Multimodal Identification of Vaccine Content Stance on Social Media

Surendrabikram Thapa[1], Shuvam Shiwakoti[1], Siddhant Bikram Shah[2], Kritesh Rauniyar[3], Laxmi Thapa[4], Surabhi Adhikari[5], Kristina T. Johnson[2], Ali Hürriyetoğlu[6], Hristo Tanev[7], Usman Naseem[3]
[1]Virginia Tech, USA, [2]Northeastern University, USA,
[3]Macquarie University, Australia, [4]O.P. Jindal Global University, India
[5]Columbia University, USA, [6]Wageningen Food Safety Research, Netherlands,
[7]European Commission, Joint Research Centre, Italy
[1]{surendrabikram, shuvam}@vt.edu, [2]rauniyark11@gmail.com,
[6]ali.hurriyetoglu@wur.nl, [7]hristo.tanev@ec.europa.eu

Abstract

Vaccination-related memes on social media play an increasingly influential role in shaping public perception of immunization, often spreading both supportive messaging and vaccine-critical narratives through multimodal communication. Detecting such content is challenging due to the combined use of images, embedded text, sarcasm, humor, and cultural references. This paper presents an overview of the Shared Task on Multimodal Identification of Vaccine Critical Content on Social Media, organized as part of the 9th Workshop on Event Extraction and Understanding: Challenges and Applications (EEUCA 2026) at ACL 2026. The task is based on the VaxMeme dataset, a large-scale collection of vaccination-related memes annotated into three classes: *Vaccine-critical*, *Neutral*, and *Pro-vaccine*. A total of 77 participants registered for the competition, with 25 teams submitting systems for evaluation. Participating approaches included transformer-based multimodal architectures, vision-language models, ensemble methods, and instruction-tuned large language models. The best-performing system achieved a macro F1-score of 0.8494. This shared task provides insights into the strengths and limitations of current multimodal approaches for vaccine stance detection and highlights future directions for robust public health misinformation analysis.

1 Introduction

Social media platforms have become primary arenas for public health discourse, where information about vaccines, treatments, and disease prevention spreads with unprecedented speed and reach (Shah et al., 2024b). Among the many forms of digital communication, memes—multimodal artifacts that fuse images and embedded text into compact, virally shareable units—have emerged as particularly influential vehicles for shaping public attitudes toward vaccination (Naseem et al., 2023; Ahmad et al., 2025). The COVID-19 pandemic dramatically illustrated both the promise and the peril of this medium: while memes proved effective at promoting awareness and disseminating accurate health information, they also became powerful conduits for vaccine misinformation, conspiracy theories, and skepticism that contributed to vaccine hesitancy and eroded public trust in immunization programs (Thapa et al., 2024b).

The challenge of identifying vaccine-critical content in memes is fundamentally compounded by the medium's inherent ambiguity. Vaccine-related memes frequently rely on sarcasm, irony, visual metaphor, and culturally specific references that obscure intent and complicate automated analysis (Pramanick et al., 2021). The boundary between legitimate critical commentary, satirical humor, and harmful misinformation is often deliberately blurred, with hateful or misleading content embedded within seemingly benign visual frames (Shah et al., 2024a). Single-modality approaches—whether text-only or image-only—consistently fail to capture this layered communicative intent, as the meaning of a vaccine-related meme typically emerges from the interplay between its visual and textual components rather than from either modality in isolation. This makes vaccine-critical content detection a paradigmatic case for multimodal reasoning, requiring systems to jointly interpret visual cues, embedded text, surrounding captions, and broader sociocultural context.

Despite the public health importance of this problem, computational resources for vaccine-critical meme detection remain limited. Most prior work on health misinformation has focused on text-only

Proceedings of the 9th Workshop on Event Extraction and Understanding: Challenges and Applications, pages 17–25

analysis of social media posts (Karafillakis et al., 2021). The scarcity of large-scale, publicly accessible, and richly annotated multimodal benchmarks has hindered systematic progress, particularly for capturing the diverse and evolving repertoire of vaccine-critical narratives that circulate across social media platforms.

To address these gaps, we present the **Shared Task on Multimodal Identification of Vaccine Critical Content on Social Media**, organized as part of the 9th Workshop on Event Extraction and Understanding: Challenges and Applications (EEUCA 2026) (Hürriyetoğlu et al., 2026), co-located with ACL 2026. The task is built upon the VaxMeme dataset (Naseem et al., 2023), a corpus of over 10,000 manually annotated vaccination-related memes spanning multiple platforms and timelines, designed specifically to support the development of multimodal vaccine-critical content detection systems. Participating systems are required to classify each meme into one of three categories: (1) *Vaccine Critical*: memes that criticize vaccines, contain vaccine misinformation, propagate conspiracy theories, or argue against vaccination, (2) *Neutral*: memes that report vaccine-related events or opinions objectively without taking a stance, or (3) *Pro-Vaccine*: memes that advocate for vaccination, promote awareness, or support immunization efforts. The task is evaluated using macro-averaged F1-score to ensure balanced performance across the three classes despite their natural distributional differences.

The shared task attracted a diverse range of participating teams employing both text-centric and multimodal approaches for vaccine stance classification. Submitted systems explored a variety of strategies, including transformer-based architectures, vision-language modeling, multimodal fusion techniques, ensemble methods, and instruction-tuned large language models. The diversity of submissions highlights the growing interest in multimodal public health content analysis and provides useful insights into the strengths and limitations of current approaches for vaccine-related meme understanding. This paper provides a comprehensive overview of the shared task, including a detailed description of the VaxMeme dataset and its annotation protocol, the evaluation methodology, summaries of the participating systems and their methodologies, and an analysis of the results. Through this shared task, we aim to advance the state of multimodal vaccine-critical content detection, support myth-debunking efforts, and contribute to the design of more effective public health communication strategies on social media platforms.

2 Related Works

Stance detection classifies whether a given text or image expresses support, opposition, or neutrality toward a specified target, distinguishing it from general sentiment analysis in that the inferred position is inherently relational and target-conditioned (Shiwakoti et al., 2024; Thapa et al., 2024a). Early work showed that stance can benefit from discourse-level information, since agreement and disagreement between utterances provide useful evidence beyond isolated lexical features (Thomas et al., 2006). The field has since expanded to social media, where short, informal, and context-dependent posts make stance recognition challenging due to implicit targets, sarcasm, and limited conversational context (Küçük and Can, 2020). However, most detection tasks remain primarily text-based, which limits their applicability to memes where stance may be encoded through visual framing, image-text incongruity, or culturally specific references (Küçük and Can, 2020; Kiela et al., 2020).

Research on misinformation detection has shown that misleading content can often be identified through linguistic, stylistic, and factuality-related cues, but such approaches usually focus on veracity rather than stance toward a public-health target (Rashkin et al., 2017). Vaccine misinformation is particularly consequential, as exposure via social media and coordinated online disinformation campaigns have been empirically associated with diminished vaccine confidence and elevated vaccine hesitancy (Wilson and Wiysonge, 2020). Experimental evidence further shows that exposure to COVID-19 vaccine misinformation can reduce vaccination intent, demonstrating the public-health importance of detecting harmful vaccine narratives online (Loomba et al., 2021). Nevertheless, vaccine-critical content subsumes misinformation, encompassing expressions of distrust, opposition, conspiracy framing, sarcasm, and satire that do not necessarily advance a directly verifiable factual claim (Rashkin et al., 2017; Loomba et al., 2021).

Multimodal meme analysis has established that image and text require joint interpretation, as each modality in isolation may appear benign while their combination conveys harmful or oppositional

meaning (Kiela et al., 2020; Thapa et al., 2025). Pramanick et al. (2021) demonstrated that harmful meme detection is improved by jointly modeling global meme-level context and local visual-textual cues, underscoring the necessity of fine-grained multimodal representations. In the vaccine domain, VaxMeme introduced a large manually annotated dataset of vaccine-critical memes and demonstrated that multimodal modeling is necessary for capturing the contextual and visual-textual signals present in vaccine discourse (Naseem et al., 2023). However, prior multimodal meme studies have largely focused on hate, harm, or binary misinformation labels, leaving a gap for systematic evaluation of multi-class vaccine content in social media memes (Kiela et al., 2020; Pramanick et al., 2021; Naseem et al., 2023).

Transformer-based language models such as BERT and RoBERTa established strong general-purpose representations for downstream NLP classification through large-scale pretraining and task-specific fine-tuning (Devlin et al., 2019; Liu et al., 2019). Vision-language transformers such as ViLBERT (Lu et al., 2019), VisualBERT (Li et al., 2019), and UNITER (Chen et al., 2020) extended this paradigm to multimodal learning by jointly encoding visual regions and textual tokens through cross-modal attention or image-text pretraining objectives. Contrastive and generative vision-language models such as CLIP (Radford et al., 2021), Flamingo (Alayrac et al., 2022), and BLIP-2 (Li et al., 2023) further improved transferability, few-shot adaptation, and instruction-following capabilities by scaling image-text supervision and integrating pretrained language models with visual encoders. Despite these advances, the classification task in the vaccine dataset remains non-trivial, as OCR noise, sarcasm, visual metaphor, and rapidly evolving sociopolitical narratives collectively undermine models trained on generic image-text corpora. The shared task features a multimodal dataset designed to engage the research community and encourage investigation into the identification and analysis of vaccine-critical content on social media platforms.

3 Shared Task Description

This shared task focuses on the automated understanding of vaccination-related memes through a multimodal lens, targeting the detection of vaccine stance in online content.

Task: Vaccine Stance Classification. Given a meme consisting of an image and associated textual content, participating systems must classify its stance towards vaccination into one of three categories: *Pro-vaccine*, *Vaccine-critical*, or *Neutral*. *Pro-vaccine* memes promote vaccination, highlight its benefits, or encourage positive health behaviors. *Vaccine-critical* memes express skepticism, opposition, or criticism towards vaccines, which may include misinformation, conspiracy narratives, or sarcastic undermining of vaccination efforts. *Neutral* memes present vaccination-related content without a clear stance, often conveying informational or ambiguous messages.

The task was evaluated using a macro-averaged F1-score to ensure balanced performance across all classes, particularly in the presence of class imbalance.

4 Dataset

The shared task is based on the *VaxMeme* dataset (Naseem et al., 2023; Thapa et al., 2026), a large-scale multimodal collection of vaccination-related memes. The dataset consists of 10,244 memes collected from Twitter, where each instance contains both an image and associated textual content (including OCR-extracted text when embedded in images).

4.1 Data Collection and Annotation

The dataset was constructed by collecting tweets containing both images and text between October 2020 and April 2021 using the Twitter API. Non-English content was excluded. Each meme was annotated by multiple human annotators with strong linguistic proficiency, following detailed annotation guidelines. Disagreements were resolved via majority voting with an additional annotator when required. The annotation quality is high, with substantial inter-annotator agreement (Fleiss' $\kappa = 0.85$).

Each meme is labeled into one of three stance categories: *Vaccine-critical* (0), *Neutral* (1), and *Pro-vaccine* (2).

4.2 Dataset Split

For the shared task, the dataset is split into training, validation, and test sets using an 80/10/10 ratio. The splits are designed to maintain a realistic and slightly imbalanced class distribution.

As shown in Table 1, class distribution reflects real-world conditions, where pro-vaccine content

Label	Train	Val	Test	Total
Vaccine-critical	2535	308	314	3157
Neutral	2461	327	316	3104
Pro-vaccine	3199	389	395	3983
Total	8064	1025	1024	10113

Table 1: Dataset statistics for the shared task.

is slightly more prevalent, while vaccine-critical and neutral memes appear in comparable proportions. This subtle imbalance makes the task more realistic and encourages the development of robust multimodal models.

5 Evaluation and Competition

This section describes the structure of our competition, along with the methodology used to determine ranks and other relevant details.

5.1 Evaluation Metrics

To evaluate the effectiveness of the participants' contributions, we used four metrics: macro F1-score, accuracy, precision, and recall. The participants' final ranks were determined using the macro F1-score as the primary ranking metric.

5.2 Competition Setup

We used Codabench[1] to organize our competition. The competition consisted of two phases: a development phase, where participants could familiarize themselves with the Codabench platform and develop their methods, and a test phase, where performance was used to determine the final ranking on the leaderboard. The results from the development phase were made available to participants after the phase concluded, enabling them to further refine their approaches for the test phase.

5.2.1 Registration

A total of 77 participants registered, out of which 25 teams submitted their predictions. The leaderboard is shown in Table 2.

5.2.2 Competition Timelines

The competition commenced on December 10, 2025, when training and development data were made available, marking the start of the development phase. During this phase, participants familiarized themselves with the Codabench platform and began developing their systems. The test phase began on January 15, 2026, when test data were provided without any ground truth labels. The test phase concluded on March 18, 2026. The paper submission deadline was March 29, 2026. Notification of acceptance was scheduled for April 28, 2026, with camera-ready papers due by May 12, 2026.

[1] https://www.codabench.org/competitions/12085/

6 Participants' Methods

LilyMeme (Li, 2026) built upon the MemeCLIP framework (Shah et al., 2024a), introducing a series of targeted enhancements for the VaxMeme vaccine stance detection task. The input is restructured using a `[POST]/[IMG]` template that explicitly separates post text from OCR-extracted image text, with `[NO_POST]` and `[NO_OCR]` markers for missing modalities, and the original element-wise fusion is replaced by a lightweight two-layer, eight-head cross-modal Transformer that models token-level image–text interactions. Training is further strengthened through noise-aware sample weighting, which derives per-instance confidence scores via nearest-neighbour consistency analysis and downweights ambiguous or likely mislabelled samples, and an auxiliary LLM description branch using Qwen2.5-VL-7B-Instruct that supplements memes with poor OCR quality. Inference-stage refinement combines test-time augmentation with a retrieval-augmented k-nearest-neighbour prior interpolated against the parametric model output, and the final submission ensembles multiple complementary variants trained across different cross-validation folds and visual backbones (CLIP ViT-L/14 and EVA02-L-14). The system achieved a macro F1 of 0.8494, securing 1st place overall.

CUET_SYNTHETICA (Zaman et al., 2026) proposed a gated cross-modal attention framework combining Twitter-RoBERTa for text encoding with CLIP ViT-L/14 for visual feature extraction. Textual inputs concatenated post-text with OCR-extracted meme overlay text, while both modalities were projected into a shared 512-dimensional fusion space. A learned scalar gate dynamically balanced cross-attended image representations against raw text features, suppressing uninformative visual signals. Final predictions were produced via a three-model weighted ensemble incorporating a text-only classifier, the full multimodal model, and a variant retrained on combined training and validation data. Their system achieved a test

Rank	Username	F1 Macro	Accuracy	Precision	Recall
1	lili12-637947 (Li, 2026)	0.8494	0.8517	0.8494	0.8517
2	wangxiuxian-637268	0.8389	0.8420	0.8386	0.8409
3	rishta_19-611897 (Zaman et al., 2026)	0.8357	0.8390	0.8383	0.8359
4	_alexcristea-636983 (Cristea and Ionescu, 2026)	0.8340	0.8380	0.8338	0.8351
5	sumaiya_110-594217 (Zaman et al., 2026)	0.8332	0.8361	0.8345	0.8340
6	anchy-637928	0.8308	0.8341	0.8309	0.8309
7	myname-637930	0.8308	0.8341	0.8309	0.8309
8	quasar-637336 (Chowdhury and Chowdhury, 2026)	0.8306	0.8322	0.8331	0.8324
9	wenbin-634065 (Shen, 2026)	0.8205	0.8244	0.8205	0.8218
10	naturia_beast-636958	0.8201	0.8244	0.8212	0.8209
11	vinaybabu-637935	0.8184	0.8215	0.8216	0.8190
12	ratpier-637076	0.8150	0.8176	0.8170	0.8161
13	yjwong1999-494691	0.8122	0.8137	0.8189	0.8141
14	linus-637363 (Acharya and Regmi, 2026)	0.8105	0.8137	0.8106	0.8123
15	havnis-636808	0.8067	0.8117	0.8080	0.8083
16	alishba-wazir-604227	0.8067	0.8088	0.8132	0.8071
17	zmin123-553584	0.7997	0.8039	0.8005	0.8013
18	lin123-637530	0.7994	0.8039	0.7992	0.8007
19	barkion-636765	0.7976	0.7990	0.8080	0.7986
20	merrli-636903	0.7972	0.7990	0.8058	0.7982
21	exterio-636705	0.7861	0.7912	0.7964	0.7846
22	abs123-504332	0.7846	0.7912	0.7868	0.7864
23	thatgrass-519137	0.7754	0.7844	0.7858	0.7802
24	wangkongqiang-637899 (Wang et al., 2026)	0.7552	0.7600	0.7652	0.7560
25	kannanrrk-615633	0.7436	0.7502	0.7435	0.7437

Table 2: Leaderboard ranked by Macro F1-score. All scores are presented as percentages (%). Note that this leaderboard contains the score till the test deadline and does not consider further runs done by participants as a part of the system description paper.

Macro F1 of 0.8357, ranking 3rd overall.

_alexcristea (Cristea and Ionescu, 2026) proposed a text-only early-fusion pipeline that skips visual encoders, instead extracting embedded meme text via OCR and concatenating it with the social media post before processing the unified sequence through an ERNIE-2.0-Large encoder. To reduce overfitting on noisy, label-ambiguous meme data, the standard classification head was replaced with a Multi-Sample Dropout architecture using five parallel dropout masks, acting as an implicit ensemble within a single forward pass. Trained with inverse class-weighted Cross-Entropy loss, the system achieved a Macro F1 of 0.8340, ranking 4th overall.

Quasar (Chowdhury and Chowdhury, 2026) presented a comprehensive ablation-driven system for three-class vaccine stance detection in social media memes, systematically evaluating text-only models (TF-IDF, BERT, RoBERTa, and DeBERTa variants), image-only models (ResNet-50, ViT, Swin, ConvNeXt, EfficientNet, CLIP Vision), and multimodal models (CLIP, BLIP, LLaVA) across multiple preprocessing and augmentation configurations. A domain-specific text normalisation pipeline preserves stance-indicative tokens such as emojis and hashtags, images are uniformly enhanced via contrast, brightness, and sharpness scaling, and balanced class oversampling — identified as the single most impactful intervention, adding approximately 4–5 macro F1 points across all model families — is applied to address the moderate class imbalance in the VaxMeme dataset. The final system combines DeBERTa-v3-large, RoBERTa-large, and CLIP multimodal (ViT-B/32) via soft voting with weights proportional to individual validation macro F1, achieving a macro F1 of 0.8306 and placing 8th out of 25 participating teams.

wenbin-634065 (Shen, 2026) introduced MoEs-VaxAgent, a hybrid discriminative-generative pipeline addressing both standard and boundary-ambiguous meme samples. Feature extraction draws on RoBERTa, ViT, CLIP, and Sentence-BERT, with the last encoder processing domain-relevant passages retrieved from MMCoVaR as external knowledge, producing five modality-specific expert representations dynamically aggregated via a learnable Top-2 gating network. Samples where the Mixture-of-Experts (MoE) classifier yields low-confidence predictions are subsequently re-evaluated by a

trio of LLM agents, namely a text agent, a visual agent, and a judge agent for conflict resolution, before a final label is assigned. The framework achieved a Macro F1 of 0.8205, ranking 9th overall.

Linus (Acharya and Regmi, 2026) compared text-only and multimodal late-fusion approaches for vaccine-critical meme classification using a shared three-layer feedforward classification head across all configurations. The multimodal systems combined CLIP ViT-B/32 image features with BERT-family text encoders via L2-normalized concatenation, while the text-only systems fine-tuned five encoders, namely BERT-base-uncased, RoBERTa-base, ModernBERT-base, DistilBERT-base, and DeBERTa-v3-base, on post text alone. Contrary to expectations, text-only models consistently outperformed their multimodal counterparts, with BERT-base-uncased achieving the best test Macro F1 of 0.8102, ranking 14th overall.

wangkongqiang (Wang et al., 2026) explored a wide range of supervised learning approaches for the multimodal identification of vaccine-critical content, evaluating both fine-tuned pre-trained transformer encoders and instruction-tuned large language models. The pre-trained model branch included ALBERT, BERT, ERNIE, and RoBERTa variants, with additional architectural augmentations such as RNN, CNN, and LSTM layers stacked on top of RoBERTa, and a hard voting ensemble over the four strongest variants. The LLM branch fine-tuned Qwen2-1.5B, Qwen2-7B, Llama2-7B, and Llama3-8B using the Llama-Factory framework with prompts composed of post text, image text, and selectable label types. The best-performing system was the fine-tuned Qwen2-1.5B LLM, achieving a Macro F1 of 0.8153, accuracy of 0.8185, and ranking 12th overall, demonstrating that smaller instruction-tuned LLMs can compete with larger variants when computational resources are constrained.

CSECU-Learners (Ahmad and Uddin, 2026) proposed a two-stage early fusion framework integrating three transformer-based encoders for vaccine-critical meme detection. The architecture combined Twitter-RoBERTa for textual encoding, Vision Transformer (ViT) for visual feature extraction, and Vision-and-Language Transformer (ViLT) for joint cross-modal representations. In Stage 1, the pooler outputs of RoBERTa and ViT were combined via performance-weighted summation with weights derived from validation rankings; in Stage 2, this visual-contextualized representation was concatenated with the ViLT pooler output and passed through a linear classification layer. To mitigate class imbalance, the system was trained with Focal Loss. Their approach achieved a Macro F1 of 0.8308 and accuracy of 0.8341, ranking 6th overall. Ablation studies confirmed that the Stage 1 RoBERTa-ViT fusion contributes most substantially to performance, partly by compensating for ViLT's restrictive 40-token sequence length limit.

7 Discussion

The submitted systems demonstrate the continued importance of multimodal reasoning for understanding vaccine-related discourse on social media. Most high-performing teams combined textual and visual representations through transformer-based architectures, cross-modal attention mechanisms, or ensemble strategies, highlighting that vaccine stance in memes is rarely conveyed through a single modality alone. In many cases, the interaction between image context, embedded OCR text, and accompanying captions was necessary for correctly identifying sarcasm, misinformation, or subtle stance cues.

A notable trend among top-performing submissions was the strong reliance on pretrained vision-language models and domain-adapted language encoders. Several teams incorporated CLIP-based visual representations, Twitter-domain RoBERTa encoders, or instruction-tuned large language models, suggesting that pretrained multimodal knowledge transfers effectively to vaccine-critical meme analysis. Additionally, ensemble methods and hybrid fusion strategies consistently improved robustness, particularly for ambiguous or noisy samples.

Interestingly, some text-only systems remained highly competitive, occasionally outperforming more complex multimodal architectures. This suggests that OCR-extracted textual content and associated captions contain substantial stance-related information in the VaxMeme dataset. However, purely text-based approaches may struggle in cases where stance is conveyed implicitly through visual symbolism, irony, or image-text incongruity. The results therefore indicate that while textual

information remains dominant in many instances, multimodal integration provides complementary contextual signals that improve generalization and robustness.

The competition also highlighted several persistent challenges. Vaccine-related memes often contain sarcasm, cultural references, visual metaphors, and low-quality OCR text, all of which complicate reliable classification. Furthermore, the evolving nature of online vaccine discourse means that models trained on static datasets may face distributional shifts over time. Many systems also relied heavily on large pretrained models, raising concerns regarding computational efficiency, accessibility, and reproducibility.

Future work can explore several promising directions. First, retrieval-augmented and knowledge-grounded systems may help models reason about evolving public health narratives and misinformation trends. Second, finer-grained explainability methods could improve transparency by identifying which textual or visual elements contribute most strongly to predictions. Third, multilingual and cross-cultural extensions of vaccine meme datasets would improve the applicability of these systems beyond English-speaking contexts. Finally, more robust handling of sarcasm, implicit stance, and adversarial meme constructions remains an important open research challenge for multimodal public health content analysis.

8 Conclusion

This shared task presented a benchmark for multimodal vaccine stance classification using the VaxMeme dataset and attracted a diverse range of approaches spanning transformer ensembles, vision-language models, and instruction-tuned LLMs. The results demonstrate that multimodal modeling remains highly effective for identifying vaccine-critical content, while also revealing the continued strength of carefully designed text-centric approaches. Through this shared task, we hope to encourage further research into multimodal public health content understanding, misinformation detection, and socially responsible AI systems for online discourse analysis.

Limitations

This shared task has several limitations. First, the dataset consists only of English-language memes collected from Twitter during a specific period of the COVID-19 pandemic, limiting generalizability across languages, cultures, and platforms. Second, vaccine-related memes often rely on sarcasm, humor, and cultural references that remain difficult for current multimodal systems to interpret reliably. OCR quality and noisy embedded text may also affect model performance. Finally, leaderboard metrics such as macro F1-score do not fully capture robustness, fairness, or real-world deployment challenges.

Ethical Considerations

This shared task involves the analysis of vaccine-related social media content, including misinformation and conspiracy-oriented memes. While such systems may support public health research and misinformation analysis, incorrect predictions could misclassify satire, political commentary, or legitimate criticism. The dataset contains publicly shared social media content and should be used responsibly and only for research purposes. We also acknowledge that multimodal content analysis systems may introduce societal risks if used for surveillance or automated censorship without appropriate human oversight.

References

Darwin Acharya and Sunil Regmi. 2026. Linus@eeuca 2026: Multimodal and text-only approaches to vaccine-critical meme detection. In *Proceedings of the 9th Workshop on Event Extraction and Understanding: Challenges and Applications (EEUCA)*.

Monir Ahmad and Md. Saif Uddin. 2026. Csecu-learners@eeuca 2026: Vaccine critical memes identification using two-stage early fusion of transformers. In *Proceedings of the 9th Workshop on Event Extraction and Understanding: Challenges and Applications (EEUCA)*.

Syed Talal Ahmad, Haohui Lu, Sidong Liu, Annie Lau, Amin Beheshti, Mark Dras, and Usman Naseem. 2025. Vaxguard: A multi-generator, multi-type, and multi-role dataset for detecting llm-generated vaccine misinformation. *arXiv preprint arXiv:2503.09103*.

Jean-Baptiste Alayrac, Jeff Donahue, Pauline Luc, Antoine Miech, Iain Barr, Yana Hasson, Karel Lenc, Arthur Mensch, Katherine Millican, Malcolm Reynolds, and 1 others. 2022. Flamingo: a visual language model for few-shot learning. *Advances in neural information processing systems*, 35:23716–23736.

Yen-Chun Chen, Linjie Li, Licheng Yu, Ahmed El Kholy, Faisal Ahmed, Zhe Gan, Yu Cheng, and Jingjing Liu. 2020. Uniter: Universal image-text

representation learning. In *European conference on computer vision*, pages 104–120. Springer.

Adiba Fairooz Chowdhury and MD Sagor Chowdhury. 2026. Quasar@eeuca 2026: Multimodal deep learning for vaccine stance detection in memes. In *Proceedings of the 9th Workshop on Event Extraction and Understanding: Challenges and Applications (EEUCA)*.

Alexandru-Marian Cristea and Costin Ionescu. 2026. _alexcristea@eeuca 2026: A robust early-fusion ernie pipeline for multimodal covid-19 vaccine meme classification. In *Proceedings of the 9th Workshop on Event Extraction and Understanding: Challenges and Applications (EEUCA)*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pages 4171–4186.

Ali Hürriyetoğlu, Surendrabikram Thapa, Hristo Tanev, Laxmi Thapa, and Surabhi Adhikari. 2026. Overview of the workshop on event extraction and understanding: Challenges and applications. In *Proceedings of the 9th Workshop on Event Extraction and Understanding: Challenges and Applications (EEUCA)*.

Emilie Karafillakis, Sam Martin, Clarissa Simas, Kate Olsson, Judit Takacs, Sara Dada, and Heidi Jane Larson. 2021. Methods for social media monitoring related to vaccination: systematic scoping review. *JMIR public health and surveillance*, 7(2):e17149.

Douwe Kiela, Hamed Firooz, Aravind Mohan, Vedanuj Goswami, Amanpreet Singh, Pratik Ringshia, and Davide Testuggine. 2020. The hateful memes challenge: Detecting hate speech in multimodal memes. *Advances in neural information processing systems*, 33:2611–2624.

Dilek Küçük and Fazli Can. 2020. Stance detection: A survey. *ACM Computing Surveys (CSUR)*, 53(1):1–37.

Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. 2023. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In *International conference on machine learning*, pages 19730–19742. PMLR.

Liunian Harold Li, Mark Yatskar, Da Yin, Cho-Jui Hsieh, and Kai-Wei Chang. 2019. Visualbert: A simple and performant baseline for vision and language. *arXiv preprint arXiv:1908.03557*.

Yixuan Li. 2026. Lilymeme@eeuca 2026: Multimodal vaccine meme stance detection with task-adapted memeclip and complementary ensembling. In *Proceedings of the 9th Workshop on Event Extraction and Understanding: Challenges and Applications (EEUCA)*.

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

Sahil Loomba, Alexandre De Figueiredo, Simon J Piatek, Kristen De Graaf, and Heidi J Larson. 2021. Measuring the impact of covid-19 vaccine misinformation on vaccination intent in the uk and usa. *Nature human behaviour*, 5(3):337–348.

Jiasen Lu, Dhruv Batra, Devi Parikh, and Stefan Lee. 2019. Vilbert: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks. *Advances in neural information processing systems*, 32.

Usman Naseem, Jinman Kim, Matloob Khushi, and Adam G Dunn. 2023. A multimodal framework for the identification of vaccine critical memes on twitter. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, pages 706–714.

Shraman Pramanick, Shivam Sharma, Dimitar Dimitrov, Md Shad Akhtar, Preslav Nakov, and Tanmoy Chakraborty. 2021. Momenta: A multimodal framework for detecting harmful memes and their targets. In *Findings of the association for computational linguistics: EMNLP 2021*, pages 4439–4455.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, and 1 others. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR.

Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pages 2931–2937.

Siddhant Bikram Shah, Shuvam Shiwakoti, Maheep Chaudhary, and Haohan Wang. 2024a. Memeclip: Leveraging clip representations for multimodal meme classification. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 17320–17332.

Siddhant Bikram Shah, Surendrabikram Thapa, Ashish Acharya, Kritesh Rauniyar, Sweta Poudel, Sandesh Jain, Anum Masood, and Usman Naseem. 2024b. Navigating the web of disinformation and misinformation: Large language models as double-edged swords. *IEEE Access*.

Wenbin Shen. 2026. wenbin-634065@eeuca 2026: Moes-vaxagent, a two-stage framework for multimodal vaccine critical meme detection. In *Proceedings of the 9th Workshop on Event Extraction and Understanding: Challenges and Applications (EEUCA)*.

Shuvam Shiwakoti, Surendrabikram Thapa, Kritesh Rauniyar, Akshyat Shah, Aashish Bhandari, and Usman Naseem. 2024. Analyzing the dynamics of climate change discourse on twitter: A new annotated corpus and multi-aspect classification. In *Proceedings of the 2024 joint international conference on computational linguistics, language resources and evaluation (LREC-COLING 2024)*, pages 984–994.

Laxmi Thapa, Aryaman Jain, Lakshmojee Koduru, Surabhi Adhikari, Junaid Rashid, Jungeun Kim, Surendrabikram Thapa, and Usman Naseem. 2026. Concept-grounded detection of vaccine misinformation in multimodal content using interpretable vision-language models. In *Companion Proceedings of the ACM on Web Conference 2026*.

Surendrabikram Thapa, Kritesh Rauniyar, Farhan Jafri, Shuvam Shiwakoti, Hariram Veeramani, Raghav Jain, Guneet Singh Kohli, Ali Hürriyetoğlu, and Usman Naseem. 2024a. Stance and hate event detection in tweets related to climate activism-shared task at case 2024. In *Proceedings of the 7th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2024)*, pages 234–247.

Surendrabikram Thapa, Kritesh Rauniyar, Hariram Veeramani, Aditya Shah, Imran Razzak, and Usman Naseem. 2024b. Did you tell a deadly lie? evaluating large language models for health misinformation identification. In *International Conference on Web Information Systems Engineering*, pages 391–405. Springer.

Surendrabikram Thapa, Siddhant Bikram Shah, Kritesh Rauniyar, Shuvam Shiwakoti, Surabhi Adhikari, Hariram Veeramani, Kristina T Johnson, Ali Hürriyetoğlu, Hristo Tanev, and Usman Naseem. 2025. Multimodal hate, humor, and stance event detection in marginalized sociopolitical movements. In *Proceedings of the 8th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Texts*, pages 20–31.

Matt Thomas, Bo Pang, and Lillian Lee. 2006. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In *Proceedings of the 2006 conference on empirical methods in natural language processing*, pages 327–335.

Kongqiang Wang, Peng Zhang, and Qingli Tan. 2026. wangkongqiang@eeuca 2026: Multimodal identification of vaccine critical content on social media. In *Proceedings of the 9th Workshop on Event Extraction and Understanding: Challenges and Applications (EEUCA)*.

Steven Lloyd Wilson and Charles Wiysonge. 2020. Social media and vaccine hesitancy. *BMJ global health*, 5(10).

Sumaiya Zaman, Miftahul Jannat Rishta, and Shiti Chowdhury. 2026. Cuet_synthetica@eeuca 2026: Gated cross-modal attention with domain-adapted text encoding for vaccine-critical meme detection. In *Proceedings of the 9th Workshop on Event Extraction and Understanding: Challenges and Applications (EEUCA)*.