LANGUAGE, DATA and KNOWLEDGE 2025

Proceedings of the 5th Conference on **Language, Data and Knowledge: Workshops**

UniorPress
Naples
2025

PROCEEDINGS OF THE 5TH CONFERENCE ON LANGUAGE, DATA AND KNOWLEDGE: WORKSHOPS

EDITORS:

Katerina Gkirtzou, ILSP "Athena" Research Center, Greece
Slavko Žitnik, University of Ljubljana, Slovenia
Jorge Gracia, University of Zaragoza, Spain
Dagmar Gromann, University of Vienna, Austria
Maria Pia di Buono, University of Naples "L'Orientale", Italy
Johanna Monti, University of Naples "L'Orientale", Italy
Maxim Ionov, University of Zaragoza, Spain

The proceedings are licensed under
Creative Commons Attribution 4.0 International

ISBN 978-88-6719-334-9

UniorPress — University of Naples "L'Orientale"
Via Nuova Marina 59 — 80133 Napoli (Italy)

UniorPress

Table of Contents

Fifth Workshop on Language Technology for Equality, Diversity and Inclusion

Introduction 1

SSNCSE@LT-EDI-2025:Detecting Misogyny Memes using Pretrained Deep Learning models
Sreeja K and Bharathi B 8

SSNCSE@LT-EDI-2025:Speech Recognition for Vulnerable Individuals in Tamil
Sreeja K and Bharathi B 13

CrewX@LT-EDI-2025: Transformer-Based Tamil ASR Fine-Tuning with AVMD Denoising and GRU-VAD for Enhanced Transcription Accuracy
Ganesh Sundhar S, Hari Krishnan N, Arun Prasad T D, Shruthikaa V and Jyothish Lal G 18

JUNLP@LT-EDI-2025: Efficient Low-Rank Adaptation of Whisper for Inclusive Tamil Speech Recognition Targeting Vulnerable Populations
Priyobroto Acharya, Soham Chaudhuri, Sayan Das, Dipanjan Saha and Dipankar Das 24

SKVtrio@LT-EDI-2025: Hybrid TF-IDF and BERT Embeddings for Multilingual Homophobia and Transphobia Detection in Social Media Comments
Konkimalla Laxmi Vignesh, Mahankali Sri Ram Krishna, Dondluru Keerthana and Premjith B33

Dll5143A@LT-EDI 2025: Bias-Aware Detection of Racial Hoaxes in Code-Mixed Social Media Data (BaCoHoax)
Ashok Yadav and Vrijendra Singh 38

Hope_for_best@LT-EDI 2025: Detecting Racial Hoaxes in Code-Mixed Hindi-English Social Media Data using a multi-phase fine-tuning strategy
Abhishek Singh Yadav, Deepawali Sharma, Aakash Singh and Vivek Kumar Singh 46

CVF-NITT@LT-EDI-2025:MisogynyDetection
Radhika K T and Sitara K 54

Wise@LT-EDI-2025: Combining Classical and Neural Representations with Multi-scale Ensemble Learning for Code-mixed Hate Speech Detection
Ganesh Sundhar S, Durai Singh K, Gnanasabesan G, Hari Krishnan N and MC Dhanush 61

CUET's_White_Walkers@LT-EDI 2025: Racial Hoax Detection in Code-Mixed on Social Media Data
Md. Mizanur Rahman, Jidan Al Abrar, Md. Siddikul Imam Kawser, Ariful Islam, Md. Mubasshir Naib and Hasan Murad 70

CUET's_White_Walkers@LT-EDI-2025: A Multimodal Framework for the Detection of Misogynistic Memes in Chinese Online Content
Md. Mubasshir Naib, Md. Mizanur Rahman, Jidan Al Abrar, Md. Mehedi Hasan, Md. Siddikul Imam Kawser and Mohammad Shamsul Arefin 75

CUET's_White_Walkers@LT-EDI 2025: Transformer-Based Model for the Detection of Caste and Migration Hate Speech
Jidan Al Abrar, Md. Mizanur Rahman, Ariful Islam, Md. Mehedi Hasan, Md. Mubasshir Naib and Mohammad Shamsul Arefin 82

NS@LT-EDI-2025 CasteMigration based hate speech Detection
Nishanth S, Shruthi Rengarajan and Sachin Kumar S 87

SSN_IT_HATE@LT-EDI-2025: Caste and Migration Hate Speech Detection
Maria Nancy C, Radha N and Swathika R 91

ItsAllGoodMan@LT-EDI-2025: Fusing TF-IDF and MuRIL Embeddings for Detecting Caste and Migration Hate Speech
Amritha Nandini K L, Vishal S, Giri Prasath R, Anerud Thiyagarajan and Sachin Kumar S . . . 97

NSR_LT-EDI-2025 Automatic speech recognition in Tamil
Nishanth S, Shruthi Rengarajan, Burugu Rahul and Jyothish Lal G 102

Solvers@LT-EDI-2025: Caste and Migration Hate Speech Detection in Tamil-English Code-Mixed Text
Ananthakumar S, Bharath P, Devasri A, Anirudh Sriram K S and Mohanapriya K T 107

CUET_N317@LT-EDI2025: Detecting Hate Speech Related to Caste and Migration with Transformer Models
Md. Nur Siddik Ruman, Md. Tahfim Juwel Chowdhury and Hasan Murad 112

KEC-Elite-Analysts@LT-EDI 2025: Leveraging Deep Learning for Racial Hoax Detection in Code-Mixed Hindi-English Tweets
Malliga Subramanian, Aruna A, Amudhavan M, Jahaganapathi S and Kogilavani Shanmugavadivel 118

Team_Luminaries_0227@LT-EDI-2025: A Transformer-Based Fusion Approach to Misogyny Detection in Chinese Memes
Adnan Faisal, Shiti Chowdhury, Momtazul Arefin Labib and Hasan Murad 123

Hinterwelt@LT-EDI 2025: A Transformer-Based Approach for Identifying Racial Hoaxes in Code-Mixed Hindi-English Social Media Narratives
Md. Abdur Rahman, Md. Al Amin, Sabik Aftahee and Md. Ashiqur Rahman 128

CUET_12033@LT-EDI-2025: Misogyny Detection
Mehreen Rahman, Faozia Fariha, Nabilah Tabassum, Samia Rahman and Hasan Murad 134

CUET_Blitz_Aces@LT-EDI-2025: Leveraging Transformer Ensembles and Majority Voting for Hate Speech Detection
Shahriar Farhan Karim, Anower Sha Shajalal Kashmary and Hasan Murad 140

Hinterwelt@LT-EDI 2025: A Transformer-Based Detection of Caste and Migration Hate Speech in Tamil Social Media
Md. Al Amin, Sabik Aftahee, Md. Abdur Rahman, Md. Sajid Hossain Khan and Md. Ashiqur Rahman 147

EM-26@LT-EDI 2025: Detecting Racial Hoaxes in Code-Mixed Social Media Data
Tewodros Achamaleh, Fatima Uroosa, Nida Hafeez, Tolulope Olalekan Abiola, Mikiyas Mebraihtu, Sara Getachew, Grigori Sidorov and Rolando Quintero 153

EM-26@LT-EDI 2025: Caste and Migration Hate Speech Detection in Tamil-English Code-Mixed Social Media Texts
Tewodros Achamaleh, Tolulope Olalekan Abiola, Mikiyas Mebraihtu, Sara Getachew and Grigori Sidorov 160

Hoax Terminators@LT-EDI 2025: CharBERT's dominance over LLM Models in the Detection of Racial Hoaxes in Code-Mixed Hindi-English Social Media Data
Abrar Hafiz Rabbani, Diganta Das Droba, Momtazul Arefin Labib, Samia Rahman and Hasan Murad 167

CUET_Ignite@LT-EDI-2025: A Multimodal Transformer-Based Approach for Detecting Misogynistic Memes in Chinese Social Media
Md. Mahadi Rahman, Mohammad Minhaj Uddin, Mohammad Oman and Mohammad Shamsul Arefin 179

girlsteam@LT-EDI-2025: Caste/Migration based hate speech Detection
Towshin Hossain Tushi, Walisa Alam, Rehenuma Ilman and Samia Rahman 185

CUET_320@LT-EDI-2025: A Multimodal Approach for Misogyny Meme Detection in Chinese Social Media
Madiha Ahmed Chowdhury, Lamia Tasnim Khan, Md. Shafiqul Hasan and Ashim Dey 191

Speech Personalization using Parameter Efficient Fine-Tuning for Nepali Speakers
Kiran Pantha, Rupak Raj Ghimire and Bal Krishna Bal 197

An Overview of the Misogyny Meme Detection Shared Task for Chinese Social Media
Bharathi Raja Chakravarthi, Rahul Ponnusamy, Ping Du, Xiaojian Zhuang, Saranya Rajiakodi, Paul Buitelaar, Premjith B, Bhuvaneswari Sivagnanam, Anshid K A and SK Lavanya 207

Findings of the Shared Task Multilingual Bias and Propaganda Annotation in Political Discourse
Shunmuga Priya Muthusamy Chinnan, Bharathi Raja Chakravarthi, Meghann Drury-Grogan, Senthil Kumar B, Saranya Rajiakodi and Angel Deborah S 216

Findings of the Shared Task Caste and Migration Hate Speech Detection
Saranya Rajiakodi, Bharathi Raja Chakravarthi, Rahul Ponnusamy, Shunmuga Priya Muthusamy Chinnan, Prasanna Kumar Kumaresan, Sathiyaraj Thangasamy, Bhuvaneswari Sivagnanam, Balasubramanian Palani, Kogilavani Shanmugavadivel, Abirami Murugappan and Charmathi Rajkumar 222

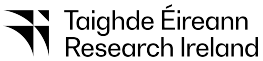
Overview of the Shared Task on Detecting Racial Hoaxes in Code-Mixed Hindi-English Social Media Data
Bharathi Raja Chakravarthi, Prasanna Kumar Kumaresan, Shanu Dhawale, Saranya Rajiakodi, Sajeetha Thavareesan, Subalalitha Chinnaudayar Navaneethakrishnan and Thenmozhi Durairaj . . . 229

Overview of Homophobia and Transphobia Span Detection in Social Media Comments
Prasanna Kumar Kumaresan, Bharathi Raja Chakravarthi, Ruba Priyadharshini, Paul Buitelaar, Malliga Subramanian and Kishore Kumar Ponnusamy 236

Overview of the Fifth Shared Task on Speech Recognition for Vulnerable Individuals in Tamil
Bharathi B, Bharathi Raja Chakravarthi, Sripriya N, Rajeswari Natarajan, Ratnavel Rajalakshmi and Suhasini S 242

Fifth Workshop on Language Technology for Equality, Diversity, Inclusion

Workshop organizers gratefully acknowledge the support from the following sponsors.



Fifth Workshop on Language Technology for Equality, Diversity and Inclusion

We are excited to welcome you to the Fifth Workshop on Language Technology for Equality, Diversity, Inclusion (LT-EDI-2025), co-located with the 5th Conference on Language, Data and Knowledge (LDK). This year, the workshop will be held in a hybrid format (both online and on-site) Workshops will take place at Palazzo S. Maria Porta Coeli on 9th September 2025, while the main venue for the conference will be Palazzo Corigliano, on 10th-11th September 2025, in Naples, Italy.

With the rapid advancement of technology, digital communication has become a central part of daily life. While many globally dominant languages have successfully transitioned into the digital era, numerous regional and low-resource languages continue to face significant technological challenges. Equality, Diversity and Inclusion (EDI) is an important agenda across every field throughout the world. Language as a major part of communication should be inclusive and treat everyone with equality. Today's large internet community uses language technology (LT) and has a direct impact on people across the globe. EDI is crucial to ensure everyone is valued and included, so it is necessary to build LT that serves this purpose. Recent results have shown that big data and deep learning are entrenching existing biases and that some algorithms are even naturally biased due to problems such as 'regression to the mode'. Our focus is on creating LT that will be more inclusive of gender, racial, sexual orientation, persons with disability. The workshop will focus on creating speech and language technology to address EDI not only in English, but also in less resourced languages.

The workshop received a total of 40 active submissions. Reviewer recruitment was highly effective, with 232 out of 249 invited reviewers accepting the invitation. Of the 270 assigned reviews, 117 were completed, resulting in a review submission rate of 43.33%. Additionally, 41.67% of reviewers (100 out of 240) completed all their assigned reviews. A majority of submissions (65%, or 26 out of 40) received at least three reviews, ensuring a robust evaluation process. Decisions were finalized for all submissions (100%), leading to an acceptance rate of 95% (38 papers). This included 6 papers (15%) accepted for oral presentations and 32 papers (80%) accepted for poster presentations. Only 2 submissions (5%) were rejected. There were no withdrawn submissions, and only one paper was desk rejected. These metrics reflect a thorough and inclusive review process, driven by active reviewer participation and a strong commitment to quality.

Bharathi Raja Chakravarthi, B Bharathi, Durairaj Thenmozhi, Miguel Ángel García Cumbreras, Alud María Jiménez Zafra and Paul Buitelaar

LT-EDI 2025 Organising Committee

Organizing Committee

Bharathi Raja Chakravarthi, University of Galway, Ireland
Bharathi B, Sri Sivasubramaniya Nadar College of Engineering, India
Paul Buitelaar, University of Galway, Ireland
Thenmozhi Durairaj, Sri Sivasubramaniya Nadar College of Engineering, India
Miguel Ángel García Cumbreras, University of Jaén, Spain
Salud María Jiménez-Zafra, Universidad de Jaén, Spain
Prasanna Kumar Kumaresan, Data Science Institute, University of Galway, Ireland
Shunmuga Priya Muthusamy Chinnan, Data Science Institute, University of Galway, Ireland

Program Committee

A. Justin Gopinath, Vellore Institute of Technology, India
Abdullah Al Nahian, Shahjalal University of Science and Technology, Bangladesh
Abdur Rahman, Shahjalal University of Science and Technology, Bangladesh
Abirami Jayaraman, Sri Sivasubramaniya Nadar College of Engineering, India
Adeep Hande, Comcast Applied AI, USA
Aishwarya Selvamurugan, Sri Eshwar College of Engineering, India
Amit Jaspal, Facebook, Inc.
Angel Deborah S, Sri Sivasubramaniya Nadar College of Engineering, India
Anusha M D Gowda, University of Mysore, India
Ariful Islam, Chittagong University of Engineering and Technology, Bangladesh
Aruna Devi Shanmugam, Sri Sivasubramaniya Nadar College of Engineering, India
Arunaggiri Pandian Karunanidhi, Micron Technology, Inc.
Asha Hegde, Mangalore University, India
Ashim Dey, Chittagong University of Engineering and Technology, Bangladesh
Ashok Yadav, Indian Institute of Information Technology, Allahabad, India
Ashraful Islam Paran, Chittagong University of Engineering and Technology, Bangladesh
Avaneesh Koushik, Sri Sivasubramaniya Nadar College of Engineering, India
Bagavathi C, Amrita Vishwa Vidyapeetham, India
Belo Abhigyan, University of Delhi, India
Bhuvaneswari Sivagnanam, Central University of Tamil Nadu, India
Boomika E, RMK Engineering College, India
Burugu Rahul, Amrita Vishwa Vidyapeetham, India
Deeptanshu Jha, IEEE, India
Dipanjan Saha, Jadavpur University, India
Dondluru Keerthana, Amrita Vishwa Vidyapeetham, India
Durga Prasad Manukonda, ASRlytics, USA
Fred Philippy, University of Luxemburg and Zortify S.A., Luxembourg
Ganesh Sundhar S, Amrita Vishwa Vidyapeetham, India
Geetha M P, Vellore Institute of Technology, India
Gersome Shimi, Madras Christian College, India
Girma Yohannis Bade, Addis Ababa University, Ethiopia
Gnanasabesan G, Amrita Vishwa Vidyapeetham, India
Hari Krishnan N, Amrita Vishwa Vidyapeetham, India
Harshita Sharma, Institute of Informatics and Communication, India
Hasan Murad, Chittagong University of Engineering and Technology, Bangladesh
Hosahalli Lakshmaiah Shashirekha, Mangalore University, India
Jayanth Jeyadevaswamy, University of Galway, Ireland
Jobin Jose, Indian Institute of Information Technology, Kottayam, India
Jyothish Lal G, Amrita Vishwa Vidyapeetham, India
Kasu Sai Kartheek Reddy, Indian Institute of Technology Tirupati, India
Keerthana Nnl, Vellore Institute of Technology, India
Lalith Kishore V P, RMK Engineering College, India
Luxshan Thavarasa, University of Moratuwa, Sri Lanka
Md. Mahadi Rahman, Chittagong University of Engineering and Technology, Bangladesh
Mahankali Sri Ram Krishna, Amrita Vishwa Vidyapeetham, India
Mahfuz Ahmed Anik, Shahjalal University of Science and Technology, Bangladesh
Mahir Absar Khan, Shahjalal University of Science and Technology, Bangladesh
Manan Buddhadev, Rochester Institute of Technology, USA

Md. Minhazul Kabir, Chittagong University of Engineering and Technology, Bangladesh
Md. Mizanur Rahman, Chittagong University of Engineering and Technology, Bangladesh
Md. Ayon Mia, Dhaka International University, Bangladesh
Md. Refaj Hossan, Chittagong University of Engineering and Technology, Bangladesh
Md. Mubasshir Naib, Chittagong University of Engineering and Technology, Bangladesh
Mikhail Krasitskii, Instituto Politécnico Nacional, Mexico
Minhaz Chowdhury, Shahjalal University of Science and Technology, Bangladesh
Minoru Sasaki, Ibaraki University, Japan
Miriam Butt, University of Konstanz, Germany
Mithun M, Sri Eshwar College of Engineering, India
Mohan Raj M A, RMK Engineering College, India
Mohan Raj, Monash University, Australia
Monorama Swain, Indian Institute of Information Technology, Kurnool, India
Moogambigai A, Sri Sivasubramaniya Nadar College of Engineering, India
Mostafa Rahgouy, Auburn University, USA
Mugilkrishna D U, Sri Sivasubramaniya Nadar College of Engineering, India
N.Nasurudeen Ahamed, United Arab Emirates University, UAE
Nida Hafeez, Instituto Politécnico Nacional, Mexico
Nishanth.S Nishanth.S, Amrita Vishwa Vidyapeetham, India
Nitin Nikamanth Appiah Balaji, Hexion Inc.
Pandiarajan D, Sri Sivasubramaniya Nadar College of Engineering, India
Payal Godhani, Oracle, India
Premjith B, Amrita Vishwa Vidyapeetham, India
Priyanka Ashokan, Sree Chitra Thirunal College of Engineering, India
Radhika K T, National Institute of Technology Trichy and Institute of Printing Technology and Government Polytechnic College, India
Rahatun Nesa Priti, Shahjalal University of Science and Technology, Bangladesh
Rajalakshmi Sivanaiah, Sri Sivasubramaniya Nadar College of Engineering, India
RajeswariRajasekar RajeswariRajasekar, Sri Sivasubramaniya Nadar Institutions, India
Raksha Adyanthaya, Yenepoya Institute Of Arts, Science, Commerce and Management, India
Ratnavel Rajalakshmi, Vellore Institute of Technology, India
Ravi Teja Potla, NVIDIA, USA
Sabik Aftahee, Chittagong University of Engineering and Technology, Bangladesh
Sai Koneru, Pennsylvania State University, USA
Sarbajeet Pattanaik, Indian Institute of Information Technology, Allahabad, India
Satya Subrahmanya Gautama Shastry Bulusu Venkata, George Mason University, USA
Saurabh Aggarwal, Autodesk, India
Sayan Das, Jadavpur University, India
Shreyas Karthik, Sri Sivasubramaniya Nadar College of Engineering, India
Shruthi Rengarajan, Amrita Vishwa Vidyapeetham, India
Shruthikaa V, Amrita Vishwa Vidyapeetham, India
Sidney Wong, University of Canterbury, New Zealand
Simran Simran, Institute of Informatics and Communication, India
Sitara K, National Institute of Technology Tiruchirappalli, India
Soham Chaudhuri, Jadavpur University, India
Somsubhra De, Indian Institute of Technology, Roorkee
Sreeja K, Sri Sivasubramaniya Nadar College of Engineering, India
Sripriya N, Sri Sivasubramaniya Nadar College of Engineering, India
Tanisha Sriram, Sri Sivasubramaniya Nadar College of Engineering, India
Tareque Md. Hanif, Shahjalal University of Science and Technology, Bangladesh
Tewodros Achamaleh, University of Gondar, Ethiopia

Tolulope Olalekan Abiola, Instituto Politécnico Nacional, Mexico
Trina Chakraborty, Shahjalal University of Science and Technology, Bangladesh
Udoy Das, Chittagong University of Engineering and Technology, Bangladesh
Uma Jothi, Amrita Vishwa Vidyapeetham, India
Vajratiya Vajrobol, Vellore Institute of Technology, India