Overview of the First Shared Task on Prompt Recovery for Misinformation Detection (PROMID 2025)

Gautam Kishore Shahi[1], Asha Hegde[2], Shrey Satapara[3], Parth Mehta[4], Sandip Modha[5], Debasis Ganguly[6], Durgesh Nandini[7], H L Shashirekha[1], Amit Kumar Jaiswal[8], Gabriella Pasi[5] and Thomas Mandl[9]

[1]*University of Duisburg-Essen, Germany*
[2]*Mangalore University, India*
[3]*Fujitsu Research, India*
[4]*Parmonic, USA*
[5]*Università degli Studi di Milano-Bicocca, Italy*
[6]*University of Glasgow, United Kingdom*
[7]*University of Bayreuth, Germany*
[8]*Indian Institute of Technology (BHU) Varanasi, India*
[9]*University of Hildesheim, Germany*

Abstract

With the increasing use of Large Language Models (LLMs) for content creation and information dissemination, the problem of understanding and alleviating misinformation and hallucinations in these systems has become an important research topic. However, the existing evaluation mechanisms do not account for the role of prompts, the effects of cross-lingual generation, and real-world events in the creation of misinformation. This was the primary motivation behind this shared task on Prompt Recovery for Misinformation Detection (PROMID), organised as a part of the 17th Forum for Information Retrieval Evaluation (FIRE) in 2025[1]. PROMID 2025 focused on three relatively unexplored problems: (i) Prompt recovery aiming at recovering the possible input prompt used for generating misinformation, (ii) Identification of factual incorrectness in machine-generated cross-lingual summaries, and (iii) Classification of misinformation in Twitter messages related to the February 2022 Russo–Ukrainian conflict. The shared task is divided into three subtasks, and we received a total of 16 submissions, with 11 teams finally submitting working notes. Out of these, task 1 received three submissions, with none of the teams submitting working notes, as all submissions were invalid. Task 2 received four submissions, with all teams submitting the working notes. Task 3 got 12 submissions, with 9 teams submitting the working notes. In this paper, we discuss the motivation behind the three tasks, their problem definitions, datasets and the participants' approaches.

Keywords

Prompt recovery, LLMs, Misinformation detection, Textual summaries

1. Introduction

In the past few years, the use of LLMs for information dissemination has increased exponentially. It is now very common for various media outlets to have at least a LLM generated summary, and in many cases entire articles are AI generated. Likewise, use of LLM in writing social media posts, blogposts, etc has become commonplace. With this rise in the use of LLMS, also rise the challenges related to unintentionally or intentionally generated false information being consumed on a large scale. However, when it comes to deriving systemic insights regarding the phenomenon of misinformation

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

gautam.shahi@uni-due.de (G. K. Shahi); hegdekasha@gmail.com (A. Hegde); shreysatapara@gmail.com (S. Satapara); parth.mehta126@gmail.com (P. Mehta); sjmodha@gmail.com (S. Modha); debasis.ganguly@glasgow.ac.uk (D. Ganguly); durgesh.nandini@uni-bayreuth.de (D. Nandini); hlsrekha@mangaloreuniversity.ac.in (H. L. Shashirekha); amit.chr@iitbhu.ac.in (A. K. Jaiswal); gabriella.pasi@unimib.it (G. Pasi); mandl@uni-hildesheim.de (T. Mandl)
0000-0001-6168-0132 (G. K. Shahi); 0000-0002-6382-4096 (A. Hegde); 0000-0001-6222-1288 (S. Satapara); 0000-0002-4509-1298 (P. Mehta); 0000-0003-2427-2433 (S. Modha); 0000-0003-0050-7138 (D. Ganguly); 0000-0002-9416-8554 (D. Nandini); 0000-0002-8078-6095 (H. L. Shashirekha); 0000-0001-8848-7041 (A. K. Jaiswal); 0000-0002-6080-8170 (G. Pasi); 0000-0002-8398-9699 (T. Mandl)

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073
published 2026-02-24

and hallucinations we are only scratching the surface. Most of the current work around combating these issues is focused on the task of detecting misinformation, akin to traditional fact checking tasks. However, there is a lack of study around the origins of such misinformation. For example, how do specific prompts result in different types of misinformation, how well-designed are the internal safeguards that are supposed to prevent an LLM from generating misinformation, etc. A further study is warranted into the effects of the generation of cross-lingual misinformation, where the source article (assuming there is one) and the resulting misleading article are in different languages. In abstractive summarization, human evaluations have revealed substantial rates of unfaithful or fabricated information, even for strong neural systems, with estimates of 20–30% of summaries containing factual errors [2, 3, 4]). Recent surveys further argue that hallucination is a structural property of LLMs rather than an isolated bug, and highlight the need for systematic benchmarks and detection methods, especially in high-stakes domains and downstream applications 2024 [5, 6]).

These concerns are amplified in multilingual and cross-lingual settings. For Indian languages in particular, the last few years have seen significant progress in building summarization datasets and models across mono, multi, and cross-lingual setups, including large-scale resources such as PMIndiaSum and related corpora that span multiple Indian language families [7]). Shared tasks like ILSUM[8] and HASOC[9] have played a key role in this ecosystem by standardizing evaluation and fostering community efforts specifically around Indo-European and Dravidian languages. The ILSUM 2023 edition, for instance, provided large-scale article–summary pairs across Hindi, Gujarati, Bengali and Indian English, and included a subtask on detecting factual errors in LLM-generated summaries [8, 10]). However, most existing benchmarks (for both global and Indic settings) primarily focus on summary quality (e.g., fluency, adequacy, ROUGE) or treat factuality as a single binary label, without offering a fine-grained view of how hallucinations manifest. At the same time, there is increasing evidence that hallucinations and factual inconsistencies behave differently in multilingual and cross-lingual summarisation than in purely monolingual English settings. Models must simultaneously perform translation, content selection, and compression, which can introduce subtle errors such as incorrect entity mappings, wrong numerical quantities, or cross-lingual semantic drift. Recent work on multilingual and cross-lingual summarisation, including for Indian languages, has highlighted these challenges and pointed to the need for specialised evaluation and mitigation strategies [7, 8]). More efforts need to be put into building benchmarks that explicitly target the factual correctness of machine-generated cross-lingual summaries for Indic languages. Especially in realistic news scenarios where such summaries are consumed by large populations and potentially contribute to the spread of misinformation. The PROMID 2025 task is designed to address this gap.

The task was organized as a part of the 17th Forum for Information Retrieval Evaluation (FIRE 2024)[1]. Traditionally FIRE has focused on shared tasks with the general cross-lingual, low-resource setting focus of FIRE[1], geared towards but not limited to south asian languages. Some of the past shared tasks include hate speech detection [11, 12, 13, 14, 15], sentiment analysis [16, 17, 18], fake news detection [19, 20, 21], machine translation [22], mixed script IR [23, 24], Indian legal document retrieval and summarization [25, 26, 27, 28, 29, 30], authorship attribution [31, 32], IR from microblogs [33], IR for software engineering [34, 35] among others. With the current shared task we aim to continue that legacy of FIRE and contribute to the broader areas of hallucination and misinformation detection. We also aim for more inclusiveness, introducing several Indo-European and Dravidian languages in research areas that are often English-centric 2020 [2, 4]).

We offer three independent tasks related to these problems. Task 1 focuses specifically on the role of prompts used to purposefully generate misinformation. It aims to explore the extent to which it is possible to predict the intention behind generating a specific title for a given news article. While determining the exact intent is a multifaceted study, we specifically focus on externalising this intent in the form of the prompt that was used to generate the misleading title. To this extent, the first task focuses on predicting the prompt that was used to generate a given misleading title from an article. This is a heavily under-researched problem, which was introduced in a Kaggle competition by Google[36]. There have been some attempts at prompt recovery[37, 38] but a more systemic study and public benchmark datasets are needed to address this problem. Task 1 of PROMID attempts at bridging that gap.

Task 2 in this track, focuses on detecting factual incorrectness in machine-generated cross-lingual summaries. Given a source article in English and a corresponding LLM-generated summary in an Indian language, systems must determine whether the summary is factually correct and, when it is not, assign one or more fine-grained error labels. We consider four broad types of factual incorrectness: misrepresentation, inaccurate quantities or measurements, false attribution, and fabrication, chosen to align with taxonomies of hallucination proposed in recent LLM surveys while remaining interpretable for downstream users and annotators (Huang et al., 2024; Sahoo et al., 2024 [5, 6]). We expand the existing ILSUM datasets to include Dravidian languages such as Kannada, Tamil, Telugu and Malayalam.

Task 3 focuses on the identification of misinformation in real life setting. This task aims to develop a model capable of classifying tweets related to the Russo-Ukrainian conflict as either misinformation (positive class) or non-misinformation (negative class). This is closer to a traditional fact-checking task, in an automated setting.

In the remainder of the paper, we first describe the dataset creation process and dataset statistics. We then outline the official evaluation setup, followed by a summary of participating systems and their performance. Finally we conclude by highlighting open challenges and directions for future work on factuality, misinformation, and prompt recovery in Indian-language LLM applications.

2. Related Work

Prompt recovery is a relatively unexplored research area, that has been slowly gaining traction in past couple of years. However, the approaches remain limited. The problem was first introduced in a Kaggle competition by Google[36]. The problem, however was not geared towards misinformation, but rather towards generating stylistic variation of texts (e.g. write this in a Shakespearean style). There have been some other attempts at prompt recovery[37, 38] but a more systemic study and public benchmark datasets are needed to address this problem. Task 1 of PROMID attempts at bridging that gap.

Compared to that the misinformation and hallucination detection problem has seen an ever-increasing amount of interest. The problem is also closely related to other tasks such as fact-checking, detecting AI-generated content, etc. To overcome this problem, a variety of models and benchmark platforms have been proposed in the last few years. Singhal et al. [39] proposed a multilingual fact-checking benchmark by filtering and binarising the X-Fact claim data for five languages (Spanish, Italian, Portuguese, Turkish, and Tamil). They also compared the performance of five large language models on various prompting techniques – zero-shot, English Chain-of-Thought, cross-lingual prompting, and their respective self-consistency methods. They employed Statistical analysis, two-way ANOVA, and correlation analysis to analyze the impact of models, methods, and language factors on performance. The work by Chikkala et al. [40] involves manually creating a high-quality, bilingual English–Telugu fact-checking dataset through claim curation, cleaning, and annotation with veracity labels, gold justifications, and multiple types of QA pairs, followed by careful translation and post-editing for Telugu. Large language models are then benchmarked under four settings: simple zero-shot prompting and three retrieval-augmented approaches (Naive RAG, Advanced RAG, which includes query rewriting, re-ranking, and prompt compression, and Automatic Scraping of up-to-date news content). Claims are verified and justifications generated by multiple LLMs; performance is evaluated in terms of F1 scores for veracity classification and through a suite of automatic metrics for justification and QA quality. This setup enables a systematic comparison of prompting versus retrieval-based methods across highand low-resource languages.

Further many shared tasks across evaluation platforms have been actively focusing on these tasks. Numerous tasks have been offered in platforms like CLEF, TREC, SemEval and FIRE. Some of the recent editions of these tasks include Checkthat Lab in CLEF[41, 42, 43], LLM Capabilities and Fact Checking and Knowledge verification themes at SemEval[44, 45], Lateral Reading Task as TREC[46] and ILSUM task in FIRE[47, 48]. ILSUM track at FIRE is perhaps the most relevant task to the PROMID task. In a way, PROMID 2025 is the spiritual successor of the ILSUM tasks, with the task 2 being a direct continuation of task 2 in ILSUM 2024.

3. Task Definition

PROMID 2025 consists of three independent subtasks[49], all related broadly to the theme of prompt recovery or misinformation detection.

3.1. Task 1: Prompt Recovery from LLM-generated Misinformative Text

In the Prompt Recovery task, participants are given a factual news article *summary* together with a *misinformation-containing title* and are asked to predict the prompt that could have been used to generate the title from the summary in an open-ended prompt generation setting. Unlike tasks that classify misinformation types, Prompt Recovery focuses on reconstructing the *instructional input* (i.e., the prompt text) that drove the transformation from a grounded summary to a misleading title.

Input and Output. Each instance contains:

- **Input:** a news summary s and a generated misinformation-containing title t,
- **Output:** a natural-language prompt p such that a generator conditioned on (s, p) could plausibly produce t.

Train/Test Setup. The training data consists of (s, t, p) triples, where p is the prompt used to produce t from s. The test set contains only (s, t) pairs, and systems must predict the missing prompt $\hat{p}$. While each test instance has a single reference prompt, since the task is open-ended multiple prompts may be semantically valid. The goal is not to generate the exact prompt, but a prompt that is semantically close to the reference prompt.

Repeated Summaries in the Test Set. In the test data, the same summary may appear in multiple instances with different generated titles. This design reflects that a single article can be reframed in multiple misleading ways, and it implies that different prompts were used to generate different titles from the same underlying summary. Systems therefore must condition on both s and t to recover the prompt, rather than relying on summary-only properties.

Task Framing. We treat prompt recovery as a conditional generation problem:

$$\hat{p} = \arg\max_{p} \; P(p \mid s, t),$$

where the goal is to generate a prompt that matches the dataset's prompting style and content closely enough to be identified as the original instruction.

3.2. Task 2: Misinformation Detection in LLM-generated Summaries

Task 2 is the continuation of the task previously offered in ILSUM 2024 and 2023 [8, 50, 51, 10]. The task aims to identify incorrectness in machine-generated summaries, which is an important step in ensuring the reliability and accuracy of information. This year the task included four Dravidian Languages - Kannada, Tamil, Telugu and Malayalam.

We focus on four types of inaccuracies for this task, same as the previous editions:

- **Misrepresentation**: This involves presenting information in a way that is misleading, or that gives a false impression. This could be done by exaggerating certain aspects, understating others, or twisting facts to fit a particular narrative.
- **Inaccurate Quantities or Measurements**: Factual incorrectness can occur when precise quantities, measurements, or statistics are misrepresented, whether through obfuscation (25 -> dozens) or through outright fudging.
- **False Attribution**: Incorrectly attributing a statement, idea, or action to a person or group is another form of factual incorrectness.
- **Fabrication**: Making up data, sources, or events is a severe form of factual incorrectness. This involves creating "facts" that have no basis in reality.

For this task, in the training data, every article has a corresponding summary associated with exactly one of the four types of incorrectness mentioned above. However, during evaluation, participants are asked to predict all possible labels associated with text summaries in test data, as one summary can have multiple types of incorrectness. More details about the dataset creation are available in the dataset paper for previous tasks[52].

3.3. Task 3: Misinformation Detection In Social Media Texts

The aim of this task is to develop a model capable of classifying tweets related to the Russo-Ukrainian conflict as either misinformation (positive class) or non-misinformation (negative class). The dataset consists of manually annotated tweets gathered through the Twitter API during the first year of the conflict, as documented in previous work [53, 54]. Data gathering was carried out using the AMUSED framework [55], which is designed for collecting social media posts from social media platforms [55, 56]. A notable characteristic of this dataset is its substantial class imbalance, making it a useful testbed for evaluating model robustness in scenarios where misinformation is comparatively rare. The misinformation subset includes tweets authored in multiple languages, all of which can be translated or processed by large language models to ensure comparability across linguistic contexts. Additionally, misinformation-labeled tweets contain supplementary metadata such as account age and bot-likelihood indicators; although these attributes are not included for the non-misinformation tweets by default, participants may extract them independently if they wish to enrich their feature set. Model performance is assessed using precision, recall, and weighted F1-score to provide a comprehensive evaluation under imbalanced conditions.

4. Datasets and Evaluation

In this section we discuss the datasets used and and the employed evaluation metric for each subtask.

4.1. Datasets

In the prompt recovery task, participants are provided 9950 training instances, each containing a summary, a prompt and a title containing misinformation generated using the provided prompt. For the test, a total of 800 test instances containing a summary and a title with misinformation were provided, making it an open-ended prompt recovery task.

Task 2 was offered in four Dravidian languages named Telugu, Tamil, Kannada, and Malayalam where participants are asked to predict one of the five categories (four misinformation categories or correct). Detailed train and test dataset statistics for task 2 are available in Table 1.

Task 3 was offered in the English language and dataset comprises 36,174 non-misinformation tweets and 778 misinformation tweets, highlighting a substantial skew toward the negative class. This imbalance is evident in both splits: the training set is even more extreme, with 34,174 non-misinformation tweets and only 364 misinformation tweets while the test set includes 2,000 non-misinformation tweets and 414 misinformation tweets.

Split	Lang.	Fab.	Misrep.	F. Attrib.	Inc. Qty.	Correct	Total
Train	Telugu	250	294	250	195	3986	4975
Train	Tamil	250	294	250	195	3986	4975
Train	Malayalam	250	294	250	195	3986	4975
Train	Kannada	250	294	250	195	3986	4975
Test	Telugu	32	25	13	10	143	200
Test	Tamil	32	25	13	10	143	200
Test	Malayalam	32	25	13	10	143	200
Test	Kannada	32	25	13	10	143	200

Table 1
Training and Test Dataset Statistics for Task 2

4.2. Evaluation

For Task 1, we report ROUGE [57] as a standard lexical-overlap metric, and additionally use BERTScore [58] to measure semantic similarity between the gold prompt used to generate the misinformation title and the recovered prompt. This is important because prompt recovery often admits valid paraphrases with low n-gram overlap, where ROUGE can underestimate performance. We therefore use ROUGE for surface-form comparison and BERTScore to better capture meaning preservation in low-overlap but semantically equivalent cases.

For Task 2, formulated as a multi-class classification problem, we use Macro-F1 as the primary metric. Since the label distribution is imbalanced—particularly between factually correct summaries and instances belonging to specific misinformation types. Macro-F1 ensures that performance on minority classes is not dominated by the majority class.

For Task 3, we use weighted F1 due to the very high label imbalance in the tweet misinformation dataset, and we compute scores via automatic evaluation hosted on Codabench[1] to ensure consistent and reproducible leaderboard ranking.

5. Results and Methodologies

In this section, we present the results for the three subtasks, as well as a summary of the approaches that the participants used for their best-performing runs.

5.1. Task 1

For task 1, all the submissions we received were invalid. Hence, there is no discussion around results or the approaches used by participants in this section.

5.2. Task 2

The results for task 2 are included in table 2. We report the macro-averaged P, R, F and Accuracy for all four languages. In total four teams participated in this task, however one team only submitted runs for Tamil and Kannada. While each team were allowed to submit up to 3 runs, we only report the best performing run for each team here.

Below we give a brief overview of the systems developed by the participated teams for task 2.

gokul [59] - propose a fine-grained misinformation detection system for LLM-generated summaries in Indian languages, targeting Subtask 2 of classifying factual inconsistencies. Fine-tuning of IndicBERTv2-

[1] https://www.codabench.org/competitions/10869

Language	Participant	P	R	F1	Acc.
Tamil	MUCS	0.53	0.37	0.42	0.70
	gokul	0.52	0.36	0.39	0.68
	wangkongqiang	0.35	0.30	0.31	0.66
	priyamsaha	0.11	0.25	0.15	0.62
Telugu	MUCS	0.60	0.44	0.50	0.73
	gokul	0.63	0.43	0.48	0.73
	priyamsaha	0.11	0.25	0.15	0.62
Malayalam	gokul	0.68	0.43	0.47	0.72
	MUCS	0.53	0.42	0.40	0.71
	priyamsaha	0.11	0.25	0.15	0.62
Kannada	gokul	0.80	0.49	0.53	0.75
	MUCS	0.59	0.42	0.48	0.72
	priyamsaha	0.81	0.38	0.45	0.73
	wangkongqiang	0.46	0.31	0.34	0.64

Table 2
Task 2 Results

MLM-only on article-summary pairs is performed by stratified sampling with optimization for macro-F1 across five categories of misinformation. For Tamil, Telugu, Malayalam, and Kannada, separate language-specific models are trained with identical architectures and hyperparameters.

wangkongqiang [60] - The authors developed multiple system variants, including a baseline Logistic Regression (LR) classifier using TF-IDF features, a Dense Neural Network (DNN) trained on distributed text embeddings, and a transformer-based architecture fine-tuned from microsoft-deberta-v3-base. They conducted extensive hyperparameter tuning and ablation studies confirming the transformer-based system consistently outperformed the other approaches across all languages in task 2.

MUCS [61] - proposed a hybrid deep learning approach for misinformation classification using BiLSTM, BiGRU, and Transformer+BiLSTM models with enhanced self-attention mechanisms to address subtask 2. Their model includes personalized subword-level tokenization and strong multilingual preprocessing to effectively preserve the morphological and syntactic differences in Indian languages. They trained the models utilizing class-weighted Cross-Entropy loss and Focal Loss, along with AdamW optimizer, powerful learning rate schedules such as OneCycleLR, CosineAnnealingLR, and mixed-precision training for better efficiency. Their proposed RNN models outperformed others by obtaining strong 1st positions in the Tamil and Telugu tasks with F1 scores of 0.42 and 0.50, respectively, and strong 2nd positions in the Kannada and Malayalam tasks with F1 scores of 0.48 and 0.40, respectively.

priyamsaha [62] - proposed a few-shot learning model for misinformation classification, which is designed to categorise errors in LLM-generated Kannada news summaries. It combines retrieval-based context selection using sentence-transformers/all-MiniLM-L6-v2, Kannada few-shot prompting, and per-label conditional log-probability scoring to assign one of the predefined misinformation categories. For robustness, predictions from Mistral-7B-Instruct and BLOOM-7B1 are aggregated using a Condorcet-style ensemble.

5.3. Task 3

The details of results obtained from Task 3 is shown in Table 3 5.3.

Below we give a brief overview of the systems developed by the participating teams for task 3.

ClimateSense [63] addresses severe class imbalance in misinformation detection by augmenting a RoBERTa-large transformer model with external Ukraine-related misinformation data from the fact-checking observatory. To mitigate overfitting, the authors employ weighted cross-entropy loss and

Participant	ID	Precision	Recall	Weighted-F1
ClimateSense [63]	430584	0.91	0.91	0.91
Sarang [64]	430731	0.90	0.91	0.90
pratikpriyanshu [65]	432337	0.92	0.91	0.89
deepish [66]	429337	0.91	0.89	0.88
priyam_saha17 [62]	431064	0.87	0.80	0.82
whiteby [67]	432078	0.82	0.84	0.82
sushma03 [68]	431997	0.82	0.80	0.85
wangkongqiang [60]	432117	0.78	0.83	0.78
gokul_n_v*	431901	0.78	0.69	0.72
shakshi57 [69]	430489	0.79	–	–
tommathew*	430009	0.71	–	–

Table 3
Final Results obtained from participants of Task 3 (* Participant has not submitted the working notes)

weighted random sampling during fine-tuning. This data-driven enhancement significantly improves recall and F1-score, showcasing the efficacy of targeted dataset expansion for underrepresented classes in transformer-based classification tasks.

Sarang [64] employs a multimodal strategy to address multilingual misinformation detection by first translating all non-English tweets into English using a Gemma-3-12B to ensure linguistic homogeneity. To counteract severe class imbalance, synthetic data augmentation is performed on the minority misinformation class via the same LLM, generating four variants per sample. Then, a DeBERTa-v3-small transformer is fine-tuned on the balanced and translated dataset to capture nuanced semantic patterns, achieving robust performance in cross-lingual settings.

pratikpriyanshu [65] employs a hybrid fusion of multilingual transformer embeddings from XLM-RoBERTa with hand-crafted linguistic features to capture both deep semantic context and surface-level stylistic patterns indicative of misinformation. To address extreme class imbalance, the system integrates class-weighted cross-entropy loss, decision threshold optimization, and stratified cross-validation. Feature concatenation is followed by a sigmoid classifier, enhanced via dropout and mixed-precision training for efficiency. This approach balances representational power with interpretability.

deepish [66] employs a fine-tuned RoBERTa-base transformer model, enhanced with a dynamic optimal thresholding strategy to maximize the F1-score on a severely imbalanced multilingual Twitter dataset. It incorporates a custom preprocessing pipeline that normalizes noise and tokenizes platform-specific features, such as URLs, mentions, and hashtags into dedicated tokens to preserve contextual signals. Class imbalance is mitigated through weighted cross-entropy loss, while training optimizations include gradient accumulation and a linear learning rate scheduler with early stopping.

priyam_saha17 [62] proposes a memory-efficient pipeline for misinformation detection that leverages a frozen RoBERTa encoder to extract contextual embeddings, which are then processed through a trainable projection head and a compact classifier. The methodology employs supervised contrastive learning to enhance representational separation between classes, using dropout to generate stochastic views for contrastive pairs without additional forward passes.

whiteby [67] introduces a hybrid deep learning framework for misinformation detection that integrates semantic embeddings from ModernBERT with hand-crafted feature engineering from X (Twitter) metadata. The model architecture fuses transformer-based text representations with engineered features from text, user profiles, and social engagement, processed through feed-forward networks. To mitigate severe class imbalance, the approach employs Focal Loss with strategic resampling and optimizes classification thresholds via grid search on validation data.

sushma03 [68] presents a fine-tuned BERT-based model for detecting and classifying misinformation in the 2022 Russo-Ukrainian conflict tweets . It employs transfer learning on a pre-trained BERT

architecture, fine-tuning it with a task-specific dataset augmented by fact-checked articles to address class imbalance. The model utilizes a hybrid training approach with defined hyperparameters, including a learning rate of 2e-5 and batch size of 10, achieving classification through weighted F1-score evaluation. Also, the method incorporates external multilingual datasets to enhance cross-domain generalization and improve detection accuracy in imbalanced data scenarios.
wangkongqiang [60] explores misinformation detection in LLM-generated and social media texts using auxiliary text supervised learning. It employs logistic regression, dense neural networks, and recurrent neural networks alongside the transformer-based DeBERTaV3 model. Enhanced through decoupled attention and relative position encoding, DeBERTaV3 is adapted for multi-class and binary classification across multiple languages. Results indicate that ensemble and pre-trained transformer approaches yield competitive performance.
shakshi57 [69] employs RoBERTa-based transformer embeddings for feature extraction, utilising TF-IDF vectorization for text representation within a highly imbalanced multilingual dataset. The system integrates an interactive web dashboard for real-time misinformation classification, providing confidence scores and performance visualizations. Evaluation shows superior weighted F1-score performance over traditional baselines, with additional validation through cross-domain fact-checking articles from PolitiFact and Boom Live.

6. Conclusion

PROMID 2025 represents an important step toward a more comprehensive understanding of misinformation and hallucinations in modern NLP systems, particularly in the context of prompt-driven generation, cross-lingual summarisation, and real-world social media discourse. By introducing novel tasks such as prompt recovery and fine-grained factual error classification for Indian languages, the shared task expands the scope of misinformation research beyond output-only analysis and English-centric benchmarks. The strong participation and diversity of submitted systems demonstrate growing community interest in addressing these challenges, while also revealing significant open problems related to ambiguity, multilingual robustness, and factual faithfulness. We hope that the datasets, evaluation frameworks, and insights provided through PROMID will serve as a foundation for future work on interpretable, reliable, and socially responsible language technologies and encourage further exploration of mitigation techniques for hallucinations and misinformation in high-impact, multilingual settings.

Declaration on Generative AI

The authors have employed Generative AI tools for writing parts of this paper. However, all AI-generated content was thoroughly reviewed and edited. The authors take full responsibility for the accuracy of the publication's content.

References

[1] P. Mehta, T. Mandl, P. Majumder, S. Gangopadhyay, Report on the FIRE 2020 evaluation initiative, SIGIR Forum 55 (2021) 3:1–3:11. URL: https://doi.org/10.1145/3476415.3476418. doi:10.1145/3476415.3476418.

[2] J. Maynez, S. Narayan, B. Bohnet, R. McDonald, On faithfulness and factuality in abstractive summarization, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020.

[3] M. Cao, Y. Dong, J. C. K. Cheung, Hallucinated but factual! inspecting the factuality of hallucinations in abstractive summarization, in: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, 2022.

[4] W. Kryściński, B. McCann, C. Xiong, R. Socher, Evaluating the factual consistency of abstractive text summarization, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, 2020.

[5] L. Huang, W. Yu, W. Ma, W. Zhong, Z. Feng, H. Wang, Q. Chen, W. Peng, X. Feng, B. Qin, T. Liu, A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions, ACM Transactions on Information Systems (2024).

[6] P. Sahoo, P. Meharia, A. Ghosh, S. Saha, V. Jain, A. Chadha, A comprehensive survey of hallucination in large language, image, video and audio foundation models, in: Findings of the Association for Computational Linguistics: EMNLP 2024, 2024.

[7] A. Urlana, P. Chen, Z. Zhao, S. B. Cohen, M. Shrivastava, B. Haddow, Pmindiasum: Multilingual and cross-lingual headline summarization for languages in india, in: Findings of the Association for Computational Linguistics: EMNLP 2023, 2023.

[8] S. Satapara, P. Mehta, S. Modha, D. Ganguly, Indian language summarization at FIRE 2023, in: D. Ganguly, S. Majumdar, B. Mitra, P. Gupta, S. Gangopadhyay, P. Majumder (Eds.), Proceedings of the 15th Annual Meeting of the Forum for Information Retrieval Evaluation, FIRE 2023, Panjim, India, December 15-18, 2023, ACM, 2023, pp. 27–29. URL: https://doi.org/10.1145/3632754.3634662. doi:10.1145/3632754.3634662.

[9] T. Mandl, S. Modha, P. Majumder, D. Patel, M. Dave, C. Mandlia, A. Patel, Overview of the hasoc track at fire 2019: Hate speech and offensive content identification in indo-european languages, in: Proceedings of the 11th Forum for Information Retrieval Evaluation, 2019.

[10] S. Satapara, P. Mehta, S. Modha, A. Hegde, H. L. Shashirekha, D. Ganguly, Indian language summarization at FIRE 2024, in: D. Ganguly, D. K. Sanyal, P. Majumder, S. Majumdar, S. Gangopadhyay (Eds.), Proceedings of the 16th Annual Meeting of the Forum for Information Retrieval Evaluation, FIRE 2024, Gandhinagar, India, December 12-15, 2024, ACM, 2024, pp. 22–25. URL: https://doi.org/10.1145/3734947.3735668. doi:10.1145/3734947.3735668.

[11] T. Mandl, S. Modha, G. K. Shahi, H. Madhu, S. Satapara, P. Majumder, J. Schäfer, T. Ranasinghe, M. Zampieri, D. Nandini, A. K. Jaiswal, Overview of the HASOC subtrack at FIRE 2021: Hatespeech and offensive content identification in english and indo-aryan languages, in: P. Mehta, T. Mandl, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2021 - Forum for Information Retrieval Evaluation, Gandhinagar, India, December 13-17, 2021, volume 3159 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. 1–19. URL: http://ceur-ws.org/Vol-3159/T1-1.pdf.

[12] T. Mandl, S. Modha, G. K. Shahi, A. K. Jaiswal, D. Nandini, D. Patel, P. Majumder, J. Schäfer, Overview of the HASOC track at FIRE 2020: Hate speech and offensive content identification in indo-european languages, in: P. Mehta, T. Mandl, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2020 - Forum for Information Retrieval Evaluation, Hyderabad, India, December 16-20, 2020, volume 2826 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 87–111. URL: http://ceur-ws.org/Vol-2826/T2-1.pdf.

[13] S. Modha, T. Mandl, P. Majumder, D. Patel, Overview of the HASOC track at FIRE 2019: Hate speech and offensive content identification in indo-european languages, in: P. Mehta, P. Rosso, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2019 - Forum for Information Retrieval Evaluation, Kolkata, India, December 12-15, 2019, volume 2517 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019, pp. 167–190. URL: http://ceur-ws.org/Vol-2517/T3-1.pdf.

[14] H. Madhu, S. Satapara, S. Modha, T. Mandl, P. Majumder, Detecting offensive speech in conversational code-mixed dialogue on social media: A contextual dataset and benchmark experiments, Expert Systems with Applications (2022) 119342.

[15] S. Modha, P. Majumder, T. Mandl, C. Mandalia, Detecting and visualizing hate speech in social media: A cyber watchdog for surveillance, Expert Syst. Appl. 161 (2020) 113725. URL: https://doi.org/10.1016/j.eswa.2020.113725. doi:10.1016/j.eswa.2020.113725.

[16] M. Subramanian, R. Ponnusamy, S. Benhur, K. Shanmugavadivel, A. Ganesan, D. Ravi, G. K. Shanmugasundaram, R. Priyadharshini, B. R. Chakravarthi, Offensive language detection in tamil youtube comments by adapters and cross-domain knowledge transfer, Comput. Speech Lang. 76 (2022) 101404. URL: https://doi.org/10.1016/j.csl.2022.101404. doi:10.1016/j.csl.2022.101404.

[17] B. R. Chakravarthi, R. Priyadharshini, V. Muralidaran, S. Suryawanshi, N. Jose, E. Sherly, J. P. McCrae, Overview of the track on sentiment analysis for dravidian languages in code-mixed text, in: P. Majumder, M. Mitra, S. Gangopadhyay, P. Mehta (Eds.), FIRE 2020: Forum for Information Retrieval Evaluation, Hyderabad, India, December 16-20, 2020, ACM, 2020, pp. 21–24. URL: https://doi.org/10.1145/3441501.3441515. doi:10.1145/3441501.3441515.

[18] B. R. Chakravarthi, P. K. Kumaresan, R. Sakuntharaj, A. K. Madasamy, S. Thavareesan, B. Premjith, S. K, S. C. Navaneethakrishnan, J. P. McCrae, T. Mandl, Overview of the hasoc-dravidiancodemix shared task on offensive language detection in tamil and malayalam, in: P. Mehta, T. Mandl, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2021 - Forum for Information Retrieval Evaluation, Gandhinagar, India, December 13-17, 2021, volume 3159 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. 589–602. URL: http://ceur-ws.org/Vol-3159/T3-1.pdf.

[19] M. Amjad, G. Sidorov, A. Zhila, Data augmentation using machine translation for fake news detection in the urdu language, in: N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis (Eds.), Proceedings of The 12th Language Resources and Evaluation Conference, LREC 2020, Marseille, France, May 11-16, 2020, European Language Resources Association, 2020, pp. 2537–2542. URL: https://aclanthology.org/2020.lrec-1.309/.

[20] M. Amjad, A. Zhila, G. Sidorov, A. Labunets, S. Butt, H. I. Amjad, O. Vitman, A. F. Gelbukh, Overview of abusive and threatening language detection in urdu at FIRE 2021, in: P. Mehta, T. Mandl, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2021 - Forum for Information Retrieval Evaluation, Gandhinagar, India, December 13-17, 2021, volume 3159 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. 744–762. URL: http://ceur-ws.org/Vol-3159/T4-1.pdf.

[21] M. Amjad, N. Ashraf, A. Zhila, G. Sidorov, A. Zubiaga, A. F. Gelbukh, Threatening language detection and target identification in urdu tweets, IEEE Access 9 (2021) 128302–128313. URL: https://doi.org/10.1109/ACCESS.2021.3112500. doi:10.1109/ACCESS.2021.3112500.

[22] J. Gala, P. A. Chitale, R. AK, S. Doddapaneni, V. Gumma, A. Kumar, J. Nawale, A. Sujatha, R. Puduppully, V. Raghavan, et al., Indictrans2: Towards high-quality and accessible machine translation models for all 22 scheduled indian languages, arXiv preprint arXiv:2305.16307 (2023).

[23] S. Banerjee, K. Chakma, S. K. Naskar, A. Das, P. Rosso, S. Bandyopadhyay, M. Choudhury, Overview of the mixed script information retrieval (MSIR) at FIRE-2016, in: P. Majumder, M. Mitra, P. Mehta, J. Sankhavara, K. Ghosh (Eds.), Working notes of FIRE 2016 - Forum for Information Retrieval Evaluation, Kolkata, India, December 7-10, 2016, volume 1737 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2016, pp. 94–99. URL: http://ceur-ws.org/Vol-1737/T3-1.pdf.

[24] R. Sequiera, M. Choudhury, P. Gupta, P. Rosso, S. Kumar, S. Banerjee, S. K. Naskar, S. Bandyopadhyay, G. Chittaranjan, A. Das, K. Chakma, Overview of FIRE-2015 shared task on mixed script information retrieval, in: P. Majumder, M. Mitra, M. Agrawal, P. Mehta (Eds.), Post Proceedings of the Workshops at the 7th Forum for Information Retrieval Evaluation, Gandhinagar, India, December 4-6, 2015, volume 1587 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2015, pp. 19–25. URL: http://ceur-ws.org/Vol-1587/T2-1.pdf.

[25] P. Bhattacharya, K. Ghosh, S. Ghosh, A. Pal, P. Mehta, A. Bhattacharya, P. Majumder, Overview of the FIRE 2019 AILA track: Artificial intelligence for legal assistance, in: P. Mehta, P. Rosso, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2019 - Forum for Information Retrieval Evaluation, Kolkata, India, December 12-15, 2019, volume 2517 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019, pp. 1–12. URL: http://ceur-ws.org/Vol-2517/T1-1.pdf.

[26] P. Bhattacharya, P. Mehta, K. Ghosh, S. Ghosh, A. Pal, A. Bhattacharya, P. Majumder, FIRE 2020 AILA track: Artificial intelligence for legal assistance, in: P. Majumder, M. Mitra, S. Gangopadhyay, P. Mehta (Eds.), FIRE 2020: Forum for Information Retrieval Evaluation, Hyderabad, India, December 16-20, 2020, ACM, 2020, pp. 1–3. URL: https://doi.org/10.1145/3441501.3441510. doi:10.1145/3441501.3441510.

[27] V. Parikh, U. Bhattacharya, P. Mehta, A. Bandyopadhyay, P. Bhattacharya, K. Ghosh, S. Ghosh, A. Pal, A. Bhattacharya, P. Majumder, AILA 2021: Shared task on artificial intelligence for legal assistance, in: D. Ganguly, S. Gangopadhyay, M. Mitra, P. Majumder (Eds.), FIRE 2021: Forum for

Information Retrieval Evaluation, Virtual Event, India, December 13 - 17, 2021, ACM, 2021, pp. 12–15. URL: https://doi.org/10.1145/3503162.3506571. doi:10.1145/3503162.3506571.

[28] V. Parikh, V. Mathur, P. Mehta, N. Mittal, P. Majumder, Lawsum: A weakly supervised approach for indian legal document summarization, CoRR abs/2110.01188 (2021). URL: https://arxiv.org/abs/2110.01188. arXiv:2110.01188.

[29] S. Ghosh, A. Wyner, Identification of rhetorical roles of sentences in indian legal judgments, in: Legal Knowledge and Information Systems: JURIX 2019: The Thirty-second Annual Conference, volume 322, IOS Press, 2019, p. 3.

[30] S. Parashar, N. Mittal, P. Mehta, Casrank: A ranking algorithm for legal statute retrieval, Multimedia Tools and Applications (2023) 1–18.

[31] P. Mehta, P. Majumder, Optimum parameter selection for K.L.D. based authorship attribution in gujarati, in: Sixth International Joint Conference on Natural Language Processing, IJCNLP 2013, Nagoya, Japan, October 14-18, 2013, Asian Federation of Natural Language Processing / ACL, 2013, pp. 1102–1106. URL: https://aclanthology.org/I13-1155/.

[32] P. Mehta, P. Majumder, Large scale quantitative analysis of three indo-aryan languages, J. Quant. Linguistics 23 (2016) 109–132. URL: https://doi.org/10.1080/09296174.2015.1071151. doi:10.1080/09296174.2015.1071151.

[33] M. Basu, S. Ghosh, K. Ghosh, Overview of the fire 2018 track: Information retrieval from microblogs during disasters (irmidis), in: Proceedings of the 10th annual meeting of the Forum for Information Retrieval Evaluation, 2018, pp. 1–5.

[34] S. Majumdar, A. Bandyopadhyay, S. Chattopadhyay, P. P. Das, P. D. Clough, P. Majumder, Overview of the irse track at fire 2022: Information retrieval in software engineering, in: Forum for Information Retrieval Evaluation, ACM, 2022.

[35] S. Majumdar, S. Paul, D. Paul, A. Bandyopadhyay, S. Chattopadhyay, P. P. Das, P. D. Clough, P. Majumder, Generative ai for software metadata: Overview of the information retrieval in software engineering track at fire 2023, arXiv preprint arXiv:2311.03374 (2023).

[36] W. Lifferth, P. Mooney, S. Dane, A. Chow, Llm prompt recovery, https://kaggle.com/competitions/llm-prompt-recovery, 2024. Kaggle.

[37] L. Gao, R. Peng, Y. Zhang, J. Zhao, Dory: Deliberative prompt recovery for llm, arXiv preprint arXiv:2405.20657 (2024).

[38] S. Liu, Y. Gao, S. Zhai, L. Wang, Stylerec: A benchmark dataset for prompt recovery in writing style transformation, in: 2024 IEEE International Conference on Big Data (BigData), IEEE, 2024, pp. 1678–1685.

[39] A. Singhal, T. Law, C. Kassner, A. Gupta, E. Duan, A. Damle, R. L. Li, Multilingual fact-checking using llms, in: Proceedings of the Third Workshop on NLP for Positive Impact, 2024, pp. 13–31.

[40] R. K. Chikkala, T. Anikina, N. Skachkova, I. Vykopal, R. Agerri, J. van Genabith, Automatic fact-checking in english and telugu, arXiv preprint arXiv:2509.26415 (2025).

[41] A. Galassi, F. Ruggeri, A. B.-C. no, F. Alam, T. Caselli, M. Kutlu, J. M. Struss, F. Antici, M. Hasanain, J. Köhler, K. Korre, F. Leistra, A. Muti, M. Siegel, M. D. Turkmen, M. Wiegand, W. Zaghouani, Overview of the CLEF-2023 CheckThat! lab task 2 on subjectivity in news articles, in: Working Notes of CLEF 2023–Conference and Labs of the Evaluation Forum, CLEF '2023, Thessaloniki, Greece, 2023.

[42] G. Da San Martino, F. Alam, M. Hasanain, R. N. Nandi, D. Azizov, P. Nakov, Overview of the CLEF-2023 CheckThat! lab task 3 on political bias of news articles and news media, in: Working Notes of CLEF 2023–Conference and Labs of the Evaluation Forum, CLEF '2023, Thessaloniki, Greece, 2023.

[43] P. Nakov, F. Alam, G. Da San Martino, M. Hasanain, R. N. Nandi, D. Azizov, P. Panayotov, Overview of the CLEF-2023 CheckThat! lab task 4 on factuality of reporting of news media, in: Working Notes of CLEF 2023–Conference and Labs of the Evaluation Forum, CLEF '2023, Thessaloniki, Greece, 2023.

[44] Q. Peng, R. Moro, M. Gregor, I. Srba, S. Ostermann, M. Šimko, J. Podroužek, M. Mesarčík, J. Kopčan, A. Søgaard, Semeval-2025 task 7: Multilingual and crosslingual fact-checked claim retrieval, in:

Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025), 2025, pp. 2498–2511.

[45] R. Vázquez, T. Mickus, E. Zosa, T. Vahtola, J. Tiedemann, A. Sinha, V. Segonne, F. Sánchez-Vega, A. Raganato, J. Libovickỳ, et al., Semeval-2025 task 3: Mu-shroom, the multilingual shared task on hallucinations and related observable overgeneration mistakes, in: Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025), 2025.

[46] D. Zhang, M. D. Smucker, C. L. Clarke, Overview of the trec 2024 lateral reading track (2024).

[47] S. Satapara, B. Modha, S. Modha, P. Mehta, FIRE 2022 ILSUM track: Indian language summarization, in: D. Ganguly, S. Gangopadhyay, M. Mitra, P. Majumder (Eds.), Proceedings of the 14th Annual Meeting of the Forum for Information Retrieval Evaluation, FIRE 2022, Kolkata, India, December 9-13, 2022, ACM, 2022, pp. 8–11. URL: https://doi.org/10.1145/3574318.3574328. doi:10.1145/3574318.3574328.

[48] S. Satapara, B. Modha, S. Modha, P. Mehta, Findings of the first shared task on indian language summarization (ILSUM): approaches challenges and the path ahead, in: K. Ghosh, T. Mandl, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2022 - Forum for Information Retrieval Evaluation, Kolkata, India, December 9-13, 2022, volume 3395 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2022, pp. 369–382. URL: https://ceur-ws.org/Vol-3395/T6-1.pdf.

[49] 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.

[50] S. Satapara, P. Mehta, S. Modha, D. Ganguly, Key takeaways from the second shared task on indian language summarization (ILSUM 2023), in: K. Ghosh, T. Mandl, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2023 - Forum for Information Retrieval Evaluation (FIRE-WN 2023), Goa, India, December 15-18, 2023, volume 3681 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2023, pp. 724–733. URL: https://ceur-ws.org/Vol-3681/T8-1.pdf.

[51] S. Satapara, P. Mehta, S. Modha, D. Ganguly, Overview of the third shared task on indian language summarization (ilsum 2024), in: K. Ghosh, T. Mandl, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2024 - Forum for Information Retrieval Evaluation (FIRE 2024), Gandhinagar, India, December 12-15, 2024, CEUR Workshop Proceedings, CEUR-WS.org, 2024.

[52] S. Satapara, P. Mehta, D. Ganguly, S. Modha, Fighting fire with fire: Adversarial prompting to generate a misinformation detection dataset, CoRR abs/2401.04481 (2024). URL: https://doi.org/10.48550/arXiv.2401.04481. doi:10.48550/ARXIV.2401.04481. arXiv:2401.04481.

[53] 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.

[54] G. K. Shahi, O. Seneviratne, M. Spaniol, Semcafe: When named entities make the difference–assessing web source reliability through entity-level analytics, in: Proceedings of the 17th ACM Web Science Conference 2025, 2025, pp. 148–157.

[55] G. K. Shahi, T. A. Majchrzak, Amused: an annotation framework of multimodal social media data, in: International Conference on Intelligent Technologies and Applications, Springer, 2021, pp. 287–299.

[56] G. K. Shahi, A. Dirkson, T. A. Majchrzak, An exploratory study of covid-19 misinformation on twitter, Online social networks and media 22 (2021) 100104.

[57] C.-Y. Lin, Rouge: A package for automatic evaluation of summaries, in: Text summarization branches out, 2004, pp. 74–81.

[58] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, Y. Artzi, Bertscore: Evaluating text generation with bert, arXiv preprint arXiv:1904.09675 (2019).

[59] N. V. Gokul, J. J. Joel, S. Gautham, J. Rajeswari, Indicbertv2-mlm-only for fine-grained misinformation analysis in south indian languages, 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.

[60] K. Wang, P. Zhang, Q. Tan, Misinformation detection in social media texts and llm generated text using auxiliary text supervised learning, 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.

[61] R. Nagaraju, H. L. Shashirekha, From misrepresentation to quantities: Labeling misinformation types in south indian language summaries, 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.

[62] P. Saha, A lightweight contrastive system for misinformation detection in social media tweets, 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.

[63] T. Ehrhart, R. Troncy, G. Burel, H. Alani, Misinformation detection in russo-ukrainian conflict tweets, 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.

[64] A. Trivedi, C. Mallikarjuna, Misinformation detection in multilingual social media texts using llm-based translation, augmentation, and deberta fine-tuning, 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.

[65] P. Priyanshu, Detecting 2022 russo–ukrainian conflict misinformation using a hybrid transformer approach, 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.

[66] D. Sharma, Y. Sharma, Misinformation detection using ml, 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.

[67] J. Peng, Z. Lin, Z. Han, A social media misinformation detection model integrating semantic and twitter features, 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.

[68] S. Kumari, Automated detection of misinformation on twitter during the 2022 russo–ukrainian conflict, 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.

[69] K. S. Charan, U. Suman, R. Jain, J. Kaur, S. Sharma, Aurora: Automated understanding and recognition of omnilingual misinformation artefacts, 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.