

International Conference on Natural Language Processing

Proceedings of the Conference

The ICON organizers gratefully acknowledge the support from the following sponsors.

PLATIUM SPONSOR



OTHER SPONSOR



ORGANIZERS







CENTER FOR DEVELOPMENT OF ADVANCED COMPUTING



Natural Language Processing Association of India

©2024 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL) 209 N. Eighth Street
Stroudsburg, PA 18360
USA

Tel: +1-570-476-8006 Fax: +1-570-476-0860

acl@aclweb.org

ISBN

Organizing Committee

Patron

Dr. R. Velraj, Vice Chancellor, Anna University

Advisory Committee Chair

Rajeev Sangal, IIIT Hyderabad

General Chair

Sivaji Bandyopadhyay, IIIT Hyderabad Shivaji Bandyopadhyay, Jadavpur University

Program Chairs

Sobha L, AU-KBC Research Centre, Chennai Karunesh Arora, CDAC, Noida

Area Chairs

Ashutosh Modi, IIT Kanpur

Asif Ekbal, IIT Patna

Dipankar Das, Jadavpur University, Kolkata

Dipti Misra Sharma, IIIT Hyderabad

Girish Nath Jha, JNU, Delhi

CV Jawahar, IIIT Hyderabad

Jyoti Pawar, Goa University

Kalika Bali, Microsoft Research India, Bengaluru

Karunesh Arora, CDAC Noida

Manish Srivastava, IIIT Hyderabad

Monojit Choudhury, AI Centre, Abu Dhabi

Niladri Shekhar Dash, ISI Kolkata

Partha Pakray, NIT Silchar

Pattabhi RK Rao, AU-KBC Research Centre, Chennai

Pawan Goyal, IIT Kharagpur

Prasenjit Majumder, DAIICT, Gandhinagar

Pushpak Bhatacharyya, IIT Bombay

Radhika Mamidi, IIIT Hyderabad

Rajeev R R, ICFOSS, Thiruvananthapuram

Sobha L, AU-KBC Research Centre, Chennai

Srinivas Banglore, Interactions LLC, USA

Sriparna Saha, IIT Patna

Sudeshna Sarkar, IIT Kharagpur

Sudip kumar Naskar, Jadavpur University, Kolkata

Tanmoy Chakraborty, IIT Delhi

Vasudeva Varma, IIIT Hyderabad

Vijay Sundar Ram, AU-KBC Research Centre, Chennai

Vishal Goyal, Punjabi University, Patiala

Doctoral Consortium Chairs

Radhika Mamidi , IIIT Hyderabad Jyoti Pawar, Goa University

Shared Task/Tool Contest

Partha Pakray, NIT Silchar Vijay Sundar Ram, AU-KBC Research Centre for Emerging Technology

Workshop/Tutorial

Sudip Kumar Naskar, Jadavpur University

Local Organizing Committee Chair

Sobha L, AU-KBC Research Centre for Emerging Technology Jayashree P, AU-KBC Research Centre for Emerging Technology Kavitha S, AU-KBC Research Centre for Emerging Technology Pattabhi RK Rao. AU-KBC Research Centre for Emerging Technology Vijay Sundar Ram, AU-KBC Research Centre for Emerging Technology

Sponsorship Chair

Vasudeva Varma, IIIT Hyderabad

Program Committee

Ashwath Rao B Ashwini Vaidya Asif Ekbal Atreyee Sharma Atul Kumar Ojha Bal Krishna Bal Baskaran Sankaran Bhradwaja Kumar G Chinmai Dharurkar **Chris Potts Constantin Orason** Corina Forascu Costanza Navarretta Dhanalakshmi V Dipankar Das Dipti Misra Sharma Diyi Yang Ekaterina Kochmar Elizabeth Sherly Elsi Kaiser Ganesh Ramakrishnan Geeetha T.V. Gihan Dias Girish Jha Hema Murthy Irina Terminikova James Anthonysamay Jawahar C.V. James Martin Julia Hockenmaier Jyoti Pawar Kalika Bali

Akilan R

Amba Kulkarni Amitava Das Anand Kumar M Anil Kumar Singh Anil Vuppala Aniruddha Ghosh Anitha Pillai

Anoop Kunchukuttan Apurbalal Senapati Arnab Bhattacharya Arulmozi Selvaraj Ashutosh Modi Kamal Kumar Choudhary

Karthik Gali

karunesh arora

Katharina Kann

Kengatharaiyer Sarveswaran

Kentaro Inui

Kishorjit Nongmeikapam

Krishna Miyapuram

Luke Zettlemoyer

Mahadeva Prasanna S.R.

Manish Shrivastava

manjira sinha

Md. Shad Akhtar

Monica Lam

Monojit Choudhury

Nagarajan T

Narayana Murthy Kavi

Niladri Chatterjee

Niladri Dash

Noah Smith

Parameswari Krishnamurthy

Partha Pakray

Pattabhi RK Rao

Pawan Goyal

Percy Liang

Pinaki Bhaskar

Prasenjit Majumder

Preslav Nakov

Priyanka Biswas

Pruthwik Mishra

Pushpak Bhattacharyya

Radhika Mamidi

Rajeev RR

Rajendran S

Rakesh Balabantaray

Renuka Devi

Ritesh Kumar

Sachin Pawar

Sandipan Dandapat

Sanjay Chatterji

Sarveswaran K

Sarveswaran Sharanyaa

Shashirekha Hosahalli Lakshmaiah

Sheeja Sudheer

Sheeja Sudhir

Shikhar Kumar Sarma

Sivaji Bandyopadhyay

Sobha Lalitha Devi

Soma Paul

Somi Ram

Srikanth Reddy

Srikumar K.

Srinivas Bangalore

Sriparna Saha

Subalalitha CN

Sudarsun S

Sudeshna Sarkar

Sudip Naskar

Sukhada Sukhada

Sukomal Pal

Suma Bhat

Surupendu Gangopadhyay

Susan Brown

Swapnil Chaudhari

Swaran Latha

Tanmoy Bhatacharya

Tanmoy Chakraborty

Tatsunori Hashimoto

Thenmozhi D.

Thoudam Doren Singh

Tim Althoff

Timothy Baldwin

Tirthankar Dasgupta

Uma mahaeswara Rao

Uma Maheswari

Umesh S

Uthayasanker Thayasivam

Utpal Garain

Utsab Barman

Vandan Mujadia

Vasudeva Varma

Vijay Sundar Ram

Vinayak Abro

Vishal Goyal

Keynote Talk: Why AI Is WEIRD and Should Not Be This Way: Towards AI For Everyone, With Everyone, By Everyone

Prof. Rada Mihalcea University of Michigan

Abstract

Recent years have witnessed remarkable advancements in AI, with language and vision models that have enabled progress in numerous applications and opened the door to the integration of AI in areas such as communication, transportation, healthcare, and arts. Yet, many of these models and their corresponding datasets are W.E.I.R.D. (Western, Educated, Industrialized, Rich, Democratic) and they are reflective of a small fraction of the population.(*) In this talk, I will show some of the limitations and lack of representation of current AI models, and highlight the need for cross-cultural language and vision models that can capture the diversity of behaviors, beliefs, and language expressions across different groups. I will also explore ways in which we can address these limitations by developing models that are re-centered around people and their unique characteristics.

(*) W.E.I.R.D. is an acronym widely used in psychology to indicate the limitation of many of the studies carried out in the field

Keynote Talk: Syntactic Blocking Effects in Hindi: a computational exploration

Prof. Rajesh Bhatt, University of Massachusetts, Amherst

Abstract

Rajesh Bhatt received his PhD in 1999 from the University of Pennsylvania, where he worked with Professor Aravind Joshi. Prior to that, he received a BTech in Computer Science from IIT Kanpur in 1993 and was advised by Professor Rajeev Sangal. His research interests involve the syntax-semantics interface, the comparative syntax of Modern Indo-Aryan languages, and Tree Adjoining Grammars. Within the syntax-semantics interface, he has written on the interaction of aspect and modality, comparatives, and implicit arguments and within comparative syntax, he has worked on long distance agreement, closest conjunct agreement, and correlative constructions. He was part of the team that built the NSF funded Hindi/Urdu Treebanks. Prior to UMass, he taught at the University of Texas at Austin. He has also taught at the LSA Summer Institutes at MIT and at Boulder, the LOT Summer School, and LISSIM.

Keynote Talk: Machine Speech Chain: From Human Auditory Feedback Principles to Language Technology Empowering Indigenous Communities

Prof. Sakriani Sakti Nara Institute of science & Technology, Japan

Abstract

"Automatic speech recognition (ASR) and text-to-speech synthesis (TTS) have enabled machines to mimic human speech perception and production. However, unlike humans, machines lack a closed-loop mechanism to monitor and adjust their output, as ASR and TTS are trained separately using large amounts of paired speech-text data. In contrast, humans learn by simultaneously listening and speaking, using auditory feedback to refine their speech. This talk introduces the machine speech chain framework, inspired by human auditory feedback, which enables semi-supervised learning by allowing ASR and TTS to teach each other with unpaired data. Applications to multilingual and multimodal settings for low-resource languages will be explored. Additionally, I will share insights from indigenous communities, highlighting challenges in ensuring language technologies effectively support under-resourced languages.

Keynote Talk: Culturally Aware Machines: Why and when are they useful?

Prof. Monojit Choudhury University of Artificial Intelligence, Abu Dhabi

Abstract

Language, culture, and technology are deeply intertwined; any technology that aims to process or generate language must have a nuanced understanding of culture. Yet, culture is elusive, often defying concrete definition, which poses a profound challenge to NLP and AI research. While the NLP community has acknowledged this fact and taken initial steps — developing culture-specific benchmarks and identifying biases in large language models (LLMs)—there remains a gap in addressing why cultural understanding matters for machines, and the real-world impact on users of culturally (un)aware machines. In this talk, I will ask and attempt to answer the following three questions: (1) What can we do better with AI systems that are culturally aware? (2) Why is modeling culture so complex, and what can we learn from psychology and anthropology come to our rescue? (3) How can LLMs, in turn, be leveraged as tools for studying and enhancing our understanding of cultures?

Table of Contents

Sumotosima: A Framework and Dataset for Classifying and Summarizing Otoscopic Images Eram Anwarul Khan and Anas Anwarul Haq Khan
Exploring Expected Answer Types for Effective Question Answering Systems for low resource language.
Chindukuri Mallikarjuna and Sangeetha Sivanesan
Precision Empowers, Excess Distracts: Visual Question Answering With Dynamically Infused Knowledge In Language Models Manas Jhalani, Annervaz K M and Pushpak Bhattacharyya
Improving Few-shot Prompting using Cluster-based Sample Retrieval for Medical NER in Clinical Text Meethu Mohan C, Sneha Shaji Punnan and Jeena Kleenankandy
Standardizing Genomic Reports: A Dataset, A Standardized Format, and A Prompt-Based Technique for Structured Data Extraction
Tamali Banerjee, Akshit Varmora, Jay J. Gorakhiya, Sanand Sasidharan, Anuradha Kanamarlapudi and Pushpak Bhattacharyya45
Identification of Idiomatic Expressions in Konkani Language Using Neural Networks Naziya Mahamdul Shaikh and Jyoti Pawar
Konkani Wordnet Visualizer as a Concept Teaching-Learning Tool Sunayana R. Gawde, Jayram Ulhas Gawas, Shrikrishna R. Parab, Shilpa Neenad Desai and Jyoti Pawar
A Systematic Exploration of Linguistic Phenomena in Spoken Hindi: Resource Creation and Hypothesis Testing Aadya Ranjan, Sidharth Ranjan and Rajakrishnan Rajkumar
A Corpus of Hindi-English Code-Mixed Posts to Hate Speech Detection Prashant Kapil and Asif Ekbal
An Aid to Assamese Language Processing by Constructing an Offline Assamese Handwritten Dataset Debabrata Khargharia and Samir Kumar Borgohain
Monolingual text summarization for Indic Languages using LLMs Jothir Adithya T K, Nithish Kumar S, Felicia Lilian J and Mahalakshmi S
Enhancing Trust and Interpretability in Malayalam Sentiment Analysis with Explainable AI Meharuniza Nazeem, Anitha R, Navaneeth S and Rajeev R. R
Automatic Sanskrit Poetry Classification Based on Kāvyaguṇa Amruta Barbadikar and Amba Kulkarni
From Data to Insights: The Power of LM's in Match Summarization Satyavrat Gaur, Pasi Shailendra, Rajdeep Kumar, Rudra Chandra Ghosh and Nitin Sharma 120
Sentiment and sarcasm: Analyzing gender bias in sports through social media with deep learning Sethulakshmi Praveen, Balaji TK, Sreeja SR and Annushree Bablani
Quality Estimation of Machine Translated Texts based on Direct Evidence Approach vibhuti kumari and Narayana Murthy Kavi

Reconsidering SMT Over NMT for Closely Related Languages: A Case Study of Persian-Hindi Pair Waisullah Yousofi and Pushpak Bhattacharyya
RoMantra: Optimizing Neural Machine Translation for Low-Resource Languages through Romanization
Govind Soni and Pushpak Bhattacharyya157
Domain Dynamics: Evaluating Large Language Models in English-Hindi Translation Soham Bhattacharjee, Baban Gain and Asif Ekbal
Synthetic Data and Model Dynamics based Performance Analysis for Assamese-Bodo Low Resource NMT
Kuwali Talukdar, Prof. Shikhar Kumar Sarma and Kishore Kashyap
Detecting AI-Generated Text with Pre-Trained Models Using Linguistic Features ANNEPAKA YADAGIRI, Lavanya Shree, Suraiya Parween, Anushka Raj, Shreya Maurya and Partha Pakray
Towards Understanding the Robustness of LLM-based Evaluations under Perturbations Manav Chaudhary, Harshit Gupta, Savita Bhat and Vasudeva Varma
Exploring Kolmogorov Arnold Networks for Interpretable Mental Health Detection and Classification from Social Media Text
Ajay Surya Jampana, Mohitha Velagapudi, Neethu Mohan and Sachin Kumar S
Human vs Machine: An Automated Machine-Generated Text Detection Approach Urwah Jawaid, Rudra Roy, Pritam Pal, Srijani Debnath, Dipankar Das and Sivaji Bandyopadhyay
Survey on Computational Approaches to Implicature Kaveri Anuranjana, Srihitha Mallepally, Sriharshitha Mareddy, Amit Shukla and Radhika Mamidi
Exploring User Dissatisfaction: Taxonomy of Implicit Negative Feedback in Virtual Assistants Moushumi Mahato, Avinash Kumar, Kartikey Singh, Javaid Nabi, Debojyoti Saha and Krishna Singh
Pronominal Anaphora Resolution in Konkani language incorporating Gender Agreement Poonam A. Navelker and Jyoti Pawar
Story-Yarn: An Interactive Story Building Application Hryadyansh Saraswat, Snehal D. Shete, Vikas Dangi, Kushagra Agrawal, Anuj Aggarwal and Aditya Nigam
End to End Multilingual Coreference Resolution for Indian Languages Sobha Lalitha Devi, Vijay SundaR Ram and Pattabhi RK Rao
LangBot-Language Learning Chatbot Madhubala Sundaram, Pattabhi RK Rao and Sobha Lalitha Devi
Towards Efficient Audio-Text Keyword Spotting: Quantization and Multi-Scale Linear Attention with Foundation Models Rahothvarman P and Radhika Mamidi
Utilizing POS-Driven Pitch Contour Analysis for Enhanced Tamil Text-to-Speech Synthesis Preethi Thinakaran, Anushiya Rachel Gladston, P Vijayalakshmi, T Nagarajan, Malarvizhi Muthuramalingam and Sooriya S

RoundTripOCR: A Data Generation Technique for Enhancing Post-OCR Error Correction in Low- Resource Devanagari Languages
Harshvivek Ankush Kashid and Pushpak Bhattacharyya27
MULTILATE: A Synthetic Dataset on AI-Generated MULTImodaL hATE Speech Advaitha Vetagiri, Eisha Halder, Ayanangshu Das Majumder, Partha Pakray and Amitava Das28
We Care: Multimodal Depression Detection and Knowledge Infused Mental Health Therapeutic Response Generation Palash Moon and Pushpak Bhattacharyya
Tarash Woon and Tushpak Bhattacharyya29
CM_CLIP: Unveiling Code-Mixed Multimodal Learning with Cross-Lingual CLIP Adaptations Gitanjali Kumari, Arindam Chatterjee, Ashutosh Bajpai, Asif Ekbal and Vinutha B. NarayanaMurthy
Aiding Non-Verbal Communication: A Bidirectional Language Agnostic Framework for Automating Text to AAC Generation Piyali Karmakar and manjira sinha
Towards Enhancing Knowledge Accessibility for Low-Resource Indian Languages: A Template Based Approach
Srijith Padakanti, Akhilesh Aravapalli, Abhijith Chelpuri and Radhika Mamidi
Value to User's Voice: A Generative AI Framework for Actionable Insights from Customer Reviews in Consumer Electronics Radhika Mundra, Bhavesh Kukreja, Aritra Ghosh Dastidar, Kartikey Singh and Javaid Nabi337
Comprehensive Plagiarism Detection in Malayalam Texts Through Web and Database Integration Meharuniza Nazeem, Parvathy Raj, Rajeev R. R, Anitha R and Navaneeth S34
MalUpama - Figurative Language Identification in Malayalam -An Experimental Study Reenu Paul, Wincy Abraham and Anitha S. Pillai
Survey of Pseudonymization, Abstractive Summarization & Spell Checker for Hindi and Marathi Rasika Ransing, Mohammed Amaan Dhamaskar, Ayush Rajpurohit, Amey Dhoke and Sanket Dalvi
Natural Answer Generation: From Factoid Answer to Full-length Answer using Grammar Correction Manas Jain, Sriparna Saha, Pushpak Bhattacharyya, Gladvin Chinnadurai and Manish Vatsa37
Assessing Assamese Suffix Productivity: A Probabilistic Study in Resource-Limited Contexts Pinky Moni Gayan and Arup K. Nath
TRO(F)LL or ROFL?: Exploring Troll Detection in Tamil Memes Aditya Krishna Ponnudurai, Swetha J and Rajalakshmi Sivanaiah
Empowering SW Security: CodeBERT and Machine Learning Approaches to Vulnerability Detection Lov Kumar, Vikram Singh, Srivalli Patel and Pratyush Mishra
Improving on the Limitations of the ASR Model in Low-Resourced Environments Using Parameter- Efficient Fine-Tuning Rupak Raj Ghimire, Prakash Poudyal and Bal Krishna Bal
Open-Source OCR Libraries: A Comprehensive Study for Low Resource Language Meharuniza Nazeem, Anitha R. Navaneeth S and Raieev R. R

A Comparative Assessment of Machine Learning Techniques in Kannada Multi-Emotion Sentiment Analysis
Dakshayani Ijeri and Pushpa B. Patil42
SansGPT: Advancing Generative Pre-Training in Sanskrit Rhugved Pankaj Chaudhari, Bhakti Jadhav, Pushpak Bhattacharyya and malhar kulkarni43
Automating Humor: A Novel Approach to Joke Generation Using Template Extraction and Infilling Mayank Goel, Parameswari Krishnamurthy and Radhika Mamidi
Chirp Group Delay based Feature for Speech Applications Malarvizhi Muthuramalingam, Anushiya Rachel Gladston, P Vijayalakshmi and T Nagarajan 44
LOC: Livestock Ontology Construction Approach From Domain based Text Documents Nandhana Prakash, Amudhan A, Nithish R and Krithikha Sanju Saravanan
Extractive Summarization using Extended TextRank Algorithm Ansh N. Vora, Rinit Mayur Jain, Aastha Sanjeev Shah and Sheetal Sonawane
Enhancing Masked Word Prediction in Tamil Language Models: A Synergistic Approach Using BERT and SBERT Viswadarshan R R, Viswaa Selvam S, Felicia Lilian J and Mahalakshmi S
DesiPayanam: developing an Indic travel partner Diviya K N, Mrinalini K, Vijayalakshmi P, Thenmozhi J and Nagarajan T
A Survey on Combating Hate Speech through Detection and Prevention in English Prashant Kapil and Asif Ekbal
Integration of Self-Attention Model with Intralingual Word Embedding for Contextual Semantic Analysis of Thirukkural Text. Shanthi Murugan, Kaviyarasu S and Balasundaram S R
A self-supervised domain-independent Named Entity Recognition using local similarity Keerthi S. A. Vasan and Uma Satya Ranjan
Profanity and Offensiveness Detection in Nepali Language Using Bi-directional LSTM Models Abiral Adhikari, Prashant Manandhar, Reewaj Khanal, Samir Wagle, Praveen Acharya and Bal Krishna Bal
PollCardioKG: A Dynamic Knowledge Graph of Interaction Between Pollution and Cardiovascular Diseases
Sudeshna Jana, Anunak Roy, manjira sinha and Tirthankar Dasgupta52
Shabdocchar: Konkani WordNet Enrichment with Audio Feature Sunayana R. Gawde, Shrikrishna R. Parab, Jayram Ulhas Gawas, Shilpa Neenad Desai and Jyoti Pawar
Analytics Graph Query Solver (AGQS): Transforming Natural Language Queries into Actionable Insights
Debojyoti Saha, Krishna Singh, Moushumi Mahato and Javaid Nabi
Emojis Trash or Treasure: Utilizing Emoji to Aid Hate Speech Detection Tanik Saikh Soham Barman Harsh Kumar Saswat Sahu and Souvick Palit

	Ceviews: A Comparative Study using various LLMs Pradnya Bhagat and Jyoti Pawar562
Sentiment Analysis for Konkani using Zero-Shot . Seamus Fred Rodrigues, Pradnya Bhagat, A	, ,
Persuasion Games with Large Language Models Shirish Karande, Santhosh V and Yash Bha	tia <u>576</u>
Vector Embedding Solution for Recommendatio Vidya P V and Ajeesh Ramanujan	n System
Multi-document Summarization by Ensembling Rajendra Kumar Roul, Navpreet . and Saif	of Scoring and Topic Modeling Techniques Nalband588
	nd Embeddings to tackle Hinglish NLP challenges Syush Mishra593
Landscape Painter: Mimicking Human Like Art Yash Gogoriya, Oswald C and Abhijith Bala	Using Generative Adversarial Networks an
Automatic Summarization of Long Documents Naman Chhibbar and Jugal Kalita	607
(GoP) measure	with DNN-HMM based goodness of pronunciation jayalakshmi P and Nagarajan T616
Severity Classification and Dysarthric Speech L Sanjay B, Priyadharshini M.K, Vijayalakshi	Detection using Self-Supervised Representations mi P and Nagarajan T621
using NMT Models	um Machine Translation of Shlokas and Sentence input629

Program

Friday, December 20, 2024

8:30 AM - 09:45 AM Registration

10:00 AM - 10:30 AM Inaugural Ceremony

10:00 AM - 11:00 AM *Keynote 1:* Culturally Aware Machines: Why and when are they

useful?

Prof. Monojit Chaudari

11:30 AM - 12:00 PM *Industry Session:*

Presentation by Digital India Bhashini Division

12:00 PM - 01:00 PM *Oral Presentation Session (Parallel Sessions)*

Session 1: QA, Information Extraction + Information Retrieval

and Text Mining

Session 2: Language Resource

Session 3: Sentiment Analysis and Summarization

02:00 PM - 03:00 PM *Oral Presentations (Parallel Sessions)*

Session 4: Machine Translation

Session 5: Machine Learning

Session 6: Discourse

03:30 PM - 04:00 PM Panel Session (Industry)

04:00 PM - 05:00 PM Keynote 2: Syntactic Blocking Effects in Hindi: a computational

exploration

Prof. Rajesh Bhutt

Saturday.	December	21.	2024
-----------	-----------------	-----	------

10:00 AM - 11:00 AM Keynote 3: Why AI Is WEIRD and Should Not Be This Way:

Towards AI For Everyone, With Everyone, By Everyone

Prof. Rada Mihalcea

11:30 AM - 12:30 PM Special Session: The Secretary, Ministry of Electronics and

Information Technology (MeitY)

12:30 PM - 01:00 PM *Poster Session*

02:00 PM - 03:30 PM *Oral Presentations (Parallel Sessions)*

Session 7: Speech Processing +OCR +Multimodality

Session 8: NLP Applications

02:30 PM - 03:30 PM Shared Task

03:30 PM - 04:30 PM Poster Booster Session

Poster Booster Presentation (10 no)

Poster Booster Presentation (10No)

04:30 PM - 05:00 PM Doctoral Consortium

05:00 PM - 06:00 PM Keynote 4: Machine Speech Chain: From Human Auditory Feedback

Principles to Language Technology Empowering Indigenous

Communities

Prof. Sakriani Sakti