AURORA: Automated Understanding and Recognition of Omnilingual Misinformation Artefacts

Kailasa Sai Charan[1], Utkrisht Suman[1], Roshan Jain[1], Jasneet Kaur[1] and Shakshi Sharma[1,*]

[1]School of Artificial Intelligence, Bennett University, Greater Noida, UP, India

Abstract

Misinformation has disseminated more quickly due to social media's explosive growth, especially during politically delicate or crisis-driven situations, such as the conflict between Russia and Ukraine. To lessen its detrimental effects on society and assist policy makers, journalists, and social media moderators, early detection of false information is crucial. In this study, we introduce **AURORA**, a generalizable, timely, and multilingual misinformation detection system that categorizes social media messages as either misinformation or non-misinformation along with a confidence score. The PROMID Task 3 dataset, which was made available at the 17th Forum for Information Retrieval Evaluation (FIRE) 2025, is used to train our method. It includes multilingual, highly imbalanced tweets about the conflict between Russia and Ukraine. We utilize several machine learning and transformer-based models after substantial preprocessing, including TF-IDF vectorization and RoBERTa-based embeddings. With a weighted F1-score of 0.81, RoBERTa outperforms traditional baselines like Logistic Regression. Furthermore, we developed a web-based interactive dashboard that shows confidence scores and performance metrics while enabling users to instantly verify any claim in order to improve accessibility and usefulness. Further analyses of current PolitiFact and Boom Live fact-checked articles show how well the approach generalizes across languages and domains. The portal can be accessed via this link: https://misinfo-4.onrender.com/.

Keywords

Misinformation, Multilingual, Code mix language, Russia Ukraine War

1. Introduction

One of the biggest threats to society is misinformation. In addition to co-opting "useful idiots," many entrenched interests have a stake in creating and disseminating false information, including by hiring paid workers to do so, inflicting chaos or damage to societies.

The response and policing strategy against misinformation must include the ability to keep an eye on the types of misinformation that are gaining traction and to quickly refute false information, especially those that have a tendency to persist, using a dedicated team of personnel or (semi-)automated tools. A step in that approach is the proposed dashboard, **AURORA**. Policymakers and "first responders" in the social media domain would be able to both (i) quickly and succinctly detect and comprehend the most common misinformation and (ii) take advantage of ready-to-use responses that were automatically generated from the social media corpus itself. This automated tool helps to timely verify the claims as quickly as they posted on social media. Moreover, this tool is not language specific tool i.e. any multilingual post irrespective of the language and topic can be utilized to verify the claim. To check its generalizability, we have further tested three test case scenarios that verify our statement.

In this work, misinformation detection model is being trained on a highly imbalanced multilingual Russia-Ukraine war dataset. The data is provided by the organizers of Shared Track named Prompt RecOvery For Misinformation Detection (**PROMID**) as a **SubTask 3**: *Misinformation Detection in social media texts*. PROMID is one of the many Shared Tracks hosted by the highly prestigious **17th Forum for Information Retrieval Evaluation (FIRE)** Conference[1] organized by **Indian Institute**

Forum for Information Retrieval Evaluation, December 17-20, 2025, India

*Corresponding author.

A24ARIU0024@bennett.edu.in (K. S. Charan); A24ARIU0001@bennett.edu.in (U. Suman); A24ARIU0013@bennett.edu.in (R. Jain); A24ARIU0015@bennett.edu.in (J. Kaur); shakshi.sharma@bennett.edu.in (S. Sharma)

https://sites.google.com/view/shakshi-sharma/home (S. Sharma)

0000-0001-8091-0781 (S. Sharma)

[1]https://fire.irsi.org.in/fire/2025/call_for_papers

of Technology (IIT) BHU, Varanasi in **2025**. Specifically, we participated on one of the three subtasks by PROMID [1, 2]. More details on the task can be found in Section 3 System Architecture.

We developed a dashboard **AURORA** from this subtask 3 of misinformation classification model and can be accessed via this link: https://misinfo-4.onrender.com/. **Please note** that due to free instance of Render API, this website might cause *delay* of approx. *2 minutes* in showing the webpage. More details of dataset wrangling, classification models and dashboard can be found in Section 3 System Architecture.

2. Related Work

Previous research [3, 4, 5] on misinformation, focused on a number of factors, such as detecting misinformation in multimodal settings [6] and in Dravidian languages as a Shared Task [7], recommendation techniques to fake news [8, 9]. In addition, there has been works that studies user-level information on social media including echo chambers detection, community detection, influential (or powerful) nodes, etc. [10, 11, 12]. Researchers have also employed more advanced techniques including combining deep learning and graph structure namely, Graph Neural Networks (GNNs) [13] and Knowledge Graphs [14].

One of the main goals of previous fake news studies has been the identification of false information mainly in English language [15, 16, 17, 18] or separate models for each language [19]; multilingual has received less attention [5, 20]. Moreover, early detection of false news is still a challenging task, thus, we developed a model that can automatically detect any recent news article irrespective of the language and topic.

3. System Architecture

This section focuses on the main architecture of this work. Specifically, we discuss the dataset used along with its preprocessing, training the models based on the dataset, and finally, an interactive web based dashboard **AURORA**, where a user can ask on the fly any news being misinformation or not.

3.1. Data acquisition & Wrangling

Data collection: We used the dataset provided by the FIRE Conference[2] 2025 Shared Track named Prompt RecOvery For Misinformation Detection (PROMID) Task 3 i.e. *Misinformation Detection in social media texts*. Precisely, this task aims to classify the tweets pertaining to Russia-Ukrainian War conflict as misinformation or non-misinformation tweets. The dataset was collected using the official Twitter API during the first year of the conflict and manually annotated by [21] following the framework [22]. This dataset contains multilingual words and highly imbalanced dataset that further make the problem complex for generalization purposes.

Data Wrangling: Organizers provided two train data files i.e. misinformation & non-misinformation tweets along with their labels & one test file for the final submission. We first merge these train set files together containing total rows 8,388. Next, in order to make the data ready for model training in the next phase, we performed tokenization, label encoding, and TF-IDF vectorization from Sklearn library[3]. We performed the same preprocessing techniques for the test set as well.

3.2. Algorithm Training & Testing Phase

Once the preprocessing is completed, we move to the next phase i.e. model training and testing. For model training, we first divide the train set into train & validation sets in an 85:15 ratio. In order to convert the words into embeddings, we utilized RoBERTa based model for embedding vector generation. Next, we employed multiple Machine Learning and Deep Learning Models. However, we found the top

[2]https://fire.irsi.org.in/fire/2025/call_for_papers
[3]https://scikit-learn.org/stable/

two best performing models for this task i.e. Logistic Regression from Sklearn library[4] and Roberta based Transformer model from HuggingFace library[5]. The hyper parameters tuning is shown in Table 2 for TF-IDF Vectorization, Logistic Regression, and roBERTa models, remaining of the hyper parameters were taken as default values.

The evaluation of the trained models has been tested using the validation set as shown in Table 1. It has been noted that RoBERTa based transformer model outperforms the other in all of the evaluation metrics i.e. Accuracy, Precision, Recall, and F1 Score. Considering the fact that the dataset is highly imbalanced, we not just rely on the accuracy which is a bad metric in such cases, we used Precision, Recall and F1 Score. The organizers also focuses on the weighted-averaged F1 Score to measure each team's performance in this shared task.

	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.91	0.62	0.72	0.65
RoBERTa based transformer	**0.97**	**0.88**	**0.77**	**0.81**

Table 1
Evaluation metrics including Accuracy, Precision, Recall, and F1 Score on top two best performing models i.e. Logistic Regression and RoBERTa based transformer models on the validation set. Bold numbers indicate the highest value of the metric.

TF-IDF	Logistic Regression	RoBERTa
max_features = 50000	class_weight = "balanced"	learning_rate = 2e-5
ngram_range = (1, 2)	max_iter = 300	num_train_epochs = 2

Table 2
Hyper parameter tuning of models.

Submissions: The best performing model in our case i.e.*RoBERTa based transformer model* has been utilized to perform the classification task for the test data released by the organizers via Coda Bench[6]. We submitted the submission.csv files first in the development phase and then in the final/test phase. The csv contains two columns only i.e. id and label. There were total 27 participants and 233 submissions for this task.

3.3. Web-based Interactive Dashboard

Our objective is not just to perform the classification on the multilingual dataset, instead, we developed an interactive web-based dashboard **AURORA** and deployed free using Render[7] API. The focus is to ask any news article on the fly using the dashboard and it provides you with the classification task along with *probability* or confidence score of the trained model as can be seen in Figure 1. Our dashboard **AURORA** can be accessed via this link: https://misinfo-4.onrender.com/. **Please note** that due to free instance of Render API, this website might cause *delay* of approx. *2 minutes* in showing the webpage.

To be more transparent, we also provided the confusion matrix including True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) in terms of bar plot to show the evaluation to the user using this dashboard in Figure 2. It can be seen that the FP and FN are quite lower leading to the better model performance.

In order to test the ***generalizability*** of this detection model, we used fact-checking websites ie PolitiFact and Boom live to check the performance of our trained model on the trending topics. We tested on various scenarios as mentioned below:

[4]https://scikit-learn.org/stable/
[5]https://huggingface.co/docs/transformers/model_doc/roberta
[6]https://www.codabench.org/competitions/10869/
[7]https://render.com/

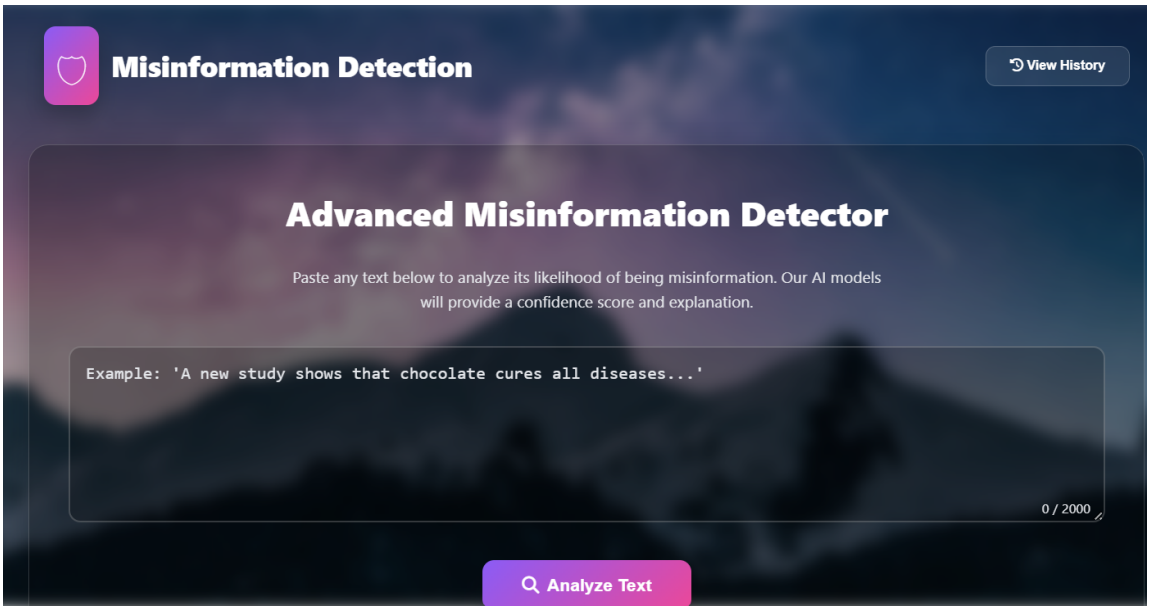

Figure 1: Landing Page of the Misinformation Detector Dashboard **AURORA**.

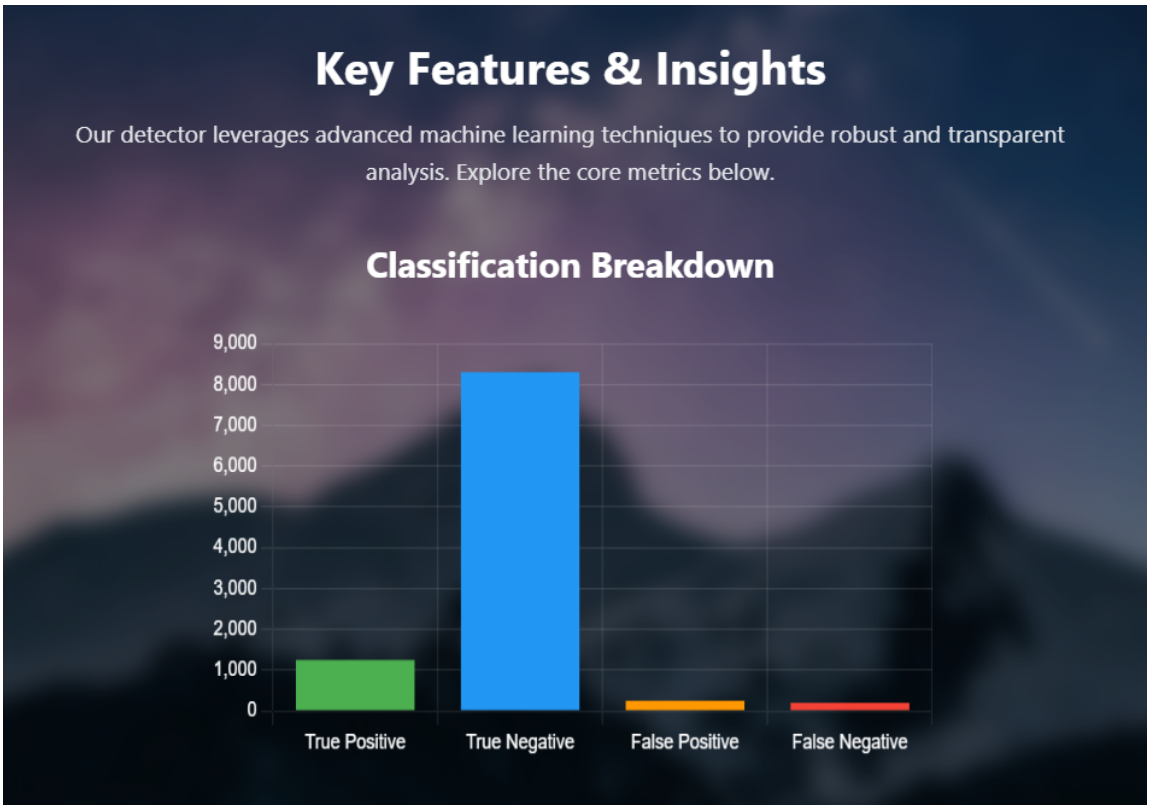

Figure 2: Model's Performance on the train set. X-axis and y-axis correspond to True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN) and count of the instances or rows from the dataset.

1. *Test Case Scenario 01:* We used the recent post of US President Donald Trump posted on 17 November, 2025 on the fact-checking website called Polifact.com[8]. The website claimed to be in *False* category. Next, in Figure 3, we copy the same claim from the fact-check website and check the prediction of our trained model which is Misinformation i.e. False which is same as

[8]https://www.politifact.com/factchecks/2025/nov/17/tiktok-posts/Donald-Trump-Bill-Clinton-Bubba-touch-crotch-video/

the fact-check article label. Moreover, the model provides the 69% confidence score stating the certainty of the predictions given by the model.

2. *Test Case Scenario 02:* Now, we test a claim that is labeled as *True* by the Politifact website. In that respect, we used topic of China's trading posted in 05 May, 2025 on the website[9]. As can be observed in Figure 4, model predicted it as True i.e. not a misinformation with 90% confidence score.
3. *Test Case Scenario 03:* Besides English, we also tested the recent news posted on 27 October, 2025 on *Hindi* language as well from Boom Live Hindi[10]. As can be seen in Figure 5, the predicted label is misinformation with 68% confidence i.e. False same as claimed in the fact-check article. Hence, the model able to handle multilingual languages as well.

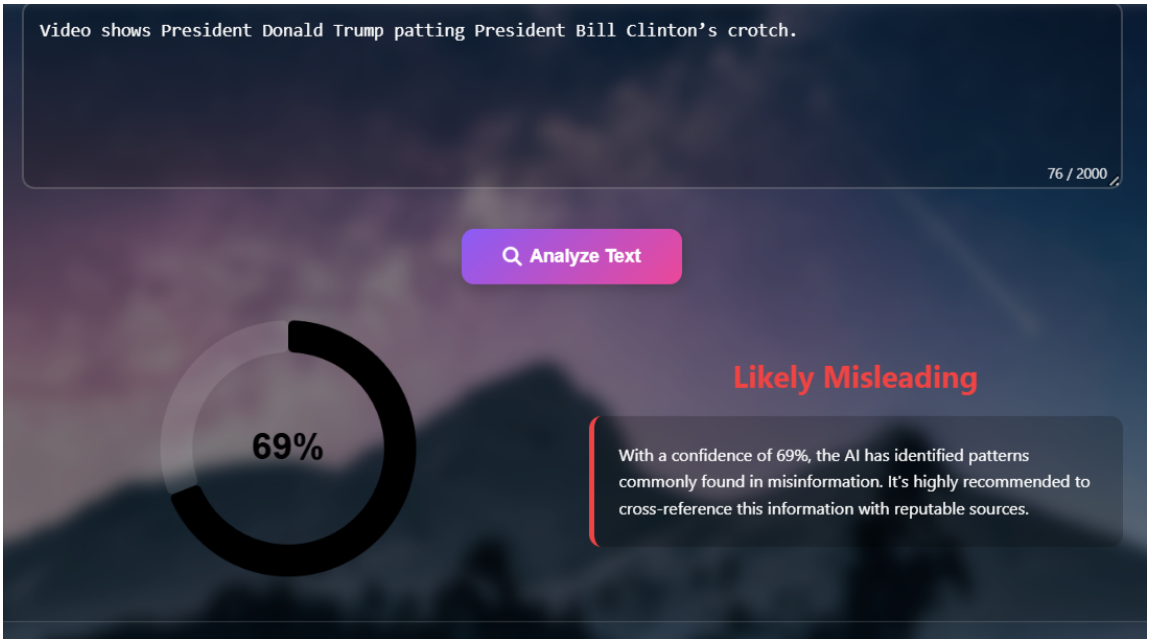

Figure 3: Test Case Scenario 01: *Fake* news article from PolitiFact fact check website.

Additional Feature: In Figure 6, we also added a feature to check the history of all the news articles predicted by the model along with the prediction and confidence score. It is possible to view each article and delete it as well in order to provide customization to user. Furthermore, user can clear all the history at once too.

4. Conclusion

In this study, we presented a multilingual misinformation detection framework that can detect fake content in a variety of languages and topics. Our approach showed significant generalizability when tested on real-world fact-checking scenarios and achieved competitive performance on a highly imbalanced dataset provided by the PROMID Shared task at FIRE Conference, IIT, BHU in 2025. By providing real-time verification capabilities and clear evaluation indicators, the interactive dashboard **AURORA** further closes the gap between research and practical deployment. Despite the system's encouraging outcomes, there are still a number of directions for future research. First, multimodal settings like images and metadata—which are increasingly employed in fake news campaigns—can be incorporated into the model to improve it. Second, robustness could be further enhanced by resolving class imbalance using sophisticated methods like contrastive learning or data augmentation. Lastly, incorporating large

[9]https://www.politifact.com/factchecks/2025/may/09/charles-blow/china-toys-christmas-goods-Trump-tariffs/
[10]https://hindi.boomlive.in/fact-check/lucknow-phoenix-mall-cheetah-ai-video-fake-claim-29837

Figure 4: Test Case Scenario 02: *True* news article from PolitiFact fact check website.

Figure 5: Test Case Scenario 03: *Multilingual* news article from Boomlive Hindi fact check website.

language model (LLM)-driven explanation features or retrieval-based fact-checking modules could give users more thorough, empirically supported explanations for each prediction.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

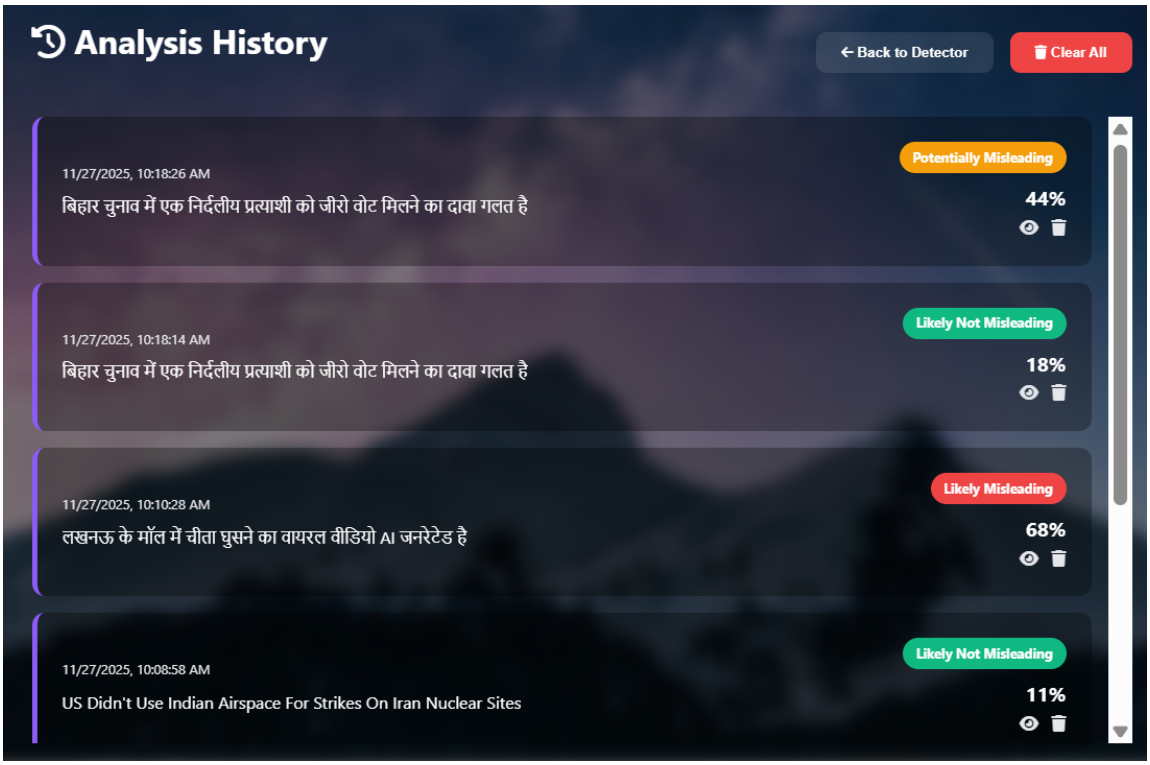

Figure 6: History of the news articles verified on the Dashboard.

References

[1] A. Hegde, G. K. Shahi, S. Satapara, P. Mehta, S. Modha, D. Ganguly, D. Nandini, H. L. Shasirekha, A. K. Jaiswal, G. Pasi, T. Mandl, Prompt recovery for misinformation detection at fire 2025, in: Proceedings of the 17th Annual Meeting of the Forum for Information Retrieval Evaluation, FIRE '25, Association for Computing Machinery, 2025.

[2] G. K. Shahi, A. Hegde, S. Satapara, P. Mehta, S. Modha, D. Ganguly, D. Nandini, H. L. Shasirekha, A. K. Jaiswal, G. Pasi, T. Mandl, Overview of the first shared task on prompt recovery for misinformation detection (promid 2025), in: K. Ghosh, T. Mandl, S. Pal, S. Majumdar, A. Chakraborty (Eds.), Working Notes of FIRE 2025 - Forum for Information Retrieval Evaluation, Varanasi, India. December 17-20, 2025, CEUR Workshop Proceedings, CEUR-WS.org, 2025.

[3] V. Balakrishnan, N. W. Zhen, S. M. Chong, G. J. Han, T. J. Lee, Infodemic and fake news–a comprehensive overview of its global magnitude during the covid-19 pandemic in 2021: A scoping review, International Journal of Disaster Risk Reduction (2022) 103144.

[4] S. Sharma, E. Agrawal, R. Sharma, A. Datta, Facov: Covid-19 viral news and rumors fact-check articles dataset, in: Proceedings of the International AAAI Conference on Web and Social Media, volume 16, 2022, pp. 1312–1321.

[5] X. Zhou, R. Zafarani, A survey of fake news: Fundamental theories, detection methods, and opportunities, ACM Computing Surveys (CSUR) 53 (2020) 1–40.

[6] L. Zhang, X. Zhang, Z. Zhou, X. Zhang, P. S. Yu, C. Li, Knowledge-aware multimodal pre-training for fake news detection, Information Fusion 114 (2025) 102715. URL: https://www.sciencedirect.com/science/article/pii/S1566253524004937. doi:https://doi.org/10.1016/j.inffus.2024.102715.

[7] M. Subramanian, P. B, K. Shanmugavadivel, S. Pandiyan, B. Palani, B. R. Chakravarthi, Overview of the shared task on fake news detection in Dravidian languages-DravidianLangTech@NAACL 2025, in: B. R. Chakravarthi, R. Priyadharshini, A. K. Madasamy, S. Thavareesan, E. Sherly, S. Rajiakodi, B. Palani, M. Subramanian, S. Cn, D. Chinnappa (Eds.), Proceedings of the Fifth

Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, Association for Computational Linguistics, Acoma, The Albuquerque Convention Center, Albuquerque, New Mexico, 2025, pp. 759–767. URL: https://aclanthology.org/2025.dravidianlangtech-1.128/. doi:10.18653/v1/2025.dravidianlangtech-1.128.

[8] S. Wang, X. Xu, X. Zhang, Y. Wang, W. Song, Veracity-aware and event-driven personalized news recommendation for fake news mitigation, in: Proceedings of the ACM Web Conference 2022, 2022, pp. 3673–3684.

[9] D. You, N. Vo, K. Lee, Q. Liu, Attributed multi-relational attention network for fact-checking url recommendation, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019, pp. 1471–1480.

[10] O. Ozcelik, C. Toraman, F. Can, Detecting misinformation on social media using community insights and contrastive learning, ACM Trans. Intell. Syst. Technol. 16 (2025). URL: https://doi.org/10.1145/3709009. doi:10.1145/3709009.

[11] A. Mahmoudi, D. Jemielniak, L. Ciechanowski, Echo chambers in online social networks: A systematic literature review, IEEE Access 12 (2024) 9594–9620. doi:10.1109/ACCESS.2024.3353054.

[12] N. Ansar, D. S. Hashmat, The politics of misinformation: Fake news, echo chambers, and public perception, International "Journal of Academic Research for Humanities" 5 (2025) 14–24. URL: https://jar.bwo-researches.com/index.php/jarh/article/view/543.

[13] H. T. Phan, N. T. Nguyen, D. Hwang, Fake news detection: A survey of graph neural network methods, Applied Soft Computing 139 (2023) 110235. URL: https://www.sciencedirect.com/science/article/pii/S1568494623002533. doi:https://doi.org/10.1016/j.asoc.2023.110235.

[14] B. Xie, X. Ma, J. Wu, J. Yang, H. Fan, Knowledge graph enhanced heterogeneous graph neural network for fake news detection, IEEE Transactions on Consumer Electronics 70 (2024) 2826–2837. doi:10.1109/TCE.2023.3324661.

[15] S. Sharma, Y. Mu, R. Sharma, N. Aletras, Bluff: Behavioral characterization of misinformation imposters and active citizens on online social media (2025).

[16] S. Sharma, A. Datta, R. Sharma, Amir: An automated misinformation rebuttal system–a covid-19 vaccination datasets-based exposition, IEEE Transactions on Computational Social Systems (2024).

[17] S. Sharma, A. Datta, V. Shankaran, R. Sharma, Misinformation concierge: a proof-of-concept with curated twitter dataset on covid-19 vaccination, in: Proceedings of the 32nd ACM international conference on information and knowledge management, 2023, pp. 5091–5095.

[18] M. Mayank, S. Sharma, R. Sharma, Deap-faked: Knowledge graph based approach for fake news detection, in: 2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE, 2022, pp. 47–51.

[19] M. Dhawan, S. Sharma, A. Kadam, R. Sharma, P. Kumaraguru, Game-on: Graph attention network based multimodal fusion for fake news detection, Social Network Analysis and Mining 14 (2024) 114.

[20] A. De, D. Bandyopadhyay, B. Gain, A. Ekbal, A transformer-based approach to multilingual fake news detection in low-resource languages, Transactions on Asian and Low-Resource Language Information Processing 21 (2021) 1–20.

[21] G. K. Shahi, Y. Mejova, Too little, too late: Moderation of misinformation around the russo-ukrainian conflict, in: Proceedings of the 17th ACM Web Science Conference 2025, 2025, pp. 379–390.

[22] G. K. Shahi, T. A. Majchrzak, Amused: An annotation framework of multimodal social media data, 2022.