

NAACL 2025

**Annual Conference of the Nations of the Americas Chapter of
the Association for Computational Linguistics**

**Proceedings of the Conference
Volume 2: Short Papers**

April 29 - May 4, 2025

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

Platinum



Gold



Bronze



Diversity and Inclusion Ally



©2025 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
317 Sidney Baker St. S
Suite 400 - 134
Kerrville, TX 78028
USA
Tel: +1-855-225-1962
acl@aclweb.org

ISBN 979-8-89176-190-2

Message from the General Chair

Message from the General Chair Welcome to the 2025 meeting of the Nations of the Americas Chapter of the Association for Computational Linguistics! I am proud to help organize the first NAA-CL conference to carry the new name of our organization, one that emphasizes inclusion for all of the Americas. I am also pleased to welcome you to Albuquerque, New Mexico, a state whose unique blend of cultural influences will make for an excellent backdrop for NAACL 2025, especially with this year's special theme on NLP in a Multicultural World.

This year's program benefited from a now mature ACL rolling review process. I would like to extend a big thank you to our ARR Editors in Chief, Viviane Moreira, Anna Rogers, and Michael White, who were very helpful not just with reviewing for the main conference, but who also shared their OpenReview expertise with chairs from our other tracks. We also benefited from the helpful advice of last year's NAACL general and program chairs: Katrin Erk, Kevin Duh, Helena Gomez, and Steven Bethard. Finally, Ryan Cotterell stepped up to help with the publications process and software, even while not serving as publications chair.

Of course, a conference of this magnitude cannot come together without some drama; in our case, we had some unexpected funding shortages from traditional government sources. We would like to extend a huge thank you to the boards of both the ACL and NAACL for filling those funding gaps, ensuring that our important D&I, volunteer, and student author support programs continue to thrive.

The virtual component of our conference is crucial to an inclusive, affordable experience for all NAACL authors and attendees. This year, we made small refinements to the hybrid format, mirroring NAACL 2024's choices to combine a virtual poster session via Gather with asynchronous online content via Underline. Our virtual poster session will be hosted on May 6, the Tuesday after the conference, in the hopes that promoting it at the conference's plenary sessions will help boost attendance. We opted not to have virtual oral presentations at the in-person conference, as those continue to be tricky to get right. To participants, virtual as well as in-person: Please let us know what worked for you and what did not, so we can continue to improve the hybrid experience.

The job of General Chair is a strange one, as it mostly involves cheering on many other people as they do amazing work. I have been fortunate to have been teamed with an excellent set of program chairs; to Luis Chiruzzo, Alan Ritter, and Lu Wang: thanks for everything, I'm very proud of what we've built together. I'd also like to extend my heartfelt thanks to Jenn Rachford (ACL) and Damira Mršić (Underline) who provide the knowledge, continuity and professionalism to bring all of this together.

Many thanks also to:

- Workshop chairs: Saab Mansour, Kenton Murray, and Alexis Palmer
- Tutorial chairs: Maria Lomeli, Swabha Swayamdipta, and Rui Zhang
- Demo chairs: Nouha Dziri, Shizhe Diao, and Sean (Xiang) Ren
- Industry track chairs: Weizhu Chen, Xue-Yong Fu, Mohammad Kachuee, and Yi Yang
- Student research workshop chairs: Abteen Ebrahimi, Emmy Liu, and Samar Haider, and their faculty advisors Maria Leonor Pacheco and Shira Wein
- Publication chairs: Arman Cohan, Manling Li, and Yichao Zhou
- Website chairs: Arya McCarthy and Vered Shwartz

- Publicity and social media chairs: Eleftheria Briakou, Tuhin Chakrabarty, and Ximena Gutierrez-Vasques
- Diversity and inclusion chairs: Akiko I. Eriguchi, Chi-Kiu (Jackie) Lo, and Niloofar Miresghallah
- Sponsorship chairs: Prithviraj (Raj) Ammanabrolu and Maha Elbayad
- Volunteers chairs: Robin Jia and David Mortensen
- Ethics chairs: Manuel Mager and Yulia Tsvetkov
- Handbook chair: Winston Wu
- Best paper committee chairs: Marine Carpuat and Anna Rumshisky
- Visa chairs: Eduardo Blanco and Parisa Kordjamshidi
- Virtual infrastructure chair: Jieyu Zhao

Whenever possible, I tried to populate each committee with someone who had served in the same role in NAACL 2024, to provide continuity, so I'll extend an extra thanks to all chairs who accepted this second year of service. Thanks also to the members of the ACL and NAACL Executive Committees for their support, feedback, and advice.

Finally, I would like to thank all authors, invited speakers and panelists, area chairs and reviewers, volunteers and session chairs, and all attendees, in-person and virtual. The conference is nothing without you.

Welcome again and enjoy the conference!

Colin Cherry

Google

NAACL 2025 General Chair

Message from the Program Chairs

Message from the Program Chairs Welcome to the 2025 Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics! NAACL 2025 marks an exciting milestone as the first conference held under our new name, reflecting our commitment to greater inclusivity across the diverse communities of the Americas. NAACL 2025 is a hybrid conference, and we are excited to have attendees and presenters join us both in person in Albuquerque and online from around the globe. We are thrilled to welcome you to what promises to be a vibrant and engaging conference.

Special Theme: NLP in a Multicultural World Current NLP tools and models, particularly large language models (LLMs), rely on vast amounts of data for training. However, this data often over-represents a small number of dominant languages, and even within those, tends to prioritize certain geographical or cultural varieties. As a result, a long tail of under-represented languages, dialects, and cultural contexts remains largely overlooked by the NLP community. For NAACL 2025, we introduced a special theme track on “NLP in a Multicultural World.” With this theme track, we sought to foster discussion and research on how NLP can better serve the linguistic and cultural diversity of the world. We encouraged contributions on topics such as cultural localization of language models, new NLP applications to support people from diverse cultures, revitalization or refunctionalization of endangered or sleeping languages, analysis of cultural biases in language models, and historical considerations and diachronic analysis. This track was dedicated to developing more inclusive, culturally aware NLP techniques that reflect and support the vibrant multicultural world we live in.

We received 71 submissions to the special theme, of which 23 were accepted for presentation at the conference. We hope these papers spark meaningful conversations and inspire future work in this important and evolving area of research.

Review Process 3,185 papers were submitted to the October ARR cycle, of which we estimate 3,099 were intended to be submitted to NAACL based on the “preferred venue” field in the submission form. We also received 147 papers from previous ARR cycles committed to NAACL. The program chairs recruited 98 Senior Area Chairs to view reviews and metareviews provided by ARR and make final recommendations on which papers to accept to both the main conference and Findings. 1,432 Area Chairs wrote metareviews for ARR, and 10,648 reviewers wrote reviews for the submitted papers.

Acceptance Rate Calculating an acceptance rate is challenging due to the multi-step ARR review process, in which papers are first submitted to ARR to get reviews, then authors commit their papers (together with reviews) to a specific *ACL conference. Of the 3,185 papers submitted to the October ARR cycle, we estimate that 3,099 intended to submit to NAACL. Based on this information, we estimate that 22% of papers submitted to the October cycle and intended for NAACL were accepted to the main conference, and another 15% were accepted to Findings, bringing the total estimated acceptance rate for papers accepted to be presented at the conference (Main + Findings) to 37%. Out of the 1,647 papers committed to NAACL with ARR reviews, 719 were accepted to the main conference, and 477 were accepted to Findings. 40 papers were desk rejected or withdrawn.

Presentation Format At NAACL 2025, papers were assigned one of three possible presentation modes: in-person participants could be assigned oral or poster presentations, while virtual participants could present posters. We selected 246 of the papers accepted to the main conference as oral presentations, and

the rest of them were assigned as posters, together with all the Findings papers. Oral presentations were assigned a 15-minute slot, with 12 minutes for presentation and 3 minutes for questions. For choosing papers as oral presentations, we first split all papers according to track, sorted them according to overall score (considering review, metareview, SAC recommendation), and took into consideration the authors' presentation preference. Then we grouped papers in sets of 6. Some tracks had very few accepted papers, so some of them were grouped together to form areas of affinity.

Program Format NAACL 2025 consists both of an in-person and a virtual conference, held on different days. The virtual part of the conference is held after the in-person one and a few days later (on May 6), so participants traveling home after the in-person conference could attend the virtual conference. The conference program includes 3 keynote speakers: Rada Mihalcea (University of Michigan), Mike Lewis (Meta), and Josh Tenenbaum (Massachusetts Institute of Technology). 260 papers are scheduled to be presented as oral presentations (also including papers from TACL, CL, and the industry track), 594 papers are scheduled as in-person posters, and 256 virtual posters.

Gratitude NAACL 2025 would not have been possible without the hard work of all people involved. We thank everyone who contributed, including:

- The General Chair, Colin Cherry.
- The ARR Editors-in-Chief of the October 2024 cycle: Viviane Moreira, Anna Rogers, Michael White.
- The OpenReview team, especially Rachel Smart.
- The 98 Senior Area Chairs.
- The 1,432 Area Chairs and 10,648 Reviewers.
- The best paper committee chairs, Marine Carpuat and Anna Rumshisky.
- The ethics chairs, Yulia Tsvetkov and Manuel Mager, and their team of reviewers.
- The website chairs, Vered Shwartz and Arya McCarthy.
- The publication chairs, Yichao Zhou, Manling Li, and Arman Cohan.
- The publicity chairs, Ximena Gutierrez-Vasques, Eleftheria Briakou, and Tuhin Chakrabarty.
- The volunteers chairs, Robin Jia and David Mortensen.
- The visa chairs, Eduardo Blanco and Parisa Kordjamshidi.
- The ACL Anthology Director, Matt Post, and his team.
- The Program Chairs of EMNLP 2024 (Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung (Vivian) Chen) and NAACL 2024 (Kevin Duh, Helena Gomez, and Steven Bethard).
- Damira Mršić and the Underline Team.
- Jenn Rachhford and the entire conference support staff.

Luis Chiruzzo, Universidad de la República, Uruguay

Alan Ritter, Georgia Institute of Technology

Lu Wang, University of Michigan

NAACL 2025 Program Committee Co-Chairs

April 2025

Table of Contents

<i>Complete Chess Games Enable LLM Become A Chess Master</i> Yinqi Zhang, Xintian Han, Haolong Li, Kedi Chen and Shaohui Lin	1
<i>Predicting the Target Word of Game-playing Conversations using a Low-Rank Dialect Adapter for Decoder Models</i> Dipankar Srirag, Aditya Joshi and Jacob Eisenstein	8
<i>Chal-TeA: A Benchmark for Evaluating Autocompletion of Interactions with LLM-based Chatbots</i> Shani Goren, Oren Kalinsky, Tomer Stav, Yuri Rapoport, Yaron Fairstein, Ram Yazdi, Nachshon Cohen, Alexander Libov and Guy Kushilevitz	18
<i>Cross-Lingual Transfer Learning for Speech Translation</i> Rao Ma, Mengjie Qian, Yassir Fathullah, Siyuan Tang, Mark Gales and Kate Knill	33
<i>Reverse Question Answering: Can an LLM Write a Question so Hard (or Bad) that it Can't Answer?</i> Nishant Balepur, Feng Gu, Abhilasha Ravichander, Shi Feng, Jordan Lee Boyd-Graber and Rachel Rudinger	44
<i>Personalized Help for Optimizing Low-Skilled Users' Strategy</i> Feng Gu, Wichayaporn Wongkamjan, Jordan Lee Boyd-Graber, Jonathan K. Kummerfeld, Denis Peskoff and Jonathan May	65
<i>Local Prompt Optimization</i> Yash Jain and Vishal Chowdhary	75
<i>Cross-lingual Transfer of Reward Models in Multilingual Alignment</i> Jiwoo Hong, Noah Lee, Rodrigo Martínez-Castaño, César Rodríguez and James Thorne	82
<i>Inference-Time Selective Debiasing to Enhance Fairness in Text Classification Models</i> Gleb Kuzmin, Neemesh Yadav, Ivan Smirnov, Timothy Baldwin and Artem Shelmanov	95
<i>Automatic Evaluation of Healthcare LLMs Beyond Question-Answering</i> Anna Arias-Duart, Pablo Agustin Martin-Torres, Daniel Hincos, Pablo Bernabeu-Perez, Lucia Urcelay Ganzabal, Marta Gonzalez Mallo, Ashwin Kumar Gururajan, Enrique Lopez-Cuena, Sergio Alvarez-Napagao and Dario Garcia-Gasulla	108
<i>STRUX: An LLM for Decision-Making with Structured Explanations</i> Yiming Lu, Yebowen Hu, Hassan Foroosh, Wei Jin and Fei Liu	131
<i>Improving Vietnamese-English Cross-Lingual Retrieval for Legal and General Domains</i> Toan Ngoc Nguyen, Nam Le Hai, Nguyen Doan Hieu, Dai An Nguyen, Linh Ngo Van, Thien Huu Nguyen and Sang Dinh	142
<i>Computational Discovery of Chiasmus in Ancient Religious Text</i> Hope McGovern, Hale Sirin and Tom Lippincott	154
<i>Characterizing the Effects of Translation on Intertextuality using Multilingual Embedding Spaces</i> Hope McGovern, Hale Sirin and Tom Lippincott	161
<i>LLM2: Let Large Language Models Harness System 2 Reasoning</i> Cheng Yang, Chufan Shi, Siheng Li, Bo Shui, Yujiu Yang and Wai Lam	168
<i>Context-Efficient Retrieval with Factual Decomposition</i> Yanhong Li, David Yunis, David McAllester and Jiawei Zhou	178

<i>Sports and Women's Sports: Gender Bias in Text Generation with Olympic Data</i>	
Laura Biester	195
<i>Alligators All Around: Mitigating Lexical Confusion in Low-resource Machine Translation</i>	
Elizabeth Nielsen, Isaac Rayburn Caswell, Jiaming Luo and Colin Cherry	206
<i>PROM: Pivoted and Regulated Optimization for Multilingual Instruction Learning</i>	
Jaeseong Lee, Seung-won Hwang, Hojin Lee, Yunju Bak and Changmin Lee	222
<i>Concept-Reversed Winograd Schema Challenge: Evaluating and Improving Robust Reasoning in Large Language Models via Abstraction</i>	
Kaiqiao Han, Tianqing Fang, Zhaowei Wang, Yangqiu Song and Mark Steedman	229
<i>Defense against Prompt Injection Attacks via Mixture of Encodings</i>	
Ruiyi Zhang, David Sullivan, Kyle Jackson, Pengtao Xie and Mei Chen	244
<i>Watching the AI Watchdogs: A Fairness and Robustness Analysis of AI Safety Moderation Classifiers</i>	
Akshit Acharya and Anshuman Chhabra	253
<i>CoRAG: Collaborative Retrieval-Augmented Generation</i>	
Aashiq Muhamed, Mona T. Diab and Virginia Smith	265
<i>Is It Navajo? Accurate Language Detection for Endangered Athabaskan Languages</i>	
Ivory Yang, Weicheng Ma, Chunhui Zhang and Soroush Vosoughi	277
<i>Don't Touch My Diacritics</i>	
Kyle Gorman and Yuval Pinter	285
<i>Pretrained Image-Text Models are Secretly Video Captioners</i>	
Chunhui Zhang, Yiren Jian, Zhongyu Ouyang and Soroush Vosoughi	292
<i>Reverse Modeling in Large Language Models</i>	
Sicheng Yu, Xu Yuanchen, Cunxiao Du, Yanying Zhou, Minghui Qiu, Qianru Sun, Hao Zhang and Jiawei Wu	306
<i>Preserving Multilingual Quality While Tuning Query Encoder on English Only</i>	
Oleg Vasilyev, Randy Sawaya and John Bohannon	321
<i>Using Contextually Aligned Online Reviews to Measure LLMs' Performance Disparities Across Language Varieties</i>	
Zixin Tang, Chieh-Yang Huang, Tsung-che LI, Ho Yin Sam Ng, Hen-Hsen Huang and Ting-Hao Kenneth Huang	342
<i>Towards Federated Low-Rank Adaptation of Language Models with Rank Heterogeneity</i>	
Yuji Byun and Jaeho Lee	356
<i>Related Knowledge Perturbation Matters: Rethinking Multiple Pieces of Knowledge Editing in Same-Subject</i>	
Zenghao Duan, Wenbin Duan, Zhiyi Yin, Yinghan Shen, Shaoling Jing, Jie Zhang, Huawei Shen and Xueqi Cheng	363
<i>STEP: Staged Parameter-Efficient Pre-training for Large Language Models</i>	
Kazuki Yano, Takumi Ito and Jun Suzuki	374
<i>Language Models Encode Numbers Using Digit Representations in Base 10</i>	
Amit Arnold Levy and Mor Geva	385

<i>A Systematic Study of Cross-Layer KV Sharing for Efficient LLM Inference</i> You Wu, Haoyi Wu and Kewei Tu	396
<i>AMPS: ASR with Multimodal Paraphrase Supervision</i> Abhishek Gupta, Amruta Parulekar, Sameep Chattopadhyay and Preethi Jyothi	404
<i>Taxi1500: A Dataset for Multilingual Text Classification in 1500 Languages</i> Chunlan Ma, Ayyoob Imani, Haotian Ye, Renhao Pei, Ehsaneddin Asgari and Hinrich Schuetze	414
<i>GameTox: A Comprehensive Dataset and Analysis for Enhanced Toxicity Detection in Online Gaming Communities</i> Usman Naseem, Shuvam Shiwakoti, Siddhant Bikram Shah, Surendrabikram Thapa and Qi Zhang	440
<i>FaithBench: A Diverse Hallucination Benchmark for Summarization by Modern LLMs</i> Forrest Sheng Bao, Miaoran Li, Renyi Qu, Ge Luo, Erana Wan, Yujia Tang, Weisi Fan, Manveer Singh Tamber, Suleman Kazi, Vivek Sourabh, Mike Qi, Ruixuan Tu, Chenyu Xu, Matthew Gonzales, Ofer Mendelevitch and Amin Ahmad	448
<i>Debate-Feedback: A Multi-Agent Framework for Efficient Legal Judgment Prediction</i> Xi Chen, Mao Mao, Shuo Li and Haotian Shangguan	462
<i>Great Memory, Shallow Reasoning: Limits of kNN-LMs</i> Shangyi Geng, Wenting Zhao and Alexander M Rush	471
<i>Repetition Neurons: How Do Language Models Produce Repetitions?</i> Tatsuya Hiraoka and Kentaro Inui	483
<i>STAR: Spectral Truncation and Rescale for Model Merging</i> Yu-Ang Lee, Ching-Yun Ko, Tejaswini Pedapati, I-Hsin Chung, Mi-Yen Yeh and Pin-Yu Chen	496
<i>Task-driven Layerwise Additive Activation Intervention</i> Hieu Trung Nguyen, Bao Nguyen, Binh Nguyen and Viet Anh Nguyen	506
<i>Scaling Multi-Document Event Summarization: Evaluating Compression vs. Full-Text Approaches</i> Adithya Pratapa and Teruko Mitamura	514
<i>Black-Box Visual Prompt Engineering for Mitigating Object Hallucination in Large Vision Language Models</i> Sangmin Woo, Kang Zhou, Yun Zhou, Shuai Wang, Sheng Guan, Haibo Ding and Lin Lee Cheong	529
<i>A Layered Debating Multi-Agent System for Similar Disease Diagnosis</i> Yutian Zhao, Huimin Wang, Yefeng Zheng and Xian Wu	539
<i>The Geometry of Numerical Reasoning: Language Models Compare Numeric Properties in Linear Subspaces</i> Ahmed Oumar El-Shangiti, Tatsuya Hiraoka, Hilal AlQuabeh, Benjamin Heinzerling and Kentaro Inui	550
<i>AlignFreeze: Navigating the Impact of Realignment on the Layers of Multilingual Models Across Diverse Languages</i> Steve Bakos, David Guzmán, Riddhi More, Kelly Chutong Li, Félix Gaschi and En-Shiun Annie Lee	562

<i>FLIQA-AD: a Fusion Model with Large Language Model for Better Diagnose and MMSE Prediction of Alzheimer's Disease</i>	
Junhao Chen, Zhiyuan Ding, Yan Liu, Xiangzhu Zeng and Ling Wang	587
<i>Transform Retrieval for Textual Entailment in RAG</i>	
Quan Guo and Xin Liang	595
<i>How do Multimodal Foundation Models Encode Text and Speech? An Analysis of Cross-Lingual and Cross-Modal Representations</i>	
Hyunji Lee, Danni Liu, Supriti Sinhamahapatra and Jan Niehues	600
<i>Explore the Reasoning Capability of LLMs in the Chess Testbed</i>	
Shu Wang, Lei Ji, Renxi Wang, Wenxiao Zhao, Haokun Liu, Yifan Hou and Ying Nian Wu	611
<i>Auto-Cypher: Improving LLMs on Cypher generation via LLM-supervised generation-verification framework</i>	
Aman Tiwari, Shiva Krishna Reddy Malay, Vikas Yadav, Masoud Hashemi and Sathwik Tejaswi Madhusudhan	623
<i>Leveraging Moment Injection for Enhanced Semi-supervised Natural Language Inference with Large Language Models</i>	
Seo Yeon Park	641
<i>A Fair Comparison without Translationese: English vs. Target-language Instructions for Multilingual LLMs</i>	
Taisei Enomoto, Hwichan Kim, Zhousi Chen and Mamoru Komachi	649
<i>Evaluating Multimodal Generative AI with Korean Educational Standards</i>	
Sanghee Park and Geewook Kim	671
<i>ScratchEval: Are GPT-4o Smarter than My Child? Evaluating Large Multimodal Models with Visual Programming Challenges</i>	
Rao Fu, Ziyang Luo, Hongzhan Lin, Zhen Ye and Jing Ma	689
<i>Interpret and Control Dense Retrieval with Sparse Latent Features</i>	
Hao Kang, Tevin Wang and Chenyan Xiong	700
<i>DART: An AIGT Detector using AMR of Rephrased Text</i>	
Hyeonchu Park, Byungjun Kim and Bugeun Kim	710
<i>Scaling Graph-Based Dependency Parsing with Arc Vectorization and Attention-Based Refinement</i>	
Nicolas Floquet, Joseph Le Roux, Nadi Tomeh and Thierry Charnois	722
<i>Language Models "Grok" to Copy</i>	
Ang Lv, Ruobing Xie, Xingwu Sun, Zhanhui Kang and Rui Yan	735
<i>Evaluating LLMs for Quotation Attribution in Literary Texts: A Case Study of LLaMa3</i>	
Gaspard Michel, Elena V. Epure, Romain Hennequin and Christophe Cerisara	742
<i>Beyond Literal Token Overlap: Token Alignability for Multilinguality</i>	
Katharina Hämmerl, Tomasz Limisiewicz, Jindřich Libovický and Alexander Fraser	756
<i>IdentifyMe: A Challenging Long-Context Mention Resolution Benchmark for LLMs</i>	
Kawshik Manikantan, Makarand Tapaswi, Vineet Gandhi and Shubham Toshniwal	768
<i>kNN Retrieval for Simple and Effective Zero-Shot Multi-speaker Text-to-Speech</i>	
Karl El Hajal, Ajinkya Kulkarni, Enno Hermann and Mathew Magimai Doss	778

<i>CORD: Balancing Consistency and Rank Distillation for Robust Retrieval-Augmented Generation</i> Youngwon Lee, Seung-won Hwang, Daniel F Campos, Filip Graliński, Zhewei Yao and Yuxiong He	787
<i>GraphLSS: Integrating Lexical, Structural, and Semantic Features for Long Document Extractive Summarization</i> Margarita Bugueño, Hazem Abou Hamdan and Gerard De Melo	797
<i>Step-by-Step Fact Verification System for Medical Claims with Explainable Reasoning</i> Juraj Vladika, Ivana Hacajova and Florian Matthes	805
<i>Developing multilingual speech synthesis system for Ojibwe, Mi'kmaq, and Maliseet</i> Shenran Wang, Changbing Yang, Michael I Parkhill, Chad Quinn, Christopher Hammerly and Jian Zhu	817
<i>Bottom-Up Synthesis of Knowledge-Grounded Task-Oriented Dialogues with Iteratively Self-Refined Prompts</i> Kun Qian, Maximillian Chen, Siyan Li, Arpit Sharma and Zhou Yu	827
<i>Sociodemographic Prompting is Not Yet an Effective Approach for Simulating Subjective Judgments with LLMs</i> Huaman Sun, Jiaxin Pei, Minje Choi and David Jurgens	845
<i>Identifying Power Relations in Conversations using Multi-Agent Social Reasoning</i> Zhaoqing Wu, Dan Goldwasser, Maria Leonor Pacheco and Leora Morgenstern	855
<i>Examining Spanish Counseling with MIDAS: a Motivational Interviewing Dataset in Spanish</i> Aylin Ece Gunal, Bowen Yi, John D. Piette, Rada Mihalcea and Veronica Perez-Rosas	866
<i>Self-Debiasing Large Language Models: Zero-Shot Recognition and Reduction of Stereotypes</i> Isabel O. Gallegos, Ryan Aponte, Ryan A. Rossi, Joe Barrow, Mehrab Tanjim, Tong Yu, Hanieh Deilamsalehy, Ruiyi Zhang, Sungchul Kim, Franck Dernoncourt, Nedim Lipka, Deonna Owens and Jiuxiang Gu	873
<i>EqualizeIR: Mitigating Linguistic Biases in Retrieval Models</i> Jiali Cheng and Hadi Amiri	889
<i>Do Audio-Language Models Understand Linguistic Variations?</i> Ramaneswaran Selvakumar, Sonal Kumar, Hemant Kumar Giri, Nishit Anand, Ashish Seth, Sreyan Ghosh and Dinesh Manocha	899
<i>Giving the Old a Fresh Spin: Quality Estimation-Assisted Constrained Decoding for Automatic Post-Editing</i> Sourabh Deoghare, Diptesh Kanojia and Pushpak Bhattacharyya	914
<i>RuleR: Improving LLM Controllability by Rule-based Data Recycling</i> Ming Li, Han Chen, Chenguang Wang, Dang Nguyen, Dianqi Li and Tianyi Zhou	926
<i>MixRevDetect: Towards Detecting AI-Generated Content in Hybrid Peer Reviews.</i> Sandeep Kumar, Samarth Garg, Sagnik Sengupta, Tirthankar Ghosal and Asif Ekbal	944
<i>DiscoGraMS: Enhancing Movie Screen-Play Summarization using Movie Character-Aware Discourse Graph</i> Maitreya Prafulla Chitale, Uday Bindal, Rajakrishnan P Rajkumar and Rahul Mishra	954
<i>Capturing Human Cognitive Styles with Language: Towards an Experimental Evaluation Paradigm</i> Vasudha Varadarajan, Syeda Mahwish, Xiaoran Liu, Julia Buffolino, Christian Luhmann, Ryan L. Boyd and H. Schwartz	966