Detection

Kongqiang Wang[1]**, Qingli Tan**[2]
[1]School of Information Science and Engineering, Yunnan University,
[2]School of College of Ecology and Environment, Yunnan University,
Kunming 650500, Yunnan, China
[1]wangkongqiang60@gmail.com, [2]tanqingli@stu.ynu.edu.cn

Abstract

This paper presents our system developed for the SemEval-2026 Task 10: PsyCoMark - Psycholinguistic Conspiracy Marker Extraction and Detection. on Subtask 1: Conspiracy Marker Extraction. on Subtask 2: Conspiracy Detection. To this end, we focus on English language use four different pre-trained languages models: models–distilbert–distilbert-base-uncased, models–distilbert–distilbert-base-multilingual-cased, models–lxyuan–distilbert-base-multilingual-cased-sentiments-student, and models–microsoft–deberta-v3-base. We experiment with (1) the training set data is analyzed visually, (2) use the gemma-3-27b-it generative model to perform data augmentation on the training dataset through prompts for Subtask 2: Conspiracy Detection, and (3) multiple numbers of single models are trained on the training set data. We further study the influence of different hyperparameters on the single model and select the best single model for the prediction of the test set. Our submission achieved the good ranking place in the test set leaderboard. For Subtask 1, the evaluation criteria for this task mainly consist of the aggregate results of the four markers: Actor, Action, Effect, and Victim, and they are measured using the Macro F1 score. For Subtask 2, this task is essentially a binary classification task for text. Performance will be evaluated using macro-averaged F1 score. In other words, this subtask evaluated using Weighted F1 score across different sentences and cultural contexts. For Subtask 1 and Subtask 2, our best approach is to obtain the results are Macro F1 score 0.1587 and Weighted F1 score 0.7411 separately. For the final ranking, Semeval organizers will use the aggregate results of Macro F1 score and Weighted F1 score. Even so, our approach has yielded good results.

1 Introduction

The official organizers of SemEval held SemEval 2026 (Ghosh et al., 2026) - PsyCoMark (Samory et al., 2026) in Task 10 in the first half of 2026. The full name of this task is "Psycholinguistic Conspiracy Marker Extraction and Detection". The purpose of this task and the ideal goal to be achieved are: To understand conspiracy thinking at its core. In the other words, understand conspiracy thinking at its roots. PsyCoMark brings together psychology and NLP to uncover how conspiracy theories are expressed in everyday conversations. This task introduces a novel dataset and challenges participants to go beyond topical detection: can your model recognize the structure of conspiratorial thought?

Online conspiracy thinking opinion is the sharp division and hostility between social, political, or identity groups. Online conspiracy thinking opinion has become a growing concern, as it often precedes hate speech, offensive discourse, and social fragmentation. In its extreme form, conspiracy can create a fragmented society where individuals or groups are unable to engage in constructive dialogue, leading to a breakdown in community cohesion and social unity. Thus, detecting and marker extraction conspiracy before it escalates is crucial to ensure safer and more inclusive online spaces (Muhammad et al., 2025b). For the first time, Semeval organizers introduce a PsyCoMark task: Psycholinguistic Conspiracy Marker Extraction and Detection, aimed at detection of online conspiracy. The task focuses on the identification of multicultural and multievent conspiracy, capturing the complexity of online discourse across diverse contexts. Participants may participate in one or more of the following two sub-tasks: Conspiracy Marker Extraction and Conspiracy Detection. To be more specific, we are encouraging to join them in tackling two synergistic subtasks. **Subtask 1**: Extract psycholinguistic markers of conspiracy

Proceedings of the The 20th International Workshop on Semantic Evaluation (2026), pages 8–14

thinking (e.g., Actors, Victims, Effects). **Subtask 2**: Detect whether a Reddit comment expresses a conspiracy belief. Based on the predict and extraction task background of conspiracy comment text, we propose the data augmentation learning method based on pre-trained language model for detecting conspiracy content.

We developed for the SemEval-2026 Task 10: Psycholinguistic Conspiracy Marker Extraction and Detection. on Subtask 1: Identify spans of text that express core conspiracy markers grounded in evolutionary psychology. Each document may include zero or multiple overlapping spans. Evaluation will use an overlap-based macro F1 score for each marker. The following provides a detailed explanation of how the five main markers are used for the extraction of psycholinguistic conspiracy markers.

- **Actor** – Mentions of individual or group agents.
- **Action** – Descriptions of what the actor is doing.
- **Effect** – Consequences of the actions.
- **Victim** – Individuals or groups being harmed.
- **Evidence** – Claims or proof used to support the theory.

on Subtask 2: Conspiracy Detection. Classify Reddit comments as either conspiracy-related or not conspiracy-related. The task is designed to benefit from the extracted markers but can be approached independently. Performance will be evaluated using macro-averaged F1 score. The code of our experiment method is available on our GitHub website.[1]

2 Related Work

SemEval in previous years has introduced tasks focusing on multi-label text classification and text binary classification (Wang et al., 2024; Su and Zhou, 2024; Tran and Tran, 2024; Brekhof et al., 2024) to evaluate internal potential elements and potential content of the text comment. These tasks provided datasets with human labeled similarity scores, which have been extensively utilized for training sentence embedding models and conducting semantic evaluations.

[1]https://github.com/WangKongQiang/Semeval-2026-Task10

2.1 Sentence Embeddings

Word embedding models such as BERT, GloVe, RoBeRTa and Word2Vec are frequently employed to assess the semantic distance between words. They are also some of the more commonly used methods in text classification tasks (Muhammad et al., 2025a). We focus on English language use four different pre-trained languages models: models–distilbert–distilbert-base-uncased, models–distilbert–distilbert-base-multilingual-cased, models–lxyuan–distilbert-base-multilingual-cased-sentiments-student, and models–microsoft–deberta-v3-base. Mainly, these pre-trained models can generate better sentence embeddings.

Sentence embeddings with a fixed length are often generated via mean/max pooling of word embeddings or employing CLS embedding in BERT. The semantic distances are commonly measured using the cosine similarity of embeddings of two expressions. Siamese or triplet network architectures are frequently employed in sentence embedding training. For example, models such as Sentence-BERT utilize a dual-encoder architecture with shared weights for predicting sentence relationships (e.g., semantic contradiction, entailment, or neutral labeling) (Belay et al., 2025) or for similarity score prediction using regression objectives, e.g., the difference between human annotated similarity score (sim) of two sentences and the cosine of two sentence embeddings.

2.2 Data Augmentation Learning

In previous studies (Brekhof et al., 2024; Tran and Tran, 2024), data augmentation learning presents several advantages. The data augmentation approach can reduce the errors from insufficient examples by expanding the learning of the text content or can make the system more robust. In our study, using the gemma-3-27b-it generative model to perform data augmentation on the training dataset through prompts to generate large amounts of similar instances while making use of information from the large data content during model pre-training. It has demonstrated that data augmentation learning can achieve remarkable success in our experiment.

In our study, we aim to use multiple pre-train learning models to assess semantic relatedness. When models are trained on diverse datasets with different architectures, they may produce varied predictions on semantic relatedness, and fine-tune them using different hyperparameters may improve final performance. We use sentence embeddings mainly from the following models: models–distilbert–distilbert-base-uncased, models–distilbert–distilbert-base-multilingual-cased, models–lxyuan–distilbert-base-multilingual-cased-sentiments-student, and models–microsoft–deberta-v3-base.

3 Methodology

3.1 Overall Architecture

For Subtask 1, we respectively used the following three different versions of pre-trained models as the backbone system model to perform marker extraction: models–distilbert–distilbert-base-uncased, models–distilbert–distilbert-base-multilingual-cased, and models–lxyuan–distilbert-base-multilingual-cased-sentiments-student. This method using simplified logic mainly involves reconstructing the character spans for the five main conspiracy indicators: Actor, Action, Effect, Victim, and Evidence. Finally, we aggregate extracts the results for each marker span and save the final aggregated predictions in JSONL format.

For Subtask 2, the pursued approach involves using a data augmentation system composed of the gemma-3-27b-it generative model and then training four different pre-train transformers. We trained several state-of-the-art natural language processing (NLP) models on a large dataset of annotated Reddit comments to create four separate of classifiers with different architectures and configurations. We then submission the predictions of these classifiers using a data augmentation system to produce the final predictions. We have used the following pre-train transformers for the predictions: models–distilbert–distilbert-base-uncased, models–distilbert–distilbert-base-multilingual-cased, models–lxyuan–distilbert-base-multilingual-cased-sentiments-student, and models–microsoft–deberta-v3-base.

For each instance, the final classification decision is based on the independent of each other of outputs of these models. The novel data augmentation system presented to improved the prediction of the pre-trained model to downstream tasks.

3.2 Implementation Step

For Subtask 1: Conspiracy Marker Extraction, our experiment mainly involes these steps:

First, pull the pre-trained model that needs to be used, get from HuggingFace. The pre-trained model can be used from the official website of HuggingFace[2].

Second, by loading a new model for each marker type, one-span training is conducted separately for each conspiracy-related type.

Third, apply each trained one-span marker extraction model to perform inference. Specifically, use simplified logic to reconstruct spans for this specific marker type. By aggregating the prediction results from all models, the predicted five main conspiracy markers were obtained and saved in JSONL format.

For Subtask 2: Conspiracy Detection, our experiment mainly involves these steps:

First, set the API key, get from Google Gemini. The API key can be applied from the official website of Google AI Studio[3].

Second, since the content of the text column provided by the train dataset is the information that the model requires as training input content, they are extracted separately, and the conspiracy column content was retained as the label. By invoking the API key of Google AI Studio using the gemma-3-27b-it generative model to perform data augmentation on the training dataset through prompts to generate large amounts of English comment texts, which is as similar as original text. Only the content of the text column has been changed. The conspiracy label of this newly generated instance remains unchanged. That is to say, it maintains its characteristic of being a text label unchanged.

Third, the transformers Python library that will be used below requires the data to be presented in a specific form. The following data rows adapt each split to contain only two columns: text and conspiracy labels, where the text is an array equal in size to the number of conspiracy labels.

Fourth, model's definition. In this section, the different pre-train transformers model that will be evaluated are separated. For this purpose, the implementation mainly relies in Trainer and TrainingArguments functions of the transformers Python library, which allows to train and test transformers within few steps.

Fifth, training. Each of the aforementioned models is trained separatedly with the entire training set or an expanded data augmentation dataset. This training is directly performed in the already defined

[2]https://huggingface.co/

[3]https://aistudio.google.com/app/api-keys

parameters for convenience.

Sixth, evaluate our method. Firstly, each pre-train transformer model is individually evaluated using the development split dataset. Subsequently, the main evaluation metrics (accuracy, F1 score, precision and recall) are stored. Secondly, the predictions of each model for the development dataset instances are derived. After calculating their metrics, it is possible to determine which model obtained the best Macro-averaged F1 score. This will be the final model used for the test dataset. Regarding the models' inputs, these are obtained through a data augmentation system. After computing the output that each of the models produces for given instances, save the results of the development dataset to a CSV file.

Seventh, selecting the best model. Once the predicted labels for each development instances are calculated for each pre-train transformer model, their metrics can be computed. In brief, it is a binary classification task, the best model will be that with a maximum Weighted F1 score.

Eighth, gain the predictions on test dataset. Finally, the model which obtained a higher Weighted F1 score can be used to predict the label of each test instance as final test result.

Further, these results will be used to portray some evaluation plots, including the confusion matrix and the ROC curve.

4 Results and Analysis

4.1 Training Dataset Analysis

The text length and conspiracy label of training set are described in Table 1. Labels here only analyzes Subtask 2: Conspiracy Detection. The rows with the conspiracy label set to "Can't tell" will be ignored. The length and quantity distribution of training text data are analyzed in Figure 1. Distribution of the size of texts for each class in Figure 2. It shows the number of percentages relative to each class for various text size cases. classes here only analyze Subtask 2: Conspiracy Detection. The rows with the conspiracy label set to "Can't tell" will be ignored.

4.2 Experimentation Configuration

For the sake of completeness and in an attempt to improve the results obtained by the transformer pre-train model. For models–distilbert–distilbert-base-uncased, models–distilbert–distilbert-base-multilingual-cased, models–lxyuan–distilbert-base-multilingual-cased-sentiments-student, and models-microsoft-deberta-v3-base, the two different hyperparameters were used respectively: See Table 2.

Table 1: The text data experiment situation and the number of labels for Subtask 2: Conspiracy Detection are described.

<table>
<tr><th>Training Set Text</th><th>Value</th></tr>
<tr><td>|count|</td><td>3531.000000</td></tr>
<tr><td>|mean|</td><td>80.064854</td></tr>
<tr><td>|std|</td><td>39.882569</td></tr>
<tr><td>|min|</td><td>24.000000</td></tr>
<tr><td>|25%|</td><td>48.000000</td></tr>
<tr><td>|50%|</td><td>69.000000</td></tr>
<tr><td>|75%|</td><td>104.000000</td></tr>
<tr><td>|max|</td><td>276.000000</td></tr>
</table>

<table>
<tr><th>Training Set Conspiracy Label</th><th>Count</th></tr>
<tr><td>|Yes|</td><td>1541</td></tr>
<tr><td>|No|</td><td>1990</td></tr>
<tr><td>|Can't tell|</td><td>785</td></tr>
</table>

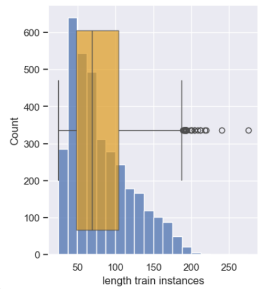

Figure 1: The length and quantity distribution of training text data are analyzed.

4.3 Development Dataset Result

The following Table 3 records the official results of SemEval-2026 Task 10: Psycholinguistic Conspiracy Marker Extraction and Detection, on Subtask 1: Conspiracy Marker Extraction, and Subtask 2: Conspiracy Detection. The metrics recorded by black bold text is the best (winning) approach in the evaluation task of the development set for English language.

4.4 Test Dataset Result

The following Table 4 records the official results of SemEval-2026 Task 10: Psycholinguistic Conspiracy Marker Extraction and Detection, on Subtask 1: Conspiracy Marker Extraction, and Subtask 2: Conspiracy Detection. The metrics recorded by black bold text is the best (winning) approach in the evaluation task of the test set for English lan-

Table 2: Experimentation configuration hyperparameters for the four different transformer pre-train models.

<table>
<tr><th>Model</th><th>Hyperparameter</th><th>Values</th></tr>
<tr><td>models–distilbert–distilbert-base-uncased</td><td>|batch_size|</td><td>16</td></tr>
<tr><td>models–distilbert–distilbert-base-multilingual-cased</td><td>|learning_rate|</td><td>2e-5</td></tr>
<tr><td>models–lxyuan–distilbert-base-multilingual-cased-sentiments-student</td><td>|num_epochs|</td><td>10</td></tr>
<tr><td></td><td>|weight_decay|</td><td>0.01</td></tr>
<tr><td>models-microsoft-deberta-v3-base</td><td>|num_train_epochs|</td><td>10</td></tr>
<tr><td></td><td>|learning_rate|</td><td>2e-5</td></tr>
<tr><td></td><td>|per_device_train_batch_size|</td><td>32</td></tr>
<tr><td></td><td>|warmup_steps|</td><td>100</td></tr>
<tr><td></td><td>|weight_decay|</td><td>0.01</td></tr>
<tr><td></td><td>|logging_steps|</td><td>100</td></tr>
<tr><td></td><td>|save_total_limit|</td><td>1</td></tr>
</table>

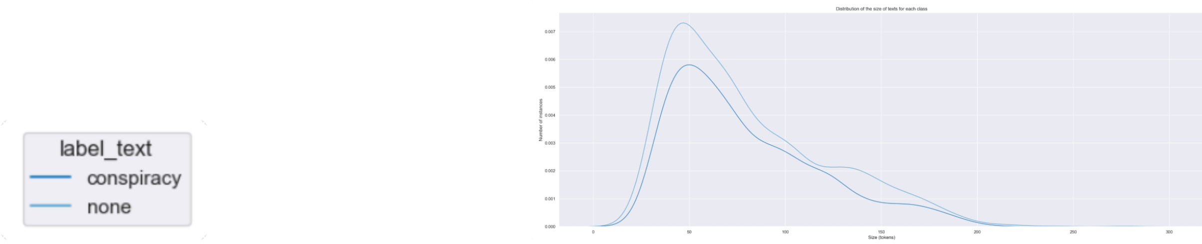

Figure 2: Distribution of the size of texts for each class. The horizontal axis represents the size of the comments text, while the vertical axis represents the percentage proportion corresponding to the size of the text.

Table 3: The development dataset experiment situation detailed results are described.

Subtask	Language	Main Technologies	F1-Score (Agg/Micro)	F1-Score (Macro)
1	|English|	models–distilbert–distilbert-base-uncased	0.1818	0.1597
1	|English|	**models–distilbert–distilbert-base-multilingual-cased**	**0.1913**	**0.1762**
1	|English|	models–lxyuan–distilbert-base-multilingual-cased-sentiments-student	0.1748	0.1602
Subtask	**Language**	**Main Technologies**	**Weighted F1 Score**	**Accuracy**
2	|English|	blind guesses are all "conspiracy": "No"	0.5113	0.6494
2	|English|	models–distilbert–distilbert-base-uncased	0.7546	0.7662
2	|English|	models–distilbert–distilbert-base-multilingual-cased	0.7000	0.7013
2	|English|	models–lxyuan–distilbert-base-multilingual-cased-sentiments-student	0.6766	0.6753
2	|English|	**models–microsoft–deberta-v3-base**	**0.7582**	**0.7662**
2	|English|	models–microsoft–deberta-v3-base + gemma-3-27b-it	0.7546	0.7662

guage.

4.5 Text Words and Labels Biased Performance

From Figure 1 of the visual analysis, we can observe that 75% of Reddit comments in training set data, either in the chart or in the previous model input column content, have no more than 100 words. This information could be useful in determining the size of a network of neurons, or when a sentence length limit needs to be set.

SemEval-2026 Task 10: Psycholinguistic Conspiracy Marker Extraction and Detection, on Subtask 1: Conspiracy Marker Extraction. This task is essentially a named entity recognition task, which requires the extraction of psycholinguistic conspiracy markers separately. Distribution of markers count per data row, as show in Figure 3. The developed model needs to extract five psycholinguistic conspiracy markers. They are respectively named Actor, Action, Effect, Victim, and Evidence. The overall quantity statistics of each type in markers for training set, see Table 5. A correct marker extraction requires the presence of the following components: $startIndex$, $endIndex$, $type$, and $text$. In the training set, there are significant differences in the occurrence frequencies of the five types of annotation elements. Among them, the number of actors is the largest (6,416 markers), indicating that conspiracy speech usually centers around a clear actor. Action comes second (4841 markers), reflecting a high coverage of behavioral expression. The number of effect and evidence is similar, indicating that the annotation of behavioral

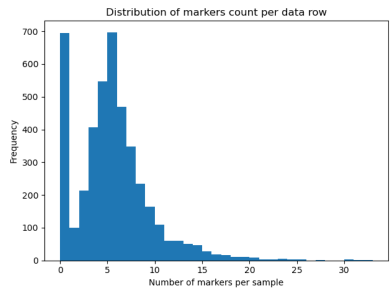

Figure 3: The markers quantity distribution of training markers data are analyzed.

consequences and evidence in the data is relatively balanced. The number of victims is the smallest (3315 markers), indicating that the victims in some texts are implicit.

on Subtask 2: Conspiracy Detection. This task is binary-label classification. Each instance can have only one category, and you need to predict which category each instance belongs to. In the specific cases (English language) we focus on, there may be up to three different conspiracy categories: Yes, No, and $Can'ttell$. For this task, we will look at the quantity distribution followed by each conspiracy category, as shown in Table 1. The rows with the conspiracy label set to "Can't tell" will be ignored. In this case, percentages can be assessed because of the labels scattered. Based on the analysis of the valid conspiracy labels in the training set, it is found that approximately 43.6% of the comment texts contain the conspiracy words, while the comment texts without conspiracy account for about 56.4%. Overall, the labels are roughly balanced.

Table 4: The test dataset experiment situation detailed results are described.

<table>
<tr><th>Subtask</th><th>Language</th><th>Main Technologies</th><th>F1-Score(Agg/Micro)</th><th>F1-Score (Macro)</th></tr>
<tr><td>1</td><td>|English|</td><td>models–distilbert–distilbert-base-uncased</td><td>0.1718</td><td>0.1587</td></tr>
<tr><td>1</td><td>|English|</td><td>models–distilbert–distilbert-base-multilingual-cased</td><td>0.1585</td><td>0.1462</td></tr>
<tr><td>1</td><td>|English|</td><td>models–lxyuan–distilbert-base-multilingual-cased-sentiments-student</td><td>0.1595</td><td>0.1476</td></tr>
<tr><th>Subtask</th><th>Language</th><th>Main Technologies</th><th>Weighted F1 Score</th><th>Accuracy</th></tr>
<tr><td>2</td><td>|English|</td><td>blind guesses are all "conspiracy": "No"</td><td>0.3773</td><td>0.5388</td></tr>
<tr><td>2</td><td>|English|</td><td>models–distilbert–distilbert-base-uncased</td><td>0.7129</td><td>0.7171</td></tr>
<tr><td>2</td><td>|English|</td><td>models–distilbert–distilbert-base-multilingual-cased</td><td>0.6800</td><td>0.6822</td></tr>
<tr><td>2</td><td>|English|</td><td>models–lxyuan–distilbert-base-multilingual-cased-sentiments-student</td><td>0.6380</td><td>0.6408</td></tr>
<tr><td>2</td><td>|English|</td><td>models–microsoft–deberta-v3-base</td><td>0.7267</td><td>0.7313</td></tr>
<tr><td>2</td><td>|English|</td><td>models–microsoft–deberta-v3-base + gemma-3-27b-it</td><td>0.7411</td><td>0.7429</td></tr>
</table>

Table 5: The overall quantity statistics of each type in markers for Subtask 1: Conspiracy Marker Extraction training set are described.

<table>
<tr><th>Markers Type</th><th>Count</th></tr>
<tr><td>|Actor|</td><td>6416</td></tr>
<tr><td>|Action|</td><td>4841</td></tr>
<tr><td>|Effect|</td><td>3739</td></tr>
<tr><td>|Evidence|</td><td>3654</td></tr>
<tr><td>|Victim|</td><td>3315</td></tr>
</table>

5 Conclusion

Our system employs a fine-tuning of pre-trained models approach to estimate semantic relatedness (Agirre et al., 2014), acquiring results from multiple pre-train models: models–distilbert–distilbert-base-uncased, models–distilbert–distilbert-base-multilingual-cased, models–lxyuan–distilbert-base-multilingual-cased-sentiments-student, and models-microsoft-deberta-v3-base. The hyperparameter is following: batch_size is 16, num_epochs is 10, learning_rate is 2e-5, weight_decay is 0.01. The dataset usage is shown in Table 6. Our findings suggest that semantic relatedness can be deduced from a variety of sources. Although some features (e.g., lexical overlap ratio, and text length) may not perform strongly in models specifically designed to obtain sentence representations. The results demonstrate that these features can outperform many individual state-of-the-art systems and achieve a better correlation with human judgment on semantic relatedness (Siino, 2024) when used in a fine-tuning pre-train model manner.

6 Limitation and Future Work

Our experiments are based on English language data sets only. Constrained by the size of the training data and the availability of pre-trained language models, it is regrettable that we did not offer insights into other Asian and African languages for Psycholinguistic Conspiracy Marker Extraction and Detection task. In future research, studies on low-resource languages (Vaidya et al., 2024) will be valuable, including tasks such as data collection, annotation, and pre-training models tailored to these languages.

Acknowledgments

We are very grateful for the assistance and discussions provided by Semeval-2026 Task10: Psycholinguistic Conspiracy Marker Extraction and Detection leaders and organizers.

References

Eneko Agirre, Carmen Banea, Claire Cardie, Daniel Cer, Mona Diab, Aitor Gonzalez-Agirre, Weiwei Guo, Rada Mihalcea, German Rigau, and Janyce Wiebe. 2014. Semeval-2014 task 10: Multilingual semantic textual similarity. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 81–91, Dublin, Ireland. Association for Computational Linguistics.

Tadesse Destaw Belay, Israel Abebe Azime, Abinew Ali Ayele, Grigori Sidorov, Dietrich Klakow, Philip Slusallek, Olga Kolesnikova, and Seid Muhie Yimam. 2025. Evaluating the capabilities of large language models for multi-label emotion understanding. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 3523–3540, Abu Dhabi, UAE. Association for Computational Linguistics.

Thijs Brekhof, Xuanyi Liu, Joris Ruitenbeek, Niels Top, and Yuwen Zhou. 2024. Groningen team D at SemEval-2024 task 8: Exploring data generation and a combined model for fine-tuning LLMs for multidomain machine-generated text detection. In *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*, pages 391–398, Mexico City, Mexico. Association for Computational Linguistics.

Debanjan Ghosh, Kai North, Ekaterina Kochmar, Mamoru Komachi, and Marcos Zampieri, editors. 2026. *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*. Association for Computational Linguistics, San Diego, United States.

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, Guanqun Cao,

Table 6: Use dataset supported by Semeval-2026 Task 10: Psycholinguistic Conspiracy Marker Extraction and Detection, on Subtask 1: Conspiracy Marker Extraction and Subtask 2: Conspiracy Detection. The style is based on raw data.

<table>
<tr><th>Dataset Input</th><th>Description</th><th>Use or Not</th></tr>
<tr><td>PsyCoMark official dataset</td><td>|datasets for English language.|</td><td>yes</td></tr>
<tr><td>other dataset</td><td>|use external or additional corpora.|</td><td>no</td></tr>
</table>

Rodrigo Tufino Cardenas, Rendi Chevi, Chiamaka Ijeoma Chukwuneke, Alexandra Ciobotaru, Daryna Dementieva, Murja Sani Gadanya, Robert Geislinger, Bela Gipp, Oumaima Hourrane, Oana Ignat, Falalu Ibrahim Lawan, Rooweither Mabuya, Rahmad Mahendra, Vukosi Marivate, Andrew Piper, Alexander Panchenko, Charles Henrique Porto Ferreira, Vitaly Protasov, Samuel Rutunda, Manish Shrivastava, Aura Cristina Udrea, Lilian Diana Awuor Wanzare, Sophie Wu, Florian Valentin Wunderlich, Hanif Muhammad Zhafran, Tianhui Zhang, Yi Zhou, and Saif M. Mohammad. 2025a. Brighter: Bridging the gap in human-annotated textual emotion recognition datasets for 28 languages. *Preprint*, arXiv:2502.11926.

Shamsuddeen Hassan Muhammad, Nedjma Ousidhoum, Idris Abdulmumin, Seid Muhie Yimam, Jan Philip Wahle, Terry Ruas, Meriem Beloucif, Christine De Kock, Tadesse Destaw Belay, Ibrahim Said Ahmad, Nirmal Surange, Daniela Teodorescu, David Ifeoluwa Adelani, Alham Fikri Aji, Felermino Ali, Vladimir Araujo, Abinew Ali Ayele, Oana Ignat, Alexander Panchenko, Yi Zhou, and Saif M. Mohammad. 2025b. SemEval task 11: Bridging the gap in text-based emotion detection. In *Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025)*, Vienna, Austria. Association for Computational Linguistics.

Mattia Samory, Felix Soldner, and Veronika Batzdorfer. 2026. SemEval-2026 task 10: PsyCoMark – psycholinguistic conspiracy marker extraction and detection. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*.

Marco Siino. 2024. All-mpnet at SemEval-2024 task 1: Application of mpnet for evaluating semantic textual relatedness. In *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*, pages 379–384, Mexico City, Mexico. Association for Computational Linguistics.

Lianshuang Su and Xiaobing Zhou. 2024. NLP_STR_teamS at SemEval-2024 task1: Semantic textual relatedness based on MASK prediction and BERT model. In *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*, pages 337–341, Mexico City, Mexico. Association for Computational Linguistics.

Bao Tran and Nhi Tran. 2024. NewbieML at SemEval-2024 task 8: Ensemble approach for multidomain machine-generated text detection. In *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*, pages 354–360, Mexico City, Mexico. Association for Computational Linguistics.

Ankit Vaidya, Aditya Gokhale, Arnav Desai, Ishaan Shukla, and Sheetal Sonawane. 2024. CLTeam1 at SemEval-2024 task 10: Large language model based ensemble for emotion detection in Hinglish. In *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*, pages 365–369, Mexico City, Mexico. Association for Computational Linguistics.

Yuxia Wang, Jonibek Mansurov, Petar Ivanov, Jinyan Su, Artem Shelmanov, Akim Tsvigun, Osama Mohammed Afzal, Tarek Mahmoud, Giovanni Puccetti, and Thomas Arnold. 2024. SemEval-2024 task 8: Multidomain, multimodel and multilingual machine-generated text detection. In *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*, pages 2057–2079, Mexico City, Mexico. Association for Computational Linguistics.