An Overview of the Misogyny Meme Detection Shared Task for Chinese Social Media

Bharathi Raja Chakravarthi[1], **Rahul Ponnusamy**[2], **Ping Du**[1], **Xiaojian Zhuang**[1], **Saranya Rajiakodi**[3], **Paul Buitelaar**[2], **Premjith B**[4], **Bhuvaneswari Sivagnanam**[3] **Anshid Kizhakkeparambil**[5], **Lavanya S.K.**[6]

[1] School of Computer Science, University of Galway, Ireland
[2] Data Science Institute, University of Galway, Ireland
[3] Department of Computer Science, Central University of Tamil Nadu, India
[4] Amrita School of Artificial Intelligence, Coimbatore, Amrita Vishwa Vidyapeetham, India
[5] Mar Thoma College, Chungathara, Kerala, India
[6] Madras Institute of Technology Campus, Anna University, India

Abstract

The increasing prevalence of misogynistic content in online memes has raised concerns about their impact on digital discourse. The culture-specific images and informal usage of text in the memes present considerable challenges for the automatic detection systems, especially in low-resource languages. While previous shared tasks have addressed misogyny detection in English and several European languages, misogynistic meme detection in the Chinese has remained largely unexplored. To address this gap, we introduced a shared task focused on binary classification of Chinese language memes as misogynistic or non-misogynistic. The task featured memes collected from the Chinese social media and annotated by native speakers. A total of 45 teams registered, with 8 teams submitting predictions from their multimodal models integrating textual and visual features through diverse fusion strategies. The best-performing system achieved a macro F1-score of 0.93035, highlighting the effectiveness of lightweight pretrained encoder fusion. This system used the Chinese BERT and DenseNet-121 for text and image feature extraction, respectively. A feedforward network was trained as a classifier using the features obtained by concatenating text and image features.

1 Introduction

The proliferation of social media and other online platforms provides users with powerful means to share visual (image and video), audio, and textual content as a form of expressing opinions and beliefs (Huang et al., 2019). Memes are typically used for expressing one's opinion through humorous or satirical images overlaid with text (Zhong et al., 2022), (Aslam et al., 2022), (Aggarwal et al., 2021). However, the dual-mode characteristic, which contains both text and images, and the speed at which the memes are spread make them a powerful tool for communication, which in turn makes them carriers of offensive and hate content, including misogyny (Jindal et al., 2024), (Chakravarthi et al., 2025). Furthermore, misogynistic memes frequently trivialize the violence against women and discriminatory attitudes by employing humor and irony (Kumari, 2021).

Misogynistic memes are difficult to detect due to the multimodal ambiguity (Hakimov et al., 2022), (Rizzi et al., 2024), (Kiela et al., 2020). Moreover, the models that are used for detecting the misogynistic memes should be capable of capturing the cultural nuances, sarcasm, and symbolism, mainly because of the diverse linguistic and social settings (Kumari et al., 2024). Much of the prior research has focused on English and European languages (Fersini et al., 2022a). For example, the MAMI shared task at SemEval-2022 targeted misogyny detection in English memes using image-text combinations, while SemEval-2024 addressed multimodal persuasion techniques in memes across multiple languages, including English and Bulgarian. Further, recent shared tasks such as LT-EDI@EACL 2024 and DravidianLangTech@NAACL 2025 have explored this problem in Tamil and Malayalam (Chakravarthi et al., 2024, 2025). Despite these advancements, misogynistic meme detection in Chinese—one of the most widely spoken languages globally (Julian, 2020)—has remained largely unexplored.

To address this gap, we introduced the shared task on misogyny meme detection in the Chinese. This task focuses on the binary classification of memes sourced from real-world Chinese social media platforms, annotated by native speakers as ei-

Proceedings of the Fifth Workshop on Language Technology for Equality, Diversity, Inclusion, pages 199–207

ther misogynistic or non-misogynistic. The memes reflect various forms of gender-based prejudice, often expressed through culturally grounded visual cues and colloquial language.

This shared task encouraged participants to develop multimodal models capable of jointly interpreting visual and textual cues. The multilingual complexity of Chinese, the presence of colloquial and coded language, and the use of culturally specific imagery added rich layers of difficulty. By bringing together researchers across NLP, computer vision, and digital humanities, the task aims to foster tools and insights that are socially aware, linguistically grounded, and technically robust. Moreover, code-mixing between Chinese and English is common on Chinese social media platforms (Zhang, 2012). The code-mixing involves vocabulary substitution, where English words, abbreviations, brand names, or proprietary terms are inserted into Chinese sentences (Guo, 2023). This creates a lot of challenges for building AI models for code-mixed Chinese social media data. The language identification and segmentation is one of the major challenges for models, particularly embedding models. The unavailability of the high-quality, large-scale corpus is another challenge for Chinese-English code-mixed data modeling. Pretrained language models, even large multilingual ones, often perform poorly on code-mixed inputs compared to monolingual or standard multilingual data (Kodali et al., 2024).

A total of 45 teams registered for the shared task, of which 8 teams successfully submitted their system outputs for final evaluation. The submitted systems employed a range of multimodal strategies, combining textual and visual features through various fusion techniques. The CUET_320 secured the first position by achieving the highest macro F1-score of 0.93035. This demonstrates the effectiveness of lightweight fusion of pretrained text and image encoders. Teams CUET_Ignite (0.91775) and Team_Luminaries_0227 (0.90355) attained the second and third positions. These teams showed that transformer-based language models and vision encoders can perform effectively when optimized with minimal compute. Despite these successes, ongoing challenges include class imbalance, noisy visual-text alignment, and the handling of code-mixed linguistic patterns in Chinese memes.

2 Related Work

2.1 Misogyny Meme Detection

Detecting misogynous memes is a challenging task due to the multimodal ambiguity. Researchers have proposed several approaches and models to address this challenge. In (Gu et al., 2022), the authors proposed a multi-modal, multi-task variational autoencoder designed to detect misogynous memes. This model learns a co-representation from both images and text together in a latent feature space to find misogynous memes and classify them into specific categories. In (Agrawal and Mamidi, 2022), the authors proposed visual linguistic models with transfer learning. These models followed a task-specific pretraining of the models. The authors (Mahesh et al., 2024) proposed an approach by combining models pretrained for text and image data in Tamil and Malayalam. The authors conducted experiments using BERT+ResNet-50, MuRIL+ResNet-50, and mBERT+ResNet-50. Among these models, mBERT+ResNet-50 and MuRIL+ResNet-50 achieved the highest macro F1-scores of 0.73 and 0.87 for Tamil and Malayalam data, respectively.

The authors (Roy et al., 2024) designed a Bidirectional Long Short-Term Memory (Bi-LSTM) with BERT embeddings network for detecting misogynistic comments in Bengali along with a comprehensive dataset. The study proposed by (Chinivar et al., 2024) used the XLM-R model along with two image transformer models, ViT and Swin, on a benchmarked meme dataset. Here, the authors concatenated the embeddings generated from text and image data and fed them into a single-layered neural network classifier for classification. In (Gitanjali et al., 2024), the authors proposed a Multimodal Multi-hop Chain-of-Thought (M3Hop-CoT) framework for misogynous meme identification. They integrated entity-object-relationship information with a multimodal CoT module, thereby providing emotional cues, target awareness, and contextual knowledge for meme analysis. Finally, a CLIP-based classifier was employed for classification.

In (Farinango Cuervo and Parde, 2022), the authors used contrastive learning for the multimodal detection of misogynistic memes. Contrastive learning ensured that memes labeled as misogynistic were grouped together by differentiating them from non-misogynistic ones. Detecting misogynistic memes in the Chinese language is a complex and emerging field. CHMEMES (Gu et al., 2024)

is a dataset for detecting harmful memes in the Chinese language. Moreover, the authors proposed a multimodal framework, semantic contrastive alignment framework (SCARE) to implement the task. The proposed framework learns both cross-modal and intra-modal information, with a cross-modal contrast alignment objective and an intra-modal contrast objective.

2.2 Related Shared tasks

Research on misogyny and hate speech detection has been significantly shaped by early shared tasks. Evalita 2018 and IberEval 2018 introduced the Automatic Misogyny Identification (AMI) challenge for English and Italian, focusing on text-based detection of gendered hate speech (Fersini et al., 2018a,b). These were among the first systematic efforts in the domain. Later, SemEval 2019 Task 5 expanded this scope to multilingual hate speech, targeting both misogyny and immigrant hate in English and Spanish. It introduced subtasks on aggression level classification, establishing the need for nuanced annotation in hate speech datasets (Basile et al., 2019).

Multimodal shared tasks further pushed boundaries by incorporating image and text jointly. The Memotion Analysis tasks at SemEval 2020 addressed meme-based sentiment and emotion classification in English, providing large annotated meme datasets and benchmarks, enabling exploration of the affective dimensions of memes (Sharma et al., 2020; Patwa et al., 2022).

The MAMI (Multimedia Automatic Misogyny Identification) task at SemEval-2022 focused explicitly on misogyny detection in English memes, with subtasks for binary and fine-grained classification of misogynistic intent using both image and text (Fersini et al., 2022b). The recent SemEval-2024 Task targeted persuasion technique detection in memes using English training data, with test sets in English, Bulgarian, North Macedonian, and Arabic, and included both text-only and multimodal subtasks (Dimitrov et al., 2024).

In line with these developments, the LT-EDI@EACL 2024 shared task introduced misogynistic meme detection in Tamil and Malayalam as part of a multitask classification challenge (Chakravarthi et al., 2024). Building on this, the DravidianLangTech@NAACL 2025 shared task extended the focus to dedicated misogyny meme detection (Chakravarthi et al., 2025). Our current shared task, organized at LT-EDI@LDK 2025, continues this progression by targeting Chinese-language misogynistic memes. This task aims to broaden multimodal hate speech research to new linguistic and cultural domains and addresses the scarcity of annotated resources in Chinese for multimodal misogyny detection.

3 Task Description

The shared task on Misogyny Meme Detection is organized as part of the LT-EDI workshop at LDK 2025[1]. This task focuses on a multimodal classification problem that targets the automatic detection of misogynistic content in memes.

Memes present a unique challenge for automated hate speech detection due to their multimodal nature, combining both images and text to convey meaning. This shared task aims to address the rising concern of misogynistic content, particularly in Chinese online social media platforms. Couple of examples from the dataset can be seen in the Figure 1. The task emphasizes the need for robust computational models that can effectively integrate textual and visual information to identify misogyny. Participants are required to develop systems that classify memes into two categories:

- Misogyny
- Not-Misogyny

This task supports ongoing research in multimodal hate speech detection and contributes to the development of ethical AI tools capable of identifying harmful content online.

3.1 Dataset and Evaluation

Dataset: The dataset utilized in this study comprises memes collected from various Chinese social media platforms. Each meme consists of two modalities: a meme image and its corresponding textual content, which is a transcription contained within the meme. The annotation provided are at the level of the entire meme. These annotations are binary categorizing each meme as either containing misogynistic content or not.

Task Phases: The task is organized into two distinct phases:

- **Training and Development Phase:** In this initial phase, participants are provided with a labeled datset. This dataset is intended for

[1] https://codalab.lisn.upsaclay.fr/competitions/21880

Figure 1: Example memes from the dataset with the translation

the training and validation of their models. It contains a collection of both misogynistic and non-misogynistic memes, each accompanied by its respective image and textual data.

- **Test Phase:** Following the training and development phase, a separate, unlabeled test set will be distributed. Participants are required to submit their system's predictions for this test set. The final evaluation of the submitted systems will be performed using this unseen data.

Evaluation Metric: The primary metric for evaluating performance in this task is the macro F1-score. This metric is chosen because it considers the performance across both the misogynistic and non-misogynistic classes equally. The macro F1-score is particularly well-suited for classification tasks where there may be an imbalance in the number of samples per class, thereby ensuring a fair and comprehensive assessment of model performance on both categories.

4 Methodology of Participants

This competition included eight entrants, each presenting a distinct approach to misogynist meme identification. Most teams employed various preprocessing methods, data augmentation strategies, and model architectures to build multimodal pipelines that integrated textual and visual features. The approaches ranged from deep learning and transformer-based models to traditional machine learning, with different fusion techniques applied to effectively combine modalities. A summary of each team's core methodology is provided in the following section.

Team_Luminaries_0227 (Adnan Faisal, 2025) used a multimodal fusion strategy to tackle the Misogyny Meme Detection task. After testing with XLM-RoBERTa (Li et al., 2021) and DistilBERT (Sanh et al., 2019), they used the Chinese BERT model (bert-base-chinese) for textual analysis. They evaluated VGG16, ViT, and ResNet in the visual domain before deciding on VGG16 because to its effectiveness. The imagess were preprocessed to 224x224 pixels and enhanced with rotations and random flips. Jieba (Zhang et al., 2019) was used to clean and tokenize the textual data. In order to merge text and image information and feed them into a classifier for the ultimate prediction, the researchers used early fusion. The Adam optimizer was used for training, with a batch size of 32, a learning rate of 2e-5, and early stop-

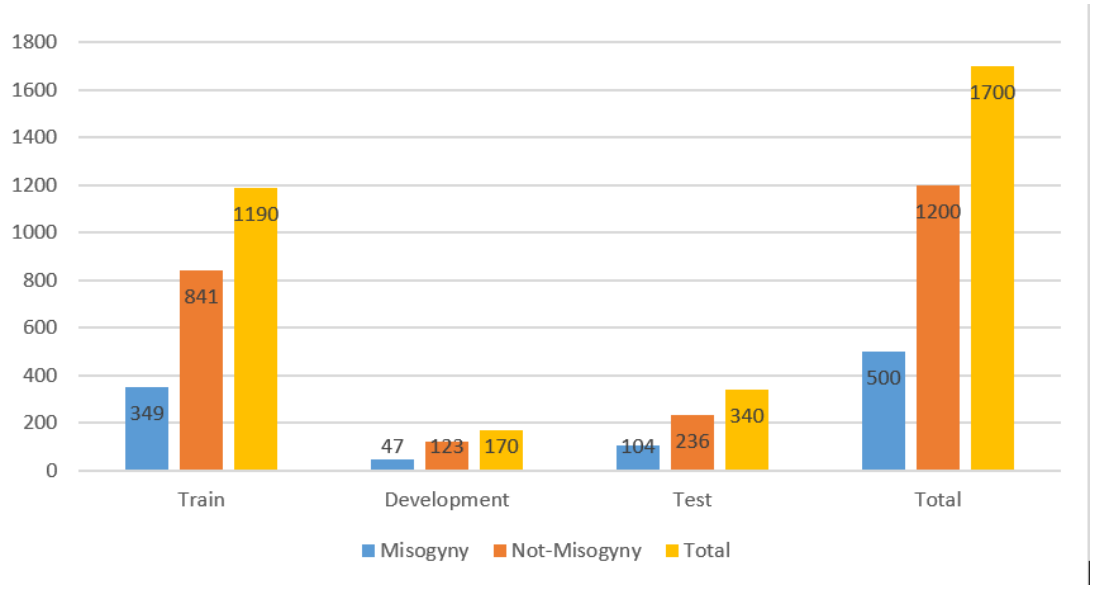

Figure 2: Dataset Statistics

ping determined by validation loss. Weighted loss functions were used to remedy class imbalance. On the Malayalam dataset, their method produced a macro F1-score of 0.87, and on the Tamil dataset, 0.73.

CUET_12033 (Mehreen Rahman, 2025) proposed a compact yet effective multimodal pipeline for detecting misogyny in Chinese memes. They preprocessed text by cleaning, converting to Simplified Chinese, tokenizing with Jieba, and transliterating using pypinyin. Images were resized to 224×224 RGB and enhanced for brightness and contrast. To address class imbalance, misogynistic samples were augmented with brightness adjustment, grayscale conversion, and posterization. For unimodal models, CharBERT-base-Chinese was enhanced with a BiLSTM, while image features were extracted using CLIP, ViT, ResNet-50, and EfficientNet-B0. Their best multimodal model fused CharBERT+BiLSTM (Ma et al., 2020) and ViT embeddings using a gated multimodal unit (GMU), followed by a classifier. Training included class-weighted loss, mixed precision, learning rate scheduling, dropout, early stopping, and gradient clipping for stability.

CUET_Ignite (MD.Mahadi Rahman, 2025) proposed a robust multimodal pipeline for misogyny detection in Chinese memes. Their preprocessing involved cleaning and transliterating text (using Jieba and pypinyin) and enhancing 224×224 RGB images. To address class imbalance, they applied augmentation only on misogynistic samples—using image transformations and back-translation. For modeling, they used CharBERT with or without BiLSTM for text and explored CLIP, ViT, ResNet-50, and EfficientNet-B0 for image features. Text and image embeddings were projected into a joint 512-dim space and fused via multi-head attention, followed by a two-layer MLP. Training used Adam/AdamW with class-weighted loss and optimization strategies like mixed precision, scheduling, and early stopping.

SSNCSE (Sreeja K, 2025) developed a multimodal classification system using frozen pretrained models for efficient misogyny detection in Chinese memes. Text was cleaned and tokenized using BERT-base, while images were resized and normalized before extracting 2048-dim features with ResNet-50. The 768-dim text and 2048-dim image embeddings were concatenated into a 2816-dim vector and passed through a lightweight feed-forward classifier with ReLU and dropout. Only classifier layers were trained using Adam (LR=1e-4, batch size=8) over 5 epochs, optimizing cross-entropy loss. Final predictions were stored in CSV format.

CUET_320 (Madiha Ahmed Chowdhury, 2025) developed an efficient multimodal system for misogyny meme detection by preprocessing text with regex cleaning, Jieba tokenization, stopword removal, and truncation, while images were resized, histogram-enhanced, and normalized. To balance classes, misogynistic samples were augmented using brightness, grayscale, and posteriza-

tion. Features were extracted via Chinese BERT, XLM-R, mBERT (text), and CNNs like ResNet-50, DenseNet-121, Inception V3 (image). Their best model combined Chinese BERT and DenseNet-121 features through concatenation and classified them using an Adam-optimized feedforward network, showing the effectiveness of simple fusion.

CVF_NITT (Radhika K T, 2025) employed Visual-Language Models (VLMs), particularly CLIP, to detect sexist elements in Chinese memes by aligning OCR-extracted text and image features in a shared 512-dimensional space. They used a lightweight logistic regression classifier on these embeddings (CLIP+LR) and compared it against traditional fusion setups like BERT+ResNet-50 and CNN+Inception V3 followed by MLPs. Their early-fusion CLIP-based approach proved more effective for cross-modal integration.

CUET's_White_Walkers (Md Mubasshir Naib, 2025) created a modular multimodal system for sexism meme identification by preprocessing text (cleaned, lowercased, tokenized to 128 tokens) and images (resized to 224×224, normalized). Images were enhanced through flipping and cropping. They used transformer-based models (mBERT, MuRIL, BERT-base-Chinese), deep (GloVe/Keras embeddings with CNN, BiLSTM-CNN, CNN-GRU), and classical (TF-IDF, BoW with ML classifiers) to extract features from images and pretrained CNNs (ResNet50, DenseNet121, InceptionV3) for text. Using an early fusion of BERT-base-Chinese and ResNet50, their optimal multimodal setup outperformed all unimodal baselines with an F1 score of 0.8541.

5 Results and Discussion

All the participant's system were evaluated using a macro F1-score. Among the participating teams, CUET_320 achieved the highest performance with a macro F1-score of 0.93035 which secured first rank with Chinese BERT and DenseNet-121 combination . This was followed by CUET_Ignite which obtained a macro F1-score of 0.91775 and ranked second. The third position was held by Team_Luminaries_0227 which attained a macro F1-score of 0.90355 with Chinese BERT and VGG16 combination. These results indicate strong overall performance and highlight the effectiveness of the multimodal approaches used by the top teams. Most teams froze the pretrained encoders (e.g., ResNet50, BERT, CLIP) and trained only the classification layers, which helped reduce overfitting and training cost. Fusion techniques such as cross-attention and multi-head attention proved to be highly effective in modeling the relationship between images and text. Several teams addressed class imbalance using weighted loss functions or data augmentation, which improved minority class recall. Teams used optimization techniques such as FP16 training and batch accumulation to efficiently utilize limited compute resources (e.g., Kaggle P100 GPUs). These methods significantly improved the misogyny detection in low-resource high-noise meme datasets.

6 Conclusion

In this shared task on misogynistic meme detection, we challenged participants to build a multimodal classification systems capable of identifying misogyny in memes by analyzing both textual and visual components. The participating teams explored a variety of deep learning architectures, including combinations of pretrained CNNs, transformer-based text encoders, and fusion mechanisms. The top performing systems demonstrated the importance of careful preprocessing and enhancement, effective multimodal fusion, and efficient training pipelines to address imbalance and limited resources. These findings confirm that tackling harmful online content like misogynistic memes requires context-aware, cross-modal learning approaches.

Future work may explore zero-shot detection using large vision-language models (e.g., BLIP-2, LLaVA), prompt-tuning for meme analysis, and generating explanations for detected misogyny to aid in interpretability and platform moderation.

Acknowledgments

The authors, Bharathi Raja Chakravarthi and Paul Buitelaar was funded by a research grant from Research Ireland under grant number SFI/12/RC/2289_P2 (Insight) and author Rahul Ponnusamy was supported by the Centre for Research Training in Artificial Intelligence under Grant No. 18/CRT/6223.

References

Momtazul Arefin Labib Hasan Murad Adnan Faisal, Shiti Chowdhury. 2025. Team_luminaries_0227@lt-edi-2025: A transformer-based fusion approach to

Team Name	Run	F1-score	rank
CUET_320_run1 (Madiha Ahmed Chowdhury, 2025)	**1**	**0.93035**	**1**
CUET_Ignite_run2 (MD.Mahadi Rahman, 2025)	2	0.91775	2
Team_Luminaries_0227 (Adnan Faisal, 2025)	3	0.90355	3
CUET's_White_Walkers (Md Mubasshir Naib, 2025)	2	0.85421	4
CVF_NITT (Radhika K T, 2025)	-	0.73622	5
CUET_12033_run1 (Mehreen Rahman, 2025)	1	0.70898	6
SSNCSE_Sreeja K (Sreeja K, 2025)	1	0.70345	7
CUET_Fog_run1	1	0.49514	8

Table 1: list of the teams participated in the shared task

misogyny detection in chinese memes. In *Proceedings of the Fifth Workshop on Language Technology for Equality, Diversity and Inclusion*, Italy. Fifth Conference on Language, Data and Knowledge (LDK2025).

Apeksha Aggarwal, Vibhav Sharma, Anshul Trivedi, Mayank Yadav, Chirag Agrawal, Dilbag Singh, Vipul Mishra, and Hassène Gritli. 2021. Two-way feature extraction using sequential and multimodal approach for hateful meme classification. *Complexity*, 2021(1):5510253.

Samyak Agrawal and Radhika Mamidi. 2022. LastResort at SemEval-2022 task 5: Towards misogyny identification using visual linguistic model ensembles and task-specific pretraining. In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, pages 575–580, Seattle, United States. Association for Computational Linguistics.

Nida Aslam, Irfan Ullah Khan, Teaf I Albahussain, Nouf F Almousa, Mizna O Alolayan, Sara A Almousa, and Modhi E Alwhebi. 2022. Medeep: A deep learning based model for memotion analysis. *Mathematical Modelling of Engineering Problems*, 9(2).

Valerio Basile, Cristina Bosco, Elisabetta Fersini, Debora Nozza, Viviana Patti, Francisco Manuel Rangel Pardo, Paolo Rosso, and Manuela Sanguinetti. 2019. SemEval-2019 Task 5: Multilingual Detection of Hate Speech Against Immigrants and Women in Twitter. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 54–63, Minneapolis, Minnesota, USA. Association for Computational Linguistics.

Bharathi Raja Chakravarthi, Rahul Ponnusamy, Saranya Rajiakodi, Shunmuga Priya Muthusamy Chinnan, Paul Buitelaar, Bhuvaneswari Sivagnanam, and Anshid K A. 2025. Findings of the shared task on misogyny meme detection: DravidianLangTech@NAACL 2025. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, pages 721–731, Acoma, The Albuquerque Convention Center, Albuquerque, New Mexico. Association for Computational Linguistics.

Bharathi Raja Chakravarthi, Saranya Rajiakodi, Rahul Ponnusamy, Kathiravan Pannerselvam, Anand Kumar Madasamy, Ramachandran Rajalakshmi, Hariharan LekshmiAmmal, Anshid Kizhakkeparambil, Susminu S Kumar, Bhuvaneswari Sivagnanam, and Charmathi Rajkumar. 2024. Overview of shared task on multitask meme classification - unraveling misogynistic and trolls in online memes. In *Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity, Inclusion*, pages 139–144, St. Julian's, Malta. Association for Computational Linguistics.

Sneha Chinivar, MS Roopa, JS Arunalatha, and KR Venugopal. 2024. Identification of misogynistic memes using transformer models. In *International Conference on Advanced Communications and Machine Intelligence*, pages 107–116. Springer.

Dimitar Dimitrov, Firoj Alam, Maram Hasanain, Abul Hasnat, Fabrizio Silvestri, Preslav Nakov, and Giovanni Da San Martino. 2024. SemEval-2024 task 4: Multilingual detection of persuasion techniques in memes. In *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*, pages 2009–2026, Mexico City, Mexico. Association for Computational Linguistics.

Charic Farinango Cuervo and Natalie Parde. 2022. Exploring contrastive learning for multimodal detection of misogynistic memes. In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, pages 785–792, Seattle, United States. Association for Computational Linguistics.

Elisabetta Fersini, Francesca Gasparini, Giulia Rizzi, Aurora Saibene, Berta Chulvi, Paolo Rosso, Alyssa Lees, and Jeffrey Sorensen. 2022a. Semeval-2022 task 5: Multimedia automatic misogyny identification. In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, pages 533–549.

Elisabetta Fersini, Francesca Gasparini, Giulia Rizzi, Aurora Saibene, Berta Chulvi, Paolo Rosso, Alyssa Lees, and Jeffrey Sorensen. 2022b. SemEval-2022 task 5: Multimedia automatic misogyny identification. In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*,

pages 533–549, Seattle, United States. Association for Computational Linguistics.

Elisabetta Fersini, Debora Nozza, and Paolo Rosso. 2018a. Overview of the Evalita 2018 Task on Automatic Misogyny Identification (AMI). In Tommaso Caselli, Nicole Novielli, and Viviana Patti, editors, *EVALITA Evaluation of NLP and Speech Tools for Italian : Proceedings of the Final Workshop 12-13 December 2018, Naples*, Collana dell'Associazione Italiana di Linguistica Computazionale, pages 59–66. Accademia University Press, Torino. Code: EVALITA Evaluation of NLP and Speech Tools for Italian : Proceedings of the Final Workshop 12-13 December 2018, Naples.

Elisabetta Fersini, Paolo Rosso, Maria Anzovino, and 1 others. 2018b. Overview of the task on automatic misogyny identification at IberEval 2018. *Ibereval@ sepln*, 2150:214–228.

Kumari Gitanjali, kirtan Jain, and Asif Ekbal. 2024. M3hop-cot: Misogynous meme identification with multimodal multi-hop chain-of-thought. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, page 22105–22138.

Tianlong Gu, Mingfeng Feng, Xuan Feng, and Xuemin Wang. 2024. Scare: A novel framework to enhance chinese harmful memes detection. *IEEE Transactions on Affective Computing*.

Yimeng Gu, Ignacio Castro, and Gareth Tyson. 2022. MMVAE at semeval-2022 task 5: A multi-modal multi-task VAE on misogynous meme detection. In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, pages 700–710.

Yuanhao Guo. 2023. The Code-Mixing of Chinese and English Among Chinese College Students: A Qualitative Study. In *2022 4th International Conference on Literature, Art and Human Development (ICLAHD 2022)*, pages 73–89. Atlantis Press.

Sherzod Hakimov, Gullal Singh Cheema, and Ralph Ewerth. 2022. TIB-VA at SemEval-2022 task 5: A multimodal architecture for the detection and classification of misogynous memes. In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, pages 756–760, Seattle, United States. Association for Computational Linguistics.

Feiran Huang, Xiaoming Zhang, Zhonghua Zhao, Jie Xu, and Zhoujun Li. 2019. Image–text sentiment analysis via deep multimodal attentive fusion. *Knowledge-Based Systems*, 167:26–37.

Nitesh Jindal, Prasanna Kumar Kumaresan, Rahul Ponnusamy, Sajeetha Thavareesan, Saranya Rajiakodi, and Bharathi Raja Chakravarthi. 2024. MISTRA: Misogyny Detection through Text–Image Fusion and Representation Analysis. *Natural Language Processing Journal*, 7:100073.

George Julian. 2020. What are the most spoken languages in the world. *Retrieved May*, 31(2020):38.

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.

Prashant Kodali, Anmol Goel, Likhith Asapu, Vamshi Krishna Bonagiri, Anirudh Govil, Monojit Choudhury, Ponnurangam Kumaraguru, and Manish Shrivastava. 2024. From human judgements to predictive models: Unravelling acceptability in code-mixed sentences. *arXiv preprint arXiv:2405.05572*.

Gitanjali Kumari, Kirtan Jain, and Asif Ekbal. 2024. M3Hop-CoT: Misogynous meme identification with multimodal multi-hop chain-of-thought. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 22105–22138, Miami, Florida, USA. Association for Computational Linguistics.

Sapna Kumari. 2021. Meme culture and social media as gendered spaces of dissent and dominance. *Journal of Visual Literacy*, 40(3-4):215–232.

Bing Li, Yujie He, and Wenjin Xu. 2021. Cross-lingual named entity recognition using parallel corpus: A new approach using xlm-roberta alignment. *arXiv preprint arXiv:2101.11112*.

Wentao Ma, Yiming Cui, Chenglei Si, Ting Liu, Shijin Wang, and Guoping Hu. 2020. Charbert: Character-aware pre-trained language model. *arXiv preprint arXiv:2011.01513*.

MD. SHAFIQUL HASAN Ashim Dey Madiha Ahmed Chowdhury, Lamia Tasnim Khan. 2025. Cuet_320@lt-edi-2025: A multimodal approach for misogyny meme detection in chinese social media. In *Proceedings of the Fifth Workshop on Language Technology for Equality, Diversity and Inclusion*, Italy. Fifth Conference on Language, Data and Knowledge (LDK2025).

Sidharth Mahesh, D Sonith, Gauthamraj Gauthamraj, G Kavya, Asha Hegde, and H Shashirekha. 2024. MUCS@ LT-EDI-2024: Exploring Joint Representation for Memes Classification. In *Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity, Inclusion*, pages 282–287.

Jidan Al Abrar Md Mehedi Hasan Md Siddikul Imam Kawser Mohammad Shamsul Arefin Md Mubasshir Naib, Md Mizanur Rahman. 2025. Cuets_white_walkers@lt-edi-2025: A multimodal framework for the detection of misogynistic memes in chinese online content. In *Proceedings of the Fifth Workshop on Language Technology for Equality, Diversity and Inclusion*, Italy. Fifth Conference on Language, Data and Knowledge (LDK2025).

Mohammad Oman Mohammad Shamsul Arefin MD.Mahadi Rahman, Mohammad Minhaj Uddin. 2025. Cuet_ignite@lt-edi-2025: A multimodal transformer-based approach for detecting misogynistic memes in chinese social media. In *Proceedings of the Fifth Workshop on Language Technology for Equality, Diversity and Inclusion*, Italy. Fifth Conference on Language, Data and Knowledge (LDK2025).

Nabilah Tabassum Samia Rahman Hasan Murad Mehreen Rahman, Faozia Fariha. 2025. Cuet_12033lt-edi-2025: Misogyny detection. In *Proceedings of the Fifth Workshop on Language Technology for Equality, Diversity and Inclusion*, Italy. Fifth Conference on Language, Data and Knowledge (LDK2025).

Parth Patwa, Sathyanarayanan Ramamoorthy, Nethra Gunti, Shreyash Mishra, S Suryavardan, Aishwarya Reganti, Amitava Das, Tanmoy Chakraborty, Amit Sheth, Asif Ekbal, and Chaitanya Ahuja. 2022. Findings of memotion 2: Sentiment and emotion analysis of memes. In *Proceedings of De-Factify: Workshop on Multimodal Fact Checking and Hate Speech Detection, CEUR*.

Sitara K Radhika K T. 2025. Cvf-nitt@lt-edi-2025:misogynydetection. In *Proceedings of the Fifth Workshop on Language Technology for Equality, Diversity and Inclusion*, Italy. Fifth Conference on Language, Data and Knowledge (LDK2025).

Giulia Rizzi, David Gimeno-Gómez, Elisabetta Fersini, and Carlos-D Martínez-Hinarejos. 2024. Pink at exist2024: a cross-lingual and multi-modal transformer approach for sexism detection in memes. *Working Notes of CLEF*.

Debopriya Deb Roy, Israt Moyeen Noumi, and Md Aminur Rahman. 2024. Towards Better Misogyny Detection in Bangla: Improved Dataset and Cutting-Edge Model Evaluation. In *2024 6th International Conference on Sustainable Technologies for Industry 5.0 (STI)*, pages 1–6. IEEE.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Chhavi Sharma, Deepesh Bhageria, William Scott, Srinivas PYKL, Amitava Das, Tanmoy Chakraborty, Viswanath Pulabaigari, and Björn Gambäck. 2020. SemEval-2020 task 8: Memotion analysis- the visuolingual metaphor! In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 759–773, Barcelona (online). International Committee for Computational Linguistics.

Bharathi B Sreeja K. 2025. Ssncse@lt-edi-2025:detecting misogyny memes using pretrained deep learning models. In *Proceedings of the Fifth Workshop on Language Technology for Equality, Diversity and Inclusion*, Italy. Fifth Conference on Language, Data and Knowledge (LDK2025).

Wei Zhang. 2012. Chinese-English code-mixing among China's netizens: Chinese-English mixed-code communication is gaining popularity on the Internet. *English Today*, 28(3):40–52.

Xianwei Zhang, Peng Wu, Jiuming Cai, and Kun Wang. 2019. A Contrastive Study of Chinese Text Segmentation Tools in Marketing Notification Texts. *Journal of Physics: Conference Series*, 1302(2):022010.

Qi Zhong, Qian Wang, and Ji Liu. 2022. Combining knowledge and multi-modal fusion for meme classification. In *International Conference on Multimedia Modeling*, pages 599–611. Springer.